



Marcel Bruno Pereira Braga

Escala de Proficiências em Concepções Térmicas: Diagnóstico Psicométrico de Estudantes em Portugal e Brasil

Vol. I: Fundamentos e Análises

Tese de Doutoramento em Ensino das Ciências, Ramo de Ensino da Física, em regime de cotutela entre a Universidade de Coimbra e a Universidade Federal de Santa Catarina, orientada pelos professores doutores Décio Ruivo Martins, Maria Augusta Vilalobos Filipe Pereira do Nascimento e José André Peres Angotti e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologias da Universidade de Coimbra

Junho de 2018



UNIVERSIDADE DE COIMBRA

Marcel Bruno Pereira Braga

Escala de Proficiências em Concepções Térmicas:
Diagnóstico Psicométrico de Estudantes em Portugal e Brasil

Vol. I: Fundamentos e Análises

Tese de Doutoramento do Programa de Doutoramento em Ensino das Ciências, Ramo no Ensino de Física, orientado pelo Senhor Professor Doutor Décio Ruivo Martins, Senhor Professor Doutor José André Peres Angotti e pela Senhora Professora Doutora Maria Augusta Vilalobos Filipe Pereira do Nascimento e apresentada na Faculdade de Ciências e Tecnologia da Universidade de Coimbra.



UNIVERSIDADE DE COIMBRA



**UNIVERSIDADE FEDERAL
DE SANTA CATARINA**
Centro de Ciências Físicas
e Matemáticas - CFM

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Braga, Marcel Bruno Pereira

Escala de Proficiências em Concepções Térmicas[tese]:
Diagnóstico Psicométrico de Estudantes em Portugal e Brasil
/Marcel Bruno Pereira Braga; orientador, Décio Ruivo
Martins, orientador, José André Peres Angotti,
coorientador, Maria Augusta Vilalobos Filipe Pereira do
Nascimento - Florianópolis, SC, 2018.

2v.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro de Ciências Físicas e Matemáticas, Programa
de Pós-Graduação em Educação Científica e Tecnológica,
Florianópolis, 2018.

Inclui referências.

Trabalho elaborado em regime de cotutela entre a
Universidade de Coimbra e a Universidade Federal de Santa
Catarina.

1. Educação Científica e Tecnológica. 2. Teste Diagnóstico.
3. Proficiências e Concepções Térmicas. 4. Teoria Clássica dos
Testes e Teoria da Resposta ao Item. 5. Portugal e Brasil. I.
Martins, Décio Ruivo. II. Angotti, José André Peres. III.
Nascimento, Maria Augusta Vilalobos Filipe Pereira do. IV.
Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Educação Científica e Tecnológica.
V Título.



UNIVERSIDADE DE COIMBRA
ADMINISTRAÇÃO

Doutoramento em Ensino das Ciências, Ramo de Ensino da Física em regime de cotutela entre a Universidade de Coimbra e a Universidade Federal de Santa Catarina, de Marcel Bruno Pereira Braga

ATA DA PROVA DE DOUTORAMENTO

Aos vinte e dois dias do mês de junho de dois mil e dezoito, na sala dos Capelos na Reitoria da Universidade de Coimbra sob a presidência de José António de Carvalho Paixão, Professor Catedrático da Faculdade de Ciências e Tecnologia da Universidade de Coimbra, por despacho de delegação de competências do Reitor da Universidade de Coimbra, número 2514/2016, publicado no Diário da República, segunda série, n.º 34, de 18 de fevereiro, reuniu o júri da prova de doutoramento acima mencionada.

Compareceram como vogais: Marta Feijó Barroso, Instituto de Física da Universidade Federal do Rio de Janeiro, Brasil; José André Peres Angotti, Universidade Federal de Santa Catarina, Brasil; David António da Costa, Universidade Federal de Santa Catarina, Brasil; Lucilia Maria Pessoa Tavares dos Santos, Universidade de Aveiro; Maria José Barata Marques de Almeida, Faculdade de Ciências e Tecnologia da Universidade de Coimbra; Manuela Ramos Marques da Silva, Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Aberta a sessão e verificando-se a existência de quórum legal, deu-se início à prova, a qual consistiu no seguinte:

Apreciação e discussão da tese, em língua portuguesa, apresentada pelo candidato intitulada: Escala de Proficiências em Conceções Térmicas: Diagnóstico Psicométrico de Estudantes em Portugal e Brasil.

Intervieram como arguentes principais Lucilia Maria Pessoa Tavares dos Santos e Marta Feijó Barroso. Intervieram ainda Maria José Barata Marques de Almeida, David António da Costa, Manuela Ramos Marques da Silva e José André Peres Angotti.

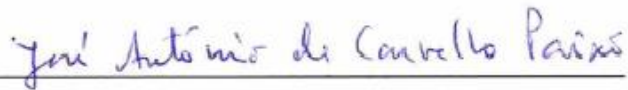
Seguidamente, reuniu o júri para proceder à votação nos termos legais. Atendendo à diversidade de áreas dos assuntos abordados e à profundidade com que o são, bem como à integração das diferentes vertentes, atendendo ainda à forma segura como o candidato respondeu às questões que lhe foram colocadas durante a discussão, o mesmo júri deliberou, por unanimidade, atribuir a classificação final de Aprovado com Distinção a Marcel Bruno Pereira Braga.

Nada mais havendo a tratar, foi a sessão encerrada.

Para constar se exarou a presente ata, que eu Leonor Simões dos Santos Quitério, técnica superior secretariei e redigi e que depois de lida em voz alta e aprovada, vai ser assinada por todos os membros do júri presentes e por mim.

3 Paixão
M. Feijó
J. Peres
D. Costa
L. Maria
M. José
M. Ramos
M. Barata
M. Ramos
D. António
M. Ramos
J. André

Presidente do Júri,

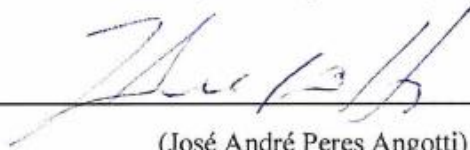


(José António de Carvalho Paixão)

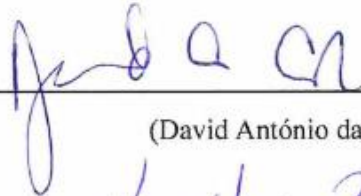
Vogais,



(Marta Feijó Barroso)



(José André Peres Angotti)



(David António da Costa)



(Lucília Maria Pessoa Tavares dos Santos)

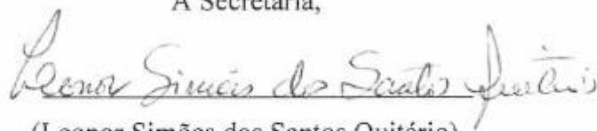


(Maria José Barata Marques de Almeida)



(Manuela Ramos Marques da Silva)

A Secretária,



(Leonor Simões dos Santos Quitério)

Marcel Bruno Pereira Braga

**Escala de Proficiências em Concepções Térmicas:
Diagnóstico Psicométrico de Estudantes em Portugal e Brasil**


Esta Tese foi julgada adequada para obtenção do Título de “Doutor (a)” e aprovada em sua forma final pelo Programa de Pós-graduação em Educação Científica e Tecnológica – UFSC, projeto Cotutela Faculdade de Ciências e Tecnologia - UC

Coimbra - Florianópolis, 22 de junho de 2018



Prof. (a) José António Paixão
Professor Catedrático
Presidente – FCTUC

Banca Examinadora:



Prof. José António Paixão
Professor Catedrático
Presidente - FCTUC



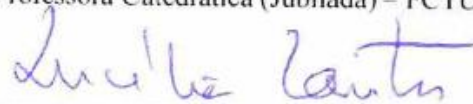
Prof. José André Peres Angotti
Professor Titular (Aposentado) – UFSC



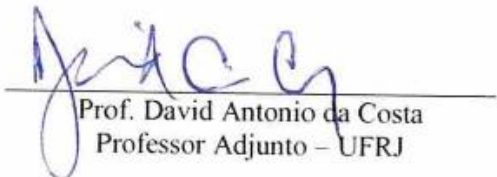
Prof.a Maria José Barata de Almeida
Professora Catedrática (Jubilada) – FCTUC



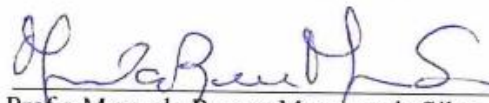
Prof.a Marta Feijó Barroso
Professora Associada – UFRJ



Prof.a Lucília Maria Pessoa Tavares dos Santos
Professora Associada – UA



Prof. David Antonio da Costa
Professor Adjunto – UFRJ



Prof.a Manuela Ramos Marques da Silva
Professora Auxiliar – FCTUC

Tese de Doutorado

Título | Escala de Proficiências em Concepções Térmicas: Diagnóstico Psicométrico de Estudantes em Portugal e Brasil

Ano | 2017

Autor | Marcel Bruno Pereira Braga

Orientação | Professor Doutor Décio Ruivo Martins (UC)/ Professor Doutor José André Peres Angotti (UFSC)

Co-orientação | Professora Doutora Maria Augusta Vilalobos Filipe Pereira do Nascimento (UC)

Domínio científico | Física

Especialidade | Ensino da Física

Instituição | Universidade de Coimbra - Faculdade de Ciências e Tecnologias da Educação/ Universidade Federal de Santa Catarina - Programa de Pós-Graduação em Educação Científica e Tecnológica (Regime de Cotutela)

Todas as atividades desenvolvidas e apresentadas na presente Tese de Doutorado foram realizadas no âmbito de financiamento por Bolsa de Doutorado concedida pela Fundação para a Ciência e Tecnologia do Ministério da Ciência e do Ensino Superior (SFRH/BD/91547/2012) em Portugal, e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (203571/2013-4) no Brasil.



DEDICATÓRIA

À minha amada mãe, Maria Lídia, que me alfabetizou
e ao meu pai, João Braga, que sempre me incentivou em toda a minha caminhada...

AGRADECIMENTOS

Agradeço à Deus, por ter me dado as oportunidades e a força necessária para chegar ao fim dessa jornada, assim como todas as pessoas que de forma direta ou indiretamente me ajudaram a seguir em frente, mesmo sem terem consciência disso. Em especial aos meus orientadores, Prof. Dr. Décio Martins, por sempre ter demonstrado acreditar e confiar na minha dedicação, sendo atento, leal e eficiente diante das necessidades e dificuldades. Ao Prof. Dr. José Angotti, por ter aceitado o desafio, pela amizade, respeito, correções, e incentivo, além das aulas de humanismo freireano permanentemente que correm em suas veias.

Ao Prof. Dr. Dalton Andrade pela gentileza de me receber no PPGE/UFSC e pelas aulas iniciais sobre os modelos da *Teoria da Resposta ao Item*, e também aos colegas envolvidos em seu grupo de pesquisa. Ao PPGE/UFSC, que me possibilitou novas vertentes de leitura e perspectivas teóricas para a *Educação Científica e Tecnológica*. Aos colegas e estudiosos da Psicometria, Danilo Pereira (UnB), Luis Anunciação (PUC-RJ), Euclides Mendonça (UFRGS) e Víthor Franco (UnB), pelos diálogos, contribuições e auxílios durante essa caminhada.

Ao Prof. Dr. Fernando Lang (UFRGS) pelas valorosas conversas virtuais e em sua casa, além de ter contribuído diretamente nesse trabalho. Aos professores da UA-Aveiro, Prof. Dr. Fransilê Neri, Prof^a. Dr^a. Dayse Neri e Prof. Dr. Pedro Costa, pelos cursos de pesquisa qualitativa, reconhecimento, criticidade e incentivo.

Pela fraterna recepção na UNED-Madrid com a Prof^a. Dr^a. María Cacheiro-González, e em especial, ao Prof. Dr. Antonio Rivilla, que me possibilitou um aprofundamento diante das perspectivas sobre o conceito de *competências* na Educação. A Prof^a. Dr^a. María Conde por viabilizar uma troca de ideias com os professores do *Programa de Doctorado Formación en la Sociedad del Conocimiento* (USAL-Salamanca). A amizade e troca de experiências com os colegas da UAB-Barcelona (Projeto Revir), Prof. Dr. Victor López, Prof^a. Dr^a. Digna Couso, entre outros, e pelas conversas e amizade com o Prof. Dr. Agustin Adúriz-Bravo (UBA-Buenos Aires), ampliando minha perspectiva para novos campos dentro da *Didática das Ciências*.

“Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid”.

“Todo mundo é um gênio. Mas se você julgar um peixe pela sua habilidade de subir em árvores, ele viverá o resto de sua vida acreditando que é um idiota”.

— Albert Einstein

LISTA DE FIGURAS

Figura 1: Perspectiva estrutural da investigação.	37
Figura 2: Comparação dos modelos de variáveis latentes da TCT e TRI.	49
Figura 3: <i>Design</i> da investigação – Abordagem mista sobre o panorama teórico-metodológico e as principais operacionalizações realizadas.	55
Figura 4: Quantitativo da produção no Brasil de 1995 a 2009 com a abordagem na TRI.	71
Figura 5: Ambiente virtual do repositório institucional da UFSC.	72
Figura 6: Ambiente virtual do repositório institucional da UC.	76
Figura 7: Ambiente virtual do repositório científico de acesso aberto de Portugal.	77
Figura 8: Relação entre saberes e conhecimentos.	90
Figura 9: Relação dos saberes entre os diferentes domínios do conhecimento.	91
Figura 10: Esquematização dos conteúdos, das capacidades e das competências em função do eixo do tempo e do eixo das situações.	93
Figura 11: Perspectiva de alinhamento curricular entre o teste TCE e o programa curricular.	121
Figura 12: Etapas de incorporação dos conteúdos seguindo uma abordagem ausubeliana .	132
Figura 13: Esquema original da representação proposta de traço latente.	138
Figura 14: Visão elementar e estrutural do traço latente.	141
Figura 15: Suposição de assimilação subsunsa durante o processo de resolução de problemas.	145
Figura 16: Exemplo de item teórico considerando apenas a alternativa correta.	178
Figura 17: Exemplos tipológicos na AGI de acordo com diferentes combinações de características psicométricas.	181
Figura 18: Gráfico g versus %<pré-teste>.	184
Figura 19: Tamanho do Efeito de Cohen.	187
Figura 20: Tamanho do Efeito de Cohen para uma cauda.	188
Figura 21: Comparativo das diferenças entre ganhos normalizados de um pré-teste e pós-teste segundo o "d de Cohen".	188
Figura 22: S-C <i>plot</i>	196
Figura 23: Comparação entre TCT e TRI.	204
Figura 24: Relações da Influência e Utilização da Matriz de Referência.	214
Figura 25: “Régua” ilustrativa do percentual de respondentes em uma escala de proficiência.	230
Figura 26: Ilustração do processo de construção e interpretação de uma escala de proficiência.	232
Figura 27: Relações das funções mais enfatizadas relativas às pessoas, a um sistema e ao conhecimento.	250
Figura 28: Perspectivas quanto a validade de construto.	256
Figura 29: Procedimentos para adaptação transcultural de instrumentos psicológicos.	266
Figura 30: Diferença entre fiabilidade e validade.	273
Figura 31: Proporção da amostra portuguesa com referência aos níveis de escolaridade. ...	279

Figura 32: Quantitativo e percentual de alunos por escola no contexto português, juntamente com os livros didáticos de Física adotados.....	280
Figura 33: Proporção de alunos nas escolas portuguesas de acordo com os livros didáticos de Física adotados.....	280
Figura 34: Proporção da amostra brasileira com referência aos níveis de escolaridade e turnos.....	285
Figura 35: Quantitativo de alunos por escola no contexto brasileiro, juntamente com os livros didáticos de Física adotados.....	286
Figura 36: Proporção de alunos nas escolas brasileiras de acordo com os livros didáticos de Física adotados.....	286
Figura 37: Gráfico quantitativo dos TCLE dos encarregados de educação em Portugal.....	289
Figura 38: Percentual na convicção da chave-de-correção entre 5 avaliadores.....	300
Figura 39: Coerência interna quanto ao equilíbrio de concepções dentro de cada grupo, ou seja, total de concepções distribuídas nos 26 itens e total de itens que possuem concepção dentro de cada grupo (Comparação entre os autores originais e as sugestões nesse estudo).....	322
Figura 40: Frequência de Concepções por Item, de acordo com os autores no trabalho original, comparados com a sugestão nesse estudo.....	323
Figura 41: Incidência das concepções no teste.....	325
Figura 42: Esquema da distribuição dos itens entre os grupos de concepções, de acordo com os autores do TCE.....	326
Figura 43: <i>Projeto de Pré-teste e Pós-teste de Grupo Único</i> (distribuída em detalhes para cada as quatro escolas portuguesas e suas turmas).....	338
Figura 44: Gráfico %<G> do TCE versus %<pré> no contexto do teste-piloto em relação aos itens, as médias de cada escola, e da média entre as quatro escolas secundárias portuguesas.....	346
Figura 45: Histograma do índice de dificuldade no pré-teste, ganhos percentuais brutos no pós-teste e ganhos normatizados dos itens.....	348
Figura 46: Especulação prévia dos distratores em destaque e das concepções correspondentes associadas a cada item, sendo, no caso, indícios preliminares de concepções térmicas resistentes após as instruções.....	349
Figura 47: Gráfico S-C <i>plot</i> para o pré-teste.....	351
Figura 48: Gráfico S-C <i>plot</i> para o pós-teste.....	352
Figura 49: Gráfico S-C <i>plot</i> para o pré-teste e pós-teste.....	352
Figura 50: Gráfico S-□ <i>plot</i> para os Grupos do Pré-teste e Pós-teste.....	355
Figura 51: Gráfico da distribuição da densidade das respostas em função das habilidades dos estudantes no pré-teste e pós-teste.....	358
Figura 52: Gráfico dos diagramas de caixa para o pré-teste e pós-teste, de acordo com os níveis de habilidade.....	359
Figura 53: <i>Design</i> do teste TCE - projeto de comparação de grupo estático não equivalentes em Portugal e no Brasil.....	366
Figura 54: Distribuição percentual de respondentes brasileiros de acordo com seu respectivo escore.....	369

Figura 55: Distribuição quantitativa de respondentes brasileiros por escores de acordo com os turnos.	369
Figura 56: Percentual acumulado de respondentes <i>versus</i> Escore bruto – Brasil.	371
Figura 57: Distribuição percentual de respondentes portugueses de acordo com seu respectivo escore.....	374
Figura 58: Percentual acumulado de respondentes <i>versus</i> Escore bruto – Portugal.....	375
Figura 59: Média dos percentuais dos distratores para os níveis de escore (Xi) dentro dos intervalos ou zona dos diferentes grupos considerados para o item y1.	376
Figura 60: Frequência de respondentes brasileiros e portugueses dentro de cada nível de escore.....	377
Figura 61: Comparação da dificuldades (percentual de acertos) dos itens entre Brasil e Portugal, tendo como referência as linhas de corte de admissão, segundo os critérios de "guessing", percentual de predominância absoluta (50%), e do percentual ideal mínimo a ser alcançado.	382
Figura 62: Representação gráfica de dispersão dos itens, segundo os índices de discriminação e de índice de facilidade (ou dificuldade) na amostra da Grande Florianópolis (Brasil).	384
Figura 63: Representação gráfica de dispersão dos itens, segundo os índices de discriminação e de índice de facilidade (ou dificuldade) na amostra da Região do Centro (Portugal).	385
Figura 64: Média percentual dos escores brutos dos principais grupos normativos do estudo no contexto Brasil.	389
Figura 65: Variação percentual dos acertos das escolas em relação à média da amostra no contexto Brasil.....	390
Figura 66: Média percentual dos escores brutos dos principais grupos normativos do estudo no contexto Portugal.	391
Figura 67: Variação percentual dos acertos das escolas em relação à média da amostra no contexto Portugal.....	392
Figura 68: Gráfico %<G> do TCE versus %<pré> no contexto Brasil, comparando o 1º ano com os ganhos do 2º e 3º ano.	393
Figura 69: Índices de Dificuldades do 2º e 3º ano - Brasil.....	395
Figura 70: Detalhamento dos ganhos percentuais do 2º e 3º ano (pós-teste) em relação ao percentual do 1º ano (pré-teste).....	396
Figura 71: Gráfico %<G> do TCE versus %<pré> no contexto Portugal, , comparando o 10º ano com os ganhos do 11º e 12º ano.....	397
Figura 72: Ganhos percentuais normatizados do 11º e 12º ano - Portugal.....	398
Figura 73: Gráfico S-C <i>plot</i> para o 1º, 2º e 3º ano (Brasil).....	400
Figura 74: Gráfico S-C <i>plot</i> para o 10º, 11º e 12º ano (Portugal).	401
Figura 75: Gráfico S-C <i>plot</i> comparativo para Portugal e Brasil, considerando todos os alunos.	402
Figura 76: Perfil de concepções térmicas de Brasil e Portugal de acordo com o «peso das incidências» das concepções para os grupos ACIM.....	409
Figura 77: Perfil de concepções térmicas com base na relação entre o total de concepções dentro de cada subconstruto e o peso das incidências das concepções evidenciadas nas respostas do grupo superior (ACIM) em cada contexto.....	412

Figura 78: Perfil de concepções térmica do contexto do Brasil e de Portugal conforme a sua representatividade nos subconstrutos.....	415
Figura 79: Gráfico para a análise do <i>Scree plot</i> e autovalores (“ <i>eigen values</i> ”) para retenção de fatores da matriz de correlação tetracórica.....	421
Figura 80: Curva do percentual acumulado <i>versus</i> escore bruto – Brasil e Portugal.....	438
Figura 81: Curva Característica do Teste na TCT.....	440
Figura 82: Curva de proporção de acertos ou de pontuação de domínio na TCT.....	440
Figura 83: Curva de Informação do Teste (Brasil) através do modelo de 3PL, estimado pelo BILOG-MG.....	441
Figura 84: Curva de Informação do Teste (Portugal) através do modelo de 3PL, estimado pelo BILOG-MG.....	441
Figura 85: Mapa item-pessoa. Distribuição das habilidades das pessoas e das dificuldades dos itens pelo modelo 4PL.....	443
Figura 86: Função de Informação do Teste e erro padrão de medida ao longo da escala de proficiência.....	444
Figura 87: Curva Característica do Item y1 para o modelo geral com 20 itens, ou seja, estando inclusos os dois contextos e as respectivas informações na escala de proficiências térmicas.....	458
Figura 88: Curva Característica do Item y1 para cada contexto, de acordo com o modelo não-restringido.....	458
Figura 89: Esquema ilustrativo das principais perspectivas comparativas na investigação:	462
Figura 90: Comparação dos resultados da pesquisa com o trabalho original do TCE em 2001.....	474
Figura 91: Exemplo comparativo de resultados clássicos do item y3 entre Brasil e Portugal.....	531
Figura 92: Análises de distorção conceptual diante aos critérios de plausibilidade, no que diz respeito a coerência e consistência dos indícios de dependência local entre os itens y1, y2 e y3.....	535

LISTA DE QUADROS

Quadro 1: Métodos e técnicas para explorar conteúdos e detectar concepções dos alunos.....	9
Quadro 2: Exposição de motivos para o condicionamento de aceitação do problema científico.....	38
Quadro 3: Comparação entre os aspectos de pesquisa que se configuram como avaliação e investigação, incluindo o indicativo nas ênfases das tipologias de pesquisa quanto ao produto, destacando (em negrito) a perspectiva do atual estudo.....	42
Quadro 4: Organização dos grupos de variáveis em caráter amplo e sistemático para o estudo de acordo com as categorias de concepções relacionado aos itens, juntamente com a declaração dos processos operacionais utilizados.	50
Quadro 5: Síntese de notações para os <i>designs</i> mistos.....	53
Quadro 6: Alguns estudos das últimas décadas relacionadas às dificuldades na compreensão dos conceitos térmicos.	62
Quadro 7: Tipos de conteúdo, metas e visão de ciência no currículo de ciências.....	84
Quadro 8: Reivindicações a serem consideradas diante das concepções alternativas.	85
Quadro 9: Processos e mecanismos explicativos sobre as origens das concepções alternativas.	88
Quadro 10: Diferenças entre um modelo de ensino genérico (baseado em capacidades) e um ensino operacional (baseado em competências).	94
Quadro 11: Concepções térmicas mais frequentes.	103
Quadro 12: O processo típico para desenvolver um inventário de conceitos.	106
Quadro 13: Práticas recomendadas gerais para administrar inventários de conceitos.....	107
Quadro 14: Critérios de alinhamento curricular para testes conceituais.	122
Quadro 15: Processo inicial de assimilação por subsunção, seguida pelo processo de assimilação obliteradora.....	129
Quadro 16: Comparação organizacional da abordagem da Termodinâmica pelo método convencional e ausubeliano.	131
Quadro 17: Recomendações para o ensino na concepção de Ausubel.	135
Quadro 18: Sugestões de etapas para um ensino ausubeliano favorável ao desenvolvimento de competências.	136
Quadro 19: Algumas ideias sobre o conceito de traço latente na Psicometria.	139
Quadro 20: Concepção ausubeliana nas interpretações pedagógicas dos itens.....	144
Quadro 21: Sentido atual e ideal no suposto processo interpretativo dos itens pelo respondente.....	147
Quadro 22: Ideias centrais da psicologia educacional de Ausubel e algumas possíveis contribuições para as análises psicométricas.	148
Quadro 23: Contexto histórico da Psicometria.....	154
Quadro 24: Classificação para o Índice de Dificuldade.	162
Quadro 25: Técnicas do Índice de Dificuldade.	165
Quadro 26: Classificação para o Índice de Discriminação.....	167
Quadro 27: Interpretação a ser feita na discriminação dos sujeitos.....	169
Quadro 28: Classificação para os coeficientes de correlação.....	171

Quadro 29: Níveis de confiança para o escore verdadeiro.....	172
Quadro 30: Determinando o valor do <i>guessing</i> do teste.....	173
Quadro 31: Comparação entre o <i>Fator de Hake</i> e a Média dos Ganhos.	185
Quadro 32: Classificação para o tamanho do efeito do "d de Cohen".	188
Quadro 33: Diretrizes para apresentar e interpretar os <i>Tamanhos Do Efeito</i> (TDE).....	191
Quadro 34: Demonstração do Fator de Concentração (C).....	193
Quadro 35: Funções de 1, 2, 3, 4 e 5PL da TRI.....	202
Quadro 36: Vantagens e desvantagens entre a TCT e a TRI.	205
Quadro 37: Potencialidades no uso de Escalas de Proficiência.	208
Quadro 38: Modelo Dreyfus de competências.	217
Quadro 39: Descrição de alguns estudos taxonômicos envolvendo objetivos educacionais dentro do domínio cognitivo.	221
Quadro 40: Descrição detalhada das dimensões da Taxonomia de Bloom (Revisada).....	227
Quadro 41: O uso de testes e o efeito da concorrência em resultados educacionais.	239
Quadro 42: Contexto históricos dos exames nacionais em Portugal.	242
Quadro 43: Critérios de credibilidade da investigação diante das principais análises e processos realizados de acordo com os enfoques-metodológicos.....	245
Quadro 44: Tipo de instrumento em função do tipo de variável.	246
Quadro 45: Tipos de avaliação e os respectivos critérios avaliativos relacionados.....	248
Quadro 46: Critérios em nível da ação dentro da pesquisa.	251
Quadro 47: Critérios e normas referenciadas em nível da pessoa (grupos normativos) e as implicações operacionais fundamentais das análises realizadas.	253
Quadro 48: Fontes de evidências de validade recomendadas no uso de testes psicológicos.....	257
Quadro 49: Diretrizes de Guillemín, Bombardier, & Beaton (1993) para a adaptação transcultural de medidas.....	262
Quadro 50: Diretrizes de Borsa, Damásio, & Bandeira (2012) para adaptação transcultural de medidas.....	264
Quadro 51: Validade do TCE Creditada pela AAPT.....	269
Quadro 52: Tipos de fidelidade.....	272
Quadro 52: Universo e população do estudo.....	277
Quadro 53: Codificação e descrição das escolas públicas na amostra na Região do Centro (Portugal Continental) que foram aplicados o TCE, assim como os respectivos LDFs (código) adotados.	278
Quadro 54: Resumo descritivo sobre principais atividades no processo de recolha das informações.	290
Quadro 55: Descrição das etapas iniciais típicas para os procedimentos de análise.....	292
Quadro 56: Itens correspondentes a níveis de consistência no julgamento da chave-de-correção.....	300
Quadro 57: Aproximações entre algumas concepções alternativas.....	312
Quadro 58: Análise do construto - Reelaboração na tradução, comparação e novas sugestões de correspondências entre as concepções e os itens/alternativas.....	316
Quadro 59: Discordâncias e sugestões para o quadro de tipicidade de concepções térmicas do TCE.	320

Quadro 60: Aspectos delimitadores na elaboração das sentenças descritoras da matriz de competências científicas.	327
Quadro 61: Matriz das habilidades cognitivas no Teste TCE de acordo com as descrições nas análises taxonômicas.....	333
Quadro 62: Análise taxonômica cognitiva dos objetivos específicos dos itens no teste TCE, baseado em Anderson et al. (2001, p. 28).	336
Quadro 63: Identificação dos modelos mentais (corretos e/ou incorretos) em evidência no percentual das respostas, de acordo com a análise do <i>Fator de Concentração</i> (C).	353
Quadro 64: Informações descritivas das convicções científicas dos alunos no TCE em destaque, de acordo os resultados da análise quantitativa do pós-teste (em ordem decrescente, segundo o fator de concentração).....	354
Quadro 65: Identificação de grupos de itens dentro de cada contexto de acordo com os critérios de dificuldade (índice de dificuldade – ID), assim como itens coincidentes....	382
Quadro 66: Identificação de grupos de itens dentro de cada contexto de acordo com os critérios de discriminação (índice de discriminação – DISCR), assim como itens coincidentes.....	383
Quadro 67: Identificação comparativa de grupos de itens dentro de cada contexto de acordo com os critérios de dificuldade (ID) e de discriminação (DISCR).....	383
Quadro 68: Critérios relativos de maior consistência quanto ao número de incidências das concepções térmicas identificadas no grupo superior (ACIM).	409
Quadro 69: Identificação comparativa de grupos de itens dentro de cada contexto de acordo com os critérios de dificuldade (b) e de discriminação (a).	425
Quadro 70: Descrição dos níveis de distorção conceptual entre os itens.	508
Quadro 71: Identificação complementar para o inventário de concepções do TCE.	518
Quadro 72: Análise parcial de interpretação implícita dos níveis de distorção conceptual e na relação de condicionamento na dependência local dos itens y2 e y3.	519
Quadro 73: Resultados parciais da rede de concepções e análise de dependência local entre os itens y1, y2 e y3, segundo os critérios de coerência e consistência para os níveis de distorção conceptual.	529
Quadro 74: Níveis de Distorção Conceptual do Item y1.	533
Quadro 75: Níveis de Distorção Conceptual do Item y2.	533
Quadro 76: Níveis de Distorção Conceptual do Item y3.	534

LISTA DE TABELAS

Tabela 1: Percentuais ideais para o índice de dificuldade, segundo Frederic Lord, 1952.....	164
Tabela 2: Tabela-exemplo explicativa do <i>guessing</i> do item e do teste.....	174
Tabela 3: Exemplos de concentração de respostas sem pico (tipo I), com dois picos (tipo II) e com 1 pico (tipo III).	193
Tabela 4: Esquema de codificação de três níveis para pontuação e fator de concentração. ..	195
Tabela 5: Padrões típicos de resposta ao usar o sistema de codificação de três níveis.....	195
Tabela 6: Três exemplos de valores de acertos (A), erros (E) e omissões (O).	200
Tabela 7: Meta 7 – Fomentar a qualidade da educação básica em todas as etapas e modalidades, com melhoria do fluxo escolar e da aprendizagem de modo a atingir as seguintes médias nacionais para o IDEB.	240
Tabela 8: Estratégia 7.11 – Melhorar o desempenho dos alunos da educação básica nas avaliações da aprendizagem no <i>Programa Internacional de Avaliação de Estudantes (PISA)</i> , tomado como instrumento externo de referência, internacionalmente reconhecido, de acordo com as seguintes projeções.	241
Tabela 9: Codificação e descrição dos LDFs recomendados pelo MEC/Portugal e quantitativo de escolas da Região Centro que adotaram na vigência 2008-2014.	278
Tabela 10: Quantidade de escolas estaduais em Santa Catarina e na Grande Florianópolis, em 2015.....	282
Tabela 11: Quantidade de escolas estaduais nos municípios da Grande Florianópolis.....	282
Tabela 12: Codificação e descrição dos LDFs recomendados pelo PNLD-2014 e quantitativos distribuídos no Estado de Santa Catarina (SC/Brasil) na primeira remessa da vigência 2014-2016.	283
Tabela 13: Codificação e descrição das escolas públicas (de acordo com as modalidades de ensino) na Grande Florianópolis (Brasil) que foram aplicados o TCE, assim como os respectivos LDFs (códigos) adotados – refere ao ano de 2015.....	284
Tabela 14: Quadro descritivo da amostra quanto ao distrito, concelho, ordenação, alcance de professores, confirmações de apoio, quantidade de <i>kits</i> entregues, quantidade de turmas e Livro Didático de Física (LDF) adotado nas escolas.....	287
Tabela 15: Correções de equívocos na base de dados.	296
Tabela 16: Elevadas incidências consecutivas de mesma resposta para os itens.	296
Tabela 17: <i>Missing</i> por item.....	297
Tabela 18: <i>Missing</i> por teste.	297
Tabela 19: Comparação da distribuição de concepções nos itens do teste entre os autores originais e o atual estudo, no caso se tem a <i>Frecuencia de Concepciones por Grupo (FCG)</i> e <i>Total de Itens (TI)</i> que incidem nessas concepções dentro de cada Grupo (G).....	324
Tabela 20: Índícios de dependência local no TCE.....	331
Tabela 21: Média e desvio-padrão amostral do pré-teste e pós-teste com base nos valores dos escores brutos.....	343
Tabela 22: Percentuais no pré-teste, ganhos percentuais dos escores brutos, tamanho do efeito, ganhos normatizados e od livros didáticos de Física correspondentes que foram adotados nas escolas.	343

Tabela 23: Informações descritivas aos distratores e as concepções térmicas que encontram-se mais evidentes entre os alunos no TCE, de acordo os resultados da análise quantitativa entre o pré-teste e o pós-teste (em ordem crescente dos itens).....	356
Tabela 24: Valores dos parâmetros estatísticos descritivos para o pré-teste e pós-teste com base nos valores da TRI.	357
Tabela 25: Matriz-Resposta Dicotômica (ilustração parcial das respostas) dos Alunos Brasileiros do 1º, 2º e 3º ano do Ensino Médio na Grande Florianópolis.	362
Tabela 26: Ganhos normalizados com relação as médias gerais do 2º e 3º ano, tendo como referência a média geral do 1º ano.	393
Tabela 27: Ganhos normalizados com relação as médias gerais do 2º e 3º ano, tendo como referência a média geral do 1º ano.	396
Tabela 28: Item y9 como exemplo na análise dos distratores mais atrativos através dos picos mais significativos dentro dos escores, segundo a AGI.....	408
Tabela 29: Contraste entre os Perfis de Concepções Térmicas para Brasil e Portugal, conforme o percentual de cobertura da concepção no seu subconstruto, das intensidades nos pesos de incidências e aspectos comuns.	416
Tabela 30: Índices de ajuste, parâmetros e variância explicada dos modelos Rasch, 1PL, 2PL, 3PL, 4PL.	422
Tabela 31: Os 4 parâmetros dos 26 itens associados a proficiência em conhecimentos introdutórios da Termodinâmica, segundo o TCE.	423
Tabela 32: Os 4 parâmetros dos 20 itens associados a proficiência em conhecimentos introdutórios da Termodinâmica, segundo o TCE (em ordem crescente de dificuldade).	424
Tabela 33: Resultados da análise de funcionamento diferencial dos itens utilizando o método Mantel-Haenszel.....	427
Tabela 34: Classificação do tamanho do efeito de DIF, segundo a ETC.....	430
Tabela 35: Análise de DIF dos 20 itens por País pelo Modelo Rasch.....	432
Tabela 36: Cruzamento no tamanho do efeito de itens com DIF nas duas análises realizadas para o grupo focal (Brasil) e grupo de referência (Portugal).	433
Tabela 37: Comparação no processo de ajuste entre o modelo multigrupo restringido e o não restringido.	434
Tabela 38: Estimativa dos parâmetros logísticos para o modelo não-restringido diante da análise multigrupo entre a amostra Brasil e Portugal.....	434
Tabela 39: Identificação de grupos de itens dentro de cada contexto de acordo com os critérios de discriminação da TRI, assim como itens coincidentes dentro de cada grupo e para os dois em geral.	435
Tabela 40: Identificação de grupos de itens dentro de cada contexto de acordo com os critérios de dificuldade da TRI, assim como itens coincidentes dentro de cada grupo e para os dois em geral.	436
Tabela 41: Identificação comparativa de grupos de itens dentro de cada contexto e para os dois juntos, de acordo com os critérios e valores estimados para os parâmetros de dificuldade (a) e de discriminação (b) para TRI.....	436
Tabela 42: Relações entre θ , V e π	439
Tabela 43: Valores médios para os parâmetros a, b, g e u.	439

Tabela 44: Escala de Proficiências Térmicas.	448
Tabela 45: Exemplo de subamostra de alunos da escola 10BR no contexto brasileiro que erraram o item y1.	457
Tabela 46: Comparação de resultados entre México, Líbia, Brasil e Portugal.	480
Tabela 47: Comparação dos resultados da pesquisa com o estudo no Brasil em 2009.....	484
Tabela 48: Tabela comparativa dos Resultados do TCE entre a Coréia do Sul, Portugal e Brasil.....	491

LISTA DE EQUAÇÕES

Equação 1: Fator de Hake (ganho médio normalizado).	183
Equação 2: Média dos ganhos.	185
Equação 3: Mudança normalizada.....	186
Equação 4: Relação geral da mudança normalizada.	187
Equação 5: Parâmetro responsável pela mudança [normalizada].	187
Equação 6: "d de Cohen" para pré-teste e pós-teste com grupos correspondentes.....	189
Equação 7: Desvio-padrão relacionado a dois grupos não correspondentes.	191
Equação 8: Método EAP.	233
Equação 9: Método da Quadratura Gaussiana.	233

SUMÁRIO

VOLUME I: Fundamentos e Análises

LISTA DE FIGURAS	XI
LISTA DE QUADROS	XV
LISTA DE TABELAS.....	XIX
LISTA DE EQUAÇÕES.....	XXIII
RESUMO.....	XXXI
ABSTRACT.....	XXXV
LISTA DE ABREVIATURAS	XXXVII
INTRODUZINDO E SITUANDO AS PERSPECTIVAS DA INVESTIGAÇÃO.....	1
CONTRIBUIÇÕES E CONTRAPONTO NO USO DE TESTES DIAGNÓSTICOS CONCEITUAIS.....	1
Conhecimento Prévio e o Movimento das Concepções Alternativas	6
Delimitações em Testes Conceituais.....	11
MOTIVAÇÕES - REFERENCIAL DE VALORES DO INVESTIGADOR	19
CAPÍTULO 1 - PARÂMETROS ESTRUTURADORES	27
1.1 PROBLEMA	27
1.1.2 Especificando o Problema	35
1.2 TIPO DE PESQUISA	40
1.2.1 Delimitações e Concepções Filosóficas Subjacentes	44
1.2.2 Definição das Variáveis e dos Critérios Operacionais	46
1.2.3 <i>Design</i> Geral da Investigação	52
1.3 HIPÓTESES.....	55
1.4 OBJETIVO MISTO.....	57
1.5 FASES DA PESQUISA.....	59
1.6 REVISÃO DE LITERATURA.....	62
1.6.1 Produções Acadêmicas Baseadas no Uso de Testes de Aferição do Conhecimento Físico na Educação	62
1.6.2 Centros de Pesquisa que utilizam a Teoria da Resposta ao Item no Brasil.....	70
1.6.3 Universidade Federal de Santa Catarina	72
1.6.4 Universidade de Coimbra	76
CAPÍTULO 2 - REFERENCIAL TEÓRICO-METODOLÓGICO.....	81
2.1 CONTRIBUIÇÕES DA DIDÁTICA DAS CIÊNCIAS E IMPLICAÇÕES CURRICULARES	82

2.1.1 Fundamentos da Pedagogia da Integração.....	88
2.1.2 Fundamentos dos Testes Conceituais.....	96
2.2 CONTRIBUIÇÕES DA PSICOLOGIA EDUCACIONAL.....	126
2.2.1 Fundamentos da Teoria da Aprendizagem Significativa de David Ausubel.....	126
2.2.2 Implicações Ausubelianas para os Testes.....	137
2.3 CONTRIBUIÇÕES DA PSICOMETRIA.....	152
2.3.1 Contexto Histórico.....	152
2.3.2 Teoria Clássica dos Testes.....	159
2.3.3 Teoria da Resposta ao Item.....	201
2.3.4 Escalas de Proficiência: Comentários e Fundamentos.....	206
CAPÍTULO 3 - METODOLOGIA.....	243
3.1 FASE PRÉ-DESCRITIVA OU EXPLORATÓRIA - Índícios de Credibilidade Investigativa, Avaliação do Instrumento e o Teste-Piloto.....	243
3.1.1 Delimitações Sobre os Aspectos de Credibilidade.....	244
3.1.2 Procedimentos no Tratamento dos Dados.....	291
3.1.3 Introdução à Avaliação de Conceitos Técnicos: Resultados Preliminares.....	297
3.1.4 <i>Teste-piloto</i> : Pré-teste e Pós-teste.....	337
CAPÍTULO 4 - RESULTADOS.....	360
4.1 FASE DESCRITIVA - Análises da Teoria Clássica dos Testes.....	360
4.2 FASE ANALÍTICA E INFERENCIAL - Análises da Teoria da Resposta ao Item.....	417
4.3 UMA SÍNTESE DOS RESULTADOS.....	462
CAPÍTULO 5 - DISCUSSÃO.....	472
5.1 Dialogando com Outras Pesquisas: Comparando Resultados e Apontando Limitações.....	472
5.1.1 Austrália, 2001.....	473
5.1.2 Estados Unidos, 2006.....	475
5.1.3 Tailândia, 2004.....	477
5.1.4 México, 2009 e Líbia, 2009/2010.....	478
5.1.5 Brasil, 2009 e 2010.....	482
5.1.6 Coréia do Sul, 2012.....	489
5.1.7 Turquia, 2006.....	493
5.1.8 Nigéria, 2015.....	493
5.1.9 Singapura, 2017.....	494
5.2 Limitações na Investigação.....	494
5.3 Perspectivas Futuras e Complementares de Análise.....	499

5.3.1 Contribuições na Função de Desenvolvimento de Competências em Livros Didáticos	499
5.3.2 Análise Complementar de Aprofundamento: Níveis de Distorção Conceptual do Item e Rede de Dependência Local de Concepções.....	502
CONSIDERAÇÕES FINAIS.....	536
REFERÊNCIAS.....	542

VOLUME II: Apêndices e Anexo – Informações Detalhadas dos Itens

LISTA DE FIGURAS	V
LISTA DE TABELAS.....	VII
LISTA DE QUADROS	IX
APÊNDICES.....	1
A – Itens Comentados.....	1
Considerações sobre os Itens Comentados.....	1
A1: ITEM y1	7
A2: ITEM y2	17
A3: ITEM y3	27
A4: ITEM y4	35
A5: ITEM y5	43
A6: ITEM y6	51
A7: ITEM y7	57
A8: ITEM y8	65
A9: ITEM y9	71
A10: ITEM y10	81
A11: ITEM y11	91
A12: ITEM y12	99
A13: ITEM y13	105
A14: ITEM y14	113
A15: ITEM y15	121
A16: ITEM y16	129
A17: ITEM y17	137
A18: ITEM y18	145
A19: ITEM y19	153
A20: ITEM y20	161
A21: ITEM y21	169
A22: ITEM y22	177

A23: ITEM y23	187
A24: ITEM y24	195
A25: ITEM y25	203
A26: ITEM y26	209
B – População e Amostra.....	217
B1: População – Região Centro, Portugal.	217
B2: População – Região da Grande Florianópolis, Brasil.....	221
B3: Quantitativo de alunos da amostra por escola e turno – Brasil.....	225
B4: Controle no Processo de Recolha de Dados – Portugal.....	226
C – Validade no <i>Design</i> da Pesquisa	227
C1: Análise de Riscos sobre a Validade Interna: Estudo Piloto	227
C2: Análise de Riscos sobre a Validade Externa: Estudo Piloto	230
C3: Análise de Riscos sobre a Validade Interna – Portugal e Brasil.....	233
C4: Análise de Riscos sobre a Validade Externa – Portugal e Brasil.....	236
D – A Física Térmica nos Programas Curriculares.....	241
D1: Portugal	241
D2: Brasil.....	252
E – Detalhes nos Resultados Obtidos	259
E1: Teste-Piloto (quatro escolas secundárias em Portugal)	259
E2: Índice de Dificuldade dos Itens nas amostras do Brasil e de Portugal.	265
E3: Parâmetros Clássicos da amostra no Brasil e em Portugal	266
E4: Escala de desempenho (escore bruto) dos alunos brasileiros e portugueses	268
E5: Teste-t <i>student</i> comparando diferenças significativas entre as médias do 1º, 2º e 3º ano com relação ao turno matutino.	270
E6: Teste-t <i>student</i> comparando diferenças significativas entre as médias dos grupos de escolas conforme os Livros Didáticos de Física (ou Manuais Escolares) em comum, tendo como referência o grupo com maior média.	270
E7: Teste-t <i>student</i> comparando diferenças significativas entre as médias dos diferentes níveis escolares nos dois contextos, assim como entre as médias gerais entre Brasil e Portugal.....	271
E8: Fator de Concentração nas Análises dos Itens – Brasil.....	271
E9: Fator de Concentração nas Análises dos Itens – Portugal.....	272
E10: Fator de Concentração nas Análises dos Itens (Visualização Geral) – Teste-Piloto, Portugal e Brasil.....	273
E11: Análise dos Distratores e as Respectivas Concepções Térmicas em Destaque dentro dos Grupos Superiores (ACIM).....	274
E12: Matriz de Correlações Tetracóricas dos Itens – Amostra Geral (Brasil e Portugal).275	

E13: Identificação dos Itens-âncoras na Escala de Proficiências Térmicas para o Modelo 4PL (20 itens) – Brasil e Portugal.....	276
F – Sintaxes.....	278
F1: Alguns Resultados da Análise Clássica no BILOG-MG – Brasil/2015.....	278
F2: Alguns Resultados da Análise Clássica no BILOG-MG – Portugal/2014.....	279
G – Documentos Formais.....	280
G1: Consentimento da Autora do Teste TCE via <i>e-mail</i>	280
G2: Carta aos Diretores (Modelo Brasil).....	281
G3: Carta de Apresentação aos Professores (Modelo Portugal)	282
G4: Carta de Apresentação aos Professores (Modelo Brasil)	283
G5: TCLE – Encarregado de Educação.....	284
G6: TCLE – Professores (Modelo Portugal).....	285
G7: TCLE – Professores (Modelo Brasil)	286
G8: TCE – Orientação aos Professores.....	287
H – Instrumentos.....	288
H1: Tradução do TCE – Versão Portugal.....	288
H2: Tradução do TCE – Versão Brasil.....	292
ANEXOS.....	298
1: CNPD – Autorização Formal para a Realização da Pesquisa	298

RESUMO

O estudo explora procedimentos de análise que fornecem diversas informações dentro de um *continuum* de desempenhos, sendo múltiplas representações sobre as evidências que constituem uma escala de proficiências. Essas análises possibilitaram realizar um diagnóstico de concepções típicas de 5542 alunos do último ciclo da educação básica em 20 escolas de Portugal (em 2014), e 18 no Brasil (em 2015), diante aos conceitos básicos e introdutórios da Física Térmica. Para isso, foi utilizado um teste de sondagem denominado *Thermal Concept Evaluation* (TCE), elaborado pelos pesquisadores australianos Yeo Shelley e Marjan Zadnik em 2001, sendo utilizado amplamente na literatura. Adotam-se procedimentos de análise específicos para avaliar *Inventários de Conceitos* (*Fator de Hake; Fator de Concentração de Análise de Lei Bao e Redish*), juntamente com modelos de análise psicométrica conhecidos (*Teoria Clássica dos Testes; Teoria da Resposta ao Item; e Funcionamento Diferencial do Item*) que contemplam uma perspectiva investigativa de educação comparada, seja identificando necessidades pedagógicas comuns e diferenciadas entre os contextos, estimando em parte a eficiência do ensino em Termodinâmica nos sistemas educacionais da rede pública em cada país dentro do recorte preestabelecido, e destacando as condições no uso de testes que favorecem um alinhamento curricular de forma representativa para as populações envolvidas. Consideram-se as contribuições da *Didática das Ciências* na análise de coerência interna do instrumento, e da *Psicologia Cognitivista* de David Ausubel como «condições de contorno» para as implicações diante as evidências que possam favorecer uma «aprendizagem significativa». Os perfis de proficiência térmica se configuram em diversas composições de grupos de alunos com base em seus desempenhos, possibilitando um direcionamento instrucional diferenciado para aqueles com maiores dificuldades e para os com alta performance. O teste-reteste (estudo piloto) em 4 escolas portuguesas ($n \approx 200$, grupos não equivalentes) contribuiu para fornecer maior precisão ao instrumento, apontar fraquezas, demarcar um regime de validade interna e externa, e com isso, apresentar subsídios preliminares para uma investigação mais ampla, detalhada, profunda e consistente. Na investigação ampla, o modelo da TRI de 4 parâmetros logísticos (discriminação, dificuldade, acerto ao acaso,

e erro ao acaso) foi o mais ajustado para explicar as análises dos dados (78,2%). Os resultados gerais da investigação apresentam uma boa consistência ($\alpha = 0,75$) e revelam um teste difícil para ambos os contextos, isso significa que as evidências tendem a discriminar com mais eficiência em um dos extremos da distribuição, no caso, entre os respondentes mais capacitados. O *fator de Hake* sugere ganhos de baixo impacto em ambos os contextos, estando Portugal sutilmente em condições mais promissoras. O *fator de concentração de análise* mostra que os respondentes brasileiros se concentraram na região de acerto aleatório, mantendo os mesmos equívocos ao longo das séries de ensino para quase todos os itens, com exceção de 3 itens que sugerem a existência de um *modelo mental correto* e outro *incorreto*. Em Portugal, essa análise revela ganhos gradativos nas capacidades cognitivas dos sujeitos de acordo com o avanço nos níveis de escolaridade. O rastreamento na *Análise Gráfica dos Itens* (AGI) permitiu identificar uma forte atração de distratores entre os alunos mais proficientes, sugerindo dificuldades na compreensão do *conceito de equilíbrio térmico*, em *diferenciar quente e frio* como extremidades opostas de um *continuum*, e de admitir *diferentes temperaturas para o ponto de ebulição da água*. A construção da *escala de proficiências térmicas* advinda da TRI estimou probabilidades de acertos empíricos consistentes para os grupos normativos brasileiros em um nível abaixo da média geral (níveis de -1 a 0 da escala), indicando uma compreensão relacionada a *“prever a temperatura do gelo dentro do congelador”*, revelando uma ausência de itens-âncoras mais fáceis, e assim, impondo limitações para a escala avaliar sujeitos menos proficientes, pois há uma falta de informações na extremidade inferior da escala. Os alunos portugueses se concentraram nos níveis de 0 a 1 da escala, com um domínio acumulativo que inclui habilidades cognitivas relacionadas a mais 5 itens. Como parte inovadora, tem-se uma perspectiva de análise original em desenvolvimento que identifica preliminarmente redes de concepções entre os itens com indícios de dependência local, fundamentada na coerência das respostas e consistência interna dos itens, determinando níveis de distorção de concepções entre as alternativas apresentadas, considerando uma sub-escala de dimensionalidade nas unidades de medida, com base em *níveis de plausibilidade* das respostas/alternativas que foram concebidas durante o processo de construção dos itens.

Palavras-chave: Teste Diagnóstico; Proficiências; Concepções Térmicas; Teoria Clássica dos Testes; Teoria da Resposta ao Item; Portugal; Brasil.

ABSTRACT

The study explores analytical procedures that provide a variety of information within a performance continuum, being multiple representations about evidences that constitute a scale of proficiency. These analyzes made possible to carry out a diagnosis of typical conceptions of 5542 students of the last cycle of basic education in 20 Portuguese schools (in 2014) and 18 in Brazil (in 2015), in face of the basic and introductory concepts of Thermal Physics. For that matter, a test called the *Thermal Concept Evaluation* (TCE) was used, developed by the Australian researchers Yeo Shelley and Marjan Zadnik in 2001, being used widely in the literature. Specific analysis procedures are used to evaluate *Conceptual Inventories* (*Hake Factor*; *Lei Bao and Redish Analysis Concentration Factor*) along with known psychometric analysis models (*Classical Theory of Tests*, *Item Response Theory*, and *Differential Item Functioning*) that contemplate a comparative education research perspective, either by identifying common and differentiated pedagogical needs between contexts, estimating in part the efficiency of teaching in thermodynamics in the public network educational systems in each country within the pre-established cut, and highlighting the conditions in the use of tests that favor a curricular alignment in a representative way for the populations involved. The contributions of the *Didactics of Sciences* in the analysis of internal coherence of the study and of the *Cognitivist Psychology* of David Ausubel are considered as "boundary conditions" for the implications in the face of the evidences that can favor a "meaningful learning". The profiles of thermal proficiency are configured in several compositions of groups of students based on their performances, allowing a different instructional orientation for those with greater difficulties and for those with high performance. The test-retest (pilot study) in 4 Portuguese schools ($n \approx 200$, non-equivalent groups) contributed to provide greater precision to the study, to point out weaknesses, to delimited a regime of internal and external validity, and with this, to present preliminary subsidies, to a broader, deeper, more thorough consistent investigation. In the broad investigation, the TRI model of 4 logistic parameters (discrimination, difficulty, *guessing*, and random error) was the best fit to explain the data analysis (78,2%). The overall results of the research show a

good consistency ($\alpha = 0,75$) and show a difficult test for both contexts, which means that the evidence tends to explain more efficiently at one end of the distribution, in the case, between the respondents more skilled. The *Hake factor* suggests low impact gains in both contexts, with Portugal subtly under more promising conditions. The concentration factor of analysis shows that Brazilian respondents focused on the random hit region, maintaining the same misunderstandings throughout the teaching series for almost all items, except for 3 items that suggest a correct mental model and other incorrect. In Portugal, this analysis reveals gradual gains in the cognitive abilities of the subjects according to the progress in the levels of schooling. Tracking in the *Graphical Analysis of Items* (AGI) allowed us to identify a strong attraction of distractors among the most proficient students, suggesting difficulties in understanding the concept of thermal equilibrium, in differentiating hot and cold as opposite ends of a continuum, and in admitting different temperatures to the boiling point of the water. The construction of the TRI thermal proficiency scale estimated the probabilities of consistent empirical correctness for the Brazilian normative groups at a level below the general average (levels of -1 to 0 of the scale), indicating an understanding related to "predicting the ice temperature within the freezer", revealing an absence of easier anchor items, and thus imposing limitations for the scale to evaluate less proficient subjects, as there is a lack of information at the lower end of the scale. Portuguese students focused on levels 0 to 1 of the scale, with a cumulative domain that includes cognitive abilities related to 5 more items. As an innovative part, there is a perspective of original analysis in development that preliminarily identifies networks of conceptions among the items with evidence of local dependence, based on the coherence of the answers and internal consistency of the items, determining levels of conception distortion among the presented alternatives, considering a sub-scale of dimensionality in the units of measure, based on plausibility levels of the answers/alternatives that were conceived during the process of construction of the items.

Keywords: Diagnostic Test; Proficiencies; Thermal Concepts; Classical Theory of Tests; Item Response Theory; Portugal; Brazil.

LISTA DE ABREVIATURAS

AA	REGIÃO DE ALTA DIFICULDADE E ALTO VALOR PARA O FATOR DE CONCENTRAÇÃO
AAPT	AMERICAN ASSOCIATION OF PHYSICS TEACHERS
ABAIX	GRUPO DE DESEMPENHO INFERIOR OU CORRESPONDENTE A 27 % DOS QUE ESTÃO ABAIXO
ACIM	GRUPO DE DESEMPENHO SUPERIOR OU CORRESPONDENTE A 27 % DOS QUE ESTÃO ACIMA
AGI	ANÁLISE GRÁFICA DO ITEM
AIC	AKAIKE INFORMATION CRITERION
AICc	CORRECTED AKAIKE'S INFORMATION CRITERION
ANA	AVALIAÇÃO NACIONAL DE ALFABETIZAÇÃO
ANEI	AVALIAÇÃO NACIONAL DA EDUCAÇÃO INFANTIL
APA	AMERICAN PSYCHOLOGICAL ASSOCIATION
APC	ABORDAGEM POR COMPETÊNCIA
BA	REGIÃO DE BAIXA DIFICULDADE E ALTO VALOR PARA O FATOR DE CONCENTRAÇÃO
BB	REGIÃO DE BAIXA DIFICULDADE E BAIXO VALOR PARA O FATOR DE CONCENTRAÇÃO
BIC	BAYESIAN INFORMATION CRITERION
BM	REGIÃO DE BAIXA DIFICULDADE E VALOR MÉDIO PARA O FATOR DE CONCENTRAÇÃO
BNCC	BASE NACIONAL COMUM CURRICULAR
BNI	BANCO NACIONAL DE ITENS
BR	BRASIL
CAED	CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO
CAFES	COMUNIDADE ACADÊMICA FEDERADA
CAPES	COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR
CAs	CONCEPÇÕES ALTERNATIVAS

CAT	COMPUTER ADAPTIVE TEST
CCI	CURVA CARACTERÍSTICA DO ITEM
CCT	CURVA CARACTERÍSTICA DO TESTE
CEP-UEA	COMISSÃO DE ÉTICA EM PESQUISA DA UNIVERSIDADE DO ESTADO DO AMAZONAS
CFI	COMPARATIVE FIT INDEX
CII	CURVA DE INFORMAÇÃO DO ITEM
CIT	CURVA DE INFORMAÇÃO DO TESTE
CITO	NATIONAL INSTITUTE FOR EDUCATIONAL MEASUREMENT
CNPD	COMISSÃO NACIONAL DE PROTECÇÃO DE DADOS
CTS	CIÊNCIA, TECNOLOGIA E SOCIEDADE
CTSA	CIÊNCIA, TECNOLOGIA, SOCIEDADE E AMBIENTE
DC	DIDÁTICA DAS CIÊNCIAS
DCC	DESENVOLVIMENTO DE CAPACIDADES E COMPETÊNCIAS
DeSeCo	PROJETO DEFINIÇÃO E SELEÇÃO DE COMPETÊNCIAS-CHAVE
DGE	DIRECÇÃO GERAL DE EDUCAÇÃO
DIF	FUNÇÃO DIFERENCIAL DO ITEM
DISCR	ÍNDICE DE DISCRIMINAÇÃO
DNCEB	DIRETRIZES NACIONAIS CURRICULARES PARA A EDUCAÇÃO BÁSICA
EAP	MÉTODO ESTATÍSTICO EXPECTED A POSTERIORI
EBE	EDUCAÇÃO BASEADA EM EVIDÊNCIAS
EC	ENSINO DE CIÊNCIAS
ECI	ENERGY CONCEPT INVENTORY
EF	ENSINO DE FÍSICA
ENADE	EXAME NACIONAL DE DESEMPENHO DE ESTUDANTES.
ENAMEB	EXAME NACIONAL DE AVALIAÇÃO DO MAGISTÉRIO DA EDUCAÇÃO BÁSICA
ENCCEJA	EXAME NACIONAL PARA CERTIFICAÇÃO DE COMPETÊNCIAS DE JOVENS E ADULTOS

ENEM	EXAME NACIONAL DO ENSINO MÉDIO NO BRASIL
ETO	ESTUDOS TAXONÔMICOS DE OBJETIVOS
ETS	EDUCATIONAL TESTING SERVICE
EU	UNIÃO EUROPEIA
FA	LIVRO DIDÁTICO OU MANUAL ESCOLAR DE FÍSICA A
FCI	INVENTÁRIO DE CONCEITOS DE FORÇA
FQ	DISCIPLINA DE FÍSICO-QUÍMICA
GC	GRUPO CONTROLE
GE	GRUPO EXPERIMENTAL
Gf	INTELIGÊNCIA FLUIDA
GMAT	GRADUATE MANAGEMENT ADMISSION TEST
GPs	GRUPOS DE PESQUISA
HCTE	HEAT AND TEMPERATURE CONCEPTUAL EVALUATION
HECI	HEAT AND ENERGY CONCEPT INVENTORY
HTCI	HEAT TRANSFER CONCEPT INVENTORY
IAVE	INSTITUTO DE AVALIAÇÃO EDUCATIVA
IBAP	INSTITUTO BRASILEIRO DE AVALIAÇÃO PSICOLÓGICA
IC	INVENTÁRIO DE CONCEITOS
ID	ÍNDICE DE DIFICULDADE DO ITEM
IDEB	ÍNDICE DE DESENVOLVIMENTO DA EDUCAÇÃO BÁSICA
IF	ÍNDICE DE FIABILIDADE
INEP	INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA
IPD	ÍNDICES PERCENTUAIS DE DISCRIMINAÇÃO
ITC	INTERNATIONAL TEST COMMISSION
IV	ÍNDICE DE VALIDEZ DO ITEM
KR20	COEFICIENTE DE FIABILIDADE DE KUDER-RICHARDSON
LD	LIVRO DIDÁTICO

LDF	LIVRO DIDÁTICO DE FÍSICA
MBT	MECHANICS BASELINE TEST
MCA	MOVIMENTO DAS CONCEPÇÕES ALTERNATIVAS
MDC	MODELOS DE DIAGNÓSTICO COGNITIVO
MH*	MATRIZ DE HABILIDADES COGNITIVAS E COMPETÊNCIAS TÉRMICAS. *O TERMO TAMBÉM FAZ REFERÊNCIA AO MÉTODO DE MANTEL-HAENZEL PARA ANÁLISE DE <i>FUNÇÃO DIFERENCIAL DO ITEM (DIF)</i> .
MIME	MONITORIZAÇÃO DE INQUÉRITOS EM MEIO ESCOLAR
MM	REGIÃO DE MÉDIA DIFICULDADE E VALOR MÉDIO PARA O FATOR DE CONCENTRAÇÃO
MMC	MODELO DE MUDANÇA CONCEITUAL
NAEP	NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
NCLB	NO CHILD LEFT BEHIND ACT
OCDE	ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO
OTI	OBJETIVO TERMINAL DE INTEGRAÇÃO
PCN+	INFORMAÇÕES ADICIONAIS AOS PARÂMETROS CURRICULARES NACIONAIS
PCNEM	PARÂMETROS CURRICULARES NACIONAIS DO ENSINO MÉDIO
PISA	PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT
PL	PARÂMETRO LOGÍSTICO, NO CASO, O MODELO PODE SER DE 1, 2, 3, 4 E 5 PARÂMETROS LOGÍSTICOS (1PL, 2PL, 3PL, 4PL E 5PL) DENTRO DA TEORIA DA RESPOSTA AO ITEM
PNE	PLANO NACIONAL DE EDUCAÇÃO
PNLD	PLANO NACIONAL DO LIVRO DIDÁTICO
PPG	PROGRAMA DE PÓS-GRADUAÇÃO
PPGECT	PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO CIENTÍFICA E TECNOLÓGICA
PPO	PEDAGOGIA POR OBJETIVOS
PT	PORTUGAL
PTC	PEDAGOGIA NA TRANSMISSÃO DE CONTEÚDOS
QEQ	QUADRO EUROPEU DE QUALIFICAÇÕES

QNO	QUADRO NACIONAL DE QUALIFICAÇÕES
RCAAP	REPOSITÓRIO CIENTÍFICO DE ACESSO ABERTO DE PORTUGAL
SABIC	SAMPLE SIZE ADJUSTED BAYESIAN INFORMATION CRITERION
SADEAM	SISTEMA DE AVALIAÇÃO EDUCACIONAL DO AMAZONAS
SADEAM	SISTEMA DE AVALIAÇÃO DO DESEMPENHO EDUCACIONAL DO AMAZONAS
SAEB	SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA
SARESP	SISTEMA DE AVALIAÇÃO DE RENDIMENTO ESCOLAR DO ESTADO DE SÃO PAULO
SBF	SOCIEDADE BRASILEIRA DE FÍSICA
SC	SANTA CATARINA
SEM	STANDARD ERROR OF MEASUREMENT (OU ERRO PADRÃO DE MEDIDA)
SFA	SABER-FAZER AFETIVO
SFC	SABER-FAZER COGNITIVO
SFG	SABER-FAZER GESTUAL
SIMAVE	SISTEMA MINEIRO DE AVALIAÇÃO DA EDUCAÇÃO PÚBLICA
SINAEB	SISTEMA NACIONAL DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA
SOLO	STRUCTURE OF THE OBSERVED LEARNING OUTCOME
STPFASL	SURVEY OF THERMODYNAMIC PROCESSES AND FIRST AND SECOND LAWS
TAI	TESTES ADAPTATIVOS INFORMATIZADOS
TCE	THERMAL CONCEPT EVALUATION
TCI	THERMODYNAMICS CONCEPT INVENTORY
TCLE	TERMO DE CONSENTIMENTO DE LIVRE E ESCLARECIDO
TCS	THERMODYNAMIC CONCEPT SURVEY
TCT	TEORIA CLÁSSICA DOS TESTES
TIC	TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO
TLI	TUCKER-LEWIS INDEX
TRC	TESTES REFERENCIADOS POR CRITÉRIO
TRI	TEORIA DA RESPOSTA AO ITEM

TRN	TESTES REFERENCIADOS POR NORMA
TTCI-T	THERMAL AND TRANSPORT CONCEPT INVENTORY
TTI	THERMODYNAMIK TEST INVENTAR
UA	UNIVERSIDADE DE AVEIRO
UC	UNIVERSIDADE DE COIMBRA
UFAM	UNIVERSIDADE FEDERAL DO AMAZONAS
UFMG	UNIVERSIDADE FEDERAL DE MINAS GERAIS
UFRGS	UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
UFSC	UNIVERSIDADE FEDERAL DE SANTA CATARINA
UNESCO	ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA
USF	UNIVERSIDADE DE SÃO FRANCISCO
USP	UNIVERSIDADE DE SÃO PAULO

INTRODUZINDO E SITUANDO AS PERSPECTIVAS DA INVESTIGAÇÃO

CONTRIBUIÇÕES E CONTRAPONTO NO USO DE TESTES DIAGNÓSTICOS CONCEITUAIS

"O fator isolado mais importante influenciando a aprendizagem é aquilo que o aprendiz já sabe. Descubra isso e ensine-o de acordo" (Ausubel, Novak, & Hanesian, 1980)

O estudo defende que as tomadas de decisões na educação sejam baseadas em evidências empíricas sob os critérios de *relevância*, *suficiência* e *veracidade* (Thomas & Pring, 2004, p. 5), adotando uma concepção filosófica pragmática. O termo “baseado em evidências” implica busca, aplicação e avaliação crítica das evidências disponíveis em pesquisas, implementação das evidências na prática e avaliação das informações disponíveis (Silva, 2014). No caso, o estudo estabeleceu como método a aplicação do teste denominado *Thermal Concept Evaluation* (TCE), e as evidências se configuram como *respostas de superfície* sobre um determinado domínio específico e introdutório da Física explorando diversas concepções dos alunos. Esse teste foi elaborado em 2001 por dois pesquisadores australianos, Shelley Yeo e Marjan G. Zadnik (Yeo & Zadnik, 2001), sendo conhecido e utilizado amplamente na literatura como instrumento de detecção de evidências empíricas, isso possibilita um diálogo com outros resultados disponíveis, sem descartar estudos correlatos voltados para instruções de ensino e eventuais implementações práticas que adotam como base essas informações. Foi identificado na literatura um teste com denominação e finalidade semelhante (Talaeb & Wattanakasiwich, 2010).

Com isso, apesar de algumas limitações identificadas nesse estudo, justifica-se a adoção do teste TCE por algumas vantagens diante a outros testes, como:

- A linguagem utilizada no teste explora um diálogo entre jovens com faixa etária que se aproxima do modo de comunicação e pensamento dos respondentes no estudo;
- As situações físicas exploram contextos típicos do cotidiano dos respondentes,

nomeadamente em recintos domiciliários como atividades rotineiras na cozinha, como pegar um refrigerante ou cubos de gelo dentro da geladeira, por exemplo. Entende-se que isso possibilite uma maior familiaridade e entendimento, favorecendo a apreensão, interesse e a significância, pois esses são elementos fundamentais em uma pedagogia de integração na qual se vislumbra uma progressão mais consistente do processo da aprendizagem na perspectiva de um desenvolvimento de competências científicas;

- Possibilidade do diálogo de resultados dentro de outros contextos com diferentes influências culturais para a interpretação de termos e conceitos que envolvem o enunciado do item e suas alternativas;
- Verificação mais precisa das possíveis causas de consistência/inconsistência nas respostas, bem como da coerência/incoerência na sequenciação de respostas com dependência local em relação a um mesmo contexto (p. ex.: em um item, retira-se o gelo do congelador e em outro mistura-se supostamente cubos de gelo advindos da mesma situação, instigando continuamente sobre suas temperaturas e conceito de equilíbrio térmico envolvido).

A finalidade prioritária está em identificar, analisar e comparar as dificuldades conceituais que os alunos (e diferentes grupos) possuem em **conhecimentos básicos e introdutórios da Física Térmica** no último ciclo da educação básica em dois contextos (Brasil e Portugal), tendo como base a aplicação de um teste de sondagem. A comparação dessas dificuldades tem a intenção de pôr *lado a lado* um perfil das concepções em termos comuns e diferenciados para cada contexto, mas também possibilitar uma comparação na mesma métrica, ou seja, identificar não apenas as diferenças como também posicionar em uma mesma “régua” os níveis de proficiência de cada grupo. Entende-se dentro dessa perspectiva que explorar um perfil de concepções térmicas deva não apenas representar as necessidades e diferenças de uma *população*, mas também demonstrar preocupação e possibilitar uma inferência para um *perfil individual*, apesar de que, para isso, existem condições que dependem do nível de consistência dos dados, de acordo com critérios estatísticos.

Tem-se imbricada nesse processo, sendo função secundária e subordinada a finalidade da pesquisa, **construir uma escala de proficiências em concepções térmicas** com vistas no aprimoramento do instrumento e da escala, avançar em novas perspectivas de análise sobre as concepções em destaque que já foram anteriormente mapeadas, possibilitar estudos longitudinais comparativos e internacionais, além de disponibilizar a escala a literatura, juntamente com o microdados. Para isso, a aplicação do teste explora concepções diante de conceitos específicos nesse campo de conhecimento, e assim possibilitando caracterizar e comparar cada contexto através das interpretações pedagógicas de indicadores psicométricos advindos da *Teoria Clássica dos Testes* (TCT) e *Teoria da Resposta ao Item* (TRI). As evidências empíricas são medidas de «traços latentes» referentes ao conhecimento prévio dos alunos em uma perspectiva estatística. Esse conhecimento evidenciado pode ser configurado como detentor de *concepções alternativas* quando se tratar dos distratores, e um *subsunçor*¹ de acordo a *Teoria da Aprendizagem Significativa* de Ausubel, quando evidenciar sua preexistência com a alternativa correta. Entende-se como desdobramentos, entretanto, que uma associação com os *conhecimentos potencialmente significativos* não se dá exclusivamente com as respostas corretas nos testes, e nem que essas respostas estejam desprovidas de equívocos. Contudo, acredita-se que as evidências possam sustentar argumentos instrucionais em caráter diagnóstico voltados para a proposição de melhorias no processo de ensino-aprendizagem. Isso se dá porque eles apontam para aspectos específicos conceituais que caracterizam os reais problemas pedagógicos. Dessa forma, pode-se dizer que a investigação se caracterize como diagnóstica, exploratória e comparativa baseada em uma avaliação de desempenhos entre alunos brasileiros e portugueses. Ressalta-se que esse tipo de avaliação, ainda que em uma perspectiva mais ampla de alcance populacional e dentre os procedimentos de análise

¹ Subsunçor é a denominação portuguesa adotada para o termo inglês "*subsumer*", proposto por Ausubel em 1968. Segundo Ausubel, os subsunçores são estruturas de conhecimento específico que podem ser mais ou menos abrangentes de acordo com a frequência com que ocorre aprendizagem significativa em conjunto com um dado subsunçor. Também pode se dizer que seja um conhecimento prévio capaz de subsumir um novo conhecimento; subsumir significa *tomar, acolher, aceitar* (Moreira, 2011). A *assimilação subsunçiva* é o processo na qual o subsunçor se relaciona com um novo conceito mais inclusivo.

explorados, ainda seja apenas um ponto de partida para que se busque reais mudanças e melhorias no campo pedagógico, como afirma Nigel Brooke (Serpa, 2011).

Espera-se que o conhecimento prévio seja o conhecimento minimamente necessário e disponível na estrutura cognitiva dos alunos antes e no momento da aplicação do teste, que por sua vez irá sugerir a natureza da evidência manifestada e aferida. Nesse sentido, um teste de sondagem de uma camada (múltiplas respostas, por exemplo) possui certas limitações de fiabilidade diante das evidências que devem ser consideradas, em que são minimizadas quando se exploram outras camadas de verificação do conhecimento, na qual os itens são acompanhados por perguntas abertas sobre o mesmo «traço latente», ou escalas ordinais referentes a unidade de medida para certificar o grau de domínio ou de convicção nas respostas.

Quando a evidência sinaliza haver certo domínio sobre o conhecimento envolvido por meio de uma resposta correta, supõe-se que o «subsunçor»² esteja previamente disponível na estrutura cognitiva. Caso contrário isso deverá sugerir a ausência desses conhecimentos ou a existência de ideias falsas sobre os conceitos e uma má interpretação diante dos fenômenos térmicos (no caso), e isso pode ser observado indiretamente na força atrativa dos distratores, ou o percentual absoluto diante as outras opções de resposta. Entende-se que conhecimento envolvido e necessário para responder com sucesso o teste, poderá eventualmente recorrer a outros recursos cognitivos considerados equívocos diante dos conceitos que estão inerentes, ou simplesmente haver um total desconhecimento, fazendo os respondentes optarem de forma aleatória. É importante ressaltar que a aferição do conhecimento prévio possui uma perspectiva estatística (e não individual) dentro da Psicometria quando se trata de médias, como ocorre geralmente, pois é sustentado pela probabilidade de sua existência (ou não) em determinados grupos normativos dentro da amostra. Apesar disso, são sempre estimados valores individuais para o seu desempenho, mas que as interpretações válidas nesse caso dependem do nível de consistência dos dados.

² A definição e relação entre os termos “traço latente”, “subsunçor” e “concepções alternativas”, no que diz respeito ao conhecimento prévio, serão explicados com detalhes mais a diante.

Pode-se dizer que a compreensão do conhecimento prévio nesse tipo de análise se encontra mais dentro de uma *concepção cognitivista*, entretanto, os procedimentos adotados possuem naturalmente uma *influência comportamentalista* ao utilizar o «paradigma fatorial» como base das análises para prever capacidades de compreensão e ação em potencial³, bem como a utilização de *critérios referenciados* no teste, e auxílio na taxonomia revisada de Bloom (Anderson et al., 2001) para elaboração das sentenças descritoras na matriz de habilidades e competências. Essa é uma perspectiva que também ocorre no PISA (Keevy & Chakroun, 2015) e no *Exame Nacional do Ensino Médio no Brasil* (ENEM) (Brasil, 2015), entre outros, e com isso, se concebe uma correspondência entre essas perspectivas no estudo.

Recorre-se a uma análise taxonômica pelo fato do conteúdo ou «saber» contido no teste estar associado em diferentes níveis de compreensão ou categorias classificatórias hierarquicamente, como fatos, conceitos, relações ou proposições entre conceitos, processos, estruturas e sistemas. Assim, esse foi um processo realizado devido a necessidade de verificar preliminarmente a natureza do conhecimento envolvido, e com isso alcançar maior precisão e confiança na sequência das análises e interpretações.

Nesse processo, as respostas dadas entre os diferentes contextos configuram as evidências coletadas e manifestadas pela suposta influência das preexistentes **concepções dos sujeitos que se associam ao conhecimento exigido no teste**. Entende-se que essa influência parte das condições que lhe são postas, isso envolve o enunciado da «situação-problema» de cada item e suas alternativas, pois o fenômeno da tomada de decisão do aluno se configura na relação entre aquilo que o teste pretende “medir” e o conhecimento prévio do respondente, que pode estar sendo influenciado por um fator externo, entre outras, a própria formulação do item. Isso torna os conhecimentos textualizados que estruturam o teste, assim como as respostas, parte do **objeto de estudo a ser explorado**.

³ Como potência no sentido de que a compreensão esteja latente na mente do sujeito, e conseqüentemente, espera-se que ele possa agir de modo correspondente diante a situações semelhantes.

No sentido de alcançar uma compreensão minimamente satisfatória e conjectural sobre essa relação, considera-se necessário e apropriado recorrer aos aportes teóricos da psicologia educacional tendo como **unidade de análise a estrutura cognitiva**. Com isso, foi considerado adequado adotar as perspectivas da psicologia educacional de David Ausubel, que fornece um suporte dentro dessa unidade de análise. Acredita-se que assim possa atender as necessidades de interpretar e reformular os itens, seja identificando os possíveis *subsunçores*, verificando-os supostamente através das evidências que se encontram nas respostas, como propondo instruções mais eficientes para o ensino ao serem sugeridos uma abordagem que contemple *conhecimentos potencialmente significativos*.

Sabe-se que a capacidade na resolução de problemas de itens em um teste conceitual está relacionada com ideia de «competência ou proficiência científica», estando, entretanto, delimitado em uma perspectiva mais complexa dentro desse paradigma, ou seja, sendo um nível secundário de competência, situado dentro do domínio cognitivo, e correspondente a um campo específico do saber. Pode-se dizer que está situado mais precisamente dentro das competências básicas, sendo comuns e fundamentalmente indispensáveis para uma progressão da aprendizagem, e com isso, acarretam serem transferíveis e comparáveis entre diferentes contextos. As análises comparativas das capacidades de compreensão em um sentido mais amplo concebem essa perspectiva.

Conhecimento Prévio e o Movimento das Concepções Alternativas

Praticamente, em todos os modelos de ensino, o *conhecimento preexistente* nos alunos é uma das evidências fundamentais e consensuais a serem detectadas e valorizadas no cenário educacional.

A preocupação com conhecimento prévio foi desenvolvida inicialmente e com mais propriedade por Jean Piaget (1896 – 1980) na década de 1920, e posteriormente se destacando com o conceito de *subsunçor* desenvolvido por David Ausubel (1918 – 2008) na *Teoria da Aquisição e Retenção Significativas de Conceitos* na década de 1960.

Atualmente a sua importância é inquestionável dentro da literatura, e segue dentro de diversos outros campos teóricos afins, como a da *Teoria dos Campos Conceituais* de Gérard Vergnaud, discípulo de Piaget. Vale ressaltar que os trabalhos de Piaget (Piaget, 1952) inspiraram diversas perspectivas na ciência intuitiva dentro da psicologia educacional. Piaget, porém, afirmou que o desenvolvimento cognitivo progride em uma direção através de um conjunto rígido de estágios, cada um envolvendo esquemas cognitivos radicalmente diferentes (Pellegrino, Chudowsky, & Glaser, 2001). Uma grande verdade é que a maioria das crianças não vão às escolas preparadas para aprender da mesma maneira, pois eles não seguem os mesmos processos de crescimento e mudança, e essa é inclusive, uma percepção defendida atualmente por Vergnaud, um dos maiores especialistas atuais nesse campo. Vergnaud considera que as crianças adquirem novos procedimentos lentamente e ao longo de vários caminhos, avançando em especificidade diante das condições essenciais para uma adaptação e transformação do conhecimento durante a sua aquisição. Para isso, ele destaca a necessidade de compreender o processo operacional e de desenvolvimento da aprendizagem em termos de competências, dando *relevância ao papel das situações e contextos* em que os alunos são submetidos. De certa forma, entende-se que os tipos de situações-problemas que os testes apresentam contribuem nesse processo.

Sabe-se que, a preocupação com a natureza desse tipo de conhecimento (prévio) influenciou o desenvolvimento de pesquisas de levantamento e mapeamentos das diferentes concepções que as crianças possuem diante os conceitos científicos, muitas delas carregadas de crenças e equívocos. Na década de 70, estudos bem fundamentados (Viennot, 1978) consideraram que essas concepções prévias seriam fatores estáveis e que estariam a bloquear a efetividade do ensino, sendo concebidas desafortunadamente por alguns como sendo *erros conceituais* (Lahera & Forteza, 2006). A ideia de *erro conceitual* foi posteriormente admitida predominantemente como ideias esquematizadas sobre temas científicos, anteriores à instrução escolar, sendo então designadas como *ideias prévias* ou *concepções prévias*, e também de *esquemas conceituais alternativos* (Driver, 1983), *ciência das crianças* (Osborne & Wittrock, 1983), *concepções* ou

construtos (Giordan & Vecchi, 1987). É bem conhecido que essas concepções muitas vezes persistem após a instrução, podendo de modo geral serem denominadas de *concepções dos alunos* ou *concepções alternativas*⁴.

O avanço desses estudos configurou o desenvolvimento do *Movimento das Concepções Alternativas* (MCA) iniciado no fim dos anos 70 com os trabalhos de Laurence Viennot e Rosalind Driver, sendo impulsionado no início da década de 80 pelo artigo clássico de Posner et al. (1982) na qual propuseram condições de acomodamento (insatisfação, inteligibilidade, plausibilidade e fertilidade) que fundamentaria e reivindicaria um *Modelo de Mudança Conceitual* (MMC). Na década seguinte, esse modelo recebeu críticas, incluindo algumas dos próprios autores originais (Strike & Posner, 1992), sendo também contestado por Mortimer (1996), na qual defende que as concepções alternativas não precisariam de serem eliminadas, mas sim inculturadas e utilizadas conscientemente dentro de cada contexto. Ainda assim, muitos esforços ainda vem sendo feitos para compreender qual instrução promove essa mudança (ou dificulta o progresso desejado) (Jorde & Dillon, 2012).

Ao analisar o teor atual da *teoria da mudança conceitual*, Vosniadou (no prelo) observou que houve um debate contínuo sobre a natureza do que é que muda e a elaboração extensiva de possíveis mecanismos que permitem processos de mudança conceitual. (Tytler & Prain, 2010, p. 4, *tradução livre, grifo nosso*)

Devido ao impulso do MCA, as *Concepções Alternativas* (CAs) foi considerada uma das linhas mais relevantes dentro da Didática das Ciências até o início do século XXI, possibilitando repositórios alargados sobre as concepções que as crianças tinham diante de conceitos científicos em diversas áreas (Carrascosa, 1985; Carrascosa, 1983; Driver, 1985; Montero & Hierrezuelo, 1988; Osborne & Wittrock, 1983; Wandersee, Mintzes, & Novak, 1994). Em meio a isso, diversos métodos foram utilizados para identificar as concepções dos alunos, como por exemplo, o exame ou *entrevista clínica* (Piaget, 1947), entrevistas coletivas, painel (de ideias prévias) de grupo, testes (Ward, Roden, Hewlett, & Foreman, 2009), análises multidimensionais (Santos & Moreira, 1979b) e de agrupamentos hierárquicos (Santos & Moreira, 1979a) ao mapeamento

⁴ *Misconceptions.*

cognitivo de conceitos físicos, entre outros. Uma síntese de alguns métodos e instrumentos com esse propósito pode ser observado:

Quadro 1: Métodos e técnicas para explorar conteúdos e detectar concepções dos alunos.

Instrumentos		
Levantamento da informação através de registo escrito	Levantamento através da informação de respostas dadas por um grupo.	Entrevistas individuais
<ul style="list-style-type: none"> • Provas: respostas abertas ou fechadas; • Desenhos interpretativos; • Dever de casa: listas de problemas, monografias, relatórios, etc.; • Trabalhos escritos, etc. 	<ul style="list-style-type: none"> • Laboratório; • Demonstrações; • Leitura de textos, vídeos, filmes seguida de discussão crítica e relato do consenso grupo, escrito ou oral (p.ex., Pereira & Barros, 2010). 	<ul style="list-style-type: none"> • Entrevista clínica (Piaget, 1933, 1947, 1973); • Entrevista conceitual; • Entrevistas coletivas; • Entrevistas sobre ocorrências (eventos).
Técnicas		
<ul style="list-style-type: none"> • Mapeamento de conceitos: estabelecem relações de significância entre conceitos (Novak & Gowin, 1984) – podendo ser aplicados para avaliar as estruturas dos conceitos e concepções dos alunos após uma entrevista em uma organização hierárquica (p.ex.: Teoria > Princípios > Conceitos). <i>Alguns trabalhos relacionados:</i> (Lacerda, Martins, & Reis, 2009; Moreira, 1998, 2006; Nunes & Del Pino, 2008; Pacheco & Damasio, 2009; Pelizzari, Kriegl, Baron, Finck, & Dorocinski, 2002; Tavares, 2007) • Análise Proposicional de Conceitos – podendo ser aplicada em entrevistas (Novak & Gowin, 1984, p. 156). • “Vê” de Gowin – podendo ser aplicada em entrevistas e servir de base organizacional para mapas conceituais⁵ (Novak & Gowin, 1984, p. 159) • Previsão, Observação e Explicação (POE) (Gunstone & White, 1981) – semelhante a metodologia proposta por Nedelsky (1961) para o ensino e avaliação no laboratório de Física; • Linhas de fortuna – permitem estabelecer escalas subjetivas (gostar; preferir, acreditar, valer, aprender, sentir) sobre aquilo que o aluno avalia ao longo de um intervalo de tempo prefixado, verificando sua resposta afetiva a situações de aprendizagem escolar; • Associação de palavras, conceitos – permitem verificar como um estudante percebe as associações entre conceitos de um dado conjunto. Facilitam a aprendizagem de vocabulários técnicos e habitam os estudantes a correlação conceitual mais aprofundada de assuntos que não parecem estar diretamente. Sua maior vantagem é a simplicidade de preparação e leitura, podem ser aplicados a grande número de estudantes e fornecem informações relevantes quanto ao grau de compreensão que os estudantes possuem num dado instante (Preece, 1976). • Dramatização e role-play – Estas metodologias são ainda pouco utilizadas em sala de aula. Requerem professores preparados para criar/aplicar/avaliar os instrumentos necessários e ainda existe pouca bibliografia no assunto. Estas técnicas são especialmente úteis para a discussão de tópicos abrangentes, como por exemplo aqueles relacionados com ciência/ tecnologia/sociedade (Rodrigues, 1993). 		

Fonte: Adaptado de Barros (1994).

Em 1929, Jean Piaget comparou o uso de métodos de testes, observação e de entrevista clínica com o propósito de detectar essas concepções. Destacou que os testes possuem vantagens indiscutíveis em diagnósticos, assim como quando aplicadas em larga escala: *“para a psicologia geral também as estatísticas resultantes geralmente fornecem informações úteis”* (Piaget, 1973, p. 4). Entretanto, apontou problemas relacionados a *falta de contextualização e inclinação mental* das crianças que os testes podem fornecer (Ibid, 1973, p. 4). A *entrevista clínica* de Jean Piaget foi detalhada e reconhecida por muitos pesquisadores como a melhor forma de identificar as concepções alternativas (Domínguez, 1985; Pines, Novak, Posner, & VanKirk, 1978), porém, é considerada *“uma técnica que exige muito treino do pesquisador, é morosa e a análise de dados obtidos é*

⁵ Os autores apresentam um exemplo diante da temática da estrutura molecular da matéria.

extremamente difícil”, além de ser pouco viável para fins de utilização na prática docente (Silveira, Moreira, & Axt, 1986).

Silveira, Moreira, & Axt (1986) comentam que diversas técnicas e métodos de análise do *mapeamento cognitivo* foram revisados por Preece (1978), e muitas delas se baseiam em identificar as “distâncias semânticas” entre os conceitos. A exemplo disso se tem o mapeamento cognitivo explorado por testes de associação de palavras e técnicas de escalonamento multidimensional, assim como de análises de agrupamento hierárquicos, na qual foram criticadas por não garantirem que haja uma relação com a estrutura cognitiva dos sujeitos (Stewart, 1979).

Na década de 90, esses levantamentos feitos durante o MCA possibilitaram a elaboração de *Inventários de Conceitos* (IC) ou de concepções alternativas a serem utilizados para diversos campos de pesquisa na área do ensino, reforçando o destaque para a utilização em testes conceituais como instrumento de recolha e produção de dados que já vinham sendo realizados nas décadas anteriores. Ainda assim, esse tipo de instrumento vem recebendo críticas quanto a sua eficiência do ponto de vista da qualidade dos itens elaborados e dos procedimentos realizados durante a sua aplicação (Hoffmann, 1961), principalmente pelas funções os testes exercem diante de avaliações externas e as implicações para a prática pedagógica. Os “*Inventários de Conceitos*” foi um movimento entre os métodos de avaliação na área do *Ensino de Ciências* (EC) que tomou raízes a partir do trabalho de Hestenes, Wells, & Swackhamer (1992) ao criarem o *Inventário de Conceitos de Força* (FCI) (Nelson, Geist, Miller, Streveler, & Olds, 2007), sendo anteriormente sido desenvolvido um trabalho similar (Halloun & Hestenes, 1985a).

De modo geral, os levantamentos realizados sobre as concepções científicas dos alunos nas décadas de 70 e 80 foram bastante utilizados como base de conhecimento dentro de unidades temáticas para a construção de testes, tornando promissora o uso de testes em larga escala para contribuir em diagnósticos sobre as dificuldades conceituais, verificação de mudança e em fornecerem informações pedagogicamente relevantes sobre o conhecimento prévio. Muitas dessas análises realizadas se

concentraram na estatística descritiva, porém sem aplicar ao menos o *fator de Hake* em pré-testes e pós-teste, que seria um ganho para essas análises, além de um procedimento fácil de replicar em salas de aula pelos professores (Hake, 2008), apesar de terem sido realizadas algumas ponderações quanto a sua finalidade de instruir, entre outros processos de avaliar ganhos médios normalizados (Miller et al., 2010). Ainda assim, foi constatado algumas vantagens no uso do *fator de Hake* para avaliar tópicos em cursos de engenharia (Azevedo & Galhardi, 2013). Por outro lado, procedimentos de análises mais robustos e rigorosos dentro da «Psicometria» estão sendo pouco exploradas entre pesquisadores que se fundamentam na psicologia educacional para avaliar a aprendizagem com base no desempenho. Essas análises possibilitam inferir com maior qualidade o conhecimento (domínio cognitivo) dos sujeitos e podem fornecer informações sobre os processos mentais, consideradas ricas pedagogicamente, entretanto, estão em um lugar marginalizado diante da predominância de pesquisas qualitativas no *Ensino de Ciências* (EC) no contexto brasileiro, sustentada, de modo geral, por uma aparente aversão equivocada de pesquisadores educacionais com relação ao paradigma positivista (Matthews, 2004).

Delimitações em Testes Conceituais

Testes que exploram conceitos básicos e introdutórios ajudam o professor a responder a seguinte questão: *Em quais aspectos eu posso mudar e melhorar meu ensino?* Obviamente que os testes não devam substituir outros mecanismos diagnósticos e nem outros instrumentos de ensino que tenham essa finalidade, mas sim contribuindo ao fornecerem indicadores complementares nesse âmbito, porém eles se destacam pela praticidade e por se fundamentarem em uma das questões centrais para otimizar o processo de ensino-aprendizagem, no que refere a identificar e valorizar o conhecimento prévio. Apesar de eventuais críticas, essa perspectiva se distancia em grande parte dos efeitos perversivos relacionados ao mal uso de testes em larga escala. Os *Inventários de Conceitos* (ICs) não devem exercer uma função de alto-risco (função avaliativa certificativa, utilizado para determinar o sucesso ou fracasso), e nem tenderem a promover distorções ou reducionismos curriculares se forem bem

projetados, administrados adequadamente e **exercerem uma função diagnóstica**, seja para orientar ou regular.

Como documento de referência aos estudos que envolvem testes, o *International Test Commission* (ITC) (Cheung, Iliescu, & Bartram, 2014c) descrevem alguns princípios gerais para a administração de testes psicológicos em pesquisas, na qual os ICs estão de certa forma relacionados dentro dos diferentes tipos de testes educacionais.

[...] procedimentos ou métodos que investigam ou **determinam a presença de um fator ou fenômeno que compreende um conjunto de itens padronizados** (por exemplo, questões, estímulos ou tarefas) que são corrigidos de uma maneira padronizada e usados para investigar e eventualmente avaliar diferenças individuais (por exemplo, **habilidades, competências**, temperamentos, atitudes e emoções) (Anastasi & Urbina, 1997; American Psychological Association, 2006; Cronbach, 1990). **Esta definição inclui testes psicológicos e educacionais** em todas as formas de desdobramento (por exemplo, livretos para resposta em papel com lápis, testes computadorizados *online*, amostras de trabalho, jogos interativos. Duas formas de testes são descritas: os que possuem direitos autorais, portanto, de domínio privado bem como **os que não possuem direitos autorais, portanto, de domínio público**. Muitos dos princípios discutidos a seguir se aplicam à ambas as formas. (Cheung, Iliescu, & Bartram, 2014, p. 3, *grifo nosso, tradução feita por*⁶ Cheung et al., 2014)

Apropriadamente, testes baseados em ICs são testes diagnósticos relacionados à educação, também denominados como *Testes Referenciados pelo Critério - TRC* ("*criterion-referenced measures*" ou "*criterion-referenced test*"), cunhado originalmente por Robert Glaser em 1963, juntamente com os *Testes Referenciados pela Norma - TRN* ("*norm-referenced measures*" ou "*norm-referenced test*") (Glaser, 1963). Esses testes já vem sendo utilizados desde as décadas de 60 e 70 (Weiss & Davison, 1981), e foram definidos com a finalidade de serem usados "*para verificar o status de um indivíduo (o escore do domínio do indivíduo) em relação a um domínio de comportamento bem definido*" (Popham, 1975, p. 130; Hambleton, Swaminathan, Algina, & Coulson, 1978, p. 2, *citados por* Weiss & Davison, 1981, p. 12, *tradução livre*). Também coexistem outros semelhantes como a *testagem referenciada por domínio* ("*domain-referenced testing*") (Hively, 1973, 1974b, 1974a; Hively & Reynolds, 1975), com alguns trabalhos mais atuais (Baerisch &

⁶ O mesmo documento foi traduzido por pesquisadores da PUC-Campinas (São Paulo).

Hasselbring, 2010); *testagem definida pelo universo* (“*universe-defined testing*”) (Hively & Patterson, 1968; Osburn, 1968).

Aparentemente a ênfase dada no segmento metodológico de utilização de testes no Brasil foi sutil e mais intenso nas décadas de 80 e 90, sendo dissipado pelo forte acultramento das abordagens qualitativas para o enfrentamento dos fenômenos educacionais, agravado supostamente pela falta de interesse e de domínio entre pesquisadores neste campo, dentro e fora dos *Programas de Pós-Graduação* (PPGs) no *Ensino de Ciências*⁷, na qual continua a ser uma realidade nos dias atuais, como revelam as produções em algumas revistas especializadas (Santos & Greca, 2013).

No início do século XXI, o MCA no contexto da Didática das Ciências já havia perdido a força como linha prioritária de investigação devido ao grande volume de levantamentos feitos. Ainda assim, mesmo as CAs sendo um condicionante indispensável para o EC, infelizmente não vinham apresentando um impacto nas salas de aula (Köhnelein & Peduzzi, 2002), muito à custa da ausência de estudos mais rigorosos dentro da psicologia cognitiva e que não informam aspectos práticos que supostamente estariam influenciando a realidade dos fenômenos do ensino e aprendizagem (Duit & Treagust, 2003).

O levantamento mostra que a grande parte dos trabalhos (p. ex.: Custódio, 2012) desenvolvidos no contexto brasileiro sobre testes conceituais de múltipla escolha foram predominantemente concentrados em percentuais de acerto e métodos clássicos de análise, na qual “*trazem poucas informações sobre os processos mentais e geralmente culminavam nas diferenças observadas no desempenho*” (Primi, 2017). Além disso, testes complementares de Física escritos de verificação vinham (e ainda continuam) apenas atribuindo notas, e assim exercendo uma função avaliativa somativa ao invés destacar o exercício de uma função diagnóstica (Barros, 1994), que no caso, poderia ser útil tanto para orientar a prática docente como para uma autorregulação do aluno em seu processo de aprendizagem. Nesse sentido, é preocupante saber que essa função avaliativa (diagnóstica) e a etapa de integração das aquisições durante as

⁷ Atualmente as denominações dos PPGs de “Ensino de Ciências” tendem a migrar para a designação “Educação em Ciências”, buscando assim abarcar e diferenciar macro e microfenômenos educacionais.

aprendizagens, apesar de serem consideradas relevantes, continuam notoriamente encontrando dificuldades de serem realizadas entre os professores, inclusive de serem valorizadas em materiais instrucionais, como no caso dos *Livros Didáticos* (LDs).

Apesar dos testes conceituais de múltipla escolha se mostrarem instrumentos promissores na constatação de concepções de grupos de alunos, a sua utilização seguindo os pressupostos da *Teoria Clássica dos Testes* (TCT) possui limitações em analisar com precisão essas concepções individualmente e comparativamente em termos de desempenho e probabilidade de acerto/erro. Isso ocorre por ser um modelo que se concentra na informação do *score bruto* (soma dos acertos de um teste) de cada sujeito, ou seja, em informações de cada sujeito diante do teste como um todo. Esse tipo de procedimento também não permite acompanhar e comparar o desenvolvimento do aprendizado numa sequência de etapas de ensino ao serem obtidos resultados advindos de testes diferentes, pois cada teste possui características próprias, estando em métricas distintas, impossibilitando comparações diretas de grupos ou indivíduos, pois as pontuações das pessoas nesse caso dependem do teste (Primi, 2017).

Os modelos logísticos da *Teoria da Resposta ao Item* (TRI) complementaram e superaram algumas limitações encontradas na TCT, além de apresentarem análises mais robustas, possibilitando mais vantagens do ponto de vista pedagógico. Entre as vantagens, pode-se dizer que ela possui a capacidade de: *aferir habilidades cognitivas em probabilidades de acerto; medir a cognição em patamares de proficiência; construir de séries históricas de desempenho utilizando testes diferentes; posicionar em uma mesma métrica dificuldades específicas relacionadas a determinados itens e sujeitos; sugerir conteúdos relevantes com base no desempenho; discriminar grupos de alto e baixo desempenho (dentro de uma sala de aula, entre diferentes turmas e escolas, etc.); analisar comparativamente populações diferentes de modo mais fiável e válido; diferenciar grupos distintos (o gênero, a raça, a condição socioeconômica, etc.) de acordo com a funcionalidade do item, entre outros.* É importante ressaltar que a avaliação de desempenho em testes conceituais é um segmento entre outros testes educacionais, e seus resultados podem ser designados como indicadores a contribuir dentro de uma perspectiva mais ampla de uma avaliação educacional.

Entende-se que muitas dessas informações possam vir a ajudar professores, pedagogos e gestores a tomarem decisões educacionais e darem um tratamento diferenciado e mais focalizado diante das necessidades e particularidades dos alunos ou grupos específicos. Obviamente, é preciso que haja uma adequada política de gestão *accountability* que desenvolva e priorize um *feedback* dos resultados de estudos nesse âmbito assumindo um compromisso com a aprendizagem.

Como foi dito, um teste conceitual elaborado e aplicado por um especialista da educação não é formalmente um típico teste psicológico, pois esse é geralmente voltado para a psicologia experimental como um instrumento padronizado e de uso exclusivo do psicólogo⁸. Ao contrário disso, testes previamente não-padronizados vem sendo elaborados e analisados entre pesquisadores educacionais, principalmente entre aqueles atraídos dentro de uma linha investigativa voltada para a avaliação, mas que aparentemente não demonstram seguirem recomendações da psicologia para o uso desse instrumento no âmbito de avaliação em larga escala, possivelmente apresentando riscos de vieses interpretativos e uma necessidade rigor científico-metodológico dentro de investigações focadas em microfenômenos no âmbito da «avaliação educacional».

Na psicologia geral se recomenda que os «testes psicológicos» tenham um caráter restrito, diferenciando de uma «avaliação psicológica» (Hutz, Bandeira, & Trentini, 2015), pois o teste psicológico deve ser considerado apenas um instrumento entre outros para a avaliação psicológica: *“principais técnicas de avaliação psicológica: entrevistas, observação, testes psicológicos, dinâmicas de grupo, observação lúdica, provas situacionais e outras”* (Machado & Morona, 2007). O rigor recomendado para uma **avaliação psicológica** no campo experimental talvez sinalize algumas das críticas existentes quanto a aplicação de testes educacionais em larga escala impostas pelas políticas de governo, na qual se tem a pretensão demasiada de estabelecer uma

⁸ Os testes psicológicos são instrumentos de avaliação ou mensuração de características psicológicas, constituindo-se um método ou uma técnica de uso privativo do psicólogo, em decorrência do que dispõe o § 1º do art. 13 da lei no 4.119/62. (Resolução CFP 002/2003).

avaliação educacional de um sistema buscando satisfazer a três propósitos: *comparar, instruir e avaliar*.

Além de limitações existentes na atribuição de propósitos, o *uso e abuso* desses recursos são apontados como causadores de prejuízos a educação por causarem constrangimentos e mal-estar devido ao *ranking* entre escolas, professores e alunos (Afonso, 2009), indo na contramão de evidências em boas práticas de gestão como ocorre na Finlândia, segundo especialistas (Abraham, 2012; Darling-Hammond & Rothman, 2011; Partanen, 2011; Sahlberg, 2011a, 2011c, 2011b, 2013), provocando não apenas um estreitamento curricular, mas uma subversão na prática escolar. O **papel instrutivo** de testes padronizados aplicados em larga escala vem sendo defendidos por alguns especialistas em *design* instrucional na educação (Tankersley, 2007), e por vezes em níveis mais restritos da função que lhe é atribuída, devendo ser mais apropriada atendendo a padrões em contextos mais específicos (Wormeli, 2006), e por outros, na condição ampla para a tomada de decisões em salas de aula sendo considerado **inapropriado** (Popham, 2016).

No entanto, testes construídos principalmente para comparações não são adequados para fins de instrução ou avaliação de qualidade instrucional na educação. Esses testes fornecem aos professores poucas informações de instrução e tipicamente levam a avaliações imprecisas da qualidade de instrução de um professor. (Popham, 2016, p. 1-2, *tradução livre*)

Algumas das principais implicações negativas quanto a utilização perversa de testes educacionais padronizados em larga escala vem sendo apontado durante anos por diversos avaliadores, os argumentos de ineficiência geralmente estão relacionados a: *relação de proximidade com interesses neoliberais e mercado de trabalho (contra os argumentos para o modelo de competências educacionais adotado); ineficiente para avaliar competências em um sentido amplo (por se restringir ao domínio cognitivo, defender um pressuposto de unidimensionalidade); e no sentido particular (por não ser diversificado no atendimento as diferentes potencialidades individuais dos sujeitos, de acordo com a Teoria das Múltiplas Inteligências de Gardner, p. ex.); por eliminar itens pedagogicamente importantes em testes devido à baixa consistência; por serem elaborados itens de péssima qualidade, além de "forçar" uma contextualização e interdisciplinaridade, como afirmam alguns críticos contra o*

ENEM⁹ (Silveira, 2013; Silveira, Barbosa, Silva, 2015); *por provocar um reducionismo curricular ao enfatizar padrões de conteúdos e desempenhos, e como consequência, deformar a prática pedagógica, entre outros.*

Muitas dessas críticas baseadas nesses argumentos tem mobilizados grupos de pais, professores e profissionais da educação em um posicionamento contrário ao uso dado aos testes educacionais, nos EUA, e Espanha, por exemplo. Estando o atual governo em meio à crise política, diversas instituições brasileiras (ANFOPE, 2016; ANPEd, 2017; CNTE, 2017; Parceria MEC/SEB/COEDI & UFMG, 2017; UFSC, 2017) vêm se manifestando contrárias às propostas das últimas versões da *Base Nacional Comum Curricular* (BNCC) quanto ao modelo de competências e padrões de conteúdo, afirmando um desajuste com o modelo de estrutural das diretrizes curriculares, do envolvimento de interesses empresarias, da exclusão de outras modalidades de ensino (especial, indígena, quilombola,...), de questões relacionadas aos gênero e orientação sexual, entre outras. A tendência é que a padronização de 60% dos conteúdos que se pretende (Brasil, 2015), seja um aspecto que irá refletir muito provavelmente dentro da reestruturação dos exames nacionais, priorizando os conteúdos na nova matriz de referência em detrimento dos 40% flexíveis, forçando um estreitamento curricular, e favorecendo o crescimento no uso de testes educacionais em larga escala por todo o país, como já estão implementados dentro das políticas de governo nos diversos sistemas estaduais de avaliação externa (SEAPE, SADEAM, SPAECE, entre outros). Essa perspectiva estabelece uma conexão com o estudo ao influenciar a aplicação de inventários de conceitos dentro das escolas, pois também são aplicados em larga escala, porém com uma função avaliativa diferente dos exames nacionais.

Apesar de algumas lacunas e necessidades apontadas serem coerentes, uma verdade quase certa é que o uso de testes irá continuar sendo uma parte da forma como as escolas e os professores serão avaliados, além dos alunos. Torna-se relevante com isso que os educadores e pesquisadores, em especial, aqueles ligados ao EC, devam conhecer e entender melhor alguns processos de padronização que permanecerão mantidos, assim como das utilidades que podem oferecer, e com isso serem discutidas

⁹ *Exame Nacional do Ensino Médio, Brasil.*

possibilidades mais fundamentadas de melhoria em sua perspectiva avaliativa. Essas discussões são comuns e ocorrem de forma semelhante nos EUA (Tankersley, 2007), e em muitos outros contextos.

Apesar de críticas aos modelos da TRI que são geralmente utilizados, deve-se levar em conta que esses modelos são amplamente reconhecidos e aplicados em diversas áreas da ciência moderna (neurociência, medicina, etc.), estando o uso desses testes padronizados em uma abordagem educacional se limitam na avaliação exclusivamente de *conhecimentos*, sendo um olhar diferenciado dos psicólogos experimentais que enfocam nos *processos mentais*.

Não deixa de ser em grande parte uma verdade afirmar que muitos dos críticos, possuem quase nenhum domínio dentro da Psicometria (Erthal, 1987), e desconhecem a existência no tratamento da convergência multidimensional e técnicas que superam o pressuposto da dependência local de itens por meio de *testlets* e métodos de *testes adaptativos computadorizados* (Haladyna, 2004). Além disso, vêm sendo desenvolvidas alternativas que podem e devem ser exploradas em pesquisas educacionais, como os *Modelos de Diagnóstico Cognitivo - MDC* ("*Cognitive Diagnostic Modeling*"), na qual enfatiza informações qualitativas (Huebner, 2010), aspecto reivindicado por muitos educadores diante dos testes em larga escala como recurso instrucional.

Interesse em modelos psicométricos referidos como *modelos de diagnóstico cognitivos* ("*cognitive diagnostic models*" - CDMs) vem crescendo rapidamente nos últimos anos, motivado em grande pelo apelo por mais avaliações formativas feitas pelo *No Child Left Behind Act of 2001* (No Child Left Behind, 2002). Ao invés de atribuir aos examinados uma pontuação em uma escala contínua representando uma habilidade latente amplamente definida, como os modelos comuns da *Teoria de Resposta ao Item* ("*item response theory*" - IRT) fazem isso de forma eficaz, os CDMs têm como objetivo fornecer aos peritos informações sobre se eles dominaram ou não cada um de um grupo de habilidades, ou atributos específicos, discretamente definidos. Essas habilidades são muitas vezes binárias, o que significa que os examinados são classificados como mestres ou não mestres de cada habilidade. (Huebner, 2010, p. 1, *tradução livre, grifo nosso*)

Análises com testes em larga escala também possibilitam contribuir em evidências de causalidade dentro de macrofenômenos educacionais relacionados a variáveis externas que influenciam a aprendizagem. Os resultados de bom

desempenho em testes aplicados em larga escala vem demonstrando uma relação com variáveis externas, como o rendimento familiar (Schwartzman, 2015) e em políticas escolares que ressaltam o envolvimento dos pais na aprendizagem (Wormeli, 2003, p. 145). De modo geral, eles não devem ser deliberados como única ferramenta de diagnóstico para um aluno individual, sendo preferível que sejam voltados a indicarem tendências e padrões de escolas ou distritos e para serem incluídos como uma das muitas fontes de informação sobre um aluno (Wormeli, 2006).

Educational Testing Service e outros fabricantes de testes padronizados são os primeiros a informar os educadores de que seus testes nunca são designados como a única ferramenta de diagnóstico para um aluno individual. Eles devem indicar tendências e padrões para uma escola ou distrito e para serem incluídos como uma das muitas fontes de informação sobre um aluno. Sim, eles podem fornecer um indicador inicial de realização ou aptidão em relação a um aluno, mas eles sempre devem ser usados em conjunto com outras ferramentas de avaliação para tomar decisões importantes em relação a uma criança individual. Isso se aplica à avaliação de professores desse aluno individual também. (Wormeli, 2006, p. 30, *tradução livre*)

Contudo, sabe-se que os modelos da TCT e TRI fazem parte da Psicometria, e que devem estar a serviço da psicologia de modo geral, com isso são adotados nesse estudo procedimentos de análise com a intenção de destacar as diferenças entre as concepções dos alunos e algumas possíveis contribuições no enfrentamento das dificuldades de aprendizagem relacionadas a aquisição de conceitos básicos, assim como indicadores para a avaliação e acompanhamento no desenvolvimento de competências.

MOTIVAÇÕES - REFERENCIAL DE VALORES DO INVESTIGADOR

Um alinhamento curricular entre os conteúdos enfatizados dentro de um programa disciplinar e as necessidades de aprendizagem é uma das condições *sine qua non* para um bom funcionamento e legitimidade de um sistema educacional, independente do contexto, e com isso, o caráter avaliativo na busca de evidências sobre as dificuldades em conceitos introdutórios é um caminho a contribuir, sendo considerado motivante neste estudo.

A escolha da temática situada numa introdução ao conhecimento da Física Térmica segue no interesse do investigador desde o mestrado, justificado em parte pela carência de estudos diante das outras temáticas em revistas especializadas no ensino da Física no Brasil (Braga, 2010, p. 32). Em Portugal, alguns tópicos específicos da Física Térmica vêm sendo cobrados continuamente em exames nacionais¹⁰ para os alunos do ensino secundário, sendo parte obrigatória a ser abordada nas escolas.

Quanto ao enfoque metodológico, a necessidade de evidências empíricas sobre o conhecimento prévio é um fator relevante a ser considerado, podendo fornecer particularidades de grupos específicos de alunos, e assim fornecer subsídios para novos estudos longitudinais, continuados e mais rigorosos, sendo úteis em abordagens mistas, na qual se reivindica. Diante disso, tem-se a perspectiva de contribuições pedagógicas advindas da Psicometria visando o desenvolvimento de competências científicas para a Física Térmica, assim como para outras áreas, tendo em vista estudos mostram que suas aplicações extrapolam o campo da Educação.

A ênfase dada no contexto brasileiro se dá pelos reflexos de intervenção futura nesse contexto. Em uma perspectiva mais ampla, os fatores motivacionais predominantes neste processo investigativo buscam destacar não apenas a relevância temática desenvolvida, mas também vislumbrar aspectos favoráveis para impactos/contribuições sociais posteriores:

- *Quanto à temática desenvolvida:* A temática encontra-se no ramo da Física Térmica, e se refere aos conceitos elementares e estruturadores da termodinâmica, que são: *o calor, a temperatura e entropia*, que se complementam numa compreensão sobre suas *propriedades térmicas* e os *processos termodinâmicos* que o corpo se envolve. Porém, o interesse pela natureza do conteúdo não é associado ao contexto na qual ele deve ser inserido e daquilo que se pretende fazer como esse tipo de conhecimento. A motivação nesse sentido está em alcançar a finalidade de promover a aquisição, o desenvolvimento e ampliação da compreensão diante dos conceitos térmicos, assim como nas capacidades e competências científicas envolvidas para resolver problemas dentro desse contexto.

¹⁰ Foi realizado um levantamento de todos os exames nacionais que se referem às temáticas experimentais previstas no programa curricular, ou seja, desde 2006 até 2013.

- *Quanto às competências do investigador para desenvolver o tema:* Segue numa continuidade investigativa dentro da mesma unidade temática (conhecimento térmico) e dos mesmos fundamentos teóricos da psicologia da educação de Ausubel, buscando coerência na preocupação com um processo pedagógico «significativo» para os sujeitos e potencialize o processo de aprendizagem com base no conhecimento prévio. Quanto aos procedimentos de análise, foi realizado um esforço para alcançar um domínio mínimo necessário para enfrentamento do problema proposto e delimitação do escopo do trabalho, para isso priorizados estudos desenvolvidos durante âmbito do regime de cotutela, curso de especialização em Psicometria realizado na Espanha, participação de grupos de estudos, além da troca de ideias com psicometristas em diversos grupos de pesquisa. Nesse processo foi realizado uma familiarização preliminar com softwares apropriados para a realização das análises, como o BILOG-MG, R, Winsteps.
- *Quanto ao caráter reivindicatório,* tem-se: (1) a um alinhamento curricular fundamentado e elaborado tendo como base em evidências diante das dificuldades conceituais dos alunos, levando em consideração aspectos regionais e contextuais, além de destacar que essas dificuldades são necessidades legítimas dos alunos; (2) a uma maior ênfase nos procedimentos de análises psicométricas dentro dos Programas de Pós-Graduação em Educação em Ciências, contribuindo assim na elaboração de diagnósticos podendo ser útil em diversas perspectivas, sobretudo em abordagens de pesquisas mistas para enfrentamento dos fenômenos educacionais.
- *Quanto à adequação do suporte teórico da Psicologia Educacional:* As dificuldades conceituais se concentram na detecção e análise do conhecimento prévio. Com isso se dá ênfase dentro de uma concepção ausubeliana para as condições de contorno de enfrentamento. Os pressupostos da psicologia educacional *ausubeliana*¹¹ estabelecem critérios que atendem anseios da pesquisa, seja pela correspondência entre o conceito de *subsunção* e as CAs, como na forma de abordar os conhecimentos, no caso partindo de um conhecimento mais inclusivo e estruturante dentro da temática (Física Térmica) que são basicamente as *Leis da Termodinâmica* e o *Modelo Cinético-Molecular*. Para isso,

¹¹ David Ausubel.

é irremediável uma mobilização e desenvolvimento de conhecimentos específicos fundamentais para garantir uma progressão adequada da aprendizagem, que no caso, o *calor*, a *temperatura* e a *entropia* são os conceitos estruturadores dentro dessa concepção (Moreira & Masini, 2006, p. 64). Exploram-se possibilidades de análise psicométrica com base nos pressupostos da psicologia de Ausubel, bem como sugestões para análises futuras, ou de interesses diversos. Tem-se algumas conjecturas sobre o processo de interpretação dos itens pelos respondentes, na qual emergem alguns riscos e necessidades a serem valorizadas. Acredita-se que isso possibilite uma melhor compreensão sobre os processamentos mentais durante a realização da atividade em responder o teste, sob um enfoque ausubeliano. Acredita-se ainda que as *condições de contorno* da psicologia educacional para uma «aprendizagem significativa» se alinhem com as possibilidades em promover um melhor desenvolvimento de capacidades e competências científicas.

- *Quanto à adequação do suporte teórico da Didática das Ciências:* Baseia-se no esclarecimento sobre as habilidades e competências científicas, e de seu papel no cenário educacional, em especial, nas delimitações quando elaborados, propostos e avaliados com base em testes conceituais. Nesse âmbito se dá ênfase nas análises das situações-problema propostas no teste, e dos seus distratores. Acredita-se que o modelo pedagógico da *integração* ou de *desenvolvimento de competências* forneça subsídios que possibilitam interpretar e sugerir aspectos favoráveis a otimizar a metodologias de ensino e promover uma autonomia dos alunos no processo de sua aprendizagem.

- *Quanto ao interesse acadêmico e profissional com implicações favoráveis do ponto de vista econômico, pedagógico e social:* Acredita-se que as análises psicométricas baseadas em testes conceituais seja um caminho que permita vislumbrar a elaboração de um observatório de dificuldades conceituais, preliminarmente na Universidade Federal do Amazonas (UFAM), eventualmente em parceria internacional com a Universidade de Coimbra (UC), podendo expandir entre a educação básica e superior. Acredita-se ainda que esse observatório possa servir de plataforma colaborativa entre outros cursos, departamentos, e instituições, podendo ampliar em perspectivas de análises, incluindo outros aspectos relacionados as dificuldades de aprendizagem no campo

socioafetivo, relato docente sobre o desenvolvimento da aprendizagem em campos temáticos específicos, assim como na criação de repositórios de vídeos e *bancos de itens*, abrindo possibilidades para um desenvolvimento sistemático com a implantação de *Testes Adaptativos Informatizados*¹² (TAI) baseados na *Teoria da Resposta ao Item*. A ambição desse projeto visa ainda ser potencializado com implementações futuras na modalidade de educação a distância, na qual ainda é uma necessidade. No *Ensino Superior*, esse processo pode partir inicialmente da elaboração e adaptação de testes em vários ramos da Física, avaliando alunos ingressantes, intermediários e finalistas dentro de diversos cursos, e dessa forma fornecer subsídios para ajustes nos respectivos programas curriculares, práticas pedagógicas mais focadas em evidências de boas práticas e de necessidades, além de materiais instrucionais mais eficientes sendo construídos pelos próprios docentes, podendo contribuir ainda na formação docente de alunos de licenciatura, e se estender a contribuições no ensino básico.

Na *Educação Básica*, em parceria com a Secretaria de Educação, podem ser aplicados testes diagnósticos no intermédio entre os exames de avaliação externa (em nível estadual e nacional), contribuindo localmente com políticas regulatórias de alinhamento curricular baseadas em boas práticas de gestão e de evidências empíricas. Numa perspectiva estratégica, aliada a inovadoras práticas de ensino, os testes podem contribuir para um melhor desempenho no SADEAM¹³, e externamente para o SAEB¹⁴, e em nível superior para o ENADE¹⁵, com expectativas de refletir favoravelmente no IDEB e o cumprimento das metas estabelecidas pelo PNE 2014-2024 vigente. Particularmente no Brasil os exames nacionais são um dos maiores do mundo, muito criticado pelos enormes investimentos, influência do setor econômico e dos desafios logísticos envolvidos diante da restrições em avaliar competências em um macro sistema e das implicações deformadoras no currículo efetivamente realizado. Entretanto, defende-se que esses testes possam exercer uma função diagnóstica e regulatória, contribuindo na eficiência em diversas outras políticas públicas

¹² *Computer Adaptive Test* – CAT.

¹³ Sistema de Avaliação Educacional do Amazonas.

¹⁴ Sistema de Avaliação da Educação Básica.

¹⁵ Exame Nacional de Desempenho de Estudantes.

educacionais com base em boas práticas de políticas *accountability*, tornando justificável os elevados fatores econômicos envolvidos nesse âmbito.

- *Quanto a contribuição em outras pesquisas:* A construção da escala de proficiência de concepções térmicas e disponibilização *online* da base de dados, sintaxes, arquivos de calibração dos itens, etc., visam contribuir como subsídios para outras pesquisas através da construção de séries históricas de desempenho, estudos comparativos na performance entre diferentes grupos e populações, bem como avaliar a eficiência de propostas metodológicas de ensino. Esses testes podem ser ampliados para os diversos campos de atuação da Física, e aplicados em diversos níveis de escolarização ou períodos/semestres da formação no ensino superior. Um maior rigor científico-metodológico diante da elaboração de instrumentos como testes e questionários, podem contribuir em pesquisas que estão sendo desenvolvidas nos PPGs na área do EC, assim como na proposição do uso psicométrico em abordagens mistas e longitudinais, na qual se configura como uma lacuna nesse contexto investigativo no Brasil.

- *Quanto a contribuições em livros didáticos*¹⁶: É uma perspectiva complementar e futura diante as análises, em que a utilização de testes para a detecção de dificuldades conceituais possa fundamentar proposições de melhorias dentre seus encargos pedagógicos. Para isso, os conhecimentos relevantes com base nessas dificuldades devam servir de subsídios na elaboração de *situações-problemas* “alvo” (Roegiers, 2006) numa perspectiva intradisciplinar e integradora dentro dos *Livros Didáticos de Física* (LDFs). Nesse campo investigativo, acredita-se na eficiência da combinação de diversas perspectivas, como por exemplo, adotar um *eixo gerador temático* transversal para serem propostas situações-problemas significativas, e ainda ser promissor em abordagens interdisciplinares, ajudando na articulação e desenvolvimento das competências, mobilizando recursos e saberes, e sendo terreno fértil para as discussões e finalidades do ensino. Também se acredita na possibilidade de conexão com

¹⁶ Os termos “Livros Didáticos” e “Manuais Escolares” devem ser tratados como equivalentes, considerando que o primeiro é usual contexto brasileiro e o segundo em Portugal, entre outros países. Escolheu-se a terminologia de *Livro Didático* na intenção de evitar uma associação ou aproximação com o modelo pedagógico tecnicista, podendo o termo “manual” inclinar o leitor a uma interpretação de seu uso dentro de um ensino mecanicista, condutivista, nos moldes de uma concepção positivista reacionária.

pesquisas sobre os Livros Didáticos que *visam a sua qualidade científica-pedagógica*, seja no processo de concepção, avaliação, seleção e utilização. A elaboração de instrumentos avaliativos mais robustos podem ser desenvolvidos por meio de testes em larga escala de modo a caracterizar de forma representativa as necessidades dos respectivos destinatários, tendo em vista que esse processo geralmente possui um peso maior na avaliação de especialistas ou entidades avaliadoras certificadas. Sabe-se que o processo de avaliação e certificação de qualidade para os Livros Didáticos (ou manuais escolares) encontra-se em um “terreno de disputas”, porém, o interesse nesse caso estaria em discussões dentro de um enfoque reivindicatório em que se privilegiam os *saberes acadêmicos* (pesquisadores universitários) em detrimento dos *saberes profissionais* (professores) e das necessidades reais manifestadas pelos alunos. Convém lembrar que a atual reforma curricular em curso¹⁷ aponta que todos os materiais didáticos no Brasil assim como os cursos de licenciatura deverão passar por reestruturações, para garantir a abordagem de conteúdos essenciais e o desenvolvimento de competências específicas contempladas na *Base Nacional Comum Curricular* (BNCC), sendo essa uma das condições essenciais que interliga uma logicidade entre as contribuições complementares motivadoras e sugeridas.

- *Quanto à matriz epistemológica da investigação*: Envolve as concepções behaviorista na elaboração da matriz de habilidades dentro de uma APC. As APC são os paradigmas adotados nos sistemas educacionais do Brasil, de certa forma em Portugal. Defende-se que essa concepção visa atender não apenas interesses advindos de um modelo profissional e trabalhista, considerada uma questão controversa, mas que possui implicações ricas pedagogicamente e favoráveis para o desenvolvimento da autonomia e pensamento crítico, sobretudo atendendo as necessidades contemporâneas para o exercício da cidadania em um mundo moderno. As tendências internacionais (em especial, na Europa) buscam consolidar um *quadro de qualificações* baseado em competências gerais para o exercício do trabalho e compromisso no âmbito educacional. Nele são incorporadas contribuições da pedagogia por objetivos educacionais e do desenvolvimento de capacidades, contudo, se tem como desafio um

¹⁷ <<https://novaescola.org.br/conteudo/4784/32-respostas-sobre-a-base-nacional-comum-curricular>>.

equilíbrio entre um currículo generalista (baseado em capacidades) e operacional (baseado em competências). De uma forma ou de outra se busca um alinhamento curricular como condição *sine qua non* para a sua legitimidade no sistema de ensino em cada contexto. Com isso, torna-se motivador pela relevância do objeto de investigação nos diversos níveis estruturais do sistema, seja na mediação de ensino-aprendizagem, como nas múltiplas representatividades de interesse.

- *Quanto à integração em linhas de investigação na Didática das Ciências:* De acordo com a temática deste estudo, pode-se dizer que há um consenso de pesquisadores indicando uma existência de dificuldades na distinção entre os conceitos de *calor* e *temperatura* para os alunos neste nível escolar e ramo do conhecimento em vários contextos, sendo uma das causas dos prováveis obstáculos cognitivos para o processo de aprendizagem. Ao aplicar *testes conceituais*, analisar sua estrutura (enunciado, itens, contextos, traço latente...), e conceber como processo de avaliação de desempenho a TRI, considera-se que dentro da *Didática das Ciências* (DC) surgem incidências no cruzamento ou aproximação de três linhas de investigação: *Concepções Alternativas* ou *Concepções dos Alunos* no teste conceitual; *Resolução de Problemas* na análise e sugestões para a elaboração de situações-problema; e na *Avaliação educacional* com referência na comparação de desempenhos entre diferentes grupos.

CAPÍTULO 1 - PARÂMETROS ESTRUTURADORES

1.1 PROBLEMA

Há um esforço de mais de três décadas para promover mudanças e melhorias diante as dificuldades dos alunos para apreensão de conhecimentos em Termodinâmica (Chu, Treagust, & Yeo, 2012), dificuldades essas não apenas restritas a alunos de ensino médio, mas também na educação infantil, educação fundamental, ensino superior, e até entre professores e cientistas.

Apesar dessas dificuldades terem sido exaustivamente levantadas, sendo algumas típicas a serem identificadas em determinados grupos específicos com relação ao nível de escolaridade e faixas etárias, existem outros fatores que supostamente estão associados e que devem ser examinados com mais atenção, como as diferenças existentes nos contextos regionais internos e externos entre sistemas de educacionais, diferenciação entre programas curriculares com relação ao conteúdo térmico a ser desenvolvido, identificação de níveis relativamente estáveis para o grau de dificuldade diante de conceitos específicos, dificuldades comuns entre alunos com melhor performance, etc. Entende-se que esse tipo de mapeamento mais refinado possa propiciar subsídios relevantes e mais aprofundados diante os critérios na escolha de conteúdos específicos a serem desenvolvidos dentro da Termodinâmica, além de outras possíveis implicações pedagógicas.

Os diversos desafios a serem enfrentados para um melhor ensino da Termodinâmica passam obviamente por questões estruturais, como o *papel do professor* em sala de aula, o uso de *metodologias* de ensino mais eficientes, a promoção de *recursos instrucionais* com maior qualidade científico-pedagógica e o foco em *conteúdos* necessários enfatizando recortes criteriosos do ponto de vista epistemológico do conhecimento envolvido, mas que sobretudo **possam atender as dificuldades reais dos alunos**. A mediação docente para diagnosticar essas dificuldades e intervir de forma eficiente, torna-se uma tarefa fundamental para que se possa desenvolver um andamento favorável da aprendizagem nas salas de aula, e ainda pode ser mais

desafiador quando se exige ou se busca novos mecanismos de ensino, além do uso de recursos didáticos inovadores, diferenciados e/ou tecnológicos.

O *problema do conteúdo* a ser ensinado não se restringe a Física Térmica, obviamente, mas de modo geral remete a necessidade do exercício de competências docentes que se adequem a grupos sociais que possuem um acesso globalizado da informação, além das mudanças socioculturais nas diversas formas de comunicação que acompanham esse processo. É natural que diante disso hajam novos paradigmas pedagógicos a serem enfrentados, pois o problema atual nesse sentido já não é mais obter e nem repassar informações, mas sim: *Quais são as informações prioritárias? O que fazer com essas informações? Como abordar uma sequenciação de ensino que favoreça uma progressão na aprendizagem, contribuindo no desenvolvimento de um pensamento crítico e maior autonomia dos alunos, sobretudo, diante de questões socioambientais e econômicas, além da compreensibilidade diante aos fenômenos térmicos do cotidiano?*

Entende-se que para isso, o professor deva saber *priorizar os esforços do processo pedagógico em pontos cruciais e necessários*, e que sejam **pertinentes às condições reais dos seus alunos dentro de cada campo temático**, na qual possuem raízes epistemológicas específicas. Com isso, a preocupação investigativa volta-se para que hajam proposições didáticas que devam ser adequadamente situadas, enfatizando conteúdos que sejam relevantes e necessários a serem desenvolvidos. Nesse sentido, uma avaliação diagnóstica para identificar esses conteúdos torna-se o alvo investigativo, levando em consideração algumas delimitações quando se trata única e exclusivamente de um teste de sondagem como instrumento.

O conteúdo a ser ensinado deve levar em consideração que para o exercício da cidadania se torna necessário compreender o funcionamento da produção moderna e tecnológica. Nisso, sabe-se que o conhecimento térmico está presente em todos os contextos do cotidiano das pessoas, seja para compreender o funcionamento de máquinas, como prever e entender o clima, a temperatura dos corpos, de ambientes, etc., e isso remete a implicações para diversas áreas relacionadas ao campo social, da saúde, ambiental e tecnológico. O conhecimento térmico é importante por ser um ramo

fundamental na estrutura da Física Clássica, e com isso, precisa ser incorporado na formação básica inicial de todos, e em todos os níveis como requisito cultural/científico e tecnológico ao longo de suas vidas, e também para os estudos futuros daqueles que vão prosseguir em uma formação e carreira na área específica ou afins. Conceituações abstratas e universais de impacto ímpar e alto valor à humanidade como energia e entropia - invariância, transformações e degradação energética, estão em jogo na sala de aula e em demais cenários e escalas, com isso, envolvem uma necessidade de compreensibilidade irremediável diante da relação dos fenômenos naturais e questões humanas.

Diante desse contexto, o estudo investe na possibilidade de extrair contribuições através de uma análise diagnóstica sobre as dificuldades dos alunos no conhecimento introdutório da Física Térmica a partir de análises psicométricas, e define como *objeto de investigação* o conhecimento envolvido no teste aplicado, bem como nas diferentes percepções dos alunos diante das situações-problemas que o constitui. Nesse processo, há de forma implícita uma preocupação em justificar o bom uso de testes em larga escala na área da educação, em especial, dentro da pesquisa no EC, explorando e propondo diversas perspectivas e possibilidades. De forma mais explícita e concreta, o estudo se preocupa em identificar evidências relativamente consistentes que possam emergir de um teste conceitual sobre as percepções de grupos e subgrupos normativos quanto aos conceitos/concepções térmicas dentro das amostras de Portugal e Brasil, enfatizando as concepções existentes mais destacadas diante os conceitos básicos em unidades temáticas, além de algumas proposições sobre o seu papel no processo da progressão da aprendizagem.

Explora-se o papel de *comparar* e de classificar não com propósito de competir e afirmar quem está melhor ou pior, mas de identificar um estado de progresso e desenvolvimento dando destaque as necessidades de melhoria para cada contexto. O propósito de instruir se entende como algo que possa transitar no limiar entre o efêmero/circunstancial com o substancial/essencial, devido às mudanças históricas dos sujeitos, das condições de trabalho, sem descartar a possibilidade de que algumas dificuldades sejam válidas independentemente do contexto regional e histórico. Já a

função de *avaliar* como um processo mais longo a ser dado, se restringindo mais em medir, porém não significa uma limitação em extrair riquezas e inferências, sobretudo na sua contribuição, ainda que delimitada em avaliar e favorecer o processo desenvolvimento da capacidades e competências científicas de alunos da educação básica. E é preciso entender que o uso de testes em larga escala através dos exames nacionais não irão desaparecer a médio prazo, ainda que exista resistência, pois as políticas públicas sinalizam o fortalecimento dessas avaliações sendo preciso que se invista na compreensão sobre a melhoria desse processo, sem que haja prejuízo aos alunos e ao sistema como um todo, buscando melhores formas na elaboração de itens, identificando suas contribuições desde a gestão até a prática educativa, aprofundando a compreensão de toda a Ciência que envolve essa prática. Diante disso, tem-se como problema a seguinte questão:

“Quais as «diferenças e similaridades entre as concepções térmicas de alunos» do último ciclo da «educação básica», a partir da avaliação de seus desempenhos em um «teste conceitual diagnóstico»... (Situando o objetivo e objeto de análise – função primária da investigação)

... que envolve um processo de «construção de uma escala de proficiências» com base em «análises psicométricas», (Produto central da investigação e processos que articulam aquilo que se pretende com o objeto – função secundária da investigação)

... visando com isso apresentar «informações comparativas e representativas de grupos», que sejam favoráveis à «aquisição significativa de conceitos específicos» e ao desenvolvimento de «competências básicas»... (Especificidade na fundamentação de análises complementares)

... entre os contextos de «Portugal e do Brasil»? (Locais de referência)

O processo de elaboração da escala visa não apenas disponibilizá-la¹⁸ a literatura para comparações diretas em estudos correlatos, mas extrair subsídios que possibilitem articular as análises para que se alcancem os objetivos e manipulem o

¹⁸ Será disponibilizado as sintaxes e base de dados *online*, sob controle de conceção e responsabilidade.

objeto de análise. As interpretações e comparações possuem um caráter pedagógico que advêm do processo e produtos subjacentes da análise psicométrica, caracterizando assim um estudo exploratório que busca apresentar subsídios relativamente consistentes para uma investigação complementar, mais ampla e aprofundada, entre outros possíveis interesses. A partir da interpretação das medidas e do conteúdo nos itens, se pretende constatar evidências que justifiquem um entendimento sobre a preexistência dos subsunçores disponíveis nos alunos, em consonância com a concepção ausubeliana. Diante das evidências, análises e avaliações interpretativas, pressupõe-se qual seria o conhecimento potencialmente significativo necessário para que seja associado ao subsunçor e assim favorecer a ocorrência de uma aprendizagem significativa (condição para uma aquisição significativa de conceitos). A ideia de desenvolvimento de competências está associada na mobilização desses conhecimentos dentro de um contexto ou conjunto de situações familiares a um mesmo problema, podendo estar presente no teste, ou além dele. Dessa forma, a **identificação das diferentes concepções entre alunos** é construída a partir de interpretações diante de análises que possibilitem uma:

- *Comparação das concepções com base nos parâmetros psicométricos dos itens – visando identificar e comparar diferentes grupos de itens quanto aos **índices de dificuldades e discriminação** para cada contexto, relacionando com as concepções térmicas envolvidas;*
- *Comparação das concepções com base no desempenho de diferentes grupos na escala de proficiências – visando as **comparar níveis de proficiências e concepções envolvidas** entre diferentes grupos normativos da amostra com base no desempenho;*
- *Comparação das concepções com base nos conteúdos necessários para diferentes grupos de desempenhos na escala de proficiências – visando destacar a **necessidade na aquisição de determinados conteúdos e habilidades específicas** dentro do desenvolvimento de competências científicas.*

De uma maneira geral, podem existir influências sobre as dificuldades conceituais relacionadas ao próprio instrumento. Diante disso, considera-se necessário identificar aspectos internos nos itens que possam estar influenciando as evidências

empíricas, e como consequência, se tornar um risco na avaliação das diferentes concepções térmicas encontradas entre os dois contextos. Para que haja um controle de variáveis moderadoras sobre essas questões foi definido uma análise focada nas:

- *Interpretações pedagógicas das habilidades e competências que relacionam o conteúdo dos itens* – visando uma análise interpretativa da natureza taxonômica do conhecimento envolvido com base na taxonomia de Bloom revisada em 2001, e assim elaborar uma **matriz de habilidades e competências para o teste**.
- *Interpretações pedagógicas das concepções térmicas que relacionam o conteúdo dos itens* – visando verificar e confirmar uma **coerência no Inventário de Conceitos (IC)** ou concepções térmicas sugerida pelos autores originais, que acompanha o TCE, e ampliar o seu nível de precisão e detalhamento, bem como identificar indícios de relações dependência ou de influências nas respostas entre os diferentes itens no teste.

A aquisição significativa de conceitos remete a valorizar os fundamentos da psicologia educacional que tratam o conhecimento prévio, para isso, foi adotado como sustentação teórica a concepção de David P. Ausubel visando contribuir na compreensão sobre as dificuldades conceituais e de **implicações favoráveis para que ocorra uma aprendizagem significativa**. Quanto ao desenvolvimento de competências se refere aquilo que o teste visa aferir entre os itens, considerando-se relevante a contribuição alguns fundamentos da Didática das Ciências para uma **melhor apreensão da natureza tipológica da competência envolvida (em desenvolvimento)**, e de contribuições na interpretação pedagógica da escala de proficiências. Diante das análises e evidências empíricas, apresentam-se algumas implicações complementares relacionadas a propostas instrucionais que possam contribuir no processo de aquisição significativa de conceitos térmicos e no desenvolvimento de competências científicas para o conhecimento térmico. Para isso, consideram-se que essas análises complementares possam apresentar:

- *Implicações pedagógicas favoráveis a ocorrência de uma aquisição significativa de conceitos* – visando identificar subsunçores e conhecimentos potencialmente significativos, e assim sugerir atividades e processos que possam **favorecer a ocorrência de uma aprendizagem significativa situada nas evidências**;

- *Implicações pedagógicas favoráveis ao desenvolvimento de competências científicas básicas em Termodinâmica* – visando elaborar **propostas pedagógicas que envolvem famílias de situações-problemas** como encargo pedagógico articulador para o desenvolvimento de competências científicas, de acordo com os subsunçores e conhecimentos potencialmente significativos considerados. Esse seria um aspecto especulativo para projetar uma análise complementar mais aprofundada.

O alvo da pesquisa foca em inferências comuns e generalizadas sobre as dificuldades e necessidades entre os dois contextos. Para isso, acredita-se que o conjunto de análises psicométricas diante das evidências empíricas possam produzir informações relativamente consistentes sobre um construto predominante explicativo. Nesse sentido se busca identificar uma justaposição analítica das informações e comparações realizadas.

A *normatização da amostra* se refere a **alunos do último ciclo da educação básica** em cada contexto, buscando assim atender ao critério de uma escala diversificada e bem distribuída, esperando que se tenha itens fáceis, medianos e difíceis. Apesar disso, serão realizadas análises com subamostras dentro de cada contexto visando comparar desempenhos e proficiências, e assim emergirem informações mais detalhadas sobre a escala de proficiências, respeitando as peculiaridades dos diferentes grupos de respondentes que compõe a base de dados dentro de cada contexto.

Através das análises preliminares da *Teoria Clássica dos Testes (TCT)*, buscam-se informações iniciais quanto aos índices de discriminação e dificuldade de grupos e da amostra em geral, bem como *análises gráficas* percentuais de grupos de respondentes para cada item. Essas informações servirão de base interpretativa na comparação com os resultados da *Teoria de Resposta ao Item (TRI)*. As informações mais consistentes estão na TRI, e se referem inicialmente a construção de uma «*escala de proficiência sobre as concepções térmicas*», seguido da análise da *Funcionamento Diferencial dos Itens (DIF)* quanto as *dificuldades entre os contextos*, e de interpretações pedagógicas advindas das análises das *Curvas Características dos Itens (CCIs)* e das *Curvas de Informação dos Itens (CCIs)*, bem como do teste em geral.

O *objeto de estudo* volta-se para a análise do instrumento, nomeadamente quanto ao conteúdo e as respectivas concepções detectadas diante de conceitos básicos da Física Térmica, nisso se tem como **critério avaliativo de relevância** as «*dificuldades conceituais*» individuais e de diferentes grupos de alunos do ensino médio (secundário) nos dois contextos.

Apesar do caráter comparativo, é proposicional uma atenuação para o contexto brasileiro devido a baixos indicadores da qualidade do sistema educacional diante de Portugal, com isso, adota-se predominantemente o contexto português como parâmetro de referência. Considera-se relevante ao estudo que a *escalas de proficiência* elaborada atenda suficientemente aos critérios de validade para ambos os contextos, possibilitando assim estudos posteriores comparativos de necessidades, construção e análises de séries históricas de desempenho, além da identificação de aspectos comuns e diferenciados de itens nos dois contextos. Essa perspectiva de análise permite **expressar propriedades dos itens em termos de parâmetros invariantes a respeito dos sujeitos e da natureza do item**, tornando disponível para que psicólogos e pesquisadores educacionais interessados nesse âmbito investigativo possam escolher o tipo de item mais adequado para os seus objetivos de investigação (Muñiz, 1997, p. 140). Propriamente, a relevância estabelecida no «conteúdo científico» se refere às *dificuldades conceituais* da população a partir de uma amostra relativamente significativa nesse âmbito.

Entre algumas perspectivas de análises complementares a esse estudo de forma aplicada e mais concreta, vislumbra-se proposições de situações-problemas em Livros Didáticos de Física (LDFs), ainda que não haja um compromisso investigativo mais aprofundado por estar além do escopo desse estudo. Para isso, tem-se um olhar na contribuição da função de *Desenvolvimento de Capacidades e Competências* (DCC) desse recurso instrucional e didático, e dentro disso, uma análise das «situações-problemas» pertinentes ao estudo, e de proposições a esse encargo pedagógico de modo que valorizem dificuldades conceituais relativamente estáveis para o conhecimento térmico de determinados grupos, e adaptações para instruções individuais, de acordo com os desempenhos e das características desses grupos (gênero, idade, contexto

regional, nível de escolaridade, nível socioeconômico, etc.). É possível que esses resultados possam contribuir na adequação e ajustamento de encargos pedagógicos presentes nas diversas «situações didáticas» realizadas no cenário pedagógico. A intenção estaria em potencializar os impactos sociais pela disponibilização de evidências empíricas, e pela proposição do uso adequado de testes conceituais como instrumento a auxiliar o professor em intervenções metodológicas diferenciadas e promissoras em salas de aula.

Apresenta-se ainda uma análise complementar original e em desenvolvimento que acrescenta algumas informações sobre a natureza do instrumento, e conjecturas sobre uma associação de concepções interligadas que podem eventualmente influenciar os respondentes. Essa é uma análise especulativa e anterior a verificação das evidências empíricas, e baseia-se na interpretação do enunciado e das alternativas da *situação-problema* que compõem cada item, que nesse caso são os objetos de estudo nessa análise qualitativa preliminar. Busca-se assim que esses objetos possam ser explorados para melhor compreender a natureza do problema, o contexto e a situação em que a habilidade cognitiva a ser aferida deve se articular, nisso se busca também identificar relações de **supostas dependências entre os itens**, assim como a **coerência existente na sequência das escolhas** subsequentes e de sua **consistência interna**, de acordo com critério de verdade, plausibilidade e rigor-científico.

1.1.2 Especificando o Problema

A pertinência no estudo de modo geral se situa dentro de uma problemática mais ampla, se relacionando com questões sobre o processo de *avaliação de desempenhos* e de *investigação sobre a eficiência* de um teste conceitual como instrumento diagnóstico que envolve o conhecimento térmico. Como a perspectiva de análise é de larga escala, ela se assemelha às avaliações educacionais externas na qual as preocupações voltam-se para a necessidade de qualidade e eficiência que buscam atender o paradigma de um «alinhamento curricular». Nesse campo existem diversos problemas complexos a serem enfrentados, tipo: *Quais são as contribuições efetivas (na prática) que os testes em larga escala podem trazer para um sistema educacional? Eles contribuem de que alguma forma?*

Qual o papel ou funcionalidade que os testes devem exercer sem que haja distorções ou reducionismos no currículo? Como administrar adequadamente os testes para que isso ocorra? Deve-se abandonar completamente o uso de testes em larga escala para avaliar a educação? Eles avaliam apropriadamente de fato algum tipo de competência ou apenas parte do processo de desenvolvimento de competências? Os testes estão contribuindo no mapeamento de competências científicas relevantes para o EC? Qual(is) tipo(s) de competência(s)? Quais deveriam ser? De que forma as informações psicométricas podem contribuir para a tomada de decisões na gestão das escolas? Podem contribuir em microfenômenos na sala de aula? Pode auxiliar o ensino? Pode estar à serviço da aprendizagem? De que forma?

No âmbito das pesquisas ou investigações na área da educação científica, se tem: *Como eles podem contribuir para as pesquisas no EC? De que forma eles estão sendo usados pelos pesquisadores? Em caso contrário, por que estão sendo rejeitados? Está sendo feito algum esforço de compartilhamento entre os estudos psicométricos de neurocientistas e psicólogos experimentais (focam em processos mentais) com os psicólogos educacionais (focam no conhecimento)? Quais poderiam ser as contribuições em uma perspectiva investigativa ampla nesse sentido? Os pesquisadores e especialistas na área do EC estão interessados no domínio desses procedimentos? É possível que os procedimentos de análises advindos da Psicometria possam contribuir no aprofundamento da compreensão sobre o conhecimento prévio e de novas implicações pedagógicas investigativas?*

Em finalidade última esses estudos devem buscar uma compreensão e contribuições sobre o *como*, o *que* e o *porquê* o aluno deve aprender e o professor deve ensinar. Para isso, o estudo busca identificar o conteúdo da Física Térmica e a relação da concepção de competência científica que está sendo proposto/exigido/legitimado dentro de cada sistema, de acordo os documentos oficiais e exames nacionais.

Para uma pertinência nas discussões, as dificuldades conceituais são estabelecidas como critérios de relevância ao conteúdo a ser priorizado, implicando necessidades tanto no campo formativo e profissional do ensino quanto a legitimação do currículo. Nessa perspectiva, acredita-se que o estudo possa fornecer indícios e atender às necessidades de ajuste levando em consideração aspectos relativamente

estáveis sobre as dificuldades na aprendizagem, além de favorecer uma legitimidade do currículo e eficiência do ensino.

Uma compreensão diante do que seja fácil ou difícil para os alunos são apenas conjecturas lógicas sobre o conhecimento prévio identificado e sugerido nos índices de dificuldades dos itens, bem como na interpretação sobre o nível de compreensibilidade que se exige nas situações-problema dos itens.

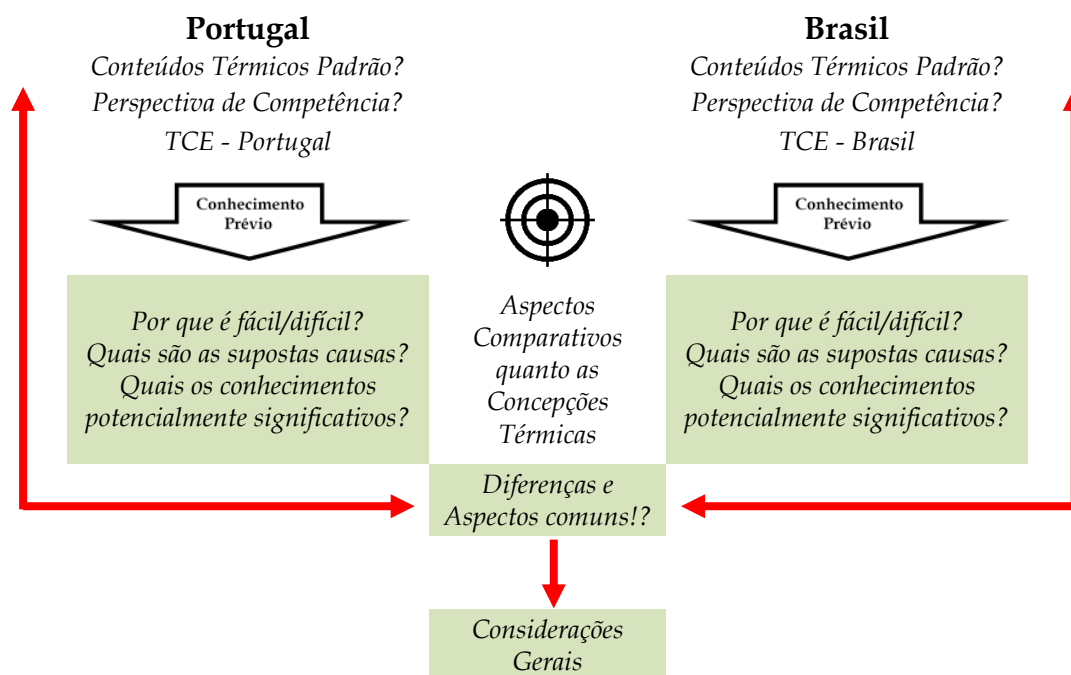


Figura 1: Perspectiva estrutural da investigação.

Na aplicação de testes, os alunos são solicitados a manifestarem os traços latentes do conhecimento disponível. As respostas, quando respondidas conscientemente são reflexos desse conhecimento ou de concepções existentes, com isso, essas respostas se caracterizam naturalmente como evidências empíricas, e uma interpretação coerente e bem fundamentada deve remeter a aproximações da realidade. Nesse sentido, os modelos logísticos da TRI representam o estado atual da medida psicológica, com isso, buscam atender as necessidades de uma interpretação aproximada da realidade sobre os processos mentais desenvolvidos e conhecimentos adquiridos, dessa forma, esse foi um dos critérios escolhidos para se obter informações mais precisas e uma melhor interpretação sobre o grau de dificuldade encontrado pelos alunos nos dois contextos.

Espera-se que os aspectos comparativos possam sinalizar algumas necessidades do ponto de vista das práticas escolares realizadas em cada sistema, onde as evidências positivas estariam a sugerir uma eficiência do sistema e boas práticas de ensino desenvolvidas. Ao serem comparados aspectos comuns entre os contextos, podem-se identificar parâmetros de itens relativamente estáveis representando informações psicométricas que caracterizam uma mesma funcionalidade. Apesar da necessidade de estudos longitudinais para uma maior consistência das conclusões obtidas, muitos resultados já foram obtidos com o uso teste TCE, podendo assim ser atribuído um aspecto de credibilidade para uma maior precisão nas análises e das conclusões.

Um mapeamento com conteúdo e habilidades enfatizados dentro de cada contexto é um aspecto necessário para situar a pesquisa, e com isso verificar sua relevância dentro de cada contexto. Procura-se com isso identificar de forma mais ampla o paradigma curricular que fundamentam as competências, e de que forma as necessidades de aprendizagem apontadas no estudo estão sendo atendidas dentro dos programas, assim como produzir argumentos que possam ser úteis numa reivindicação para ajustes de objetivos de ensino e aprendizagem.

Para justificar as contribuições do estudo, segue uma exposição de motivos para o condicionamento de aceitação do problema científico na investigação.

Quadro 2: Exposição de motivos para o condicionamento de aceitação do problema científico.

O problema é relevante?	<p>A relevância está em vários pontos:</p> <ul style="list-style-type: none"> • Nunca foi realizada uma investigação nessa perspectiva internacional entre os dois contextos, apesar de existirem diversos tratados internacionais de cooperação científico-cultural e educacional e com elevada mobilidade de estudantes em todos os níveis de escolaridade entre os países; • A perspectiva comparativa do estudo poderá fornecer indicativos de melhoria em ambos os contextos, assim como na identificação de aspectos comuns entre os contextos; • Não há uma escala de proficiência de concepções térmicas que possa ser utilizada na comparação de desempenho entre alunos brasileiros e portugueses, ou qualquer outro instrumento similar na literatura; • O uso da escala construída, ou da elaboração de outras similares utilizando os mesmos procedimentos, poderá contribuir na comparação de desempenho de escolas e distritos ao longo dos anos e com isso exercer uma função avaliativa reguladora em diversas intervenções de ensino assim como em tomadas de decisão na gestão escolar; • A identificação de <i>Funcionamento Diferencial do Item</i> (DIF) quanto ao gênero poderá contribuir em estudos comparativos e na tomada de decisão sobre aspectos pedagógicos em estudos reivindicatórios/emancipatórios;
--------------------------------	--

	<ul style="list-style-type: none"> • As evidências poderão servir de base para pesquisas complementares que busquem fundamentar uma avaliação educacional mais ampla sobre as dificuldades conceituais de alunos em conceitos introdutórios e suas respectivas concepções diante do conhecimento técnico; • Os indícios contribuem na consistência em outros estudos, bem como em práticas com base em evidências empíricas; • Os procedimentos de análise adotadas no estudo são escassas nas metodologias de pesquisa na área da educação em ciências, caracterizando assim, uma obra de referência a estudos afins que visem valorizar e preencher essa “lacuna” epistemológica investigativa, bem como difundir essa perspectiva metodológica; • Evidenciar necessidades legítimas de um alinhamento curricular com implicações para o ensino e reelaboração de materiais instrucionais;
<p>O problema conduz a novos problemas?</p>	<p>A investigação fornece subsídios com fortes indícios de validade e significância que podem ser úteis:</p> <ul style="list-style-type: none"> • Na melhoria de materiais instrucionais como os livros didáticos (ou manuais escolares) de física preocupados com uma abordagem que valorize as dificuldades reais de alunos em termodinâmica; • Em ajustes no <i>alinhamento curricular</i> entre os conteúdos no programa curricular e prática docente; • Na definição de padrões de desempenho em exames nacionais dentro da temática da termodinâmica; • Na elaboração de objetos educacionais laboratoriais e virtuais e voltados para as dificuldades típicas e específicas dos alunos; • Em análises taxonômicas dos objetivos educacionais apresentados em cada contexto; • Em estudos comparativos entre os mesmos aspectos considerando amostras de escolas da rede privada; • Em estudos exploratórios com a base de fundamentação teórica e investigativa em outros campos do conhecimento do programa curricular do ensino secundário (médio) e/ou do ensino superior; • Na função de diagnóstica de dificuldades conceituais em termodinâmica comparadas a outras temáticas da física.
<p>O problema é investigável?</p>	<p>O problema é investigável pelo fato de que:</p> <ul style="list-style-type: none"> • O objeto da investigação se concentra na natureza textual de um instrumento previamente elaborado, com isso é de fácil acesso para as implicações e manipulação das variáveis; • O instrumento foi elaborado por especialistas na área educacional, e recomendada a sua ampla utilização, além de ser considerada válida pela <i>American Association of Physics Teachers (AAPT)</i>; • A língua nativa dos dois contextos é a mesma, possuindo assim forte aproximação de correspondência semântica, estando a tradução do instrumento auxiliadas por orientadores experientes, minimizando os riscos; • O instrumento foi testado avaliado continuamente em diversas épocas e contextos, sendo obtidos resultados diversos dentro de amostras normativas correspondentes, possibilitando assim investigar, comparar e dialogar com outros pesquisadores e resultados obtidos; • Os documentos oficiais dentro de cada contexto estão disponíveis de fácil acesso; • O conhecimento prévio e as concepções alternativas relacionadas a Física Térmica já foi amplamente discutida, e existem diversos suportes teóricos dentro da psicologia educacional e da didática das ciências que dão todas condições necessárias de análise, reflexão e tomada de decisão; • Foi viabilizada uma amostra representativa dentro do recorte em cada contexto alcançando cerca de 20% da população, assim como a computação e a realização das análises com auxílio de <i>softwares</i> adequados;

	<ul style="list-style-type: none"> • As informações coletadas foram viáveis de serem obtidas por estarem devidamente regularizados pelos órgãos competentes, obedecendo normas éticas de livre consentimento e proteção de dados.
<p>O problema é adequado ao investigador?</p>	<p>Considera-se adequado pelo fato do investigador:</p> <ul style="list-style-type: none"> • Ser licenciado em física, professor desde 1999, especialista em gestão escolar, mestre no ensino de ciências aplicados ao ensino da física do ensino secundário (médio); • Ter sido responsável direto pela reelaboração curricular na componente de física pela secretaria de estado de educação e qualidade de ensino do Amazonas (Estado Federativo Brasileiro) em 2011; • Com cursos extras em metodologia qualitativa e quantitativa somando 804 h de atividades durante o doutoramento, ressaltando estudos introdutórios da <i>Teoria de Resposta ao Item</i> no âmbito de cotutela com a UFSC, além de curso de 300 h de especialização pela instituição certificada e reconhecida na Europa denominada Euroinnova na Espanha, bem como diálogos e participação de grupos de estudo com psicometristas experientes existentes nesses contextos; • Considerando que o problema no estudo não necessita de explorar recursos avançados da Psicometria moderna, considera-se ser um especialista no ensino da física estando ao alcance dos objetivos desta investigação de doutoramento.

1.2 TIPO DE PESQUISA

Uma investigação pode ser definida “*como um processo sistemático e intencionalmente orientado e ajustado tendo em vista inovar ou aumentar o conhecimento num dado domínio*” (De Ketele & Roegiers, 1993). Nisso, um tipo de pesquisa realizada exclusivamente em um *processo de avaliação de desempenhos* fica à margem de um contexto de investigação (Ibid, 1993, p. 105), pois, supostamente representa apenas uma aplicação de uma informação já concebida para o seu efeito. O estudo incide sobre um processo de avaliação de desempenho, porém com implicações exploratórias diante do instrumento, das análises das respostas e suas interpretações, buscando emergir novas hipóteses tanto sobre o processo de construção e elaboração de itens e testes conceituais como das contribuições em práticas educacionais com base nessas evidências. Com isso, a pesquisa realizada pode ser denominada como sendo uma **Investigação Científica Exploratória**, focando, entretanto, em uma *avaliação de desempenhos*.

Assim, **investigador** e **avaliador** podem fazer incidir a sua atenção no mesmo objeto (a melhoria do desempenho ou do conceito de si, por exemplo: *estes exemplos mostram bem que não podemos escapar à questão do valor*). [...] São dois papéis diferentes, mas que não podem ser considerados como contraditórios. [...] permiti-nos dizer, como Worthen (1968), que se é legítimo e até desejável distinguir bem investigação e avaliação como dois processos de investigação diferentes, **existem, no entanto,**

muito mais similitudes do que diferenças entre investigação e avaliação, se nos colocarmos ao nível das escolhas das técnicas de coleta de informação. (De Ketele & Roegiers, 1993, p. 126, *grifo nosso*)

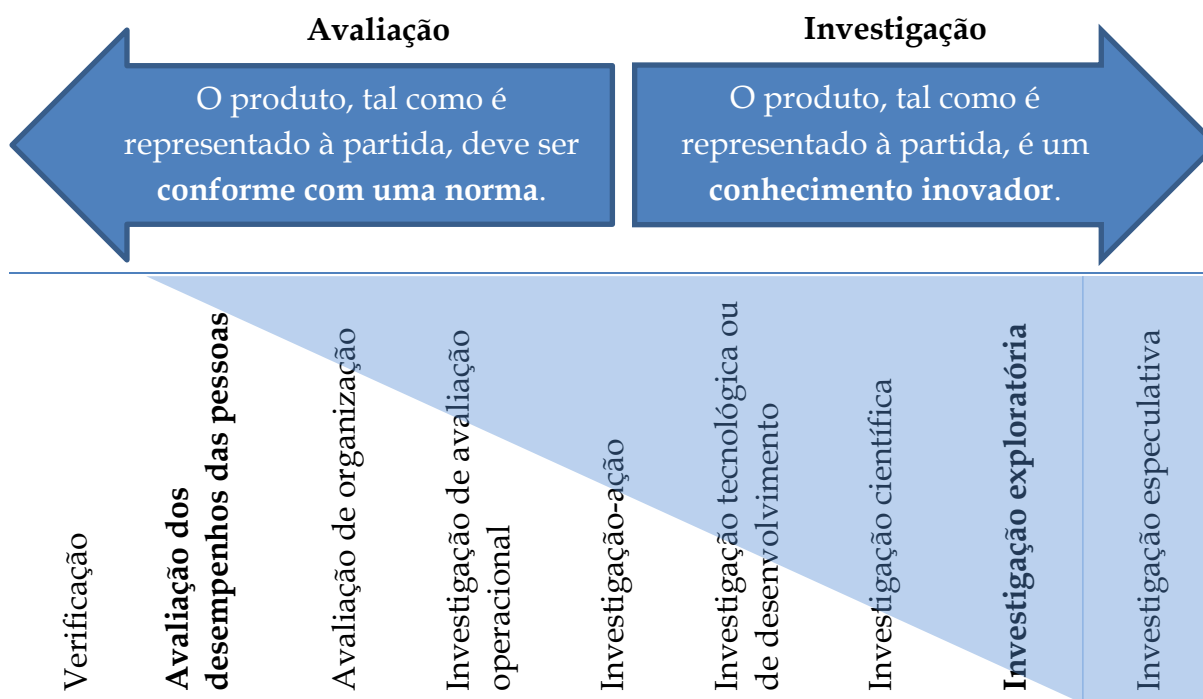
Na *fase heurística* são feitas observações (aplicação do teste TCE) e análises preliminares em um teste-piloto sendo aplicado em parte de uma amostra da população-alvo, acarretando reflexões, uma análise mais ampla dos dois contextos, e por fim, buscando gerar hipóteses (análise crítica, reflexiva, inferencial e interpretativa sobre as evidências empíricas) em uma perspectiva comparativa no desempenho de alunos de Portugal e do Brasil.

A investigação remete diversas preocupações com as informações contidas nos itens, como: *Qual a relação ao tipo de conhecimento que ele exige ao respondente? Existem equívocos no texto que pode influenciar negativamente o respondente? Há outras concepções alternativas associadas nos itens que não foram declaradas no inventário de conceitos? Qual é o tipo de habilidade que se exige? Qual é o subsunçor relacionado ao traço latente que se pretende medir (e avaliar!)? Há alguma similaridade entre as dimensões taxonômicas dos conhecimentos exigidos entre diferentes itens? Existem fatores a influenciar uma resposta dada por item em relação a outro, ou seja, existe de fato independência local entre os itens? As dificuldades ou concepções equivocadas são comuns independentemente de populações? Essas dificuldades podem estar relativamente estáveis em níveis ou patamares de complexidade distintas entre elas?*

Diante dessa perspectiva da investigação implica dizer que um sistema de conhecimentos internos pertinentes ao processo de avaliação de desempenhos não podem estar desassociados. Nesse sentido, pode-se considerar que existem duas perspectivas de análise mais amplas no estudo, sendo intrínsecas e relacionáveis, na qual se defende que podem coexistir, combinarem e explorar níveis de complexidade para uma melhor compreensão das contribuições advindas dos procedimentos de análise psicométrica.

Quadro 3: Comparação entre os aspectos de pesquisa que se configuram como avaliação e investigação, incluindo o indicativo nas ênfases das tipologias de pesquisa quanto ao produto, destacando (em negrito) a perspectiva do atual estudo.

Avaliação	Investigação
O avaliador está implicado em medir o desempenho e comparar os resultados.	O investigador esta intrigado em avaliar a natureza do instrumento e possíveis influencias nos respondentes.
Necessariamente orientada para a decisão sobre a qualidade dos itens, da seleção entre os respondentes de baixa e elevada proficiência, necessidades na aprendizagem; Idiográfica (tendência de descrição do particular)	Busca-se uma orientação para a emissão de conclusões em uma perspectiva diagnóstica e fundamentada na psicologia educacional; Nomotética (visa a formulação de leis).
O “porquê” e o “como” não são indispensáveis, mas sim os resultados de performance e classificação de diferentes grupos normativos de desempenho	A preocupação do “porquê” e do “como” é primordial, em especial, diante das dificuldades identificadas e relações entre elas.
A utilidade é primordial, pois o foco está na generabilidade dos resultados e construção de uma escala que possa estar disponível em outros estudos	A utilidade social é secundária, pois o foco está em investigar e compreender as microcomplexidades e supostas causas.
Os limites são estabelecidos na perspectiva de guiar aquele que deve tomar a decisão, ou seja, de identificar necessidades quanto ao currículo.	Os limites estabelecidos na perspectiva de alargar um universo de conhecimentos, ou seja, ampliar o olhar sobre as particularidades.
As referências a um sistema de valores são primordiais, pois as normas (padrões) preconizadas no contexto são importantes	As referências a um sistema de valores são secundárias, pois as normas (padrões) preconizadas no contexto não são importantes



Fonte: Baseado em De Ketele & Roegiers (1993, p. 127).

Nesse tipo de investigação, se tem como *“finalidade principal não ver o que se passa, o que é verdadeiro, provar alguma coisa, mas ver o que poderia passar-se, o que poderia ser*

verdadeiro” (Patry, 1981, p. 39, citado por De Ketele & Roegiers, 1993). Nesse sentido não são considerados de forma conclusiva fatores de causalidade, mas sim conjecturas diagnósticas fundamentadas tendo como base evidências empíricas, com isso, são exploradas algumas técnicas de análise sobre a natureza do conhecimento envolvido, e estabelecendo planos experimentais relativamente rigorosos quanto a validade, fiabilidade, pertinência, entre outros critérios de credibilidade para assim emergirem hipóteses. Segue-se uma lógica indutiva nas análises comportando microprocessos de dedução e verificação. A *indução*, no caso, se concentra em análises qualitativas e interpretativas durante todo o processo, partindo da análise do instrumento, comparações dos resultados obtidos (nesse estudo e de outros) e das conclusões, enquanto os processos de *dedução* e *verificação* se concentram mais na calibração de itens, geração de índices e parâmetros psicométricos. Vislumbra-se com o estudo exploratório uma complementariedade de novos estudos de aprofundamento e inovação quanto as aplicabilidades no cenário pedagógico (intervenções de ensino mais focalizadas, melhoria na qualidade de recursos, etc.) e na metodologia da pesquisa (uso da Psicometria em abordagens mistas, explorar e associar conjecturas dos processos mentais com as formas de aquisição de conhecimento e entendimento sobre o conhecimento prévio, etc.).

Como critérios demarcatórios da *investigação exploratória*, se tem como referencial existente à partida, uma fundamentação teórica diante das concepções dos alunos com base na Didática das Ciências, tendo como auxílio na compreensão do conhecimento prévio e como condições de contorno implicações relacionadas a psicologia educacional de David. P. Ausubel, buscando acompanhar e ser adaptado durante o percurso do estudo. A Psicometria também é um procedimento de análise estabelecido *a priori* para buscar alcançar contribuições pedagógicas a partir do desempenho dos alunos entre os contextos, tendo como valor prioritário as dificuldades conceituais de acordo com as concepções e equívocos associados, e uma intenção de generabilidade para as populações delimitadas.

1.2.1 Delimitações e Concepções Filosóficas Subjacentes

Na delimitação do estudo realiza-se um conjunto de recortes na complexidade do objeto investigativo, na qual se considera uma triangulação entre diferentes perspectivas e abordagens, em especial, da influência comportamentalista na estrutura instrucional do instrumento com relação as expectativas avaliativas de predição e diagnóstico, como das análises mais fundamentadas na psicologia cognitiva diante do conhecimento preexistente entre os alunos examinados. De modo geral, considera-se uma coerência pragmática no estudo, e em linhas gerais delimita-se:

1. Quanto ao **conteúdo** relacionado ao objeto investigativo (conceitos e concepções térmicas): *concepções alternativas dentro da Física Térmica relacionadas aos conceitos de calor, temperatura, propriedades térmicas e processos termodinâmicos, sendo alguns dos conceitos fundamentais para o corpo de conhecimento mais estruturado da Termodinâmica (ou Termologia), assim como nas influências das concepções dos alunos diante do processo de aquisição de conceitos científicos.*
2. Quanto a **abordagem metodológica** investigativa: *caracterizada como pesquisa mista com ênfase quantitativa no método de coleta e procedimentos de produção e análise dos dados e qualitativa em análises interpretativas pedagógicas preliminares e posteriores sobre um mesmo instrumento¹⁹.*
3. Quanto a **fundamentação teórica**: (1) *Teoria da Aquisição e Retenção Significativa de Conceitos de David Ausubel – fundamentando a identificação de subsunçores, proposição de conhecimentos potencialmente significativos e de sugestões metodológicas de ensino (organizadores avançados para diferentes unidades temáticas, organização hierárquica de conteúdo, mecanismos de retenção de conhecimento, sequenciação de ensino e tarefas lógicas favoráveis aos processos de diferenciação progressiva e reconciliação integrativa, entre outros) como condições de contorno diante das evidências de dificuldades conceituais;* (2) *Didática das Ciências – concepções de um pedagogia de integração para argumentações e interpretações*

¹⁹ O instrumento foi traduzido para o português do Brasil e de Portugal. Esse processo pode acarretar uma mudança funcional do instrumento em cada contexto.

analíticas relacionadas aos conceitos de competência científica; propondo aspectos essenciais para otimizar os itens e atendendo a aspectos que caracterizam as situações-problema.

4. Quanto ao caráter e descrição das **concepções filosóficas (concepção pragmática)** subjacentes diante das **escolhas e fundamentações teórico-metodológicas** associadas:

- De caráter **reivindicatório** pela preocupação em proposições de melhorias do processo pedagógico de diferentes grupos partindo de evidências sobre dificuldades reais na aprendizagem: *apresentando necessidades e propostas pedagógicas específicas para diferentes grupos (de baixo desempenho, alto desempenho, gênero, e contextos geográficos); apresentando comparações qualitativas sobre aspectos comuns quanto ao paradigma de competências enfatizados nos currículos e exames entre os contextos de Brasil e Portugal; valorizando a perspectiva qualitativa dos itens para determinação da chave-de-correção do teste com base na opinião de cinco juízes (avaliação de pares); destacando a necessidade de métodos mistos em pesquisas no EC, destacando o papel e possíveis contribuições da Psicometria.*
- De caráter **pós-positivista** quanto ao *design* diante do método de recolha de dados e dos procedimentos de análise enfatizados na abordagem quantitativa (uso de modelos introdutórios da Psicometria Clássica e Moderna): *analisando previamente a validade interna e externa do design da pesquisa com o uso do instrumento; aplicando um teste conceitual em larga escala como método de recolha e instrumento investigativo; caracterizando e comparando diversos perfis conceptuais sobre os conceitos de calor, temperatura, propriedades térmicas e processos termodinâmicos a partir de índices e parâmetros psicométricos discriminativos, de dificuldades, de acertos e erro ao acaso, além de técnicas de DIF (Função Diferencial do Item) entre diferentes grupos de desempenho; elaborando análises comparativas quantitativas entre os contextos de Brasil e Portugal; determinando métodos quantitativos para determinação da chave-de-correção do teste com base na opinião de cinco juízes (avaliação de pares).*
- Na mesma visão pós-positivista, se tem uma ênfase **behaviorista/conducionista** quanto a utilização de análises taxonômicas de itens, proposição sistematizada de tarefas e organização sequencial de conteúdo, inventário de conceitos, interpretando os resultados psicométricos como expectativas de comportamento ou competências

(em potencial cognitivo) diante de probabilidades de acerto relacionados a habilidades (θ) de resolver situações-problemas em fase introdutória da Termodinâmica: *analizando a taxonomia do conhecimento envolto aos itens dentro do domínio cognitivo; estabelecimento de critérios referenciados para o teste; construindo uma matriz de habilidades para o teste; analisando previamente e qualitativamente indícios de dependência local entre os itens (se a escolha na alternativa de um item depende ou tem suposta influência na escolha a partir da escolha de uma alternativa em um item anterior).*

- De caráter **construtivista** quanto as perspectivas de (re)elaboração de itens, testes, construtos e escalas, tendo como base as percepções dos alunos, bem como assumindo um compromisso no favorecimento do desenvolvimento de competências científicas partindo e valorizando os conhecimentos prévios e contextos favoráveis a aprendizagem centradas nos alunos: *escolhendo um teste que tenham itens numa linguagem cotidiana adequada (critério de legibilidade quanto a faixa etária) e que explore diferentes percepções de alunos (de modo simulado e fundamentado previamente) quanto ao conhecimento térmico; traduzindo o teste item-a-item; apresentando sugestões de reelaboração/aprimoramento dos itens com base nas dificuldades dos alunos; analisando o construto do quadro de tipicidade do teste, considerando a coerência na correspondência entre as concepções alternativas e os respectivos distratores dos itens envolvidos; descrevendo as informações pedagógicas de itens e do teste TCE a partir dos resultados psicométricos; interpretando qualitativamente os gráficos; dando sentido as dificuldades conceituais relacionadas aos distratores mais atrativos; construindo e interpretando pedagogicamente a escala de proficiência de concepções térmicas sendo referência na comparação de grupos de desempenho, habilidades de sujeitos, dificuldades e comportamento funcional de itens entre grupos focais e de referência; fundamentando em uma pedagogia integradora para as competências científicas, na qual se apoia pela articulação intradisciplinar e interdisciplinar na abordagem dos conteúdos e das formas de ensino, valorizando o desenvolvimento da autonomia e o contexto cotidiano dos alunos.*

1.2.2 Definição das Variáveis e dos Critérios Operacionais

O estudo se concentra no instrumento TCE. A natureza do seu conteúdo se configura como o objeto de estudo, mais precisamente a relação entre a natureza da

complexidade que remete as situações-problemas propostas nos itens com as concepções dos alunos diante do conhecimento e conceitos envolvidos, a análise em última instância se dá na tentativa compreender esse conhecimento e os fatores que influenciam os processos mentais que levam o aluno a tomar a decisão de escolha, bem como os equívocos que se destacam como processo e resultado dessa escolha. Tem-se como subsídios para isso, evidências empíricas que as medidas psicométricas fornecem, na qual os modelos logísticos desenvolvidos por psicólogos e neurocientistas vêm fazendo um esforço para se aproximar da realidade da dinâmica cognitiva. É importante ressaltar que esses modelos, apesar de modernos e sofisticados, e de medirem com certo rigor e precisão os traços latentes, sempre estão presentes erros, ou seja, a influência de variáveis externas que remetem a necessidade de serem confirmadas por meios de outros estudos e métodos. As influências na tomada de decisão de respostas consideradas certas e erradas fazem parte desses erros e também são aspectos especulativos diante do objeto de estudo.

Diante disso, e do propósito do estudo, as “variáveis” irão se concentrar em conjecturas sobre as evidências identificadas, ou seja, em avaliações conclusivas preliminarmente com base em evidências empíricas dos traços latentes, sendo fundamentadas dentro das perspectivas da psicologia educacional de Ausubel como subsunçores, mas que não devem ser consideradas um prova verdadeira e universal, mas diagnósticos avaliativos fundamentados dentro do campo de suposições que devem ser verificados e reforçados na observação com outros estudos e práticas, porém, com um olhar voltado para uma aplicabilidade relativamente coerente, delimitada e válida.

No contexto geral, não há um compromisso de verificação propriamente dita sobre a relação de causalidade entre as variáveis, mas sim de expressar a estrutura lógica e sistemática do texto e da sequência das ações, sinalizando, entretanto, as possíveis influências diretas e indiretas entre elas no estudo. Já no contexto específico das análises estatísticas é natural que se compreenda essa relação de causalidade e efeitos.

As **variáveis dependentes** foram delimitadas de acordo como sugere a estrutura do instrumento pelo *critério de coerência interna*, sendo os itens e concepções a que exploram organizadas em um quadro de tipicidade ou inventário de conceitos de acordo com as concepções de calor, temperatura, processos termodinâmicos e propriedades térmicas dos materiais, e com isso, contribuições estruturalmente dentro de cada ramo de concepções. Apesar disso, o estudo irá mostrar que o instrumento também explora outros grupos de concepções que não foram declarados pelos autores originais. As análises diante das contribuições dentro de cada ramo de concepções será feita em uma perspectiva comparativa entre os contextos, buscando ao fim identificar aspectos comuns das dificuldades relacionadas as concepções térmicas dos alunos, tendo como critérios os percentuais assinalados, ganhos, e a verificação de parâmetros psicométricos consistentes e relativamente estáveis. Após a identificação desses itens e as interpretações pedagógicas dos parâmetros psicométricos, será feita uma correspondência com o quadro de tipicidade das concepções do teste e da matriz de habilidades cognitivas desenvolvida com base nas respostas corretas. Os padrões de respostas dentro de cada item sugerem modelos mentais diante do conhecimento térmico, e com isso pretende-se mostrar as supostas dificuldades e proficiências dos diversos grupos normativos definidos na análise.

A consistência das informações se baseia em critérios estatísticos, na escala de proficiências se ocorre pela seleção de itens-âncoras, visando apresentar indícios de validade para os dois contextos. Na escala de proficiências serão posicionados os itens do instrumento de acordo com a base de calibração e estimação dos parâmetros, tendo como referência o parâmetro de dificuldade. Na mesma métrica, também serão posicionados diversos grupos de desempenho relacionados ao nível de escolaridade, turnos (Brasil), escolas. Como parâmetro de corte no design considerando forma de tratamento entre as observações daqueles que antecedem a abordagem ou instrução dentro da termodinâmica (10º ano Portugal, e 1º ano do Brasil) e dos que já viram, se estabeleceu o livro didático de Física dentro de cada contexto que aborda a unidade temática, não havendo, entretanto, comprometimento de avaliar essa forma de tratamento por falta de consistência quanto que sustentem um perfil docente de utilização ou quanto aos hábitos de aprendizagem diante desse recurso, ainda assim,

48

acabam por ser evidências das práticas que são desenvolvidas nas salas de aula dentro de cada escola.

As **variáveis independentes** foram definidas tendo como critério os processos geradores das informações, ou seja, os critérios estatísticos na escolha de técnicas e procedimentos de análise que fornecem as evidências dentro do enfoque hipotético-dedutivo e dos mecanismos e fundamentos no enfoque interpretativo-hermenêutico. Já as **variáveis moderadoras** declaradas se referem aquelas que supostamente estão influenciando o direcionamento conclusivo das análises, que no caso são os: *documentos oficiais; programas curriculares; livros didáticos; e subjetividade inerente nas escolhas metodológicas e interpretações*. Naturalmente, sabe-se que existem diversas outras variáveis moderadoras (Z) que possivelmente estejam influenciando as variáveis independentes (X) ou diretamente as variáveis dependentes (Y).

Na análise do conteúdo do instrumento, os itens são tratados como categorias qualitativas de análise de acordo com as concepções envolvidas, taxonomia dos conhecimentos e relação entre eles, que se associam a categorias quantitativas de análise (ou variáveis) no que se referem aos respectivos índices e parâmetros psicométricos. Entretanto, deve-se considerar que o modelo psicométrico de análise da TCT, e do modo similar os modelos da TRI, admitem como premissa uma relação de causalidade, na qual o escore observado (E), e os não-observáveis diretamente sendo o *pontuação ou escore verdadeiro* (V) e *erro de medida* (e) da forma: $E = V + e$.

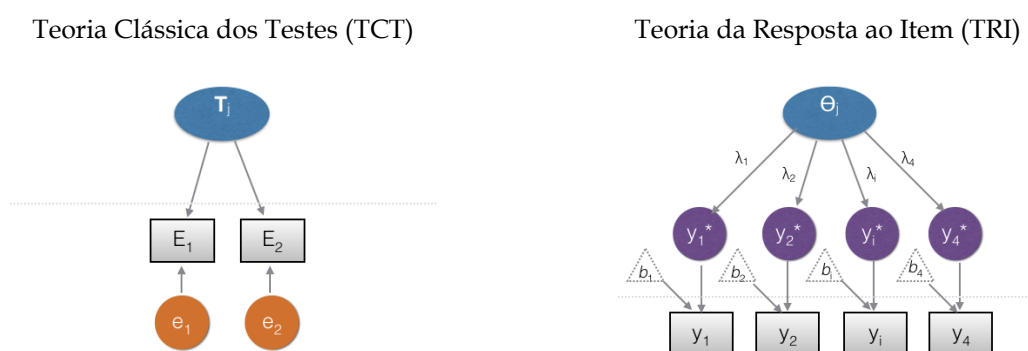


Figura 2: Comparação dos modelos de variáveis latentes da TCT e TRI.

Fonte: Primi (2017).

Diante disso, o E (sendo os itens y_1, y_2, \dots , os correspondentes na TRI) representam as evidências empíricas obtidas e o erro de medida a fatores externos

diversos que buscam uma compreensão próxima da realidade (o correspondente do escore verdadeiro V , no inglês “true” T , seria a habilidade θ como probabilidade de acerto) sobre o conhecimento cognitivo dos alunos e os processos mentais, na qual o estudo busca atribuir por análise interpretativa fundamentada uma compreensão sobre alguns dos possíveis aspectos (relacionado o “erro de medida”) que levam os alunos a optarem por escolhas que representam “supostamente” ao conhecimento prévio ou parte dele.

Quadro 4: Organização dos grupos de variáveis em caráter amplo e sistemático para o estudo de acordo com as categorias de concepções relacionado aos itens, juntamente com a declaração dos processos operacionais utilizados.

Codificação das variáveis	
De acordo com categorização dos grupos de itens	
Y	Variáveis Dependentes (VD): Perspectivas sobre as concepções térmicas
Y1	Alternativas dos Itens relacionadas ao subconstruto A no inventário – Concepções de calor.
Y2	Alternativas dos Itens relacionadas ao subconstruto B no inventário – Concepções de temperatura.
Y3	Alternativas dos Itens relacionadas ao subconstruto C no inventário – Concepções sobre processos termodinâmicos.
Y4	Alternativas dos Itens relacionadas ao subconstruto D no inventário – Concepções quanto às propriedades térmicas.
X	Variável Independente (VI): Interpretações pedagógicas com base nas evidências empíricas levantadas na análise do teste TCE
	Critérios Operacionais*
X1	<p>Intepretações pedagógicas gerais sobre a tipologia dos itens de acordo com parâmetros psicométricos da TCT e da TRI de cada contexto e entre eles.</p> <ul style="list-style-type: none"> • TCT utilizando o <i>software</i> Excel para: determinar os índices de discriminação, de dificuldade e coeficiente bisserial. • TRI utilizando o <i>software</i> R, BILOG e Excel (pacote suplementar <i>virt</i>) para: determinar os parâmetros de discriminação, dificuldade, acerto ao acaso e erro ao acaso; habilidades de grupos (e alguns exemplos de indivíduos); as <i>Curvas Características dos Itens</i> (CCI); as <i>Curvas de Informação dos Itens</i> (CII).
X2	<p>Interpretações pedagógicas a partir das análises gráficas dos itens (abordagem TCT)</p> <p>TCT utilizando o <i>software</i> Excel para:</p> <ul style="list-style-type: none"> • <i>Violação do tipo 1</i>: analisar o número de violações contra a suposição de que os aumentos das alternativas corretas devam seguir crescentemente com a pontuação total; • <i>Violação do tipo 2</i>: analisar o número de violações contra a suposição de que a proporção dos distratores (falsas alternativas) deva diminuir o aumento da pontuação global; • <i>Violação do tipo 3</i>: analisar o número de intersecções entre a alternativa correta e as falsas após o início do intervalo de discriminação; • <i>Violação do tipo 4</i>: analisar a ausência de discriminação ou de uma baixa potência de inclinação.
X3	<p>Interpretações pedagógicas comparativas das dificuldades entre os diferentes grupos normativos nos dois contextos</p> <ul style="list-style-type: none"> • TCT: Comparação de escores brutos; • TRI: Análise sobre a <i>Função Diferencial do Item</i> (DIF)
X4	<p>Classificação e interpretação pedagógica comparando grupos de desempenho, destacando a relevância na aquisição de conteúdos na escala de proficiência</p> <ul style="list-style-type: none"> • TRI: grupos de desempenho por nível de escolaridade; por turno (Brasil); escolas; nível de escolaridade; turno; e por <i>livro didático de física adotado na escola</i>²⁰

²⁰ Parâmetro de comparação intraescolar, porém sem relação de causa e efeito no desempenho e nas dificuldades encontradas.

Z**	Variáveis Moderadoras (VM): Influências declaradas de diferentes perspectivas sobre as possíveis contribuições pedagógicas no estudo, bem como os riscos associados.	CrITÉrios Operacionais
Z1	Perspectiva dos documentos oficiais e programas curriculares que fundamentam o paradigma de competência em cada contexto, com ênfase nas competências essenciais e básicas consideradas por cada sistema, em especial, dentro da temática de termodinâmica	Análise de documentos oficiais; descrição das competências essenciais e básicas; identificação de conteúdo padrão dentro da termodinâmica; frequência de itens relacionados a termodinâmica nos exames; e comentários críticos do currículo dentro da temática em cada contexto
Z2	Perspectiva dos livros didático vigentes no contexto da aplicação do instrumento (10º em Portugal e 2º ano no Brasil) quanto a abordagem dos conceitos de calor, temperatura, processos termodinâmicos e propriedades térmicas.	Análise das definições dos conceitos térmicos abordados e aspectos relacionados a função de desenvolvimento de capacidades e competências científicas em termodinâmica.
Z3	Perspectiva do investigador nas análises	Análise de riscos de validade sobre <i>design</i> em cada etapa da investigação; verificação de riscos durante o processo de recolha das informações; análise taxonômica dos itens e definição das proficiências; revisão sobre os processos realizados no tratamento dos dados e análise dos dados; fundamentação nas tomadas de decisão sobre os procedimentos de análise e interpretações; declaração prévia sobre os critérios avaliativos e operacionais.
Z4	Perspectiva quanto ao professor (aplicador do teste)	Contato pessoal sobre as orientações de abordagem e fornecimento de texto explicativo quanto aos procedimentos na aplicação do teste, com recomendações para evitar indiligência.
Z5	Perspectiva quanto ao respondente	Recomendações explicativas para os procedimentos de realização do teste, entretanto, corre-se o risco de descompromisso e abandono do teste.

Nota: *Os critérios operacionais podem ser classificados do ponto de vista avaliativo aos níveis da ação e da pessoa (ou grupo). Esses aspectos mais detalhados e complementares encontram-se no capítulo 3 – Critérios e Normas Referenciados na Pesquisa; **A adoção do termo “variáveis” exerce mais a função como “categorias de análise” para Z1 e Z2.

- *Interpretações pedagógicas dos itens com base em seu conteúdo* – para isso uma análise interpretativa da natureza taxonômica do conhecimento envolvido utilizando a taxonomia de Bloom revisada em 2001, e assim elaborar uma matriz de habilidades para o teste. Juntamente busca-se verificar e confirmar uma coerência no inventário de conceitos sugerido pelos autores originais, que acompanha o TCE, bem como identificar indícios de relações dependência ou de influências nas respostas de diferentes itens no teste

Um aspecto complementar das análises enfoca em propostas instrucionais ou evidências que sinalizem um favorecimento na aquisição significativa de conceitos térmicos e no desenvolvimento de competências científicas para o conhecimento térmico, para isso consideram-se necessárias análises relacionadas em:

- *Implicações pedagógicas favoráveis a ocorrência de uma aquisição significativa de conceitos* – visando identificar subsunçores e conhecimentos potencialmente

significativos, e assim sugerir atividades e processos que possam favorecer a ocorrência de uma aprendizagem significativa;

- *Implicações pedagógicas favoráveis ao desenvolvimento de competências científicas em termodinâmica* – com base nos subsunçores e conhecimentos potencialmente significativos identificados são propostos famílias de situações-problemas como encargo pedagógico articulador para o desenvolvimento de competências científicas.

1.2.3 Design Geral da Investigação

O estudo apresenta abordagem mista, com um peso relativamente maior para a abordagem quantitativa devido ao tempo gasto no *levantamento dos dados empíricos* e nas *análises* realizadas. Entende-se que a perspectiva teórica global está envolvida com o paradigma da complexidade no campo das competências científicas, entretanto, se delimita na avaliação de desempenhos como um recorte na complexidade dentro dessa visão, vislumbrando uma complementariedade investigativa mais ampla e integradora posterior entre as abordagens metodológicas qualitativas e quantitativas, complementando-se na exploração e aprofundamento.

No caso, estando os resultados fundamentalmente voltados e comprometidos em perspectiva na projeção de boas práticas educacionais, sempre tendo como base necessidades legítimas levantadas nas evidências empíricas.

A elaboração do *design* investigativo adota as notações²¹ recomendadas (Quadro 5), em que Creswell (2010, p. 245) reformula uma notação²² para *designs* mistos adaptando de diversos estudos (Creswell & Clark, 2007; Morse, 1991; Tashakkori & Teddlie, 1998). Dessa forma, as diferentes ênfases nas abordagens se distribuem em 4 fases.

²¹ As notações quanto as incorporações de dados foram intencionalmente adaptadas para uma melhor visualização da estrutura e sequencia das análises.

²² Os *designs* de pesquisas puramente quantitativas usam-se tradicionalmente as notações clássicas de Campbell & Stanley (1963, p. 6). Os *designs* de pesquisas exclusivamente qualitativas estão representados nos *designs* mistos.

Quadro 5: Síntese de notações para os *designs* mistos.

Aspectos básicos	Letras maiúsculas	“Letras maiúsculas” indicam o peso, uma prioridade dos dados, uma ênfase de uma abordagem ou método, seja na análise ou na interpretação dos dados (QUAN ou QUAL). Nos estudos mistos pode ser dada a mesma ênfase para as duas abordagens.	
	Quan e Qual	Essas abreviações representam os termos “quantitativo” e “qualitativo” com a mesma quantidade de letras para indicar igualdade entre as formas dos dados. Nos <i>designs</i> esses termos irão aparecer de duas formas: QUAN ou QUAL (dando maior ênfase) e <i>quan</i> ou <i>qual</i> .	
Mixagem	Combinar	QUAN + qual QUAL + quan	“+” indica uma forma simultânea ou concomitante de coleta de dados, com os dados quantitativos e qualitativos coletados ao mesmo tempo em um único banco de dados (quando tendem a um dos extremos) ou não (quando tendem a parte central do extremo), e na sequência são fundidos.
		QUAN + QUAL	
		QUAN → qual	
	QUAL → quan		
	Incorporar	QUAN/qual QUAL/quan	Indicam uma incorporação da abordagem de menor ênfase sobre a de maior peso. A coleta de dados ocorre simultaneamente se concomitante e separado em duas fases se forem sequenciais.
		<div style="border: 1px solid black; padding: 5px; margin-bottom: 5px; display: inline-block;"> <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-bottom: 2px;"><i>quan</i></div> QUAL </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px; display: inline-block;"> <div style="border: 1px solid black; padding: 2px; display: inline-block; margin-bottom: 2px;"><i>qual</i></div> QUAN </div>	As “caixas” indicam a coleta e análise de dados (1ª fase), que antecipa a fase seguinte relacionada aos resultados comparativos e interpretações finais de todo o processo. A caixa menor (mais interna) é de menor ênfase, assim sua coleta e análise de dados ocorre de forma influenciada e dentro da perspectiva da abordagem predominante. Pode-se ainda afirmar que a menor é incorporada a maior, ou seja, QUAL/quan ou QUAN/qual .

Fonte: Adaptado de Creswell (2010, pp. 245 e 246).

Na Fase 1 são realizados procedimentos de análise interpretativos, e se concentram na: tradução do teste do inglês australiano para o português de Portugal e do Brasil atendendo a alguns critérios de adaptação; análise do conteúdo quanto a confirmação da chave-de-correção e verificação de ambiguidades de itens, entre outros eventuais equívocos e incompletudes; tradução, verificação e análise do *inventário de concepções térmicas* (IC) ou mapa de tipicidade concepção-item, acrescentando com a identificação/interpretação dos distratores; análise taxonômica item-a-item elaborando uma *matriz de habilidades cognitivas e competências térmicas* (MH), buscando

identificar um contexto comum que se associe a característica de uma família de situações-problemas.

A Fase 1 se conecta com a Fase 2 e 3 pelo *design exploratório sequencial* (**QUAL** → **quan**), no caso, conduzindo sequencialmente de um banco de dados a outro, sendo o primeiro com base na interpretação do conteúdo do instrumento, enquanto os outros formados pelos resultados da análise psicométrica.

A sequência da Fase 1 para 2 sendo mais especulativa, verificando preliminarmente o comportamento dos itens (TRI), eficiência em medir ganhos normatizados de desempenho (entre o pré-teste e pós-teste), análise de riscos de *design*, sendo aplicado a uma subamostra normativa, com menos indivíduos, tendo uso de poucas técnicas de análises, por isso um peso menor quantitativo. Na conexão com a Fase 3 já é mais conclusiva e consistente, utilizando um banco de dados mais alargado dentro de cada contexto separadamente e juntos, realizando múltiplas técnicas de análises da TCT e TRI, incluindo a análise de DIF. Na fase 3 também ocorre a construção da escala de proficiências de concepções térmicas.

A passagem da Fase 3 para a 4 é realizada simultaneamente, estando presentes características de dois *designs*. O *design transformativo concomitante* se caracteriza pela transformação de dados quantitativos em qualitativos, sendo quantitativamente orientado (de acordo os critérios estatísticos), com um núcleo quantitativo e componente incorporativo qualitativo sendo adicionado, ou seja, em interpretações pedagógicas dos itens com base nos índices e parâmetros para cada contexto. Pode-se dizer que corre a princípio sem uma orientação privilegiada e sem um núcleo prioritário na abordagem, pois as componentes ocorrem de forma simultânea (Creswell, 2010).

Também inclui características de *design incorporado concomitante* por adicionar uma forma secundária de dados para dar apoio e ajudar a comparar de forma separada os bancos de dados lado a lado. Especificamente se refere ao processo de interpretação dos desempenhos na escala, estando *lado a lado* com a descrição das habilidades cognitivas dos itens, conforme a matriz de referência do teste. Na Fase 4 também

ocorre um breve diálogo sobre as implicações do conteúdo e das habilidades diante os conteúdos, habilidades e competências dentro da Termodinâmica que são preconizados nos programas curriculares e cobrados nos exames nacionais. A etapa final se concentra em uma síntese das análises interpretativas buscando enfatizar aspectos gerais diferenciados e comuns entre os dois contextos, em especial sobre a avaliação dos desempenhos e das interpretações de um único construto explicativo sobre as concepções térmicas.

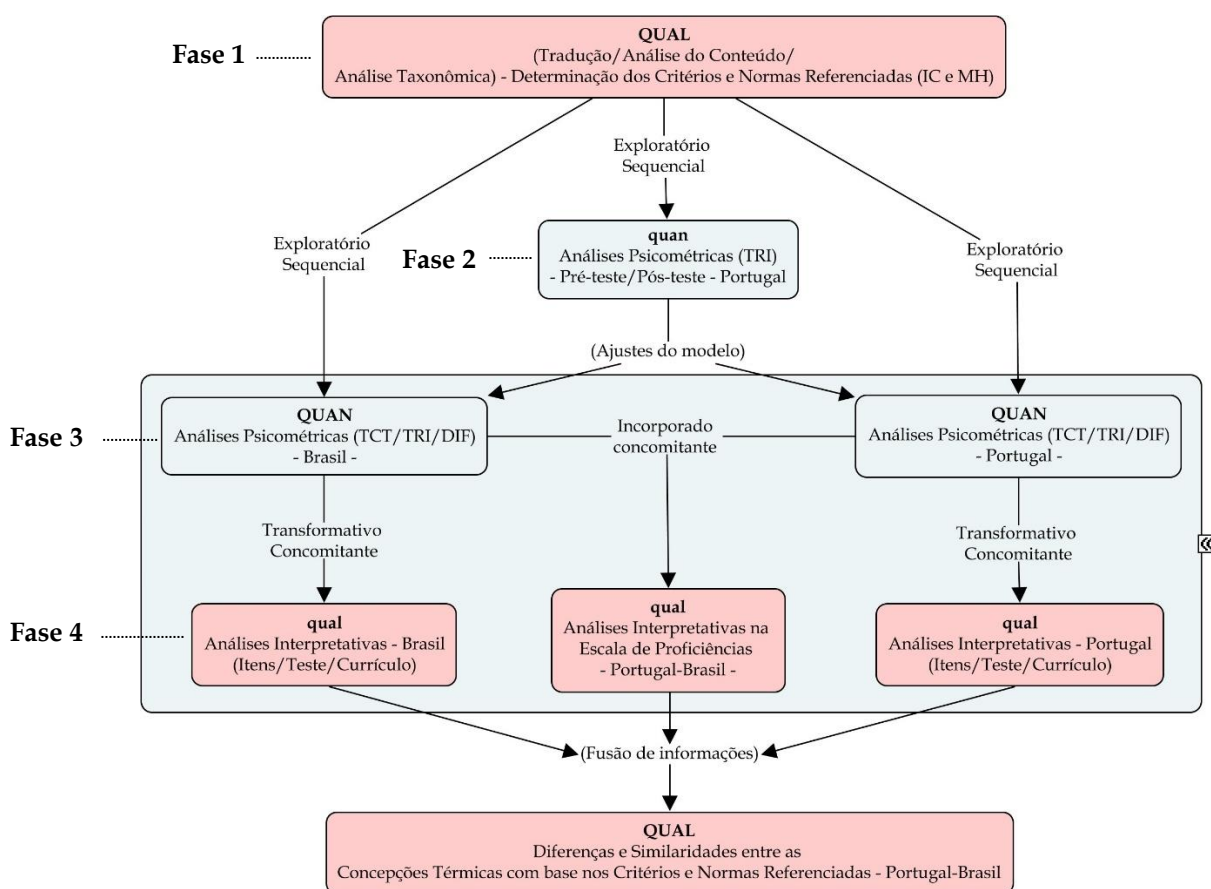


Figura 3: *Design da investigação - Abordagem mista sobre o panorama teórico-metodológico e as principais operacionalizações realizadas.*

1.3 HIPÓTESES

Espera-se que o estudo possa:

- **Proporcionar** uma *melhor compreensão sobre o instrumento TCE*, otimizando sua precisão e destacando novas perspectivas para a validade interna e externa;

- **Inferir** *ganhos percentuais de acerto* para os grupos normativos (níveis de escolaridade, escolas, livros didáticos adotados) em cada contexto, bem como os *modelos mentais* científicos associados que estão em destaque (corretos e incorretos);
- **Constatar** *maiores dificuldades* na compreensão de conceitos e processos básicos em Termodinâmica entre diferentes grupos de desempenho de alunos brasileiros comparados com os portugueses;
- **Comparar** perspectivas de proficiências visando destacar *necessidades comuns e diferenciadas* entre os alunos brasileiros e portugueses diante as dificuldades na compreensão e aquisição de conceitos básicos e introdutórios da Física Térmica;

Acredita-se que esse conjunto de informações possam ser exploradas por diversos procedimentos de análise, tanto da TCT como da TRI, visando construir e sintetizar os aspectos mais consistentes e pertinentes na elaboração de perfil de proficiências e concepções térmicas dos estudantes em cada contexto. É natural que existam no percurso investigativo microproblemas a serem enfrentados devido aos processos de transição que são identificados e considerados necessários de serem superados. Estes acabam por abrirem novas perspectivas de avanço na investigação em um sentido complementar posterior, que no entanto serão considerados como *processos secundários incidentes* e de importância relativa, porém enriquecedores diante a necessidade de alcançar aos objetivos gerais propostos com êxito, como por exemplo:

- Verificar²³ novas concepções no instrumento sugeridas tanto no enunciado como nas alternativas de múltipla escolha que não foram previstas e consideradas pelos autores originais, mas que podem influenciar em uma interpretação e avaliação diagnóstica sobre as necessidades de aquisição de conceitos;
- Identificar indícios na relação de dependência nas respostas de itens, ainda que se considere um construto predominante e explicativo para o instrumento.

²³ A verificação é apenas um subconjunto da avaliação [...] funciona mais num sistema fechado [...] é válida sobretudo para objetivos no sentido estrito e para regulamentos. (De Ketele & Roegiers, 1993, p. 75)

1.4 OBJETIVO MISTO

O objetivo deste estudo de métodos mistos de **4 fases**, sendo inicialmente *exploratório sequencial* depois *transformativo e incorporado concomitante*, é **identificar as diferenças e similaridades entre as concepções térmicas de alunos brasileiros e portugueses**.

A passagem da 1ª fase para a 2ª fase será uma exploração qualitativa analisando o instrumento como documento textual, buscando com isso caracterizar natureza taxonômica do conteúdo nos itens, as habilidades envolvidas e possíveis influências nas respostas dos alunos, sobretudo, definir os critérios referenciados do teste. A razão de se analisar os dados qualitativos inicialmente é para verificar a adequação do instrumento, para que os itens possam ter um quadro que expresse de forma clara as suas finalidades avaliativas (habilidades), que sejam caracterizados em uma taxonomia de objetivos cognitivos, para que em uma fase posterior os resultados possam ser interpretados com maior confiança, ou seja, ampliando a capacidade e fiabilidade das análises quantitativas.

A 2ª fase será *exclusivamente quantitativa* e caracterizada como um teste-piloto, especificamente um pré-teste/pós-teste do instrumento TCE coletando dados para avaliar o desempenho de alunos do 10º ano, antes e após a instrução em termodinâmica, em 4 escolas de Portugal, que no caso, correspondem a uma subamostra da normatização do estudo. Os resultados desta fase quantitativa serão utilizados para testar a validade preditiva do instrumento, riscos de validade interna e externa do *design*, verificar a eficiência em ganhos normatizados comparando os escores entre o pré-teste e o pós-teste, e com isso verificar eventuais relações previstas na análise qualitativo, bem como observar se o comportamento dos itens para a subamostra corresponde com os da amostra da população do estudo.

Na passagem da 3ª fase para a 4ª ocorrem simultaneamente dois processos, *transformativa concomitante e incorporação concomitante*.

O processo que se refere na *transformação concomitante* visa transformar dados quantitativos em qualitativos fundindo os resultados da aplicação do teste TCE. Busca-se avaliar as diferenças e similaridades **entre os valores dos parâmetros com as concepções térmicas de alunos brasileiros e portugueses**. A razão de fundir dados quantitativos em qualitativos é entender melhor as diferenças e similaridades entre as concepções térmicas dos alunos nos dois contextos com base nos índices e parâmetros da TCT e TRI, avaliando uma relativa correspondência e consistência entre grupos de itens, e assim, convergindo os dados quantitativos (tendências numéricas amplas relacionadas aos indicadores e parâmetros dos itens) em dados qualitativos (concepções térmicas e as características desse conhecimento na matriz de habilidades) para realizar as análises. Nesse processo também estão inclusas as interpretações pedagógicas item-a-item com base nas análises da TCT e TRI.

O processo que se refere na *incorporação concomitante* visa adicionar as habilidades cognitivas da matriz de referência dos itens *lado a lado* com o posicionamento de itens e das habilidades estimadas entre os diferentes grupos de desempenho na escala de proficiência de concepções térmicas. Nessa fase do estudo, são utilizadas análises e técnicas para a construção da escala de acordo com os itens calibrados no modelo explicativo. Após esse estudo, são feitas medições de desempenho entre os diferentes grupos de Brasil e Portugal, e posicionados na escala juntamente com os itens e a descrição das habilidades cognitivas correspondentes, que correspondem as interpretações na escala. A razão de se emparelhar os dados quantitativos e qualitativos é entender melhor aspecto do problema de pesquisa convergindo tanto dados quantitativos relacionados das habilidades estimadas θ (ou Z , escore padronizado) quanto os qualitativos que se referem as concepções térmicas correspondentes e interpretações dos traços latentes avaliados (em certa parte, os subsunçores evidenciados), para assim realizar uma comparação entre as diferenças e similaridades de domínio sobre as concepções térmicas conforme suas competências (em patamares de proficiências), na qual se baseiam em probabilidades de acerto.

A etapa final visa organizar uma interpretação geral sobre as diferenças e similaridades com base em uma justaposição das interpretações, identificando

possíveis indicadores intercambiáveis relacionados as concepções térmicas em uma perspectiva diante de um único construto explicativo para os dois contextos.

1.5 FASES DA PESQUISA

Atendendo a uma *revisão na literatura* nos enfoques desta investigação, configura-se como um *estudo exploratório* com perspectiva comparativa na *avaliação de desempenhos* especificamente diante de concepções térmicas entre Brasil e Portugal.

O estudo comparativo está estruturado em três fases (Ferreira, 2001): **Pré-descritiva** ou *Exploratória*, **Descritiva** ou *Interpretativa*, e **Comparativa** ou *Analítica*. Considera-se que o estudo tenha uma *abordagem mista*, com **maior peso quantitativo** devido ao *design* na coleta de dados dos alunos, do tipo de bases de dados construídos e analisados.

As ações investigativas da **Fase Pré-Descritiva** ou *Exploratória*, se referem:

1. Na *delimitação do objeto* através da revisão sistemática na literatura sobre a temática nos contextos diante da perspectiva do estudo, valorizando de certa forma o contexto local do pesquisador. Para isso foi realizado inicialmente uma busca no repositório da Universidade de Coimbra e da Universidade Federal de Santa Catarina de acordo com as mesmas palavras-chave de busca. Foram destacados aqueles que adotaram abordagens metodológicas e procedimentos de análises equivalentes, além da temática.
2. No *mapeamento curricular* buscando identificar padrões de conteúdos da Física Térmica pelos critérios incidências em exames nacionais e no programa curricular, nomeadamente, a *Base Nacional Comum Curricular* – BNCC. Também está incluído aspectos relacionados as *condições de alinhamento curricular*, tendo como referência o uso de testes conceituais.
3. Na *delimitação metodológica* tem-se a definição das variáveis de acordo com os instrumentos de coleta: um «teste conceitual» para diagnosticar o conhecimento térmico dos alunos (TCE). A partir disso foram estabelecidos critérios avaliativos

quanto às operacionalizações e manipulação das variáveis no tratamento das informações coletadas, produzidas e dos processos de análise.

4. Quanto a *preparação do instrumento* TCE foi realizada sua *tradução*²⁴ para o português do Brasil e Portugal. Segue na etapa de *recolha de informação* empírica no campo de estudo indo às escolas, contactando os diretores e professores (por e-mail, telefone móvel, telefone fixo da escola, etc.), estando com todos os documentos legais que viabilizam a pesquisa, negociando um cronograma de visita e conversa inicial com os professores, com marcação de entrevista, e data de entrega e distribuição de material a serem aplicados e posteriormente recolhidos.

5. A *aplicação dos instrumentos* que restringe ao teste TCE. As análises psicométricas visam estabelecer parâmetros para uma caracterização em cada contexto como para um estudo comparativo diante das necessidades sobre diagnósticos que retratam as dificuldades conceituais dos alunos no conhecimento térmico.

6. Quanto aos *recursos* utilizados se tem a aplicação de teste em «folha de papel» (e não digital ou por meio computadorizado), registro de «fotos» no ambiente escolar (salas de aula e laboratórios), aplicação de testes de verificação das concepções alternativas dos alunos quanto à temática de Termodinâmica.

7. Quanto a *delimitação da área de estudo* realizou-se uma comparação internacional baseadas em amostras intranacionais (dentro de um corte do contexto nacional), onde a amostragem em Portugal estão escolas secundárias da Região do Centro, e no Brasil, escolas do ensino médio da Grande Florianópolis, no Estado de Santa Catarina (SC), sendo todas públicas.

As implicações da **Fase Descritiva** ou *Interpretativa*, que se caracteriza em reflexão, interpretação e conclusões analíticas comparativas. Entre as ações estão:

1. A *recolha, computação e tratamento dos dados* obtidos pelo teste-piloto. Posteriormente repete-se o processo na sequência da aplicação do TCE em cada contexto organizando inicialmente em bases separadas.

²⁴ A versão traduzida para do teste TCE no português do Brasil encontra-se no repositório da PhyspPort no site: <<https://www.physport.org/assessments/assessment.cfm?I=17&A=TCE>>.

2. A *interpretação dos dados e conclusões analíticas* a partir de um conjunto de evidências que ilustrem o progresso dos alunos nos diferentes níveis de escolaridade, em suas probabilidades de acerto, e no domínio de conteúdo, tendo como referência uma escala de proficiências térmicas. Dessa forma verificam-se as dificuldades conceituais dos alunos em diversas perspectivas, levando em consideração uma perspectiva complementar entre os métodos estatísticos que envolvem a *Teoria da Resposta ao Item (TRI)* e a *Teoria Clássica dos Testes (TCT)*, dando ênfase nas evidências mais consistentes que possibilitem contribuições pedagógicas.

As implicações para a **Fase Comparativa** ou *Analítica* se estabelecem no refinamento e fundamentação maior das informações e conclusões analíticas iniciais, entre as ações estabelecidas, tem-se:

1. Para a *formulação de hipóteses comparativas* serão manipulados diversos aspectos relacionados às variáveis da amostra quanto aos resultados das dificuldades dos alunos, tendo como fundamento estatístico a TRI e TCT.

2. A *justaposição de dados e de conclusões analíticas* se dará diante da confrontação dos dados obtidos dentro de parâmetros equiparáveis preestabelecidos e o uso de técnicas relacionadas a TCT e TRI. Nesse último estágio se disserta sobre aspectos comparativos mais amplos diante de uma perspectiva explicativa de um único construto integrador sobre o conhecimento térmico dos alunos. Para isso são apresentadas suposições fundamentadas com base nas evidências quanto às relações de causa-efeito para as dificuldades conceituais em Termodinâmica. Esse é um aspecto conclusivo em uma perspectiva diagnóstica, delimitada e que necessita ser averiguada em estudos complementares, em especial, que vise explorar a observação em sala de aula, coletar evidências por meio de entrevistas com alunos e professores, para assim reforçarem as consistências da atual avaliação. Espera-se, contudo, que essas análises possam ser úteis em promover melhorias em outros processos educacionais, seja na tomada de decisão em novas investigações e em atividades de ensino, fazendo parte da construção de um dossiê de avaliação educacional nesta unidade temática, e assim reforçar recomendações que visem à melhoria no processo do ensino da Física na Termodinâmica em ambos os contextos, e comparativamente entre eles.

1.6 REVISÃO DE LITERATURA

1.6.1 Produções Acadêmicas Baseadas no Uso de Testes de Aferição do Conhecimento Físico na Educação

Há mais de três décadas de pesquisa relacionadas à compreensão dos conceitos térmicos pelos alunos indicaram uma variedade de concepções alternativas realizadas por estudantes de 12 a 17 anos de idade (Chu et al., 2012).

Quadro 6: Alguns estudos das últimas décadas relacionadas às dificuldades na compreensão dos conceitos térmicos.

Concepções alternativas identificadas*	Idade dos alunos	Método	Autor(es)
<i>Calor</i>			
Existem dois tipos de calor: calor quente e calor frio			
O calor é como uma onda que se eleva da estrada feito fumaça	12 anos de idade	Entrevistas	(Erickson, 1979)
<i>Temperatura</i>			
A temperatura é uma quantidade extensa	12 anos de idade	Entrevistas	(Erickson, 1979)
A temperatura da água fervente pode exceder 100 °C durante a ebulição	12-15 anos de idade	Entrevistas	(Andersson, 1980)
A temperatura é uma medida de calor	15-16 anos de idade	Entrevistas	(Kesidou & Duit, 1993)
<i>Ebulição</i>			
A matéria dentro de bolhas de água fervente é água, vapor de água, calor, ar, fumaça, oxigênio ou dióxido de carbono	12 anos de idade (além disso 6-11 anos de idade)	Testes escritos abertos	(Bar & Travis, 1991)
<i>Condutividade de calor</i>			
Os metais atraem, mantêm ou armazenam calor e frio	12-14 anos de idade (incluindo adultos e cientistas)	Entrevistas	(Eileen Lob Lewis & Linn, 1994)
A lã aquece as coisas			
<i>Equilíbrio térmico</i>			
	15 anos de idade	Entrevistas	(Clough & Driver, 1986)
A temperatura de diferentes objetos é diferente mesmo que eles tenham sido colocados no mesmo ambiente durante um longo período de tempo	17 anos de idade	Método investigativo (discussão em sala de aula)	(Harrison, Grayson, & Treagust, 1999)
	12-14 anos de idade (incluindo adultos e cientistas)	Entrevistas	(Eileen Lob Lewis & Linn, 1994)
<i>Conceitos térmicos gerais</i>			
Confusão de estudantes por causa do dualismo de termodinâmica/falta de compreensão nos processos de resfriamento (ou aquecimento) de substâncias puras/entendendo que não depende das sensações do corpo	13-18 anos de idade (incluindo professores)	Testes escritos abertos	(Sciarretta, Stilli, & Missoni, 1990)

Fonte: Chu et al. (2012, p. 1512, tradução livre). *Todas as referências no quadro foram verificadas.

Com o avanço desse campo investigativo, muitos pesquisadores se envolveram na elaboração e análise de testes sobre os conceitos básicos que envolvem a *Termodinâmica* (Baser, 1996; Erickson, 1980; Lewis, Stern, & Linn, 1993; Lewis & Linn, 1994; Midkiff, Litzinger, & Evans, n.d.; Niaz, 2000; Olds, Streveler, & Miller, 2004;

Thornton & Sokoloff, 1998; Yeo & Zadnik, 2001). Esses testes conceituais, também conhecidos como **Inventários de Conceitos**, encontram-se disponíveis em diversos sites²⁵. Na análise de revisão da literatura, os principais testes encontrados foram: *Survey of Thermodynamic Processes and First and Second Laws (STPFaSL)* (Brown, 2015); *Thermodynamic Concept Survey (TCS)* (Wattanakasiwich, Taleab, Sharma, & Johnston, 2013); *Thermal Concept Evaluation (TCE)* (Yeo & Zadnik, 2001); *Thermal and Transport Concept Inventory: Thermodynamics (TTCI-T)* (Nelson et al., 2007); *Thermodynamics Concept Inventory (TCI)* (Clark Midkiff, 2000); *Heat and Temperature Conceptual Evaluation (HCTE)* (Thornton & Sokoloff, 2001); *Energy Concept Inventory (ECI)* (Swackhammer, 2005); *Heat and Energy Concept Inventory (HECI)* (Prince, Vigeant, & Nottis, 2012); *Heat Transfer Concept Inventory (HTCI)* (Jacobi, Martin, & Mitchell, 2003, 2004); *Thermodynamik Test Inventar (TTI)* (Einhaus, 2007). Pode-se dizer que esses são alguns dos precursores dos *principais testes da Física Térmica* nas últimas duas décadas, considerando que anos seguintes foi mais um predomínio de suas aplicações, considerando ainda versões adaptadas, traduções e fusões de itens entre diferentes testes, lembrando que são instrumentos livres, e com isso podem ser difundidos, reproduzidos e manipulados.

O levantamento mostra que o TCE desde 2001 já foi aplicado em diversos contextos e culturas²⁶, contribuindo no campo da pesquisa no *Ensino de Física* (EF) para a compreensão de conceitos (Thacker, 2003, pp. 1835-1836). Entre eles tem-se a aplicação em professores de escola primária em um estudo na Turquia (Baser, 2006) e a estudantes secundários em outros dois (Cigdemoglu & Geban, 2015; Tastan & Yalçinkaya, 2008) - sendo o último em versão adaptada para a química; a estudantes secundários na Tailândia (Kruatong, Sung-ong, Singh, & Jones, 2006), e com estudantes universitários (Talaeb & Wattanakasiwich, 2010; Wattanakasiwich et al.,

²⁵ Ver em: <<https://www.physport.org/assessments/>>.

Ver em: <<https://cgi.tu-harburg.de/~zllwww/fachdidaktik/ci/>>

²⁶ Foi feita uma varredura na plataforma "ResearchGate" e identificou-se 67 citações e 23 referências quanto aos trabalhos. Dentro de cada um trabalho citado fez uma varredura interna utilizando a palavra "YEO" e "TCE" e foram descartadas todas aquelas que não fizeram uso diretamente do TCE. No Brasil foi identificado um trabalho no *Caderno Brasileiro de Ensino de Física* (CBEF) e rastreado outras publicações dos autores. O outro trabalho foi baseado na varredura do Currículo Lattes da pesquisadora Marta Feijó (UFRJ) e analisado todas as publicações a respeito do TCE.

2013); a professores e alunos do ensino fundamental em um estudo nos Estados Unidos (Luera, Otto, & Zitzewitz, 2006; Otto, Luera, & Everett, 2009), e do secundário (Schnittka & Bell, 2011); com estudantes do ensino médio no Brasil (Gonçalves Júnior, 2012; Júnior & Barroso, 2011a, 2011b, 2014; Louzada & Elia, 2015; Louzada, 2012); com calouros universitários em Sydney (Georgiou, 2014; Georgiou, Sharma, Byrne, Sefton, & Mcinnes, 2009; Georgiou & Sharma, 2010, 2012; Tanahoung, Chitaree, Soankwan, Sharma, & Johnston, 2006); no México (Gómez, & Hernández, 2010); e na Líbia (Alwan, 2011); com estudantes secundários da Coréia do Sul (Chu et al., 2012); e da Nigéria (B. Madu & Orji, 2015).

Segundo a revista *Science*, as projeções e interesses no uso apropriado de testes diagnósticos como instrumentos de auxílio ao professor já vinham sendo apontados como um campo promissor desde o início do século na comunidade científica de pesquisadores no *Ensino de Física* (EF) nos EUA, e sendo acompanhados por pesquisadores de outras disciplinas (Wood & Gentile, 2003). Essa perspectiva também vem avançando em novos campos de análise a medida em que se difundem e se tornam mais familiares os métodos da Psicometria para os pesquisadores, com isso, acredita-se ser relevante buscar acompanhar esse movimento, pois há diversas possibilidades de contribuírem dentro do EF (Engelhardt, 2009).

Um levantamento no estado da arte nos EUA dentro das pesquisas educacionais em Química mostra discussões relacionadas aos critérios de credibilidade no processo avaliativo diante das evidências psicométricas em diversos estudos de ensino, visando otimizar e ampliar esse processo nos aspectos de validade (Arjoon, Xu, & Lewis, 2013). Enquanto isso, no Brasil, o interesse no uso de boas práticas em testes conceituais na área da educação não demonstra ter havido maiores repercussões, e nem se percebe perspectivas de interesse atualmente entre os pesquisadores. Talvez isso ocorra devido ao predomínio de abordagens qualitativas no EC que vem desde os anos 70 (Nardi & Almeida, 2007), além da falta de familiaridade com a perspectiva quantitativa por grande parte de pesquisadores nesse segmento, e um distanciamento com os adeptos quantitativos, atrasando a necessidade de avançar em pesquisas mistas na educação científica (Santos & Greca, 2013).

Essa pode ser considerada uma visão simplista para as pesquisas educacionais que por vezes se reveste com uma aversão estereotipada, preconceituosa e de certa forma ingênua diante a esse tipo de abordagem, e que se agrava pela disseminação de um discurso equivocado que remete a um “positivismo tablóide”, na qual se desconhece os diversos compromissos-chave de cunho epistemológico assumidos que são igualmente compartilhados no construtivismo (Matthews, 2004).

Apesar disso foram detectados dois instrumentos relativamente estruturados dentro da *Termodinâmica*, envolvendo conceitos de *calor e temperatura* (Henrique, Nardi, & Laburú, 2010; Lang & Moreira, 1996). Também foi identificado outros *testes elaborados* no Brasil na década de 80, entre eles estão alguns preocupados em detectar concepções de *força e movimento* (Silveira, Moreira, & Axt, 1992; Silveira et al., 1986); *corrente elétrica* em circuitos simples (Silveira et al., 1989). Na década de 90 também foi encontrado trabalho preocupado com a *validade de testes* (Moreira & Silveira, 1993).

Recentemente alguns trabalhos demonstraram interesse na aplicação de testes internacionais (praticamente todos sendo traduzidos do inglês e aplicados) no contexto brasileiro que envolvem *conceitos básicos* da Mecânica (Custódio, 2012; Ferreira, Caíres, Silva, & Oliveira, 2009; Custódio & Barroso, 2012) – relacionados ao *Mechanics Baseline Test* (MBT) (Hestenes & Wells, 1992); da **Física Térmica** (Gonçalves & Barroso, 2011; Júnior, 2012; Júnior & Barroso, 2011a, 2011b, 2014; Louzada, 2012; Louzada, Elia, & Sampaio, 2015; Louzada, Elia, Sampaio, & Vidal, 2014) – relacionados ao *Thermal Concept Evaluation* (TCE) (Yeo & Zadnik, 2001).

Chama-se a atenção que a grande parte desses trabalhos não fazem referência aos procedimentos típicos de análise que foram desenvolvidos nesse âmbito investigativo, verificando, por exemplo, o “ganho normalizado padrão” ou a “mudança normalizada”, como também não buscam declarar cuidados ou recomendar a seleção de alunos correspondentes entre o pré-teste e o pós-teste, a função específica avaliativa (em que momento e condições ele deve ser aplicado? Está a serviço de quem? Qual a sua finalidade?...), como também seguir critérios específicos sugeridos na literatura sobre a administração de boas práticas no uso desses

instrumentos. Também são raros os trabalhos identificados que exploram os procedimentos da TRI nesses testes. Um deles faz críticas aos pressupostos de unidimensionalidade que devem ser seguidos (Gonçalves Júnior & Barroso, 2012).

A relevância diante das necessidades de *investigar e avaliar* os desempenhos dos alunos, entre outros aspectos intrínsecos, crescem de acordo com a expansão da avaliação externa dos sistemas estaduais, seja com um olhar na busca de melhorias nos indicadores nacionais como para políticas internas de cada estado federativo. Os desdobramentos mais recentes que antecipam versão final da BNCC é um fator que impulsiona atualmente esse interesse por valorizar a necessidade no desenvolvimento de um conjunto de habilidades, e vem recebendo críticas severas diante da matriz epistemológica desenhada para promover o desenvolvimento de competências, argumentando incompatibilidades existentes com as recentes *Diretrizes Nacionais Curriculares para a Educação Básica* (DNCEB).

Diante dos exames nacionais, vem sendo publicadas pesquisas com enfoques relacionados as críticas de *qualidade científico-pedagógica de itens* (Silveira, 2013; Silveira, Stilck, & Barbosa, 2014; Silveira, Barbosa, & Silva, 2015; Oliveira, 2014); *análises comparativas de desempenho* (Duarte, Júnior, & Barroso, 2013; Gonçalves & Barroso, 2014; Júnior, 2012a; Júnior & Barroso, 2012), críticas quanto a *dimensionalidade do exame de acordo com o modelo da TRI adotado* (Costa, 2015; Oliveira, 2014, 2015; Quaresma, 2014; Vieira, 2016), entre outras. Ainda nesse âmbito, também surgem trabalhos com interesse na *elaboração de itens* para disponibilizar em sistemas digitais de montagem de provas (Vernet, 2013).

Quanto aos fundamentos dentro da psicologia educacional, o papel do conhecimento prévio ganha amplitude de interesse diante das necessidades interpretativas de resultados de testes conceituais, principalmente quando supostamente possuem a intenção de detectar evidências de concepções formuladas diante de conceitos científicos. Diante disso, ao ser posicionado um interesse nos fundamentos de David Ausubel, pode-se observar a sua importância e pertinência

nesse âmbito, podendo servir condições de contorno no que se refere a proposição de instruções ou recomendações para o ensino a partir dos resultados nos testes.

Sabe-se que a concepção de Ausubel já foi bastante explorada, sendo *referência argumentativa para a ocorrência de uma mudança conceitual* em obras clássicas (Posner et al., 1982), além de muitos trabalhos acadêmicos. O catálogo analítico da USP, mostra que no início da década de 80 tiveram trabalhos relacionados a *influência do conhecimento prévio sobre o desempenho em física* (Cubillos, 1981); *Mapeamento cognitivo* (Farias, 1982); *laboratório estruturado versus não estruturado* (Gonçalves, 1979); *em mapas conceituais como instrumentos para investigar a estrutura cognitiva em física* (Guerra, 1983); *no ensino de eletricidade e magnetismo em nível universitário* (Peduzzi, 1981); *uma abordagem ausubeliana à organização do conteúdo em Termodinâmica e teoria cinética dos gases ao nível de física geral* (Lima, 1981); *na discussão sobre a compreensão e memorização no ensino da estatística* (Feijoo, 1986). Também se encontram artigos relacionados a *análise dos livros didáticos de Física quanto ao processo na aquisição de conceitos* (Neves & Valadares, 2004); *na organização de conteúdos* (Moreira, 1998), *no ensino investigativo da ótica* (Valadares & Fonseca, 2011); *no ensino demonstrativo investigativo para aquisição de conceitos* (Braga & Kalhil, 2011); *em instruções no uso de analogias* (Nagem, Carvalhaes, & Dias, 2001); *desenvolvimento na compreensão elementar da Termodinâmica* (Eileen Lob Lewis, 1991), entre muitos outros.

Em um levantamento de 100 trabalhos publicados na primeira década do século XXI sobre as temáticas mais abordadas em sete das principais revistas especializadas no EF no Brasil mostra que apenas 1% deles trata sobre *resolução de problemas*, 8% tratavam sobre *levantamento de concepções/dificuldades conceituais*, 22% sobre *recursos didáticos* e 5% sobre *conteúdos reelaborados* para o ensino médio (Rezende, Ostermann, & Ferraz, 2009). Podem-se destacar quatro desses trabalhos pertinentes ao interesse desse estudo, que são: *dificuldades conceituais na Termodinâmica na perspectiva de Vergnaud* (Grings, Caballero, & Moreira, 2006); *perfil conceitual diante de concepções sobre o conceito de calor* (Amaral & Mortimer, 2001); *instrumento de avaliação para os manuais da disciplina de física em Portugal baseados na concepção ausubeliana*

(Neves & Valadares, 2004); discussão conceitual sobre equilíbrio térmico a partir dos fenômenos do dia-a-dia (Cindra & Teixeira, 2004).

As concepções associadas ao conceito de calor têm sido investigadas em diversos âmbitos, incluindo as *concepções prévias* dos alunos, dificuldades conceituais associadas e propostas de ensino (Amaral & Mortimer, 2011), assim como das contribuições que os Livros Didáticos exercem no processo de aquisição de conceitos (Neves & Valadares, 2004). Diversas perspectivas podem ser exploradas para melhor perceber a compreensão que alunos possuem diante de determinado conceito. Para isso, Sepulveda & Mortimer (2013) apresentam a construção de um modelo de «perfil conceitual» (E. Mortimer, 1994, 1995) baseado em quatro zonas denominadas de *funcionalismo intraorgânico, ajuste providencial, perspectiva transformacional e perspectiva variacional*, na qual compromissos epistemológicos e ontológicos ajudam a estruturar distintos modos de pensar sobre o conceito, sendo essa uma forma alternativa de lidar e encarar com a ideia de *mudança conceitual*.

Diante de um mapeamento (Dentz & Truccolo, 2013) de 130 dissertações e 7 teses sobre os objetos de pesquisa no EC em seis *Instituições de Ensino Superior* (IES) no sul do país, dentro do período de 2004 e 2008, sendo duas privadas, destaca-se com 28,5% o maior número de incidências sobre as «*novas metodologias e propostas didáticas de ensino e análise dos seus resultados em termos de melhorias na aprendizagem dos alunos (o lúdico, metodologias para conteúdos específicos, métodos de avaliação, etc.)*», que parecem muito abrangentes, porém, e em segundo lugar, de modo mais específico tem a «*análise de livros didáticos*» com 17,5%, no entanto a preocupação com a «*resolução de problemas como estratégia de ensino*» apenas 3%, e com a «*construção conceitual dos alunos*» e os «*problemas e dificuldades de aprendizagem*», cerca de 1,5% cada um. Esse é um campo de delimitação temática nas pesquisas diante da expansão de programas na área do EC em que avançou de 42 em 2008 para 157 em 2017, e que a maioria das produções se encontram na região sul e sudeste, segundo os últimos relatórios de avaliação da área de ensino da CAPES²⁷ (Araújo-Jorge & Fonseca, 2013; Araújo-Jorge, Borba, & Sovierzoski, 2017). Apesar disso, esses levantamentos sinalizam a importância dada

²⁷ Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES.

ao Livro Didático (LD) como objeto de estudo, porém mostra-se necessário avançar em linhas de investigação que se aliem com as *dificuldades conceituais* dos alunos valorizando a forma como concebem os conceitos. Acredita-se que seja apropriado e vantajoso agrupar essas necessidades em situações-problemas que atendam uma pedagogia integradora dentro campo de resoluções de problemas como estratégia de ensino, assumindo um comprometimento de atender o desenvolvimento de habilidades e competências almejadas nos documentos curriculares.

A importância de considerar as CAs nesse estudo, assim como de sua associação dentro do âmbito da necessidade de desenvolver competências científicas também podem ser evidenciadas numa publicação de 2011, em que a *Sociedade Brasileira de Física* (SBF) comenta sobre os desafios e perspectivas da Pesquisa no EF (Carvalho, Nardi, Vianna, Almeida, & Ferracioli, 2011) para os próximos cinco anos, ou seja, até o final de 2016. Os autores sintetizam dois principais resultados básicos diante dos estudos na área ao longo das últimas três décadas: (1) a identificação das *concepções alternativas* pelos alunos, não estando por vezes contrárias à visão da Ciência; (2) e a sua extrema resistência à instrução formal. Destacam ainda que essas questões avançaram na década de 2000 para problemáticas sobre o *Letramento Científico*, onde se associam os aspectos das linguagens nos materiais didáticos e da interação dos professores e alunos em sala de aula, e do paradigma das relações entre *Ciência, Tecnologia e Sociedade* (CTS) ou *Ciência, Tecnologia, Sociedade e Ambiente* (CTSA) para a área do EC. Ressaltam ainda que para favorecer uma compreensão científica se torna necessária uma articulação de diversos saberes, tornando relevantes estudos que **priorizem o desenvolvimento de habilidades e competências no EC**. Diante disso, distinguiram dois focos para os estudos: o *processo científico* e as *competências curriculares*, sendo destacados alguns pontos:

“Sobre o processo científico podemos distinguir vários pontos, tais como: (1) reconhecimento das questões científicas a serem investigadas; (2) identificação das evidências necessárias na investigação científica; (3) proposição e avaliação das conclusões; (4) **demonstração do entendimento dos conceitos científicos** e (5) comunicação com validade o processo do produto dos conhecimentos científicos.

Sobre as competências curriculares, pode-se citar: (a) aprendizagem auto-reguladora; (b) **habilidade para resolver problemas**; (c) comunicação e cooperação.” (Carvalho et al., 2011, p. 122, *grifo nosso*).

1.6.2 Centros de Pesquisa que utilizam a Teoria da Resposta ao Item no Brasil

Existem alguns centros de investigação a desenvolver estudos com base na Psicometria, porém são praticamente inexistentes aqueles voltados para a pesquisa no EC, e mais ainda os estudos que se aprofundam nas interpretações e implicações dos exames a partir de mecanismo de análise similares. Os estudos recorrentes concentram-se basicamente nas áreas da psicologia, psicanálise, neurociência e da avaliação educacional, estando predominantemente envolvidos profissionais da estatística e psicologia geral e experimental, enquanto os da *psicologia aplicada* à área da educação estando praticamente ausentes. Observa-se também um baixo envolvimento de profissionais da saúde dentro dessa perspectiva, ao contrário do que ocorre nos países do Norte, em que esses procedimentos de análise acompanham o crescimento e desenvolvimento de pesquisas com abordagens mistas.

A Associação Brasileira de Avaliação Educacional (ABAVE) realiza anualmente o Congresso Brasileiro de Teoria da Resposta ao Item e Métodos Quantitativos em Avaliação²⁸ (CONBRATRI), na qual pesquisadores de diversas comunidades científicas apresentam trabalhos nesse seguimento. Atualmente, o site²⁹ do Instituto Brasileiro de Avaliação Psicológica (IBAP) apresenta **44 principais laboratórios** em funcionamento no Brasil, sendo **apenas um que se declara estar voltado para área da educação**, apesar de quase todos estarem desenvolvendo trabalhos relacionados: *Laboratório de Avaliação Psicológica e Educacional* – LabAPE, que está ligado ao Programa de Pós-Graduação (PPG) em Psicologia da *Universidade de São Francisco* (USF). Moreira Junior (2010) apresenta um panorama de trabalhos desenvolvidos na TRI no Brasil desde sua implantação como política no SAEB em 1995 até 2009, assim como os Grupos de Pesquisa (GPs). Identificou-se 219 trabalhos, sendo 120 em *avaliação educacional*

²⁸ Em 2018 ocorrerá o 6º congresso.

²⁹ Existem laboratório que são coincidentes denominados “LAPE”, que fazem parte da mesma universidade. Ver: <<http://www.ibapnet.org.br/index.php?cd=8&descricao=laboratorios>>.

representando 55% do total, aqueles com a *porcentagem arredondada* para 0% possuem apenas uma única incidência. Apesar dessa ênfase no campo educacional, a concentração desses centros e grupos de pesquisa sugerem que a maioria dessas produções são desenvolvidas predominantemente dentro de uma perspectiva da psicologia mais geral e experimental devido ao perfil dos profissionais que estão envolvidos nos processos de análise. Considera-se que vem avançando a adesão da TRI nos institutos de educação preocupados com a avaliação (Moreira Junior, 2010), sendo mais distante ainda do interesse entre pesquisadores do EC, apesar das implicações curriculares, nos livros didáticos e para o cenário pedagógico.

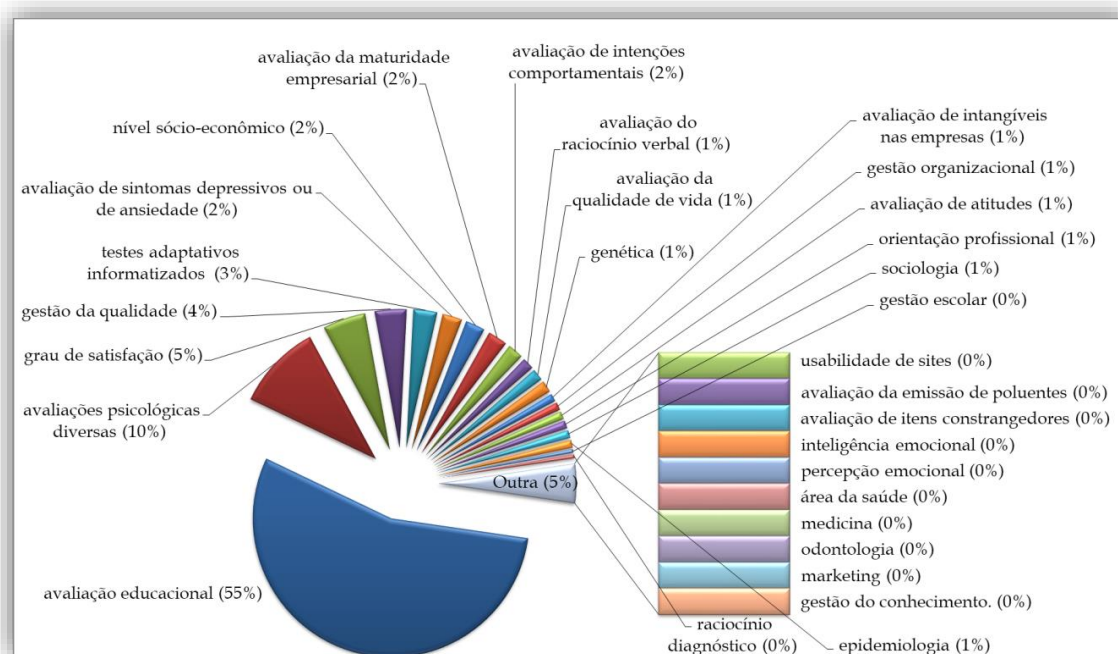


Figura 4: Quantitativo da produção no Brasil de 1995 a 2009 com a abordagem na TRI.

Fonte: Elaborado a partir dos dados de Moreira Junior (2010).

Pode-se destacar o *Centro de Políticas Públicas e Avaliação da Educação (CAEd)*, da *Universidade Federal de Juiz de Fora*, com experiência na avaliação psicométrica em alguns dos sistemas estaduais de avaliação pelo país, incluindo o *Sistema de Avaliação do Desempenho Educacional do Amazonas (SADEAM)*, porém foram identificados 9 GPs na qual trabalham apropriadamente na TRI no Brasil:

Atualmente, os grupos de pesquisa cadastrados no Diretório dos Grupos de Pesquisa no Brasil, do *Conselho Nacional de Desenvolvimento Científico e Tecnológico*

(CNPq) que trabalham com TRI no Brasil são: **Laboratório de Custos e Medidas (LCM)** e **Laboratório de Custos e Controle de Gestão**, do Programa de Pós-Graduação em Engenharia de Produção e Sistemas da UFSC; **Inferência Estatística em Modelos de Resposta ao Item**, do Departamento de Estatística da UFPA; **Métodos Estatísticos em Avaliação**, do Departamento de Estatística da UNB; **Laboratório de Pesquisa em Avaliação e Medida**, do Departamento de Psicologia Social e do Trabalho da UNB; **Núcleo de Avaliação**, Investigação e Medidas Educacionais, da UFC; **Modelos de Regressão e Aplicações**, da USP; **Modelos de Regressão**, da UFCG; **Núcleo de Instrumentos e Medidas**, da UFBA. (Moreira Junior, 2010, p. 143)

1.6.3 Universidade Federal de Santa Catarina

1.6.3.1 Teoria da Resposta ao Item

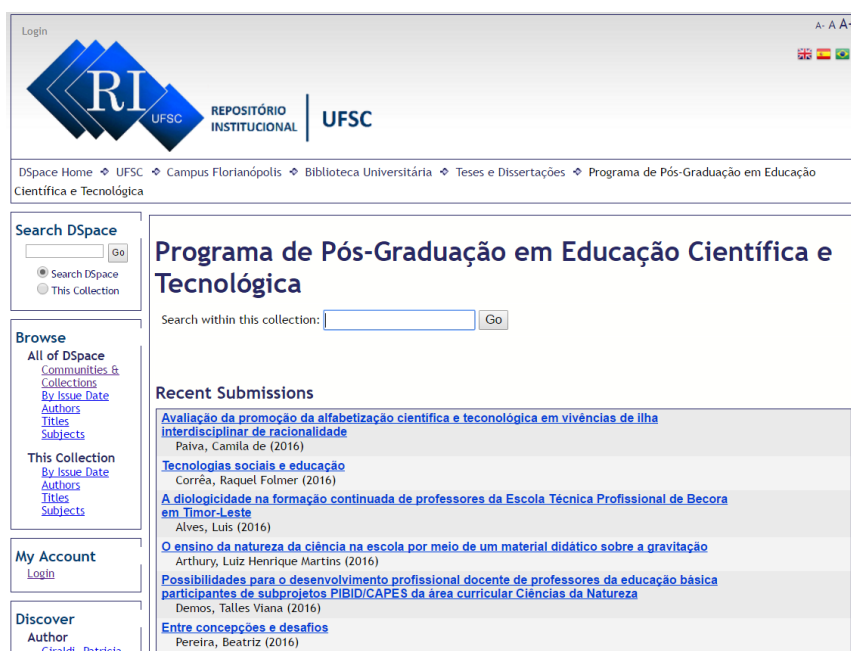


Figura 5: Ambiente virtual do repositório institucional da UFSC.

O acesso ocorreu via *Comunidade Acadêmica Federada (CAFES)*, utilizando as credenciais da UFSC, no Portal de Periódico da CAPES. Entre teses e dissertações de 2000 a 2017, há **26 produções** no *Repositório Institucional* da **UFSC** que “aplicaram” (e não apenas “citaram” no seu texto) modelos estatísticos da *Teoria da Resposta Item (TRI)*, mais precisamente se distribuem nos *Programas de Pós-Graduação (PPGs)* da seguinte forma: **16 em engenharia de produção**³⁰, **entre 12 teses** (Almeida, 2009; Junior, 2010;

³⁰ Há um total de 4334 trabalhos disponíveis em 14.04.2017. Todos os números divulgados do Repositório Institucional da UFSC servem para a mesma data.

Mafra, 2011; Menegon, 2013; Moreira Junior, 2011; Pacheco, 2014; Peixe, 2014; Tezza, 2012; Tirloni, 2013; Vargas, 2007; Vergara, 2005; Vey, 2011) e **4 dissertações** (Afonso, 2013; Alves, 2011; Testa, 2014; Tezza, 2009); 3 em *métodos e gestão em avaliação* (Costa, 2015; Gomes, 2014; Vieira, 2016); 2 em *educação física* (P. M. da Silveira, 2015; Testa, 2014); 1 em *ciência da computação* (Justino, 2007); 2 em *engenharia e gestão do conhecimento* (Almeida, 2009; Mafra, 2011); 1 em *odontologia* (Torres, 2015); 1 em *nutrição* (Gabriel, 2008). Quando a busca torna-se refinada utilizando o termo “escala de proficiência” são **13 produções** exclusivamente em *engenharia de produção* (Alves, 2011; Balbim Junior, 2010; Bortolotti, 2003; Brotti, 2004; Moreira Junior, 2011; Menegon, 2013; Peixe, 2014; Tezza, 2009, 2012; Tirloni, 2013; Vargas, 2007; Vergara, 2005; Vey, 2011).

1.6.3.2 Termodinâmica, concepções alternativas, livros didáticos e competências.

Há 8 trabalhos disponíveis do PPG de *Ensino de Física*³¹ (EF) da UFSC. Buscando a temática “*termodinâmica*” encontram-se 2, um relacionado a *modelagem computacional* (Nunes, 2016) e outro sendo uma *proposta de contextualização* (Rodrigues, 2016), **porém nenhum deles** incide sobre *concepções alternativas*.

No PPG em *Educação Científica e Tecnológica* (PPGECT) há **292 produções**. A busca por citações «dentro texto da produção» com termo “livro didático” no PPGECT ocorrem **137 resultados**, porém, aquelas que referem o termo no título foram 15, configurando assim objeto principal no estudo ou relação mais pertinente, entre eles: *Cidadania e matemática no livro didático para as séries iniciais do ensino fundamental* (Arruda, 2004); *Linguagem em textos didáticos de citologia: investigando o uso de analogias* (Giraldi, 2005); *A genética humana no livro didático de biologia* (Casagrande, 2006); *A linguagem natural e a linguagem algébrica: nos livros didáticos e em uma classe de 7ª. série do ensino fundamental* (Jacomelli, 2006); *Aproximações entre ciência-tecnologia-sociedade e os temas transversais no livro didático de matemática do ensino fundamental de 5ª. a 8ª. Séries* (Lima, 2008); *A organização praxiológica do objeto triângulo nos livros didáticos da 7ª série do ensino fundamental* (Maia, 2008); *As Figuras geométricas no ensino de matemática: uma*

³¹ Mestrado Profissional.

análise histórica nos livros didáticos (Brigo, 2010); Concepções de progresso biológico em livros didáticos de biologia aprovados pelo PNLD 2012 (Alfaya-Santos, 2013); Seleção e uso do livro didático na visão de professores de ciências: um estudo na rede municipal de ensino de Florianópolis (M. Rosa, 2013); Um livro didático na circulação de sentidos sobre energia (Caum, 2013); O livro didático de ciências: a persistência da fragmentação dos conteúdos (Gramowski, 2014); Análise da didatização do tema radiação de corpo negro sob a luz da teoria antropológica do didático (Gonnelli Netto, 2014); O professor e o uso do livro didático de biologia (Zancan, 2015); Análise praxeológica de tópicos de física moderna em livros didáticos do PNLD (Bernardo, 2015); Discursos de genética em livro didático: implicações para o ensino de biologia (Montalvão Neto, 2016); Os temas "vidros e metais" em livros didáticos de química (Tochetto, 2016). Com o auxílio do «Google Acadêmico» foram encontrados mais 3 produções: A história da matemática no ensino fundamental uma análise de livros didáticos e artigos sobre história (Peters, 2005); História da ciência: investigação do tema em livros didáticos do ensino fundamental (Batista, 2007); Literatura infantil no ensino de ciências: articulações a partir da análise de uma coleção de livros (Linsingen, 2008).

Das 292 produções no PPGECT, existem **40** que citam o termo “*termodinâmica*”, porém «combinando» com a expressão “*concepções alternativas*”, “*calor*” ou “*temperatura*” **não há nenhum resultado**, com exceção do Trabalho de Conclusão de Curso (TCC) de especialização de Köhnlein (2001), e em outra busca genérica dentro do mesmo PPG, encontra-se o trabalho: *Atividades experimentais: contribuições para o aprendizado dos conceitos de calor e de temperatura* (Farias, 2013).

Quando a busca prioriza “*concepções alternativas*” tem-se 88 trabalhos, porém **não há nenhum resultado** quando associado com os termos “*calor*”, “*temperatura*”, “*termodinâmica*” ou “*termologia*”, mas quando se refina a busca «associando» ao termo “*livro didático*” surgem **3 produções** entre os seguintes temas envolvidos: *circulação de sentidos sobre energia* (Caum, 2013); e *seleção e uso do LD por professores de ciências* (M. Rosa, 2013); *representações moleculares sobre ácidos e bases em Química* (Gorri, 2014). Em síntese, pode-se dizer o Repositório Institucional da UFSC diante da aplicação de procedimentos de análise que envolve TRI é concentrado em múltiplas áreas do

conhecimento, porém não foram desenvolvidas investigações nessa perspectiva em nenhum PPG ligado a educação ou ensino.

Buscando especificamente produção que levam no título o termo “concepções alternativas” obteve um único resultado: *Estratégias didáticas de professores do ensino fundamental ao lidar com concepções alternativas de alunos* (Macedo, 2008). Com auxílio do «Google Acadêmico» ainda a obra: *A energia no ensino fundamental: o livro didático e as concepções alternativas* (Jacques, 2008).

Verificando os trabalhos PPGECT que envolvesse a concepção de “competências”, tem-se: *Projetos de trabalho e avaliação por competências: encontros, desencontros e contribuições à educação matemática* (Soares, 2004); *Projetos de trabalho e avaliação por competências: encontros, desencontros e contribuições à educação matemática* (Ricardo, 2005). Já no âmbito mais geral considerando todos os PPGs surgem 151 resultados, porém sendo predominantes na área de gerenciamento empresarial e administração. Algumas aproximações podem ser consideradas, como: *Avaliação do desenvolvimento de competências e habilidades: uma proposta de arquitetura de sistema de controle acadêmico para feedback de aprendizagem, em curso de graduação em administração* (Carvalho, 2006), em outras circunstâncias sendo útil para mostrar a diferença conceptual com os termos utilizados entre a área da educação e de gestão/administração: *Competências essenciais: proposta de um modelo de concepção* (Pacheco, 2010). Também há trabalhos que apresentam informações úteis no sentido de esclarecer a concepção sobre o conceito de competência em diversos contextos: *Competências de pesquisadores manifestadas em teses de doutorado em psicologia e educação de universidades do Brasil, Espanha e Suécia* (Zwierewicz, 2013).

1.6.4 Universidade de Coimbra

1.6.4.1 Teoria da Resposta ao Item

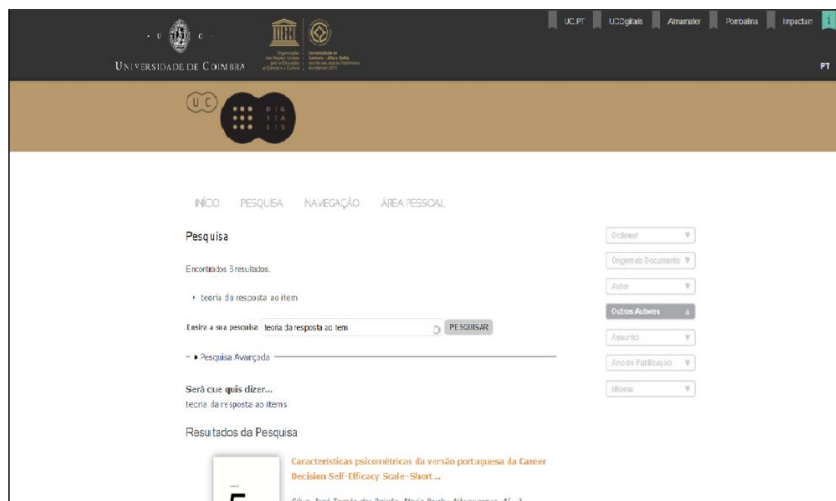


Figura 6: Ambiente virtual do repositório institucional da UC.

No *Repositório Digitalis* da Universidade de Coimbra surgem 6 artigos através de busca direta com o termo “Teoria da Resposta ao Item”, sendo basicamente referentes a aplicação de testes psicológicos padronizados: *Características psicométricas da versão Portuguesa da Career Decision Self-Efficacy Scale–Short Form (CDSE–SF)* (Silva, Paixão, & Albuquerque, 2009); *Questionário de Avaliação do Perfil Sazonal: características psicométricas da versão portuguesa do Seasonal Pattern Assessment Questionnaire (SPAQ)* (Fonte & Coutinho, 2015); *Propriedades Psicométricas da Escala de Cuidado do Questionário de Experiências de Cuidado e Abuso na Infância (Childhood Experiences of Care and Abuse–CECA. Q)* (Carvalho, Pinto-Gouveia, & Pimentel, 2011); *Coping Health Inventory for Parents (CHIP)* (Cunha, Major, & Relvas, 2016); *Resultados das Provas de Aferição e dos Exames de Matemática e de Português do 9º ano: Que conclusões tirar?* (Santos, Araújo, & Graça, 2006); *Consumo de álcool e seus efeitos no desempenho escolar* (Fonseca, 2010).



Figura 7: Ambiente virtual do repositório científico de acesso aberto de Portugal.

No *Repositório Científico de Acesso Aberto de Portugal* (RCAAP) surgem **21 trabalhos** (12 dissertações de mestrado e 9 teses de doutoramento) com o termo “Teoria da Resposta ao Item”. Não há incidência de produção na UC, todas resultaram em produções advindas de universidades brasileiras, e constata-se um predomínio de trabalhos na UFSC, porém não há aplicações na área da “educação” ou “ensino”, quando se refina a busca.

As **dissertações** são: UFSC – *Aplicação de um modelo de desdobramento graduado generalizado da Teoria da Resposta ao Item* (Bortolotti, 2003); *Desenvolvimento de ambiente computacional para implementação de métodos da Teoria da Resposta ao Item* (Justino, 2007); *Proposta de um construto para medir usabilidade em sites de e-commerce utilizando a Teoria da Resposta ao Item* (Tezza, 2009); *Criação de uma escala da qualidade de vida para profissionais de educação física por meio da Teoria da Resposta ao Item* (Testa, 2014); *Mensuração da predisposição ao comportamento sustentável por meio da Teoria da Resposta ao Item* (Afonso, 2013); PUC/RJ – *Métodos alternativos no critério Brasil para construção de indicadores socioeconômicos: Teoria da Resposta ao Item* (Pereira, 2004); UnB – *Teoria da Resposta ao Item em processo de decisão* (Araujo, 2014); UFRGS – *Teoria da Resposta ao Item na avaliação de qualidade de vida de idosos* (Guewehr, 2007); USP – (Neves, 2010) *Experimentos de microarrays e Teoria da Resposta ao Item*; PUC/RS – *Análise de atitudes de*

alunos universitários em relação à estatística por meio da Teoria da Resposta ao Item (TRI) (Turik, 2010); UFC – Um Estudo da evolução da GQT por meio do modelo de resposta gradual da Teoria da Resposta ao Item (Oliveira, 2010); UFLA – Proposta de um modelo logístico de três parâmetros a partir da Teoria da Resposta ao Item: uma aplicação em dados do vestibular da Universidade Federal de Lavras (Fernandes, 2008).

As **teses** são: UFSC – Avaliação do Ensino de Ergonomia para o Design aplicando a Teoria da Resposta ao Item (TRI) (Vergara, 2005); Medida padronizada para avaliação de intangíveis organizacionais por meio da teoria da resposta ao item (Vargas, 2007); Sistemática para a implantação de testes adaptativos informatizados baseados na teoria da resposta ao item (Moreira Junior, 2011); Modelagem multidimensional para mensurar qualidade em website de e-commerce utilizando a teoria da resposta ao item (Tezza, 2012); Mensuração de conforto e desconforto em poltrona de aeronave pela teoria da resposta ao item (Menegon, 2013); Mensuração da maturidade do sistema de gestão ambiental de empresas industriais utilizando a teoria da resposta ao item (Peixe, 2014); UFRGS – Teoria da resposta ao item: aplicação na avaliação da intensidade de sintomas depressivos (Castro, 2008); USP – Modelos longitudinais de grupos múltiplos multiníveis na teoria da resposta ao item: métodos de estimação e seleção estrutural sob uma perspectiva bayesiana (C. L. N. Azevedo, 2008); Modelagem para construção de escalas avaliativas e classificatórias em exames seletivos utilizando teoria da resposta ao item uni e multidimensional (Quaresma, 2014).

1.6.4.2 Termodinâmica, concepções alternativas, livros didáticos e competências.

No Repositório Digitalis da Universidade de Coimbra quando a busca se refere “manuais escolares” tem-se 7 resultados: *Manuais escolares: 'a ponte e a porta' da vida e da cultura* (Serrano, 2008); *"Heróis do lar, nação católica e rural": os livros únicos do ensino primário no Estado Novo* (Monteiro, 2008); *Género e cidadania nas imagens de História: estudos de manuais escolares e software educativo* (Vieira, 2008); *A (re)valorização de outras fontes históricas: a problemática dos manuais escolares* (Monteiro, 2010); *Representações pictóricas nos manuais de ensino de Ciências Naturais (7º ano) e de Geologia (10º ano)* (Parreiral, Pereira, & Gomes, 2011); *As temáticas do Ano Internacional do Planeta Terra nos manuais escolares de Geologia dos 10º e 11º anos de escolaridade do Ensino Secundário*

Português (Pacheco & Henriques, 2012); *Contextualização da aprendizagem: sua representação em manuais escolares de Estudo do Meio* (Moleiro, Damião, & Festas, 2014).

A busca por “termodinâmica” resultou em 4 trabalhos: *Termodinâmica e propriedades termofísicas: vol.2 teoria cinética e propriedades de transporte dos gases* (Lobo & Ferreira, 2005); *Termodinâmica e propriedades termofísicas: vol.1 Termodinâmica das fases* (Lobo & Ferreira, 2006); *Algumas reflexões sobre o segundo princípio da Termodinâmica ou princípio da entropia / João Maria de Almeida Lima* (Lima, 2013); *Temperatura de ponto de orvalho: um risco ou uma necessidade* (Talaia & Vigário, 2016).

Com relação a “competências” possui 132 resultados. Realizada uma seleção, foram excluídas obras que focam no âmbito trabalhista e priorizado aqueles mais pertinentes no sentido de contribuir com implicações para a educação básica ou em sua contextualização, entre eles: *Competências do século XXI, transição para o ensino superior e sucesso académico: estudo do nível de proficiência linguística em inglês e português dos estudantes do 1º ano da Universidade de Coimbra* (Soares, Pereira, & Canavarro, 2010); *Interesses e competências percebidas em crianças do ensino básico* (R. David, Paixão, & Silva, 2009); *Competência Percebida e Realização Escolar: Que Relações em Alunos do Ensino Secundário?* (Stocker, Neves, & Faria, 2010); *Da construção à validação de um referencial de competências para uma Licenciatura em Enfermagem* (Mendonça, Huet, & Alves, 2014); *Desenvolvimento de competências ou transmissão de conhecimento: acerca da necessidade de superar uma antinomia curricular no Ensino Universitário* (Damião, 2015); *Avaliação de competências de literacia mediática: instrumentos de recolha de informação e opções teórico - metodológicas* (Lopes, 2015); *Competências: um caminho educativo para novos desafios* (Almeida, 2007); *Ambiente térmico de sala de aula pode condicionar o desenvolvimento de competências e avaliação de alunos* (Talaia & Silva, 2014); *Aprender a aprender no processo de reconhecimento, validação e certificação de competências (RVCC): estudo de caso duplo* (Inácio, 2009); *Educação para desenvolvimento sustentável e práticas interdisciplinares* (Capelo & Pedrosa, 2012)

No Repositório Científico de Acesso Aberto de Portugal (RCAAP) a palavra de busca “termodinâmica” dentro das teses resultou em 111 trabalhos em áreas diversas, sendo

apenas uma identificada na área da educação: *Ensino e aprendizagem da termodinâmica: questões didáticas e contribuições da história da ciência* (Silva, 2013).

CAPÍTULO 2 - REFERENCIAL TEÓRICO-METODOLÓGICO

A fundamentação teórica do ponto de vista da Didática das Ciências para as análises interpretativas dos itens, elaboração de uma matriz de referência para o teste e de implicações pedagógicas sustentam-se em leituras e reflexões nos trabalhos desenvolvidos pelos autores Xavier Roegiers, Jean-Marie De Ketele e François-Marie Gérard (Alves & De Ketele (orgs.), 2011; Alves & De Ketele, 2011; De Ketele, J-M, 2006, 2008, 2010; Gérard & Roegiers, 1998; Gerard, Peyser, & Roegiers, 2006; De Ketele, J-M, 2004; Roegiers, 2000, 2006; Roegiers & De Ketele, 2004).

Os objetivos dos testes envolvem conteúdos e remetem a capacidades específicas para cada item e entre eles, com isso torna-se relevante compreender as relações pedagógicas envolvidas de forma apropriada no que diz respeito a natureza das situações-problema voltadas para o desenvolvimento de competências científicas de base, sendo um paradigma secundário, porém indissociável diante as implicações das análises. As *condições de contorno* dentro da psicologia educacional apoiam-se nas concepções da *Teoria da Aprendizagem Significativa* de David Ausubel, sendo um elo com a perspectiva comportamentalista, pois os procedimentos geram expectativas comportamentais diante de suas capacidades cognitivas. Além de auxílio na elaboração de instruções ou recomendações fundamentadas, a perspectiva de Ausubel se torna um parâmetro delimitador interpretativo diante de possíveis influências e riscos existentes nos escores sem levar em causa hipotéticas situações que põe em risco uma representação do construto. O suporte metodológico para as análises se concentram dentro da Psicometria, em especial nos modelos da TCT e TRI, estando destacados os seus fundamentos básicos mais utilizados, e no caso, suficiente para esse estudo.

2.1 CONTRIBUIÇÕES DA DIDÁTICA DAS CIÊNCIAS E IMPLICAÇÕES CURRICULARES

O desenvolvimento do campo da Didática das Ciências (DC) é considerado emergente por atender a três razões gerais (Cachapuz, Gil-Pérez, Carvalho, & Praia, 2005):

- A existência de uma problemática relevante, susceptível de despertar um interesse suficiente que justifique os esforços necessários ao seu estudo;
- O caráter específico dessa problemática, que impeça o seu estudo por outro corpo de conhecimentos já existente;
- O contexto sociocultural, bem como a recursos humanos – condições externas.

No final da década de 80, as *tentativas de inovação*³² e investigação na DC foram muito bem caracterizadas e justificadas nos estudos de Linn (1987, citado por Cachapuz, Gil-Pérez, Carvalho, & Praia, 2005, p. 188; Pérez, Alís, & Terrades, 1999, pp. 15-16) onde chama de “**a amnésia crônica**” o processo de reintrodução de propostas já declaradas como ineficazes e estudos pontuais, carentes de fundamentação. Nesse processo, a Psicologia da Educação assumiu-se bem antes como suporte na busca por explicações diante das dificuldades dos fenômenos educacionais, entretanto, as tentativas de aplicação de construções teóricas advindas por esse campo de conhecimento não demonstraram serem suficientes pelo fato de pedagogos tratarem de modo mais geral os processos de aprendizagens simplesmente por ignorarem o *papel central da natureza dos conteúdos*, onde Shulman (1987, citado por Cachapuz, Gil-Pérez, Carvalho, & Praia, 2005, p. 188; Pérez, Alís, & Terrades, 1999, pp. 20) criticou metaforicamente como o “**paradigma perdido**”.

Os avanços diante dessas perspectivas foram bastante relevantes com o surgimento de inúmeras revistas especializadas a partir da década de 80, e os estudos na revisão na literatura conseguiram identificar algumas das principais linhas de

³² “... foram se ajustando ao movimento browniano, produzindo certa agitação... sem, contudo provocar mudanças efetivas” (Novak, 1982, citado por Cachapuz et al., 2005, p. 188).

investigação (Martínez-Terrades, 1998): *concepções alternativas; resolução de problemas; práticas de laboratório; práticas de campo; currículo; materiais didáticos; relações ciência/tecnologia/sociedade e o papel do meio; linguagem e comunicação; avaliação; formação de professores; concepções epistemológicas dos docentes; história da ciência; questões axiológicas.*

Alguns congressos mais recentes sobre a Didática das Ciências também vêm destacando como linhas de pesquisa como: *o ensino por investigação; interdisciplinaridade, globalização e complexidade; educação ambiental; desenho, implementação e avaliação de programa e propostas didáticas; educação científica e gênero; formação de professores no ensino de ciências,* entre outros.

Como se sabe, o MCA que partiu dos anos 70 se configurou como uma das linhas investigativas mais promissoras e marcantes na emergência da DC, e sua relevância continua a manter o interesse nos dias atuais, seja pelos resultados claros e convincentes, de levantar questões sobre visões simplistas de ciência como também por colocar em xeque propostas de ensino previamente elaboradas baseadas na transmissão do conhecimento (Driver, 1973; Viennot, 1977). A abundância de estudos nessa área já havia sido levantada por Reinders Duit em 1993 (Duit, 1993), a Física sendo representada na ocasião com cerca de 66% do total da produção acadêmica, foi capaz de levantar concepções alternativas (CAs) nos diversos campos disciplinares de ciências, assim como compreender a *natureza de seu caráter* (resistência a mudanças, estabilidade relativa, entre outros), identificar suas *origens* (Pozo & Crespo, 2009), e de fortalecer um *viés construtivista* para o desenvolvimento de *designs* inovadores no cenário pedagógico. Sendo uma base promissora, investigações nessa área demonstraram uma grande capacidade integradora com outras linhas de pesquisa da DC (Gil-Pérez, 1994), na qual demonstram estarem cada vez mais integradas (Cachapuz et al., 2005, p. 203).

Acredita-se que uma integração entre as linhas das concepções alternativas e a de resolução de problemas com o auxílio de procedimentos de análises mais robustos diante de testes em larga escala, podem revelar algumas características conexionistas

entre as concepções que estruturam perfis conceituais de grupos, como diferentes patamares relativamente estáveis sobre determinadas concepções, sendo um campo ainda a ser explorado. Esse é um parâmetro que pode auxiliar a elaboração de situações de aprendizagem significativas típicas para diferentes populações de acordo com amostras representativas, sugerindo fundamentalmente a continuação da aprendizagem em níveis de maior e menor complexidade dentro do processo de aquisição de conteúdo, sobretudo, na progressão do desenvolvimento de *competências científicas*. Essa é uma preocupação sugerida nos documentos balizadores dentro do contexto brasileiro: “O que se deseja é que os estudantes desenvolvam competências básicas que lhes permitam desenvolver a capacidade de continuar aprendendo” (Brasil, 2000, p. 14).

A preocupação com a compreensão de alunos diante de conceitos básicos e uma mudança conceitual de suas concepções equivocadas podem ser situadas dentro de uma das cinco finalidades para a educação científica possivelmente viáveis de serem assumidas (Pozo & Crespo, 2009): (1) *a aprendizagem de conceitos e a construção de modelos*; (2) *o desenvolvimento de habilidades cognitivas e de raciocínio científico*; (3) *o desenvolvimento de habilidades experimentais e de resolução de problemas*; (4) *o desenvolvimento de atitudes e valores*; (5) *a construção de uma imagem da ciência*.

Quadro 7: Tipos de conteúdo, metas e visão de ciência no currículo de ciências.

Tipos de conteúdo	Mais específicos	↔	Mais gerais
Conceituais	Fatos/dados: (1)		Conceitos
Procedimentais	Técnicas (2) e (3)		Normas
Atitudinais	Atitudes (4)		Normas
Visão Coerente da Ciência (5)	É transversal aos tipos de conteúdo citados, e ajuda não somente a identificar as características do conhecimento científico, mas principalmente, a diferenciar e valorar esse saber em comparação com outros tipos de discurso e de conhecimento social.		

Fonte: Adaptado de Pozo & Crespo (2009).

Refletindo sobre essa perspectiva, a *aprendizagem de conceitos* remete a não apenas em sua aquisição, mas no seu desenvolvimento e integração com um conhecimento mais geral e estruturado. Compreende-se com isso uma necessidade para que se estabeleça um *continuum* de compreensibilidade conceitual, partindo daqueles mais específicos e concretos, porém, buscando garantir uma progressão para que se alcance os mais gerais e abstratos/complexos. Didaticamente, torna-se

necessário que haja uma *vigilância epistemológica* na condução da aprendizagem, para assim tentar garantir não apenas uma colação de aquisições conceituais, mas o desenvolvimento de competências científicas de forma mais apropriada, e assim, quem sabe, encontrar indícios de ocorrências de uma “aprendizagem significativa”. Esse tipo de aprendizagem é facilmente suposta de forma equivocada quando se observa qualquer tipo de ganho após as instruções, porém dificilmente se confirma/verifica que *realmente* ocorreu quando se deixa de lado evidências concretas a médio ou longo prazo.

Entre os aspectos fundamentais que favorecem uma progressão da aprendizagem, encontram-se o papel dos erros e equívocos dos alunos, sendo recursos considerados relevantes a serem aproveitados e explorados na tomada de decisões do ensino nesse processo. Nisso, tem-se uma atenção maior para aqueles que envolvem concepções alternativas preexistentes. Essas concepções em grande parte estão relacionadas aos conceitos básicos de um curso. Engelhard (1997, p. 9, com base nos trabalhos de Wandersee et al., 1994) propõe algumas reivindicações necessárias a serem verificadas diante a essas concepções.

Quadro 8: Reivindicações a serem consideradas diante das concepções alternativas.

Reivindicação 1	Os alunos vêm para instrução formal da ciência com um conjunto diversificado de concepções alternativas relativa objetos e eventos naturais
Reivindicação 2	As concepções alternativas que os alunos trazer à instrução formal da ciência passa transversalmente entre a idade, capacidade, gênero e fronteiras culturais.
Reivindicação 3	As concepções alternativas são tenazes e resistentes à extinção por estratégias de ensino convencionais
Reivindicação 4	As concepções alternativas muitas vezes explicam paralelamente os fenômenos naturais oferecidos por gerações anteriores de cientistas e filósofos
Reivindicação 5	As concepções alternativas têm suas origens em um conjunto diversificado de experiências pessoais, incluindo observação e percepção direta, cultura e linguagem de pares, bem como nas explicações e materiais de instrução dos professores
Reivindicação 6	Os professores muitas vezes possuem as mesmas concepções alternativas que seus alunos
Reivindicação 7	O conhecimento prévio dos alunos interage com o conhecimento apresentado na instrução formal, resultando em um conjunto diversificado de resultados de aprendizado não planejados
Reivindicação 8	As abordagens de instrução que facilitam mudanças conceituais podem ser ferramentas efetivas na sala de aula

Fonte: Engelhard (1997, p. 9)

Para o enfrentamento dessas dificuldades, existem diversos mecanismos de ensino possíveis, e o papel dos *inventários de conceitos* pode vir a auxiliar o professor em decisões que visem as melhores instruções a serem deliberadas, não substituindo os outros mecanismos, mas aliando-se a eles. Em suma, esse processo deve ser realizado aplicando o teste precisamente no início do curso, no 1º ou 2º dia de aula, e volta-se para verificar conceitos básicos em uma etapa introdutória do curso, explorando as diferentes concepções diante desses conceitos, podendo ser reaplicado posteriormente nos últimos dias, e assim verificar se houve ganhos com a instrução e decisões tomadas durante o processo.

Essas instruções se referem a elaboração de estratégias de ensino mais eficientes, podendo considerar que as CAs se caracterizam como *obstáculos* ou *auxiliares* (Thouin, 2004).

A ideia de **obstáculo** na aprendizagem foi transposta e influenciada pelos trabalhos de Bachelard em 1938 sobre os obstáculos epistemológicos (Bachelard, 1996). As principais adaptações se referem a obstáculo (Thouin, 2004): *animista, classificativo, da univocidade das relações, tautológico, subjetivista, da unicidade dos pontos de vista, antropomórfico e substancialista*. Diversos trabalhos fazem referência a alguns desses obstáculos dentro do *ensino da Física Térmica* (Braga & Kalhil, 2011; Farias, 2013; Grings, Caballero, & Moreira, 2008; da Silva, 2013); *suas influências no ensino de ciências* (Gomes & Oliveira, 2007); *em livros didáticos* (Dominguini & Silva, 2010); relacionando a *concepção empirista-indutivista no ensino de ciências* (Köhnlein & Peduzzi, 2002), entre outros.

Sendo **auxiliares** (um auxiliar é qualquer coisa que ajuda) podem exercer o papel de facilitadores das aprendizagens, com certas concepções dos alunos consideradas em parte primitivas e em parte falsas – denominadas *primitivas fenomenológicas* ou «*p-prims*», permitem coordenar entre elas diversas observações e podem servir de base válida para a construção, por parte do aluno, de conceitos científicos, mais complexos. A aprendizagem situa-se então num plano, mais do que de ruptura, de continuidade em relação às concepções dos alunos (Thouin, 2004).

A preocupação com as concepções alternativas está no fato de que esses que aspectos formam a base estruturante das *teorias implícitas* na explicação do mundo físico, e acabam por estabelecer uma compreensão errônea e/ou equivocada diante dos modelos científicos ao interpretar a realidade. Entre alguns mecanismos de elaboração das concepções alternativas (Thouin, 2004, pp. 106-107), considera-se a:

- *Inferência*: consiste em passar de uma ideia para outra que é tida como pertinente devido à sua relação com a primeira;
- *Restrição*: refere-se à aplicação de um conhecimento a um domínio mais restrito do que aquele com o qual pode realmente estar relacionado;
- *Extensão*: consiste em aplicar um conhecimento a um domínio mais vasto do que aquele com o qual pode realmente estar relacionado;
- *Estabelecimento de uma Relação Direta entre Duas Ideias*: equívoco numa compreensão de proporcionalidade direta ou inversa entre duas grandezas ou conceitos que não existe, pois são independentes (explicação pessoal);
- *Formação de Uma Categoria Mental Geral*: equívoco na transferibilidade explicativa fenomenológica a vários objetos diferentes, seres vivos ou eventos na qual seja pouco provável (explicação pessoal).

As suas origens podem ser de natureza *sensorial, cultural e/ou escolar* (Pozo & Crespo, 2009), sendo um campo mais complexo e árduo de se verificar. Essas concepções se constituem como autênticas teorias implícitas, sendo altamente organizadas, tendo funcionalidade no cotidiano, na qual os conceitos e fenômenos científicos são produtos de uma aprendizagem informal ou implícita que tem como objetivo estabelecer regularidades no mundo, torna-lo mais previsível e controlável (Ibid., 2009).

Quadro 9: Processos e mecanismos explicativos sobre as origens das concepções alternativas.

CIÊNCIA INTUITIVA «dos alunos»: As Concepções Alternativas como Teorias Implícitas		
<p>ORIGEM SENSORIAL: Concepções Espontâneas (Significação direta em atividades cotidianas – <i>causalidade linear em primeira instância</i>)</p> <ul style="list-style-type: none"> • A <i>Semelhança</i> entre causa e efeito ou entre a realidade que observamos e o modelo que explicaria. • A <i>Contingência Espacial</i> e, se for possível, o contato físico entre a causa e efeito. • A <i>Contingência Temporal</i> entre causa e efeito, que devem suceder-se de modo próximo não apenas no espaço, mas também no tempo. • A <i>Covariação Qualitativa</i> entre causa e efeito. As variáveis relevantes serão aquelas que ocorram sempre que produz o efeito. • A <i>Covariação Quantitativa</i> entre causa e efeito, de modo que um aumento da causa produza um aumento proporcional do efeito, e vice-versa 	<p>ORIGEM CULTURAL: As Representações Sociais (origem linguística e cultural – <i>verbalizados facilmente</i>)</p> <ul style="list-style-type: none"> • A <i>Assimilação Interiorizada</i> por meio de crenças compartilhadas em grupos sociais • <i>Difusão de Concepções Alternativas</i> por diferentes meios de comunicação sem filtro (transmissão oral e veículos de divulgação científica) • <i>Conceptualização Diferenciada Estabelecida</i> do significado entre a linguagem cotidiana e os modelos científicos. • <i>Processo de Esquematização</i> onde as teorias científicas reduzidas a esquemas simplificados, usualmente reduzidos a uma imagem. • <i>Processo de Naturalização</i> concebida como parte da realidade ao invés do processo de construções sociais. 	<p>ORIGEM ESCOLAR: As Concepções Analógicas (o discurso científico confundido pelo conhecimento sensorial e social)</p> <ul style="list-style-type: none"> • <i>Visão de Ciência Distorcida por Ação ou Omissão</i>: Apresentações deformadas ou simplificadas de certos conceitos levam a uma compreensão errônea, desviada, por parte do aluno, que não fazem mais do que refletir a informação ou interpretação recebida. • O <i>Erro Conceitual</i> não esta apenas nos livros didáticos ou explicação recebida, mas de um erro didático na forma como lhe são apresentados os saberes científicos. • Uma <i>Incompreensão do Discurso Científico</i>, que se mistura com seu conhecimento sensorial e social. Modelos científicos (do macro ou microcosmo) misturam-se, tornam-se difusos, naqueles âmbitos do discurso cotidiano (mesocosmo) com referenciais comuns. Ou seja, se concebe como análogos sistemas de conhecimento que são complementares, mas diferentes (Ex: a <i>estrutura microscópica da matéria é atribuída propriedades macroscópicas, e vice-versa.</i>)

Fonte: Elaborado a partir de Pozo & Crespo (2009, pp.89-95).

Os livros didáticos, as discussões e argumentos defendidos nas inter-relações sociais e o discurso docente são alguns dos meios mais imediatos a se verificar dentro do meio escolar, enquanto as atividades experimentais investigativas, nas quais se exploram previsões e explicações dos alunos podem ser promissores para observar a influência sensorial, enquanto a cultural remete a identidade social, conjunto de valores e costumes praticados e compartilhados, parecendo ser mais complexo a ser investigado.

2.1.1 Fundamentos da Pedagogia da Integração

2.1.1.1 Saberes, conhecimentos, capacidades, objetivo específico e competências.

Estabelecer aquilo que se deva ensinar é considerado um dos problemas criteriosos e estruturantes para o ensino em ciências (de Carvalho, 2004), portanto, onde começar? Já foi levantada a questão de que os conteúdos em livros didáticos são *saberes escolares* consolidados e considerados “caixas pretas” (Latour, 1998), sendo os conhecimentos concebidos e defendidos conforme se veicula por certo regime de

autoridade. Esses saberes representam uma reprodução recodificada, textualizada e didatizada do *saber científico* construído em um coletivo de sujeitos e pensamentos da vanguarda da ciência. Obviamente que essa é uma questão de natureza epistemológica do conhecimento que circula consensualmente no meio escolar, diferentemente sobre do processo cognitivo da aprendizagem relativo as influencias que levam o aluno a compreender ou que se tornam obstáculos para a aprendizagem, em especial, diante da aquisição de conceitos básicos. Observa-se que o termo “saber” e “conhecimento” possuem diversos e diferenciados campos epistemológicos de investigação, sendo aparentemente sutil e simples esses distanciamentos, porém não menos profundo e relevante é a importância de deixar claro etimologicamente o sentido atribuído a cada um desses termos.

Na área da pesquisa em educação que envolve conteúdos é importante compreender a diferença entre os saberes e os conhecimentos, e do papel que representam diante da concepção de competências. Aparentemente, alguns documentos oficiais não demonstram ajudar nesse processo. O Parecer da Comissão Nacional de Educação Nº 16/1999, por exemplo emitiu um trecho afirmando que: “O conhecimento é entendido como o que muitos denominam simplesmente saber” (Brasil, 1999), sendo um documento importante como referência para a compreensão do conceito de competências dentro do sistema educacional brasileiro, entretanto considera equivalente os termos.

Concebe-se a acepção do termo *saber* como *conteúdo*, podendo estar em diferentes domínios do conhecimento quanto as suas formas ou ações de se processar. Os *saberes disciplinares* são aqueles relacionados ao contexto pedagógico, regidos pelo currículo, enquanto os *saberes adisciplinares* são aqueles adquiridos sem o propósito e vinculação didática, sem um compromisso de ditatização formal, disponíveis e adquiridos geralmente fora do contexto escolar, em meios e situações cotidianas. O *conhecimento* seria aquilo que se encontra na estrutura cognitiva dos alunos, ou seja, um *saber adquirido* e disponível cognitivamente.



Figura 8: Relação entre saberes e conhecimentos.

Fonte: Elaborado a partir de Roegiers & De Ketele (2004).

Há muita discussão entre um modelo comprometido no ensino de saberes e aquele que enfatiza o desenvolvimento de competências. Nesse sentido, talvez a principal diferença entre saber e competências é que o saber não se apresenta e nem necessita de ser concebido por meio de um contexto, enquanto as competências estão e devem estar associadas a um contexto e situação concreta.

Para compreender minimamente algumas das principais diferenças entre os diversos saberes, é que eles se encontram em níveis diferentes de complexidade, e que podem transitar em domínios distintos do conhecimento, como por exemplo, o saber reproduzir (*savoir-reproduire*) em um *nível mais superficial de abstração*, isso não significa que seja fácil reproduzir por exemplo atitudes, comportamentos ou executar um solo de uma guitarra, já que exige destreza no campo psicomotor. O saber-fazer (*savoir-faire*) estaria em um *nível mais intermediário* de apropriação do conhecimento e o saber-ser (*savoir-être*) em um *nível mais internalizado*. Como eles são transferíveis nos domínios, eles podem corresponder exclusivamente ao domínio o cognitivo, por exemplo, *saber-fazer cognitivo* (SFC), o correspondente no psicomotor seria o *saber-fazer gestual* (SFG), e no campo socioafetivo o *saber-fazer afetivo* (SFA).

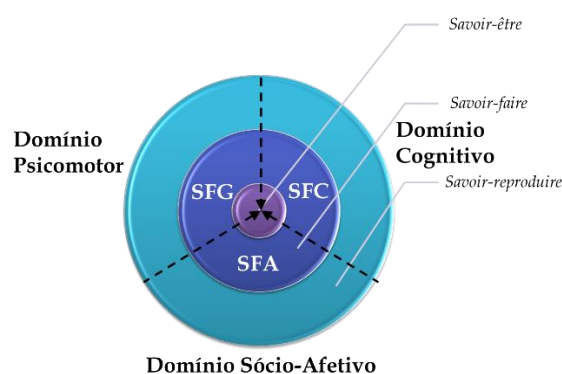


Figura 9: Relação dos saberes entre os diferentes domínios do conhecimento.

Fonte: Adaptado de Roegiers & De Ketele (2004).

As competências se fundamentam dentro de uma pedagogia de integração, e no que se refere a formação do sujeito, visam articular diferentes saberes e recursos e possibilitar uma progressão consistente e significativa do aprendizado. Para isso, diversos processos educacionais são considerados para atender essa perspectiva (organizacional dos conteúdos, estrutural da sequenciação didática e da aprendizagem, da gestão e avaliação), sendo a interdisciplinaridade uma aliada articuladora nesse processo. Existem alguns saberes que são considerados estratégicos nesse processo, como os *saberes metodológicos*, sendo transitórios, porém servem para auxiliar a aprendizagem, como regras didatizadas, artifícios matemáticos, etc., que visam contribuir na construção de *saberes finalizados* e na aquisição do conhecimento pelo sujeito, sendo úteis para abordar situações concretas vivenciadas (Roegiers & De Ketele, 2004). Esse é um argumento dentro da pedagogia de competências que denota a importância do papel de saberes básicos e específicos no processo de desenvolvimento de competências científicas, e dentro disso estão os conceitos de base dentro dos ramos do conhecimento, e juntamente, as concepções, ideias ou percepções dos sujeitos com relação a esses saberes.

A **capacidade** se define como o poder, a aptidão (habilidade) para **fazer algo**, ou seja, e a atividade que se exerce. Uma capacidade só se manifesta porque ela se aplica a conteúdos. Considera-se que as capacidades cognitivas, gestuais e socioafetivas não são mutuamente exclusivas. De modo geral as capacidades possuem quatro características principais (Roegiers & De Ketele, 2004, pp. 37-39):

transversalidade (são adisciplinares); *evolutividade* (evoluem em patamares diferenciados de aptidão e possuem maior eficácia no eixo do tempo); *transformação* (interagem e se combinam no *eixo das situações*); *não-avaliabilidade* (dificilmente se consegue avaliar devido à ausência do contexto).

O objetivo específico é uma especificação de atividade para um conteúdo, lembrando que uma mesma atividade pode ser exercida a respeito dos diferentes tipos de conteúdo. A partir de um mesmo conteúdo, é possível exercer diferentes atividades. O **objetivo específico** expressa a intenção que se tem de fazer com que o aluno exerça uma capacidade em determinado conteúdo (Roegiers & De Ketele, 2004).

Objetivo Específico «de saberes»= Capacidade + Conteúdo

Por exemplo: Aplicar (**Capacidade**) a 1ª Lei da Termodinâmica (**Conteúdo**).

Concebe-se o conceito de **competência** como a possibilidade para um indivíduo mobilizar, de maneira interiorizada, um conjunto integrado de recursos com vistas a resolver uma família de situações-problemas. De modo geral, as competências possuem as seguintes características (Roegiers & De Ketele, 2004, pp. 47-48): *mobilização de um conjunto de recursos* (conhecimentos, saberes, esquemas, automatismos, capacidades, etc.); *caráter finalizado* (não-gratuita, inseparável da capacidade de agir, função social – “portador de sentido” ou caráter significativo); *ligação com uma família de situações*³³ (restrita situações-problemas nas quais o aprendiz será solicitado a exercer a competência); *caráter frequentemente disciplinar* (relacionado ao um campo disciplinar, e por vezes, transferível quando há uma certa proximidade com outro campo); *avaliabilidade* (pode ser medida na qualidade da realização da tarefa e na qualidade do resultado)

Competência = Objetivo Específico «Capacidade x Conteúdo» + Situações

³³ Zabala & Arnau (2010, p. 30) atribui a contribuição desse conceito a Tremblay (1994) e Guy Le Boterf (2000), na qual se refere ao conjunto de possíveis situações-problema com características semelhantes às apreendidas, sendo cada situação singular, única e irrepetível.

Por exemplo: Aplicar (*Capacidade*) a 1ª Lei da Termodinâmica (*Conteúdo*) no desenvolvimento de um projeto de refrigeração geral da casa (*Situação*).

Exemplo de uma situação-problema favorável ao desenvolvimento de competências: “a partir de uma situação vivida ou de um desenho que evidencia diferentes problemas de poluição da água, do ar e sonora, o aluno deverá propor soluções apropriadas aos diferentes problemas, que será identificado previamente”.

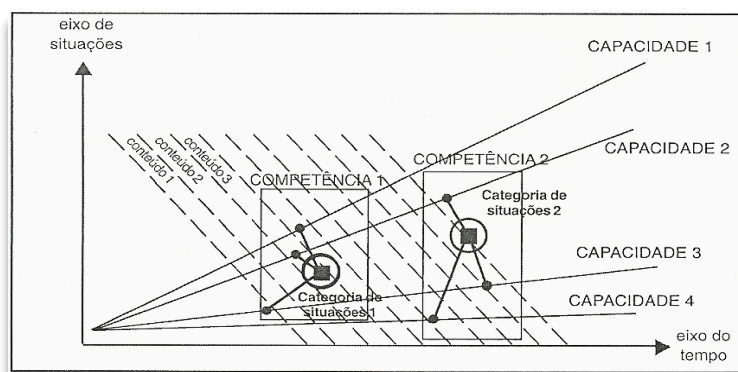


Figura 10: Esquemática dos conteúdos, das capacidades e das competências em função do eixo do tempo e do eixo das situações.

Fonte: Roegiers & De Ketele (2004).

A potencialidade de uma APC se caracteriza como um processo que estabelece uma relação mais flexibilizada entre a *construção do conhecimento* e a transposição desse conhecimento para ações cotidianas, ou seja, é uma perspectiva articuladora e integrada nos mais diversos campos de domínio do conhecimento na qual favorece e fundamenta processos educacionais interdisciplinares dando sentido ao sujeito em formação. Com isso, torna-se importante ressaltar que conhecimentos e competências são processos que se articulam, mas não se confundem, considerando como condição necessária para o desenvolvimento de competências a utilização de conhecimentos.

“Um ensino eficaz se basearia, então, em um equilíbrio sensato entre uma *aquisição de conhecimentos cuidadosamente selecionados*, um desenvolvimento de capacidades transversais por natureza, e um desenvolvimento de competências, mais direcionadas, centradas no reinvestimento, na prática de aquisições de todos os tipos: *conhecimentos, capacidades, automatismos, atitudes, atitudes e aquisições da experiência.*”

Para diferenciar capacidade de competência, consideram-se os seguintes aspectos.

Quadro 10: Diferenças entre um modelo de ensino genérico (baseado em capacidades) e um ensino operacional (baseado em competências).

Ensino generalista	Ensino Operacional
Capacidade	Competência
Desenvolve-se segundo o eixo dos tempos	Desenvolve-se segundo o eixo das situações
Evolutiva com o tempo	Interrompida em determinado momento
Relacionada a um conjunto não-limitativo de conteúdos	Relacionada a uma categoria determinada de situações
Atividade que pode ser realizada livremente	Atividade finalizada, que entra no âmbito de uma tarefa precisa
Caráter integrador, não necessariamente presente	Mobilização de um conjunto integrado de recursos , especialmente de capacidades
Especialização possível, mas em termos sensoriais e cognitivos	Especialização em termos de qualidade da execução de uma tarefa (“desempenho”)

Fonte: Roegiers & De Ketele (2004).

No âmbito das discussões sobre o problema do conteúdo dentro do processo de ensino-aprendizagem se pergunta a princípio: A escola deve desenvolver *capacidades* ou *competências*? Por um lado, existe o processo de *aquisição de saberes* arraigado a um currículo generalista voltado para uma *lógica longitudinal* onde se privilegia o desenvolvimento de capacidades e saberes metodológicos por meio da aquisição de conhecimentos e *savoir-faire*³⁴ gerais (capacidades), mas que, entretanto, abandona a questão do reinvestimento dessas aquisições, especialmente em práticas sociais. Essa lógica de aquisição dos saberes é algo redutor e ultrapassado, seja pelo lugar incontornável dos saberes ou pela quantidade astronômica de saberes necessários a serem desenvolvidos. **Defende-se** uma *Abordagem Por Competências*³⁵ (APC) dentro de uma *lógica operacional* onde o currículo seja mais específico e voltado para aquisição de competências, ou seja, preocupado no reinvestimento de tais conhecimentos e capacidades em práticas sociais portadoras de sentido para o aluno, cuja *interdisciplinaridade* constitui uma variante entre outras. Contudo, considera que os dois modelos não se opõem, mas se complementam, e o equilíbrio entre as abordagens constitui um dos maiores desafios para o sistema educacional no futuro, em especial

³⁴ Saber-fazer.

³⁵ Considerado o paradigma vigente predominante entre os diferentes sistemas e níveis educacionais, na qual advém (porém não se rompe) do paradigma da *Pedagogia Por Objetivos Educacionais*. Este modelo de abordagem que ganhou força com os trabalhos da equipe de especialistas liderados por Bloom a partir de 1956.

no ensino médio, pois no início de escolaridade trata-se mais nas aquisições de base (Roegiers & De Ketele, 2004, pp. 64-65).

As competências por sua vez podem ser diferenciadas entre competências de base e de aperfeiçoamento (De Ketele, 1996, *citado por* Roegiers & De Ketele, 2004). As **competências de base** nas aprendizagens se define como uma competência que deve ser necessariamente dominada pelo aluno para que possa entrar sem problemas em novas aprendizagens que a envolvem (Ibid, 2004). Esse tipo de competência está sempre relativo ao contexto e ao momento da formação, também pode ser denominada como *competência fundamental, competência essencial, competência mínima, competência-base, base de competências, etc.* Sugere-se que se proponha cerca de 5 ou 6 competências de base por ano e por disciplina, ou 8 a 10 competências de base por ano e por grupo de disciplina.

É importante determinar bem os patamares a serem alcançados, sobretudo nos países nos quais tenta-se diminuir os fracassos, que desencorajam os alunos e que custam muito caro ao sistema escolar. Exigir o domínio do conjunto das competências é seletivo demais; ao contrário, tomar como patamar uma média calculada sobre o conjunto das competências não garante aos alunos as aquisições necessárias para passar de ano.

Entre esses dois extremos, a exigência de êxito na obtenção de um número limitado de competências, as competências de base, parece que constitui uma solução interessante. (Roegiers & De Ketele, 2004, pp. 51-52)

Entende-se que as **competências de aperfeiçoamento** de extensão são aquelas mais abstratas, de maior profundidade. A falta de domínio não implica a incapacidade de continuar as aprendizagens, e estão mais relacionadas dentro um processo cumulativo e integrador, e pode se transformar em uma competência de base em outro contexto, dentro de uma perspectiva continuada e integradora (Ibid, 2004).

As competências, de modo geral, contribuem em três níveis (Ibid, 2004): *dar sentido às aprendizagens* (contextualizar *a posteriori* as aprendizagens; utilidade prática e concreta ao aluno); *tornar as aprendizagens mais eficazes* (aquisições bem fixadas; o essencial é ressaltado; diferentes aquisições são vinculadas umas às outras); *dar base as aprendizagens posteriores* (construção de um sistema global de aquisições sendo reinvestidas e colocadas a serviço de competências mais complexas).

Os principais conceitos-chave dentro do **processo de integração das aquisições** são: *Interdisciplinaridade; Transversalidade e Objetivo Terminal de Integração (OTI)*. O objetivo terminal foi um termo sugerido por De Ketele, e se refere a uma *macrocompetência* ou uma síntese integradora e mobilizadora de diversas competências, sendo estabelecida para o final de um ciclo de ensino-aprendizagem. As principais causas em uma APC nas aprendizagens se refere a: a *resistência a mudanças* (professores e encarregados de educação optam pelo processo tradicional); a *falta de formação* (professores, supervisores, etc.); *os aspectos institucionais* (estruturas de coordenação; “grades” de horários; modalidades de avaliação, etc.); *organização da aula* (alunos com oportunidades de exercerem suas competências) (Ibid, 2004).

2.1.2 Fundamentos dos Testes Conceituais

2.1.2.1 As concepções térmicas e os inventários de conceitos

O *Movimento das Concepções Alternativas (MCA)* iniciado na década de 70, com influência de alguns estudos como os de Laurence Viennot e Rosalind Driver, em que na década seguinte obteve impulso com Posner, Strike, Hewson, & Gertzog (1982), com repercussão em trabalhos de diversos pesquisadores renomados no Ensino da Física (além de outros campos), como, p. ex., nos de Lillian C. McDermott (Mcdermott, 1993; McDermott, Shaffer, & Constantinou, 2000; McDermott, 2006; McDermott & Redish, 1999; McDermott, 2001), entre outras produções.

Apesar desse movimento ter sido explorado de forma exaustiva quanto ao levantamento das diversas concepções entre as múltiplas áreas das Ciências, entende-se que essas concepções diante aos conceitos científicos não devam ser consideradas de forma pejorativa como “peças de museu”, exercendo a função de simples arcabouço de consulta, mas sim, de atribuir explicações e intervenções de forma lógica e consistente para o fenômeno da *mudança conceitual* em perspectiva atual, buscando ainda impulsionar novos desafios frente aos mecanismos tecnológicos de interação e socialização do conhecimento. O fato de conhecer as mais frequentes concepções não demonstra ser suficiente para refletir mudanças na realidade nas salas de aula (Köhnlein & Peduzzi, 2003), é preciso avançar identificando e contrastando diferentes

perfis conceituais, assim como o enfrentamento mais particularizado na aprendizagem, na qual envolvem questões relacionadas ao desenvolvimento de *competências docentes*, incluindo *meios e recursos adequados*, *designs* didáticos inovadores. O uso de computadores e dos *softwares* tornou possível realizar análises sofisticadas com menos bancos de dados que antes eram inviáveis, tanto qualitativamente como quantitativamente, dessa forma alguns procedimentos de análise podem ser diagnósticos mais precisos e relevantes para os processos *de aprendizagem*. Essa possibilidade não é tão recente, porém é um caminho que ainda necessita ser melhor explorado e utilizado entre pesquisadores com o uso de testes de compreensão conceitual.

De certa forma, o uso de testes tende a se concentrar na linha investigativa de resolução de exercícios escolares pelos estudantes ou de *resolver problemas*, e é considerado um mecanismo em potencial para consolidar habilidades cognitivas (Peduzzi, 1997) e de contribuir na avaliação do desenvolvimento de competências científicas. Mas é preciso delimitar, entretanto, que os testes conceituais (no que se refere a um inventário de conceitos) não são voltados para medir habilidades de resolução de problemas (Madsen, McKagan, & Sayre, 2014), mas sim para realizar sondagens de conceitos introdutórios, na intenção de que os diagnósticos possam auxiliar os professores na tomada de decisões diante as necessidades de desdobramentos para o ensino, podendo ainda estar a serviço do aluno, ajudando-o a regular e priorizar o aprendizado em determinados campos do conhecimento, e assim possibilitar um avanço dentro das *habilidades em resolver problemas*. É importante destacar que o fato do aluno desenvolver capacidades em resolver problemas, não significa que isso possa contribuir em seu desempenho diante de um inventário de conceitos (Redish & Burciaga, 2004).

Quanto ao conteúdo envolvido, algumas teorias buscam dar suporte e discutir a essa questão preocupando-se com a interface das relações entre *Ciência, Tecnologia, Sociedade e Ambiente* (CTSA), valorizando com isso temáticas transversais que favorecem o processo de articulação interdisciplinar, e assim, sendo relevante para o desenvolvimento de competências científicas básicas mais gerais, pois essas são mais

determinantes para que haja ou não um progresso nas aprendizagens (Roegiers & De Ketele, 2004).

[...] uma teoria cognitivista, que visa fornecer uma abordagem coerente e alguns princípios de base para o estudo do desenvolvimento e da aprendizagem das competências complexas, sobretudo aquelas relacionadas às ciências e à tecnologia [...]. (Fávero & Sousa, 2001, p. 153)

O modelo de Quebeque para crianças do ciclo primário exploram competências transversais como por exemplo: *Propor explicações ou soluções para problemas de ordem científica ou tecnológica* (Thouin, 2004). Esse tipo de competência é considerada muito controversa pela falta de uma delimitação de um conteúdo mais concreto para o seu exercício, tornando uma aprendizagem fragmentada (Rey, 1996).

Para uma *pedagogia baseada em competências* o **papel do conteúdo** tem seu destaque dentro de conjunto de situações-problemas que caracterizem as competências. A preocupação com o conteúdo que deva ser enfatizado tanto no ensino como na abordagem em livres textos, é uma questão estruturante em que se acredita que hajam razões lógicas para valorizar o papel das CAs nesse processo, sendo condição necessária de apropriação dentro do processo de aquisição das competências de base. Além disso, tem demonstrado sua importância ao longo dos anos e em diversos estudos, como por exemplo, na *teoria dos conteúdos escolares*, em que García Díaz (1998) define três níveis de formulação para conteúdos escolares: (1) Os dados a partir da análise científica e epistemológica dos conceitos científicos envolvidos; (2) Os dados a partir da investigação didática sobre as ideias dos estudantes e sobre as dificuldades de aprendizagem apresentadas por estes; (3) Os dados específicos sobre como evoluem as ideias de determinados alunos, obtidos no desenvolvimento do programa em sala de aula.

Pode-se notar uma preocupação com os conceitos científicos, explorar investigações sobre as CAs e identificar dificuldades, para assim priorizar conhecimentos de mais significância e necessidade, bem como vigiar essas compreensões visando uma progressão consistente na aprendizagem. Apesar de complexa, essa é uma questão superada no momento, sendo estabelecido dar uma

maior atenção sobre as CAs dentro do conhecimento térmico. O interesse em investigar o nível de compreensão de jovens diante os conceitos de calor e temperatura está presente literatura nos mais diversos contextos e níveis de escolaridade nas últimas décadas, inclusive para universitários em Portugal (Caldeira & Martins, 1990), bem como da abordagem desses conceitos nos livros didáticos (Leite, 1999), entre outros. Nisso, observa-se que os conceitos de calor e temperatura vêm sendo utilizados como parâmetros na elaboração de estratégias didáticas.

En la asignatura de Innovación Educativa se llevó a cabo la comparación de dos congresos de Didáctica de la Ciencias de 1985 y 1987 mostrando que los primeros didactas estudiaron principalmente la enseñanza de la Física dado que su explicación presenta más complejidad que la Química. Esta mayor complejidad puede ser debida a las interacciones del lenguaje cotidiano con el científico, **agravándose más en cuestiones como el Calor y la Temperatura**. Detectar esos errores así como diseñar una estrategia didáctica son los objetivos en lo que se ha basado mi proyecto de innovación. (Palma, 2012, p. 16)

Relacionado as necessidades de enfrentamento das CAs dentro do conhecimento térmico, foi proposto uma noção sobre perfil conceitual (E. Mortimer, 1995) aplicado ao conceito de calor (Amaral & Mortimer, 2001), sendo a análise do *perfil conceitual do calor*, metodologicamente consistindo nos seguintes aspectos: (1) *estudo da evolução histórica e epistemológica do conceito*; (2) *levantamento bibliográfico das concepções alternativas*; (3) *reinterpretação à luz da noção de perfil*; (4) *identificação das categorias do perfil*; (5) *pesquisa prática para validação do perfil*. São propostas as seguintes zonas para o perfil conceitual de calor (Araújo, 2014): (1) *calor como sensações térmicas*; (2) *calor animista*; (3) *calor como substância*; (4) *calor como temperatura elevada*; (5) *calor como energia*. Nesse âmbito, as diferenças entre ideias **animista** e **substancialista** podem ser feitas considerando que há apenas uma mudança dentro da mesma categoria ontológica (o calor visto como substância – matéria) quando se distingue entre *calor vivo ou inerte* (Chi, 1992, citado por Amaral & Mortimer, 2001)

O perfil conceitual mostra que existem zonas que transpassam o domínio epistemológico para o ontológico sobre as influências diante da compreensão do conhecimento científico. Sepulveda & Mortimer (2013) apresentam a construção de um modelo de «perfil conceitual» (E. Mortimer, 1994, 1995) baseado em quatro zonas

denominadas de *funcionalismo intraorgânico*, *ajuste providencial*, *perspectiva transformacional* e *perspectiva variacional*, na qual compromissos epistemológicos e ontológicos ajudam a estruturar distintos modos de pensar sobre o conceito, sendo essa uma forma alternativa de lidar e encarar com a ideia de *mudança conceitual*.

Acredita-se que esse campo de estudo possa ainda ser ampliado levando em consideração a possibilidade de elaboração de perfis conceituais com *base na análise fatorial*, juntamente com as discussões epistemológica e ontológicas associadas, sobretudo cruzando com as evidências empíricas advindas de estudos consistentes dentro das abordagens qualitativas. Dessa forma, possibilita um caminho de *confirmabilidade* e *consistência* a inferir sobre formas de pensar de diferentes grupos ou populações quanto a natureza epistemológica envolvidas nos conceitos e conhecimentos térmicos. A Psicometria sugere que alguns itens podem se comportar de forma estável para grupos, mas também pode identificar um comportamento diferenciado, tornando essas evidências importantes para a tomada de decisão diante a escolha de métodos e ênfase em determinados conteúdos. Ou seja, é uma *perspectiva generalizada* que pode ser enriquecida dentro de uma *perspectiva mais situada* dentro de um contexto escolar, levando em consideração métodos de registro de evidências apropriados e fundamentação teórica adequada.

A perspectiva metodológica de construção de escala de proficiências demonstra vantagens por permitir um aprimoramento da escala em níveis de refinamento e precisão, assim como a construção de uma análise longitudinal por meio de séries históricas de desempenhos. Nesse processo, além de possibilitar uma verificabilidade com relação as evidências anteriores, também poderá identificar eventualmente certo padrão de estabilidade de patamares de itens que representam competências específicas, que no caso, estão associadas as concepções térmicas.

A preocupação e esforço em identificar as concepções térmicas subjacentes aos modelos mentais incorretos indicado pelos distratores, está no fato de que essas concepções podem se manifestar na tomada de decisão de respondentes diante aos estímulos do teste. Sabe-se que o conhecimento e conceitos envolvidos encontram-se situados na problemática e contexto, bem como das concepções que a acompanha, e as

respostas tendem a carregar nas concepções envolvidas uma causa de natureza epistemológica que podem configurar uma característica predominante no *perfil de concepções de populações* ao serem verificadas suas incidências e atrações diante do percentual das respostas erradas, inclusive entre os níveis mais elevados de acerto em uma *Análise Gráfica do Item (AGI)*, mas de forma mais consistente através de modelos estatísticos de possibilitem estimar probabilidades de respostas e de comportamento, como é o caso dos procedimentos de análise mais rigorosos e robustos que a Psicometria oferece, que incluem os modelos da TRI. Dependendo da consistência dos resultados, essas análises também podem extrapolar para inferir sobre *perfis de concepções de indivíduos*, ampliando dessa forma as possibilidades avaliativas que auxiliam as tomadas de decisão no cenário pedagógico.

Acredita-se que as contribuições nesse campo poderão advir não apenas para o processo de construção de instrumentos e itens, como também em perspectivas de análises mais aprofundadas sobre as dificuldades dos alunos diante do conhecimento científico. Os resultados dos estudos locais também poderão ser comparados dentro de uma perspectiva mais generalizada, ao serem explorados métodos mistos para o enfrentamento dos macro fenômenos educacionais.

O processo de generalização com base na análise fatorial não é sinônimo de contrariedade ao processo de valorizar as dificuldades locais e individuais dos alunos, como acredita-se que muitos pensam. Essas análises produzem informações que possibilitam contribuir na avaliação e na instrução de grupos e alunos com características específicas, porém, deve ser encarado como um indicador, e não um decisor. Além disso, as implicações de generalidade podem ser promissoras ao possibilitarem verificações e proposições de subsídios que busquem favorecer um *alinhamento curricular*, seja do ponto de vista de políticas públicas de formação docente, como em metodologias mais apropriadas, e no processo de inclusão de conteúdos prioritários advindos das análises sobre as concepções errôneas e necessidades dos alunos, podendo obviamente ser estendido a todas as áreas do conhecimentos e disciplinas. Os resultados também podem servir para construir evidências a serem contrapostas ao *status quo* do paradigma curricular vigente, assim como servir de

forma mais pragmática as unidades de ensino quando busca atender os interesses dos projetos políticos pedagógicos nas escolas.

Constatando algumas das diversas concepções térmicas identificadas na literatura, pode-se observar (quadro) um campo de ideias primitivas que podem e devem ser exploradas dentro de novas perspectivas de análise. Para isso, foi feito um recorte criterioso na compilação de Thouin (2004) sobre as CAs térmicas que advêm de outras obras (Champagne, 1992; De Vecchi, 1992; Driver, 1989; Duit, 1991; Giordan, Astofi, Gohau, & Host, 1978; Joshua & Dupin, 1993; Laurence, 1979; Resnick, 1989; Stepans, 1996; Thouin, 2001). Foram destacadas (**negrito e sombreado**) as concepções térmicas correspondentes no teste TCE.

Quadro 11: Concepções térmicas mais frequentes.

Calor	Mecanismos de Elaboração	Conceitos científicos	Exemplos de atividades
Uma garrafa térmica contém uma fonte de calor. (0)	Inferência baseada no fato de uma garrafa térmica manter uma bebida quente durante várias horas.	Uma garrafa térmica é um péssimo condutor e retém o calor.	Observar o interior de uma garrafa térmica quebrado.
A dispersão de uma gota de colorante ocorre da mesma maneira em água fria e em água quente. (0, 1)	Formação de uma categoria mental geral para todas as dispersões.	Uma gota de colorante dispersa-se mais depressa em água quente do que em água fria.	Deixar cair algumas gotas de colorante em água fria e em água quente.
Os materiais quentes contêm bolhas de ar quente e os materiais frios contêm bolhas de ar frio (1)	Inferência baseada na impressão de que o calor é um fluido.	O calor não é nem um gás nem um líquido, mas uma forma de energia.	Observar recipientes com óleo quente e óleo frio.
O gelo derrete mais rapidamente no ar a 10 °C do que na água a 10 °C. (1)	Inferência baseada na impressão de que os gases aquecem melhor do que os líquidos.	O gelo derrete mais rapidamente na água a 10°C do que no ar a 10 °C.	Deixar derreter gelo no ar e na água à mesma temperatura.
O metal desliza melhor no gelo do que na borracha. (1,2).	Inferência baseada no fato de que os patins de gelo deslizarem melhor do que botas	O metal não desliza melhor do que a borracha, mas a lâmina dos patins desliza bem por ser muito fina e fazer derreter um pouco de gelo sob o peso do patinador, o que lubrifica a superfície	Fazer deslizar sobre o gelo um objeto de metal e outro de borracha, ambos com forma e peso quase iguais.
Sai calor de uma panela de água a ferver. (1)	Formação de uma categoria mental geral para calor e o vapor de água	É vapor de água que sai de uma panela de água a ferver	Colocar um prato frio por cima de uma panela de água a ferver
Os objetos metálicos são mais frios do que os outros. (1,2)	Inferência baseada na sensação de frio causada pelo toque nos objetos metálicos.	Os objetos metálicos não são mais frios. Parecem mais frios ao toque porque são bons condutores e deixam escapar o calor da mão.	Colocar um termômetro em diversos objetos metálicos e não metálicos situados no interior de uma mesma divisão.
Materiais como o alumínio são bons isolantes térmicos. (1,2)	Inferência baseada na sensação de frio do toque nos metais. Inferência baseada na utilização frequente de papel de alumínio para conservar alimentos frescos	O alumínio é um péssimo isolante térmico, não impedindo o aquecimento dos objetos ou alimentos que envolve	Envolver um cubo de gelo ou um objeto muito quente em diversas embalagens (papel de alumínio, papel de jornal, tecido, etc.)
Todos os sólidos conduzem o calor da mesma maneira. (1)	Formação de uma categoria mental geral para a condutividade térmica de todos os sólidos	Certos sólidos, como os plásticos são isolantes; outros, como os metais, são condutores	Tocar em colheres de madeira, plástico, aço e prata mergulhados em água muito quente.
A temperatura de um objeto depende do seu tamanho e da sua massa. (1)	Estabelecimento de uma relação direta entre o tempo que um objeto leva a aquecer ou a arrefecer e a sua temperatura.	Não há uma relação entre a temperatura de um objeto e o seu tamanho ou a sua massa.	Medir temperatura de um grande pedaço e de um pequeno pedaço de gelo tirados do mesmo congelador
A temperatura da mistura de duas quantidades de água é igual à soma das temperaturas das duas quantidades. (1)	Estabelecimento de uma relação direta entre o acréscimo de uma quantidade de água a uma outra e a soma das duas temperaturas.	A temperatura da mistura de duas quantidades de água é igual à média ponderada das duas quantidades.	Medir a temperatura da mistura de urna pequena quantidade de água a 10 °C com uma grande quantidade de água a 50 °C.
Todas as substâncias se contraem ao passarem do estado líquido para o estado sólido. (2)	Inferência baseada na observação da cera ao solidificar-se.	É verdade que quase todas as substâncias se contraem, exceto a água e o bismuto.	Medir o volume de uma quantidade de cera líquida e sólida.

Todas as substâncias se dilatam ao passarem do estado líquido para o estado sólido. (2)	Inferência baseada na observação da água ao gelar.	É verdade que quase todas as substâncias se dilatam, exceto a água e o bismuto.	Medir o volume de uma quantidade de água antes e depois de a fazer gelar.
Um balão cheio no interior de um edifício, a uma temperatura de 20 °C, conservará o mesmo volume que no exterior, independentemente da temperatura. (1, 2)	Inferência baseada na impressão de que um volume de gás permanece constante.	O volume do balão diminuirá se estiver mais frio no exterior e aumentará se aí estiver mais calor.	Encher um balão no interior de um edifício, a uma temperatura de cerca de 20 °C, e colocá-lo depois no exterior, a uma temperatura mais baixa ou mais alta.
No verão, trememos de frio após um banho porque a água da piscina está mais quente do que o ar. (1)	Estabelecimento de uma relação direta entre os tremores de frio e a temperatura do ar.	Trememos de frio sobretudo por a evaporação da água arrefecer a pele.	Atentar à sensação causada por algumas gotas de álcool isopropílico ou águas colocadas na palma da mão.
É impossível que esteja mais calor ou mais frio do que as temperaturas indicadas num termómetro doméstico. (1)	Estabelecimento de uma relação direta entre a temperatura mais quente e mais fria indicada por um termómetro e temperaturas máximas e mínimas absolutas.	Existem temperaturas muito mais frias e muito mais quentes do que as indicadas pelos termómetros domésticos,	Observar um termómetro concebido para medir temperaturas elevadas ou baixas.
O calor é uma substância que se desloca de um ponto para outro. (2)	Extensão ao calor da observação da água que corre nos tubos. Inferência baseada na linguagem corrente («fazer sair o calor»).	O calor não é uma substância, mas uma forma de energia que se manifesta através da agitação dos átomos ou das moléculas.	Ver um documentário ou ler um texto sobre a forma como o Sol aquece a Terra.
O calor comporta-se como um líquido: acumula-se num local até enchê-lo e corre depois para outros locais. (2)	Extensão ao calor da observação da água que corre nos tubos. Inferência baseada na linguagem corrente («fazer sair o calor»).	O calor não é um líquido, mas uma forma de energia.	Pesar um objeto frio, aquecê-lo e voltar a pesá-lo.
Calor e temperatura são sinónimos. (2, 3)	Formação de uma categoria mental geral para o calor e a temperatura.	O calor é uma forma de energia, enquanto que a temperatura é a medição da agitação dos átomos e das moléculas de uma substância.	Medir a temperatura de uma certa quantidade de gelo aquecido para o derreter.
Os metais atraem mais o calor do que as outras substâncias. (2)	Inferência baseada no facto de os metais serem bons condutores de calor.	Os metais não atraem calor, mas são melhores condutores do que muitos outros materiais.	Medir a temperatura de objetos metálicos e não metálicos colocados numa mesma divisão.
Um chão revestido com carpete é sempre mais quente do que um chão de ladrilhos de cerâmica. (2)	Inferência baseada na impressão de que os materiais são, por natureza, quentes ou frios.	Em muitos casos, o chão revestido com carpete não é mais quente, mas, por ser pouco condutor, permite uma menor fuga de calor dos nossos pés.	Medir, com a ajuda de um termómetro, a temperatura de diversos pisos revestidos com carpete e ladrilhos de cerâmica.
A cor e a espessura de um material influenciam a sua condutividade térmica. (2)	Estabelecimento de uma relação direta entre certas características do um material e a sua condutividade térmica	A condutividade térmica depende sobretudo da natureza do material.	Verificar como é que objetos feitos de diversos materiais, e de diversas cores e espessuras, conduzem o calor.
A dilatação de um corpo é causada pela dilatação dos átomos ou das moléculas desse corpo. (2, 3)	Extensão ao nível atómico ou molecular de um fenómeno observável ao nível macroscópico.	Os átomos ou as moléculas permanecem idênticos, mas o espaço entre eles aumenta.	Medir o volume de um balão no exterior, num dia frio, e medir depois o volume do mesmo balão no interior de uma divisão bem aquecida.
As moléculas da água modificam-se quando o gelo funde. (2)	Extensão ao nível molecular de um fenómeno observável ao nível macroscópico.	As moléculas de água permanecem idênticas, mas deslocam-se livremente umas em relação às outras quando o gelo funde.	Ler um texto sobre as alterações físicas
Numa casa, os estalidos ouvem-se devido ao envelhecimento dos materiais. (2)	Extensão baseada numa das razões pelas quais a superfície de certos materiais estala.	Os estalidos são causados pela dilatação ou contração dos materiais, causada pelas mudanças de temperatura.	Com o tempo muito frio, ouvir os ruídos de uma casa nova.
As juntas do vão de uma ponte servem para	Restrição baseada na observação de certos	As juntas permitem sobretudo a dilatação e a	Observar as juntas do vão de uma ponte

fazer encaixar as secções de um na outra. (2)	jogos de construção.	contração das secções do vão durante as mudanças de temperatura.	com o tempo quente e o tempo frio.
O aumento ou a diminuição da quantidade de calor provoca sempre um aumento ou uma diminuição da temperatura (2, 3).	Inferência baseada na observação da temperatura de água aquecida ou arrefecida a temperaturas distantes dos pontos de fusão ou ebulição.	Durante as mudanças de estado, a temperatura mantém-se a mesma, embora a substância ganhe ou perca calor.	Medir a temperatura da água prestes a ferver ou prestes a gelar.
É impossível ferver água num copo de papel. (2, 3)	Restrição baseada na temperatura de ignição relativamente baixa do papel.	Quando há água num copo de papel, a sua temperatura não passa dos 100 °C, pelo que não arde.	Colocar um copo de papel com cerca de 20 ml de água por cima de uma vela.

Termodinâmica	Mecanismos de Elaboração	Conceitos científicos	Exemplos de atividades
A quantidade de energia utilizável disponível após uma transformação de energia é a mesma que antes da transformação. (2, 3)	Inferência baseada na impressão de que não há ou há muito pouca energia dissipada sob a forma de calor.	Embora a quantidade total de energia seja a mesma após uma transformação, parte importante da mesma assume a forma de calor e não é facilmente utilizável.	Ligar um pequeno motor elétrico acionado por gerador manual idêntico ao motor
É possível fabricar dispositivos que funcionam sem parar, mesmo sem motor (dispositivos de movimento perpétuo). (2,3)	Extensão baseada na observação de pêndulos e moinhos.	Nenhum dispositivo de movimento perpétuo funciona, pois, parte da energia dissipa-se sempre sob a forma de calor.	Tentar fabricar dispositivos de movimento perpétuo.

As técnicas do calor	Mecanismos de Elaboração	Conceitos científicos	Exemplos de atividades
Qualquer aquecimento implica a existência de uma combustão. (1)	Inferência baseada na observação do fogo de uma lareira ou alimentado a gás.	Vários sistemas de aquecimento funcionam através de elementos elétricos.	Examinar um sistema de aquecimento elétrico.
Um forno de micro-ondas aquece com a ajuda de uma luz de aquecimento. (1)	Estabelecimento de uma relação direta entre o funcionamento de uma incubadora e o de um forno de micro-ondas.	Um forno de micro-ondas aquece os alimentos fazendo vibrar as moléculas de água destes últimos.	Aquecer alimentos num forno de micro-ondas sem lâmpada.
Um frigorífico possui uma potente ventoinha para arrefecer os alimentos. (1)	Inferência baseada na sensação de frio sentida perante uma ventoinha.	Num frigorífico, é a dilatação de um gás comprimido que arrefece os alimentos.	Examinar o interior e o exterior de um frigorífico.
Uma panela de pressão cozinha rapidamente os alimentos graças à sua espessa parede metálica. (2)	Inferência baseada no fato de o metal ser um bom condutor de calor.	Uma panela de pressão cozinha rapidamente os alimentos porque a pressão e a temperatura no seu interior são mais elevadas do que no interior de uma panela normal.	Observar a forma como a panela de pressão é vedada, bem como os dispositivos de segurança que possui.

Legenda: (0) mais frequente no pré-escolar (4-5); (1) mais frequentes no 1º ciclo primário (6-7); (2) mais frequentes no 2º ciclo primário (8-9); (3) mais frequentes no 3º ciclo primário (10-11). *em Quebeque. Fonte: Recorte feito a partir de Thouin (2004, tradução livre).

No MCA, os testes conceituais talvez foram um dos instrumentos mais utilizados para tentar identificar e mapear o conhecimento prévio dos alunos, assim como formular instruções, alguns mais concentrados na realização das tarefas dos alunos, outros no papel do ensino e das operacionalizações a propor e conduzir em situações didáticas, assim como no problema do conteúdo a ser ensinado. O conteúdo era um dos maiores enfoques, compreender a sua natureza epistemológica e identificar quais os conceitos estariam ligados a uma compreensão ingênua dos alunos. Os testes ganharam um destaque com o auxílio dos *Inventários de Conceitos*, sendo um movimento entre os métodos de avaliação que tomou raízes a partir do trabalho de David Hestenes (1992) e colaboradores ao criarem o *Inventário de Conceitos de Força* (FCI), um instrumento para medir a compreensão dos alunos sobre os conceitos fundamentais da força newtoniana, como foi dito anteriormente.

Testes conceituais estão relacionados com o termo “Inventário de Conceitos” (IC), sendo geralmente de múltipla escolha, e são utilizados para sondar a compreensão dos alunos sobre conceitos particulares de física (Madsen et al., 2014). Existem atualmente mais de 30 inventários de conceitos para vários tópicos introdutórios e de nível superior em física e astronomia (Ibid., 2014). Nesse âmbito é importante destacar alguns aspectos fundamentais na elaboração de IC, bem como algumas instruções para administrar sua aplicação.

Quanto à elaboração, pode se destacar as opiniões dos alunos nesse processo de modo a elaborar questões que sejam pertinentes e significativas aos sujeitos. Além disso, utilizar múltiplas formas de analisar para aumentar a fiabilidade nas conclusões.

Quadro 12: O processo típico para desenvolver um inventário de conceitos.

1º - Reúna as ideias dos alunos sobre um determinado tópico, geralmente com entrevistas ou perguntas escritas abertas.

2º - Use as ideias dos alunos para escrever questões conceituais de múltipla escolha, onde as respostas incorretas cobrem o leque de ideias incorretas mais comuns dos alunos. Use a redação atual dos alunos.

3º - Teste essas questões com outro grupo de estudantes. Geralmente, os pesquisadores usam entrevistas onde os alunos falam sobre seu pensamento para cada pergunta.

4º - Reveja as perguntas com base nos comentários dos alunos.

5º - Administrar o inventário de conceitos para um grande número de alunos. Verifique a reprodutibilidade dos resultados em cursos e instituições. Verifique as distribuições de respostas. Use vários métodos estatísticos para garantir a confiabilidade da avaliação.

6º - Reveja novamente.

Fonte: Madsen et al. (2014, baseado em Adams & Wieman, 2011; Redish & Burciaga, 2004, *tradução livre*).

Diferentemente de testes de aferição do conhecimento, na qual visam geralmente em verificar o domínio dos alunos num conjunto acumulativo de conhecimentos, como no caso dos exames nacionais, os ICs buscam se caracterizar como uma sondagem diante de conhecimentos essenciais para uma unidade temática ou ramo do conhecimento, nesse caso pode exercer uma função preditora, previsiva ou diagnóstica estando a serviço do professor. Ao ser aplicado posteriormente (entre 4 a 6 meses, de preferência) se torna um recurso de avaliação mais abrangente, podendo indicar sobre a eficiência de uma intervenção de ensino. Ainda é preciso verificar um *checklist* de riscos associados a validade para esse tipo *design* (ver no Apêndice C, p. 227).

Quadro 13: Práticas recomendadas gerais para administrar inventários de conceitos.

-
- Quando você dá a prova aos seus alunos, dê-lhe um título genérico, como "levantamento de mecânica" para evitar que os alunos busquem e encontrem as respostas.
 - Dê aos alunos o tempo recomendado para fazer o teste.
 - Consulte as diretrizes específicas do teste na avaliação através de um guia especializado e siga-as. Por exemplo, alguns testes só devem ser dados para pós-testes.
 - Para obter a participação máxima de seus alunos, faça do teste um requisito e dê um pequeno número de pontos para completar o teste. Os alunos não devem receber notas para a correção de suas respostas.
 - Suponha que esta avaliação tenha sido concebida para avaliar o instrutor e o uso do currículo do instrutor, e não a força de qualquer aluno individual. Deixe claro que seus resultados não influenciarão suas notações. Deixe-os saber que não se espera que conheçam as respostas corretas, pois a intenção é que apenas você saiba como eles pensam sobre essas questões. Eles devem fazer o melhor possível se não tiverem certeza sobre uma pergunta e devem responder a todas as questões.
 - Para fazer comparações com outras classes, o mais significativo, dar o teste na sua totalidade e com a redação original e a ordem das perguntas.
 - Certifique-se de que você dá o pré-teste antes de cobrir qualquer material de curso relevante para que você capture com precisão o conhecimento recebido dos alunos e não o que eles aprenderam nas primeiras sessões de aula.
-

Fonte: Madsen et al. (2014, baseado em Adams & Wieman, 2011; Redish & Burciaga, 2004, *tradução livre*).

Testes conceituais na configuração de inventário de conceitos podem ser considerados como testes cognitivos educacionais diagnósticos, sem a pretensão de atribuição de notas visando classificar os alunos por *ranking*. Mais precisamente um teste é projetado com critério-referenciado para ajudar a determinar se um aluno tem um conhecimento de trabalho exato de um conjunto específico de conceitos. A concepção de *Testes Referenciados por Critério* (TRC) tem sua origem nos trabalhos de Flanagan (1951) e Nedelsky (1954) ao introduzirem o conceito de *padrão absoluto e relativo* quanto

a pontuação de um teste, enquanto a denominação “*Testes Referenciados por Critério*” se deve a Ebel (1962) e sua diferenciação com relação aos *Testes Referenciados por Norma* (TRN) estabelecida pelo especialista em *design instrucional* Robert Glase em 1963 (García, Abad, & Tello, 2015). Contudo, a definição dos TRC não constitui um marco teórico dentro da teoria dos testes, mas sim um *novo enfoque* que responde a perguntas e necessidades distintas dos TRN (Ibid, 2015). Robert Glase os define originalmente da seguinte forma:

Quanto a medida de um *Teste Referenciado Por Critério*:

Subjacente ao conceito de medida de realização é a noção de um *continuum* de aquisição de conhecimento que varia de nenhuma habilidade para o desempenho perfeito. O nível de realização de um indivíduo cai em algum ponto sobre este *continuum* como indicado pelos comportamentos que ele exibe durante o teste. O grau em que sua conquista se assemelha ao desempenho desejado em qualquer nível especificado é avaliado por medidas de realização ou proficiência referenciadas pelo critério. O padrão contra o qual o desempenho de um aluno é comparado quando medido dessa maneira é o comportamento que define cada ponto ao longo do *continuum* da realização. O termo "critério", quando usado dessa maneira, não se refere necessariamente ao comportamento final de fim de curso. Os níveis de critério podem ser estabelecidos em qualquer ponto da instrução onde é necessário obter informações sobre a adequação do desempenho de um indivíduo. O objetivo é que os comportamentos específicos implementados em cada nível de proficiência podem ser identificados e usados para descrever as tarefas específicas que um aluno deve ser capaz de realizar antes de alcançar um desses níveis de conhecimento. É nesse sentido que as medidas de proficiência podem ser avaliadas por critério. (Glaser, 1963, p. 519, *tradução livre, grifo nosso*)

Quanto a medida de um *Teste Referenciado Por Norma*:

Nos casos em que a posição relativa de um estudante ao longo do *continuum* de realização é o principal objetivo da mensuração, não é necessário fazer referência ao comportamento do critério. Os exames de realização educacional, por exemplo, são administrados com frequência para solicitar estudantes em uma aula ou escola, em vez de avaliar a obtenção de objetivos específicos do currículo. Quando essas medidas referenciadas por norma são usadas, a realização de um determinado aluno é avaliada em termos de comparação entre seu desempenho e o desempenho de outros membros do grupo. Tais medidas precisam fornecer pouca ou nenhuma informação sobre o grau de proficiência exibida pelos comportamentos testados em termos do que o indivíduo possa fazer. Eles contam que um aluno é mais ou menos proficiente do que outro, mas não diz o quanto ele é competente em relação às tarefas de assunto envolvidas. (Glaser, 1963, p. 520, *tradução livre, grifo nosso*)

2.1.2.2 Situando o papel da competência científica em testes conceituais

"Os fenômenos humanos são biológicos em suas raízes, sociais em seus fins e mentais em seus meios". Jean Piaget (9 de agosto de 1896 - 16 de setembro de 1980)

A palavra competência (derivada do latim "*competentia*" - de "*competere*": estar em gozo ou no uso de) pode significar idoneidade na avaliação de algo ou alguém, aptidão de resolver problemas e cumprir tarefas, capacidade de exercer determinada função, sinônimo de cultura, conhecimento e jurisdição.

"[...] terá surgido pela **primeira vez** na língua francesa, no século XV, designando a legitimidade e a autoridade das instituições (por exemplo, o tribunal) para tratar de determinados problemas. No século XVIII amplia-se o seu significado para o nível individual, designando a capacidade devida ao saber e à experiência". (Dias, 2010, p. 74)

Atualmente o termo competência é *polissêmico*³⁶, podendo ter atribuições específicas nas diversas áreas do conhecimento, destacando similaridades e diferenças entre outros termos e construtos como capacidade, proficiência, desempenho, inteligência, habilidade e conhecimento (Condé, 2008). Contudo, a maioria dos sistemas educacionais se baseiam em um domínio curricular na perspectiva do conceito de competência, seguindo uma tendência global em estabelecer *competências gerais comuns* a serem desenvolvidas entre os diferentes sistemas educacionais visando alcançar uma integração de múltiplos interesses em comum, sobretudo, diante à dimensão **educacional** e do **mercado de trabalho**. Esses dois campos encontram-se em zona de conflito de interesses e perspectivas sobre as competências, sendo as noções estruturantes do modelo das competências no mundo do trabalho concentradas mais na flexibilidade, a transferibilidade, a polivalência e a empregabilidade (Deluiz, 2001), porém se busca congruência com a visão educacional, pois encontram-se inevitavelmente *lado a lado* em meio a uma disputa ideológica quanto ao direcionamento do processo formativo. Justifica-se que a importância deste **paradigma das competências** resulta do processo de globalização e modernização na qual cria um mundo cada vez mais diversificado e interligado (OCDE, 2005, p. 3).

³⁶ A polissemia, ou *polissêmica lexical* (do grego *poli*: "muitos"; *sema*: "significados"), é o fato de uma determinada palavra ou expressão adquirir um novo sentido.

Entende-se que o conceito de competência na educação ampliou em *novos significados*, e vem superando a resistência da concepção de competência com base na crença que seu uso levaria ao empobrecimento da formação, que ficaria reduzida à aquisição de habilidade e de automatismos pelo adestramento externo (Suñé, Araújo, & Urquiza, 2015, p. 14). Esse entendimento tem raiz histórica na crítica ao behaviorismo³⁷ como princípio formativo (Rué, Almeida, & Arantes, 2009, citado por Suñé et al., 2015).

É próprio desse tipo de confusão estabelecer uma posição oposição entre competências e conteúdos disciplinares como duas formas opostas de adquirir conhecimento, como se as competências tivessem de se desenvolver em detrimento dos conteúdos disciplinares e vice-versa. A versão atual do que é competência, entretanto, integra ambos. (Rué, Almeida, & Arantes, 2009, p. 39)

Ao restringi-lo dentro de uma perspectiva educacional no que diz respeito a avaliar o conhecimento, e não os processos mentais propriamente ditos e explorados como fazem na psicologia experimental, entende-se que testes conceituais ao serem caracterizados como *Inventário de Conceitos*, não devam ser utilizados ou recomendados para avaliar diretamente competências científicas. Apesar disso, os resultados que advêm de suas análises servem de indicadores complementares como evidências empíricas em dois pontos principais nesse âmbito:

- Em um pré-teste como **avaliação diagnóstica** sobre o domínio do conhecimento básico e das necessidades na ênfase do ensino diante de conceitos introdutórios em um campo específico do conhecimento, se caracterizando assim como uma sondagem do conhecimento prévio e pontos-chave a serem desenvolvidos dentre as competências científicas preestabelecidas;
- Em um pós-teste como **avaliação verificativa** sobre a qualidade do processo de desenvolvimento das *competências científicas de base conceitual* que supostamente tenha sido realizado durante as instruções do ensino, ou seja, sendo uma verificação se houve

³⁷ Historicamente na psicologia, o empirismo passou a ser chamado de comportamentalismo (behaviorismo) e associacionismo, em razão dos condicionamentos e das associações estímulo-resposta (Matui, 1995, p. 40).

ganhos na aprendizagem (ou não) diante da forma de tratamento ou modelo de ensino adotado.

Algumas das principais funções dos testes no que diz respeito a sua aplicação em meio ao contexto escolar se resume em (Laveault & Grégoire, 2014):

- Fazer um balanço das aquisições do aluno;
- Tomar uma decisão visando a promoção do aluno;
- Selecionar estudantes de acordo com características particulares para formar grupos;
- Identificar aspectos dentro do processo de resolução de problemas que causam dificuldades;
- Identificar transferências que tenham ocorrido ou não tenham ocorrido;
- Preparar uma revisão sobre o tópico de alguns conteúdos para os quais alguns alunos têm dificuldade;
- Conscientizar os alunos diante de alguns pontos principais relacionados ao conteúdo.

Esta lista não é exaustiva. Ele simplesmente ilustra dois grandes conjuntos de situações em que os testes desempenham um papel importante na situação escolar: *avaliação sumativa* (p. ex.: teste certificativo); a *avaliação formativa* (p. ex.: teste diagnóstico) (Laveault & Grégoire, 2014). Como foi destacado anteriormente, um inventário de conceitos se concentra dentro de uma *função diagnóstica* que pode estar inclusa tanto em uma avaliação de orientação quanto na avaliação formativa (De Ketele & Roegiers, 1993), ainda que seja reaplicado após as instruções através de um pós-teste, exercendo uma *função um pouco mais verificativa*, entretanto se restringe ao conhecimento envolvido no teste que permanece envolvendo conceitos introdutórios, e não o conhecimento acumulado durante o processo da aprendizagem. Isso não impede a possibilidade de ser elaborado, aplicado e analisado um teste em larga escala com uma *função certificativa* dentro de uma perspectiva avaliativa sumativa, sendo acumulativa, visando avaliar um ciclo de formação, geralmente aplicado na etapa final de uma aprendizagem, como é o caso dos exames nacionais e alguns internacionais. Porém, esse tipo de função não é recomendada para os testes conceituais, e se forem

realizados, que sirvam apenas como um indicador ou decisor de baixo risco.

É preciso delimitar que o uso de **testes** como método de recolha não seja apropriado em avaliar o âmbito em que se *manifestam e se desenvolvem as competências* no sentido mais lato, apesar de se possível identificar aspectos relacionados as capacidades cognitivas que se relacionam com destrezas e outras capacidades operacionais. Entende-se que os resultados de um teste aplicado em larga escala podem carregar informações implícitas e significativas que caracterizam o coletivo de pensamento de uma população com relação as suas dificuldades conceituais e concepções diante dos conteúdos, podendo ser considerados conteúdos relevantes a serem valorizados e tratados de forma diferenciada dentro das salas de aula.

A literatura atual disponibiliza as principais CAs dos alunos, e ao serem consideradas critérios referenciados na construção de um teste conceitual, como no caso do TCE, se tornam aspectos que caracterizam uma parte fundamental dentro dos *traços latentes* a ser medido, verificado, comparado, associado e discriminado em patamares diferenciados quanto aos níveis de dificuldades, demonstrando assim um aspecto de relevância. A interpretação desses aspectos representam necessidades de compreensão a serem desenvolvidas no âmbito das competências científicas como capacidades cognitivas em potencial para executar tarefas, resolver conjunto de situações-problemas específicos, além de contribuir em uma progressão da aprendizagem para níveis mais complexos. Diante disso, os modelos estatísticos da TRI se configuram atualmente como processos mais robustos na medição de *traços latentes*, e tornam-se um campo fértil para poder identificar padrões psicométricos em diferentes patamares de dificuldade que envolvem as CAs, assim como para evidenciar as correlações existentes entre elas. Esse tipo de estudo exige uma análise qualitativa na interpretação pedagógica do conteúdo nos itens do teste. Entretanto uma interpretação analítica mais integral com relação à *engenharia de construção dos itens* (Rabelo, 2013) no teste inclui o *enunciado*, o *contexto* envolvido, a descrição dos *itens corretos* e seus *distratores*, sendo um fator a contribuir para a validade interna do instrumento como também aprofundar a compreensão das análises e interpretações pedagógicas advindas dos indicadores psicométricos.

Acredita-se que esse campo investigativo deva e possa ser explorado em diversos segmentos dentro da educação científica, em *diferentes patamares de dificuldade* e de uma melhor compreensão quanto a *forma de acomodação da estrutura cognitiva* do conhecimento. A natureza da competência científica que se desenvolve está associada a um conjunto de situações-problema que o sujeito aprende a lidar em seu cotidiano pelas tarefas que faz, ou seja, a capacidade de compreender e resolver situações reais e familiares. Mesmo que se admita serem únicas e diferenciadas por força da subjetividade, é possível que haja certa consistência analógica no sentido das relações entre elas, por mais complexo que seja na construção e interpretação semântica dos termos como as funções metafóricas que elas possam assumir e associações. É preciso assumir, entretanto, um posicionamento moderador diante das contribuições que os testes podem fornecer, pois o papel de investigações e análise de testes não é a resposta aos problemas, são indicadores que devem estar a serviço da orientação e regulação dos métodos de ensino e da reformulação de recursos, assim como no auxílio a processos avaliativos mais eficientes.

O conhecimento envolvido em testes em larga escala remete a ideia de alcançar metas de desempenho, na qual existem habilidades cognitivas inerentes aos itens, e assim, a busca no alcance de padrões gerais ou básicos dentro de um recorte do conhecimento, que no caso, se concentra nas concepções alternativas, se justifica por serem campos universais e fundamentais a serem enfrentados na intenção de obter uma mudança conceitual, ou de garantir que alunos tenham uma compreensão minimamente adequada para o uso dos termos e conceitos científicos, independentemente do contexto da população em que fazem parte. É importante destacar que alguns padrões de conteúdo e desempenho devam ser propostos e desenvolvidos dentro de cada unidade escolar de acordo com suas políticas pedagógicas institucionais, buscando assim atender e se adequarem as suas realidades, essa perspectiva segue consonância com uma proposta de educação baseada em padrões, mas com instrução diferenciada (Wormeli, 2003, 2006). Apesar disso, entende-se que os padrões relacionados a esses tipos de testes sejam mais

específicos e estão sempre inclusos dentro dessas propostas devido ao seu caráter universal.

Com isso, observa-se que, apesar de ser um teste de larga escala, ele deve exercer uma função delimitada e específica no que se refere ao contexto das competências, sendo diferenciado de testes semelhantes que visam construir indicadores voltados para uma avaliação do sistema educacional, como no caso dos exames nacionais e internacionais de desempenho de estudantes. Diante da definição para o PISA, por exemplo, entende-se que esse tipo de teste atende de forma diferenciada ou bem mais restringida o conceito e expectativas de avaliar competências tendo em vista que nesse caso, esse tipo de avaliação possui uma ênfase sumativa, acumulativa de conhecimentos de todo um ciclo da educação básica.

Dentro da Física (entre outros ramos das ciências não formais), associa-se o cunho de **competência científica**³⁸ para as necessidades do processo de aprendizagem, em que, segundo o PISA, conceitua-se como:

- **A capacidade de utilizar o conhecimento científico, identificar questões científicas e concluir com base nas provas de compreender e ajudar a tomar decisões sobre o mundo natural e as alterações feitas pela atividade humana.**

É importante ressaltar que o PISA é um teste aplicado levando em consideração conteúdos acumulativos, isso representa uma perspectiva de diagnóstico mais geral, diferentemente aquela que se almeja em testes conceituais, que de certa forma podem estar a serviço dos professores no que diz respeito às regulações e redirecionamentos no ensino e em seu planejamento. Entende-se assim que os testes conceituais não sejam

³⁸ Ernst Waldfried Josef Wenzel Mach (Brno, 18 de fevereiro de 1838 – Vaterstetten, 19 de fevereiro de 1916) foi um físico e filósofo austríaco, e é considerado uma figura dominante na física e filosofia europeias nas últimas décadas do século XIX e início do século XX. A primeira de suas 500 publicações apareceu em 1859 e a última em 1921. Ele também trabalhou extensivamente com professores da escola e escreveu vários livros escolares e universitários. Ele também pode ser considerado o **fundador da disciplina da educação científica** em virtude de ter começado em 1887 o primeiro periódico no assunto *Zeitschrift für den Physikalischen e Chemischen Unterricht* (Jornal de Instrução em Física e Química) (Matthews, 2004, p. 32).

apropriados propriamente falando no sentido de avaliar competências, no entanto, exerce um papel fundamental no processo de desenvolvimento dessas competências.

Estrategicamente, no final de 1997, no mesmo ano que foi lançado o PISA, a OCDE lançou o *Projeto Definição e Seleção de Competências-Chave (DeSeCo)*, tendo os fundamentos teóricos e conceptuais com o objetivo de fornecer um concreto quadro conceitual para informar a identificação de competências-chave, e assim fortalecer as avaliações internacionais e para ajudar a definir metas globais para os sistemas educativos e a aprendizagem ao longo da vida. O uso de testes em larga escala são os instrumentos considerados para avaliar essa perspectiva de competência, sendo alvo de críticas, apesar disso destacam-se a sua capacidade em avaliar competências básicas fundamentais.

“Embora a avaliação ainda se aplique principalmente por meio de testes de papel e lápis ela centra-se em grande parte da **capacidade de interagir com a linguagem, símbolos e texto** (competência 1-A, DeSeCo), um progresso considerável tem sido feito para medir se os jovens e os adultos têm os poderes necessários”. (OCDE, 2005, p. 16, *tradução livre*)

Nesse entendimento as competências de base fazem parte de um *continuum* de desenvolvimento e progressão durante as aprendizagens, sendo uma aquisição articuladora e facilitadora para que se avance dentro desse *continuum* a medida em que são incorporados e diversificados os conhecimentos em níveis de compreensão e de comportamentos esperados.

Como referência as competências em níveis globais, de acordo com a OCDE, pode-se dizer que se encontra dentro das **competências aplicadas**, especificamente nas **competências fundamentais** que remetem a habilidades intelectuais de um conhecimento introdutório, ou seja, estando essas habilidades nos fundamentos, enfatizando e atendendo a *literacia* dos sujeitos, e por serem de base, podem ser facilmente transferíveis e comparáveis entre diferentes contextos.

A criação em 2008 do *Quadro Europeu de Qualificações (QEQ)* incluiu o ensino primário, secundário, profissional e superior. Sabe-se que esse quadro envolve vinte e sete Estados-Membros da UE, bem como nove outros países. Essas atividades têm-se

Portugal por meio da Portaria n.º 782/2009 de 23 de Julho apresentando em seu *Quadro Nacional de Qualificações* (QNQ) bem alinhado com o QEQ fazendo correspondência com os respectivos níveis de educação e formação. Portugal, em todos os seus níveis de escolarização (incluindo o ensino secundário – níveis 3 e 4) atende³⁹ ao compromisso firmado diante do documento “*QEQ para a aprendizagem ao longo da vida*”.

O conceito de competência segundo o *Quadro Europeu de Qualificações* destaca a necessidade das capacidades serem comprovada, verificadas, com isso, entende-se que esse processo deva ser vigiado não apenas durante um diagnóstico introdutório, mas um diagnóstico formativo, devendo ser vigiado epistemologicamente continuamente durante o processo de aprendizagem até atingir uma aquisição de competências mais abstratas visadas no término de ciclos de ensino. O sentido da responsabilidade e autonomia dos alunos durante a aprendizagem é um aspecto importante a ser considerado nesse processo, e isso deve ser feito através de um *feedback* sobre sua progressão em termos qualitativos e quantitativos, porém, deixando claro os objetivos e finalidades que se pretende e explorando os contextos mais familiares, se assim for pertinente e adequado ao curso.

Encontra-se relativamente entre os níveis 2 e 3 de dentro do *Quadro Europeu de Qualificações*, atendendo a capacidade de resolução de problemas diante da aquisição de conhecimentos na categoria de fatos, princípios e processos dentro das fundações da termodinâmica, apesar de não contemplar alguns outros conceitos essenciais dessa fundação, como o caso do conceito de entropia.

A aquisição de conceitos básicos e introdutórios contidos no TCE reflete uma capacidade preliminar em se apropriar dos fundamentos estruturais de um campo de conhecimento, que possibilite e favoreça o seu desenvolvimento em níveis de complexidade e diversidade na compreensão científica. Busca-se com isso utilizar esse conhecimento como parte necessária na construção de ideias e percepções mais abstratas, e com isso possibilite tomadas de decisões na vida cotidiana, podendo avançar e ser transferível a outros contextos. Considera-se um papel essencial dentro

³⁹ Ver: <http://www.dge.mec.pt/sites/default/files/Legislacao/quadro_europeu_qualificacoes.pdf>.

de um processo de alfabetização científica que contemple o exercício mínimo da cidadania diante da produção tecnológica e de um mundo globalizado. O mapeamento de **competências mínimas** demonstra ser uma preocupação crescente desde o início do século XXI (Vendramini & Dias, 2005), e vindo sendo mantida nos exames em larga escala, tendendo a ampliar em diversidade e especificidade nos campos temáticos.

Há uma preocupação crescente com habilidades pouco desenvolvidas de leitura, escrita e aritmética que motiva a busca por testes de competências e habilidades básicas como um meio de certificar a obtenção de uma competência mínima (Anastasi & Urbina, 2000).

Com relação as operações mentais, e não especificamente na aferição do conhecimento, as competências científicas demonstram situa-se em uma *Inteligência Fluida* (Gf) dentro da psicologia comportamental, na qual é entendida como “a capacidade das pessoas em aprender coisas novas que requerem relações lógicas e abstratas, assim como a capacidade em resolver problemas onde o **conhecimento prévio é incipiente**” (Gomes & Borges, 2009). O conhecimento prévio que se necessita para o sucesso em testes conceituais, é, e continuará sendo uma condição fundamental dentro da psicologia educacional, porém demonstra ser insuficiente diante de um largo conjunto de atividades que demandam processos lógico-abstratos que o mundo requer (Gomes & Borges, 2009). Nesse sentido, se percebe uma necessidade em progredir de uma perspectiva sobre aferição dos conhecimentos, e das diferentes concepções térmicas, para uma perspectiva sobre as competências necessárias e típicas que estão envolvidas, levando em consideração os processos mentais mais gerais para a compreensão de situações-problemas, especialmente aqueles voltados para a realização de tarefas específicas na qual se manifestam concretamente as competências.

2.1.2.3 Testes e o alinhamento curricular

Apesar das críticas recorrentes por alguns especialistas, continuará a ser predominante o modelo de avaliação de qualidade e eficiência do *status quo* de um sistema educacional tendo como parâmetro diversos indicadores que resultam de testes em larga escala, ao menos, a médio e longo prazo para o Brasil e em Portugal.

Sabe-se que esses indicadores juntamente com os do PISA exercem uma influência deliberativa nas decisões em políticas educacionais, estando sustentada, entre outras, no argumento da necessidade de avançar nos indicadores educacionais comparado a outros países. Alguns críticos apontam como promissor uma avaliação externa deva ser mais diversa, flexível e voltada a atender padrões locais (Wormeli, 2006), necessidades específicas do contexto da sala de aula, escolas, agrupamento de escolas ou distritos educacionais, que entretanto, são quase sempre um incômodo para os governos, inclinados a resultados mais convenientes de curto prazo, estando em meio a uma série de interesses socioeconômicos e políticos.

De uma maneira ou de outra, é importante explorar as vantagens para uma melhoria do sistema educacional a partir destes indicadores, em especial, no que diz respeito a um *alinhamento curricular vertical* (Webb, 1997). O conjunto de práticas e a utilização dos mais diversos recursos didáticos nas escolas são permeados por diferentes concepções influenciadas por questões socioeconômicas, culturais e ambientais, ainda assim, por estarem inseridas de forma mais ampla em um contexto em que exige qualidades relativamente comuns para a sua formação e qualificação, é natural que se busque certo *alinhamento curricular* destacando competências essenciais e transferíveis, preparando para a cidadania, mercado de trabalho e para a vivência em um mundo globalizado. Nesse sentido, ainda que de forma mais restrita, espera-se que haja um alinhamento de objetivos e das necessidades em contemplar os conteúdos núcleos sugeridos dentro do programa disciplinar dentro dos LDFs recomendados e adotados nas escolas, assim como as habilidades e competências prescritas, acompanhando obviamente o «paradigma» do *modelo de currículo* vigente, que no caso, se configura em um modelo de *Abordagem Por Competências*.

Alves & De Ketele (2011) destacam que um sistema educacional tende a ser negligenciado por ações que se inspiram em concepções **positivistas** (racionalismo clássico, o incrementalismo e a contingência são essencialmente movimentos *top-down*), que ações que se inspiram numa epistemologia **interpretativista** (a pós-

racionalização, o *Sensemaking-sensegiving*⁴⁰, as planificações comunicativas raramente chegam aos atores do terreno, em contato com os alunos). As críticas quanto ao excesso e precisão em preconizar demasiadamente *standards* pelo fato de contrariarem a criatividade, autonomia e contextualização ao priorizarem o operacionalizável e o mensurável, consideram que ambas as perspectivas são necessárias e contribuem numa perspectiva de alinhamento curricular (triangulação do “tecido” horizontal e vertical) dentro de uma APC, conforme seus pontos de orientação para políticas em educação (Alves & De Ketele, 2011, pp. 31-37).

Na concepção de Webb (1997) o «alinhamento curricular» pode ser tanto *horizontal* quanto *vertical*, estando o *alinhamento horizontal* mais voltado a reivindicar **mudanças** no *status quo*, enquanto o *alinhamento vertical* a se **adequar** às proposições do sistema. Focando no desenvolvimento de competências na perspectiva de um alinhamento vertical, os livros didáticos devem ser priorizados a exercerem a função de desenvolvimento de capacidade e competências (Gérard & Roegiers, 1998; Gérard, 2010; Gonçalves, 2011; Leite, Costa, & Esteves, 2008; Antunes, 2015), assim como as atividades e sequencias de ensino propostas pelos professores. Esse tipo de alinhamento vertical diz respeito a corresponder *Currículo Nacional do Ensino Secundário, Programa Curricular, e Metas Curriculares*, assim como *Políticas Públicas Educacionais*, e pode ser realizado de três modos (Webb, 1997):

1. Desenvolvimento Sequencial (“*Sequential Development*”): **Ajuste por meios de análise de testes padrões aplicados sequencialmente.**
2. Revisão especializada (“*Expert Review*”): Ajuste realizado por grupo de pesquisadores especializados na área curricular.
3. Análise de documentos (“*Document Analysis*”): Ajuste por meio de análise e reformulação de propostas curriculares vigentes que estejam inadequadas ou pouco eficazes em determinados aspectos.

Não se defende o “desenvolvimento sequencial” como uma exclusividade desse mecanismo para se estabelecer um alinhamento, mas uma contribuição para a tomada

⁴⁰ *Sensemaking* se refere a “construção do significado” e *sensegiving* a “construção do sentido” dentro da lógica institucional (Graeff, 2011)

de decisões baseado em evidências na organização em conteúdos essenciais e na proposição de situações de aprendizagem que sejam promissoras para o desenvolvimento de competências. Alguns pontos positivos e negativos desse conceito desenvolvido nos EUA e aplicado nos anos 90, que ganhou força no programa *No Child Left Behind Act* - NCLB (“Nenhuma criança deixada para trás”) com a política de George Bush podem ser destacados.

Os elementos da política, como expectativas e avaliações, são alinhados pelo *design*. Um conjunto de padrões, por exemplo, pode ser convertido diretamente em especificações para o desenvolvimento de uma avaliação. Uma vez que um elemento de política é estabelecido, ele se torna o modelo para os elementos subsequentes. Por exemplo, o Departamento de Educação da Carolina do Sul (1996) aprovou padrões em uma área de conteúdo que são usados para desenvolver padrões de realização acadêmica (resultados mensuráveis), que são usados para desenvolver instrumentos de avaliação.

Uma **desvantagem** para esta abordagem é a **quantidade de tempo** necessário para colocar um programa sequencialmente desenvolvido no lugar. Essa abordagem também **ignora o sinergismo entre os elementos políticos**: o desenvolvimento de avaliações, por exemplo, pode fornecer informações úteis para pensar sobre a instrução e o que os alunos podem aprender. Outra desvantagem dessa abordagem é que **frequentemente não reflete a realidade**. Em muitos estados, o processo para desenvolver expectativas e avaliações não é linear ou sequencial, mas é mais dinâmico. Webb (1997, p. 2)

Apesar da incidência de algumas críticas na falta de *alinhamento horizontal*, o estudo situa-se em um *alinhamento curricular vertical* numa perspectiva mais delimitada e interna, preocupado com necessidades sobre as dificuldades conceituais dos alunos em uma unidade temática específica, contudo vislumbrando processos legítimos de uma APC, incluindo uma abordagem interdisciplinaridade quanto a organização dos conteúdos e de situações de aprendizagem favoráveis. A perspectiva da complexidade, referente aos testes conceituais, ao explorar os conceitos introdutórios e necessários para a compreensão de diversos fenômenos térmicos pode promover uma ligação entre habilidades cognitivas comuns e concepções interdependentes entre diferentes ramos da termodinâmica, isso significa um caminho para uma integração interna «intradisciplinar» a ser explorada. O pensamento complexo nesse processo não perde de vista encontrar subsídios de alinhamento entre os sistemas, com base em dificuldades comuns entre os sujeitos nos dois contextos, considerando que o desenvolvimento das competências envolve questões de mesma natureza

epistemológicas relacionadas aos conteúdos, estando presentes dentro dos programas curriculares, sendo essa uma coerência com um pensamento de complexo.

Na pretensão de se estabelecer *indicadores intercambiáveis*, a identificação de características específicas e comuns sobre os traços latentes do conhecimento prévio poderá ser um viés na obtenção de outros indicadores para uma avaliação comparativa. No esquema adaptado se observa a perspectiva de alinhamento vertical que se pretende entre o desempenho dos estudantes sobre o conhecimento térmico e os aspectos priorizados e preconizados nos documentos oficiais e políticas do governo na qual se incluem os conteúdos, habilidades e competências nos programas curriculares, livros didáticos e conhecimentos exigidos em exames nacionais.

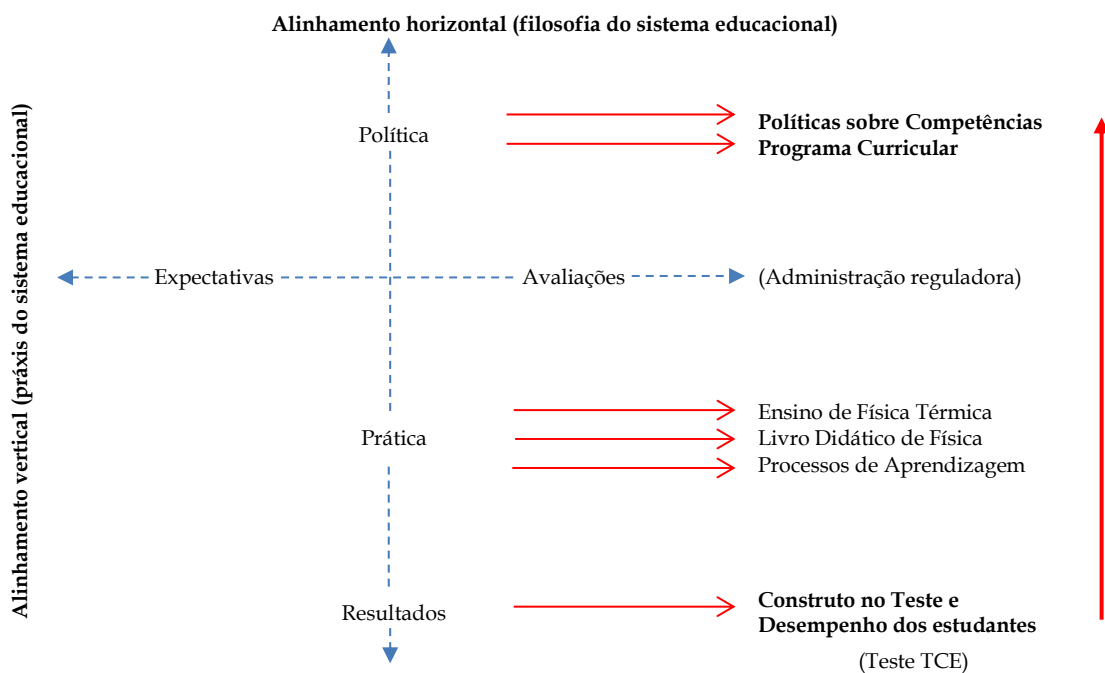


Figura 11: Perspectiva de alinhamento curricular entre o teste TCE e o programa curricular.

Fonte: Adaptado de Webb (1997).

Um *alinhamento curricular* de um sistema educacional visa uma correspondência entre os documentos oficiais reguladores e as reais necessidades educativas em sala de aula, uma relação de pertinência com os objetivos e finalidades curriculares preconizadas que intencionam o desenvolvimento de competências científicas. No ensino da Física (como em qualquer outro campo disciplinar), todos os métodos e recursos intrínsecos no processo pedagógico devem buscar esse alinhamento. No

âmbito dessa problemática há inúmeros fatores a influenciar, assim como diversas forças de interesse, entretanto, este estudo se propõe e delimita-se em evidenciar dificuldades conceituais entre dois contextos (Brasil e Portugal) e as concepções relacionadas as essas dificuldades.

Uma fundamentação sobre a possibilidade de atender a esse alinhamento por meio do *teste conceitual*, se caracteriza como um aspecto de validade interna do conteúdo, sendo são sugeridos o cumprimento de quatro critérios.

Quadro 14: Critérios de alinhamento curricular para testes conceituais.

Critérios	Definição e condições
Concordância categórica do conhecimento:	<p>É a correspondência e adequação entre os padrões de conteúdo apresentados em documentos curriculares (especialmente bases nacionais, matrizes e metas curriculares) e o campo conceitual que o teste explora.</p> <ul style="list-style-type: none"> • Ajustar de acordo com o tempo de instrução dada aos estudantes; • Identificar e ajustar o peso de temáticas-padrão comuns para diversos níveis de escolaridade e finalização de ciclo; • Tempo suficiente de duração de aplicação do teste; • Número de itens bem ajustado para o padrão que pretende avaliar em sua correspondência; • Em testes de múltipla-escolha estimar cerca de 1 min para avaliar <i>habilidades e conceitos específicos</i>; • Estimar cerca de 5 min para itens com respostas-construídas onde se avaliam desempenhos de competências mais inclusivas e estruturadas. • A inclusão de muitos tópicos-padrão em um teste administrado com uma restrição de tempo reduz a confiabilidade dos resultados. • Tempo de no máximo 60 min tendo uma <i>fiabilidade de 0,63</i> corresponde a seis tarefas (questões) por tópico de modo razoável, portanto, cerca de seis tópicos em um teste é algo admissível.
Consistência da profundidade-de-conhecimento	<p>É o ajuste para a distribuição de níveis mais elevados de abstração para a realização de atividades que envolvem a aquisição de conceitos, procedimentos e atitudes, não necessariamente sendo hierárquicas.</p> <ul style="list-style-type: none"> • Realizar <i>análise taxonômica</i> dos itens e ajusta-lo ao nível de abstração do padrão; • Ter no <i>mínimo 50 % dos itens com o mesmo ou maior nível de profundidade de conhecimento que o padrão correspondente</i>. Isto é baseado no pressuposto de que um aluno que responde pelo menos metade dos itens que medem o conteúdo relacionado com esta norma é considerado para demonstrar o nível mínimo aceitável;
Correspondência do alcance-de-conhecimento:	<p>É o alcance ou a amplitude do conhecimento medido pelo teste comparado ao intervalo de conteúdo no domínio (padrão) que está sendo avaliado.</p> <ul style="list-style-type: none"> • Dentro de no <i>mínimo 50% dos objetivos de um padrão possa ter pelo menos um item de avaliação</i> relacionado para que exista pelo menos algum alinhamento entre um teste e o padrão. Este nível é baseado no pressuposto de que o conhecimento dos alunos deve ser testado em conteúdo de mais de metade do domínio de conhecimento de um padrão. • A correspondência de alcance de conhecimento é mais difícil de alcançar se as expectativas de conteúdo forem divididas entre um maior número de padrões e com um grande número de objetivos em cada um deles.

Equilíbrio-de-representação do conhecimento:	<p>É a distribuição balanceada de itens dando ênfase a diferentes objetivos ou partições de conteúdo, como em um conteúdo padrão ou domínio de conteúdo.</p> <ul style="list-style-type: none"> • Os objetivos mais importantes ou mais abrangentes devem ser colocados em maior número de itens de avaliação que medem o conhecimento relacionado a esses objetivos; • Ter <i>mais de 50% dos itens para cada padrão ou acima do nível de profundidade de conhecimento dos objetivos da norma;</i> • Itens <i>uniformemente distribuídos</i> de tal forma que possam medir o conhecimento de conteúdo relacionado a <i>no mínimo a metade dos objetivos</i> para que se alcance o <i>equilíbrio pedagógico do conteúdo</i>.
---	---

Fonte: Elaborado a partir de Webb (2006, p. 163, *tradução livre*).

Esses critérios se aplicam a toda a forma de avaliação:

Eles se aplicam a qualquer uma das diferentes maneiras pelas quais as expectativas para a aprendizagem dos alunos podem ser expressas por tópicos de conteúdo, padrões ou estrutura, propriedades estatísticas ou por quais problemas de aprendizagem dos alunos estão sendo diagnosticados. Esses critérios fornecem algumas diretrizes práticas para a criação de um modelo para avaliação e estabelecimento dos parâmetros necessários. (Webb, 2006, p. 171, *tradução livre*)

E se relacionam a processos mais consistentes na avaliação de competências dentro da *estrutura do conteúdo* em um teste, pois deve relacionar à aquisição de conceitos e as suas capacidades de resolver problemas estando envolvidas numa perspectiva pedagógica integradora, e ao mesmo tempo sendo sustentado dentro de uma teoria de aprendizagem:

“[...] a especificação de conteúdo para um teste precisa ser desenvolvida com base em uma *teoria de aprendizagem* para a área de conteúdo. Essa teoria da aprendizagem precisa esclarecer o conhecimento dos alunos sobre a área de conteúdo, como seu conhecimento se aprofunda através da formação de relacionamentos entre os conceitos e habilidades na área de conteúdo e os caminhos que os alunos podem usar para **tornar-se competente na área de conteúdo**. Para que os testes sejam teoricamente baseados e alinhados com expectativas mais complexas, os desenvolvedores de testes devem atender não somente aos fatos, conceitos, habilidades e processo de raciocínio na área de conteúdo, mas também devem considerar como todos eles se relacionam uns com os outros e formam um sistema de conhecimento.” (Webb, 2006, p. 171, *tradução livre*)

Especificamente, no que se refere às competências básicas e necessárias que um aluno ao final da educação básica deve alcançar após a abordagem do conteúdo de Física Térmica, é possível identificar alguns com base em *padrões de conteúdo preestabelecidos* (alinhamento vertical) ou buscando atender à *recomendações fundamentadas nas teorias da Psicologia Educacional e da Didática das Ciências* (alinhamento

horizontal). Além disso, a natureza do próprio conteúdo deve ser levada em consideração, assim como as concepções e dificuldades dos alunos (conhecimento prévio). Realizando um recorte na complexidade do conteúdo da Física Térmica, pode-se estabelecer em uma dimensão horizontal de amplitude, abrangência e relevância, e na dimensão vertical, no que se refere às hierarquias taxonômicas de abstração e complexidade dos objetos de aprendizagem. Entende-se que as competências científicas necessárias devem transitar nesses dois eixos, mobilizando diferentes saberes, mobilizando e integrando diversas capacidades, destrezas e atitudes em famílias de situações-problemas preestabelecidas, tendo como ponto de partida os equívocos mais comuns já mapeados nesse campo de conhecimento. Direcionando sobre competências científicas em Física Térmica a serem desenvolvidas ao fim da educação básica, pode-se dizer que:

- Na perspectiva de um *alinhamento vertical* (atendendo aos documentos formais e exigências do sistema):

(1) A verificação das temáticas e conceitos específicos mais incidentes nos exames nacionais que exercem a função de avaliação externa do sistema. Em Portugal, constata-se que de 2006 a 2016, os exames nacionais tem incidido nas temáticas sobre “absorção/emissão de radiação” (45%), assim como na temática de “rendimento no aquecimento” (22%) e “capacidade térmica mássica” (20%). As competências científicas que se espera desenvolver, se referem a situações práticas e concretas que precisam ser definidas, explorando esses conhecimentos na perspectiva conceptual e processural (além da socioafetiva, que são transversais), ou seja, tanto quanto aos conceitos envolvidos nesses processos termodinâmicos e seus diferentes níveis de abstração, quanto aos procedimentos experimentais que se exige na competência em ação da resolução de uma situação-problema, em que se deva mobilizar diversas capacidades cognitivas e psicomotoras (destrezas).

(2) No Brasil, as matrizes de referência e a BNCC apresentam descritores que representam as habilidades cognitivas específicas para o conteúdo da Física Térmica. Nesse caso, as competências científicas que se espera desenvolver devem atender a situações preestabelecidas e pertinentes ao cenário pedagógico (ainda que existam

capacidades mais gerais e transferíveis, independente do conteúdo), mas que contemplem necessariamente os documentos oficiais vigentes. De um modo geral, vale a destacar que independentemente do contexto, existem 7 grandezas consideradas fundamentais na Física, e, portanto, fundamentais de serem assimiladas e compreendidas, duas delas se concentram relativamente na Termodinâmica: (a) temperatura termodinâmica; e (b) quantidade de matéria.

- Na perspectiva de um *alinhamento horizontal* (podendo atender a novos critérios e argumentos, diferentes daqueles que preconiza o sistema):

(1) Teoria da aprendizagem significativa: Modelo Cinético Molecular (MCM) e as Leis da Termodinâmica (Moreira & Masini, 2006). Como conhecimento básico a ser apreendido, o MCM remete a necessidade de compreender as propriedades gerais da matéria e as especificidades do comportamento das partículas as constituem, como os estados físicos de agregação e as estruturas moleculares de ligação. Nesse sentido, pode se caracterizar como um patamar de competência científica necessário a ser desenvolvido dentro no domínio cognitivo, e está relacionado a uma aprendizagem em que se percebe as relações e compreensões fenomenológica de causa dos fenômenos térmicos em níveis microscópico e mesoscópico (dimensão mais familiar às percepções sensoriais humanas), na possibilidade de extrapolar para o macroscópico. Já as Leis da Termodinâmica envolvem uma compreensão mais ampla diante os fenômenos térmicos como um todo quando se concebe as noções sobre os conceitos de calor, temperatura e entropia, seguindo para a ideia de calor específico e equilíbrio térmico. Nos gases, os conceitos básicos que torna-se elementar a sua apreensão é além da temperatura, o volume, e a pressão. De modo geral, um objetivo terminal de integração como sugestão a caracterizar uma competência básica a ser necessariamente desenvolvida ao final do ciclo básico poderia ser: a compreensão do funcionamento de máquinas térmicas, como por exemplo, os mecanismos necessários para funcionar um gerador termoelétrico, levando em consideração as condições mínimas para atender por um mês uma cidade de 50 mil habitantes.

(2) Teoria dos conteúdos escolares (García Díaz, 1998; Izquierdo, 2005): nesse caso os conteúdos prioritários devem ser definidos com base em modelos pedagógicos mais próximos da realidade dos alunos (de preferência, concretos), levando em

consideração suas dificuldades, as potencialidades individuais, os problemas relacionados ao seu cotidiano e o pensamento complexo (com destaque para as capacidades de pensar, fazer e comunicar). Nisso, na perspectiva do desenvolvimento das competências necessárias dentro da Física Térmica, sugere-se contemplar os conteúdos nucleares, passando pelos conceitos fundamentais, a apreensão de procedimentos e atitudes científicas na realização de tarefas. Temas transversais, como energia, em especial, em situações bem delimitadas dentro de um contexto, poderá possibilita articular esse processo.

2.2 CONTRIBUIÇÕES DA PSICOLOGIA EDUCACIONAL

2.2.1 Fundamentos da Teoria da Aprendizagem Significativa de David Ausubel

“ O ensino da matemática e das ciências, por exemplo, continua a basear-se muito na aprendizagem memorizada de fórmulas e de passos de procedimentos, no reconhecimento memorizado de ‘problemas tipo’ estereotipados e na manipulação mecânica de símbolos. Na *ausência de ideias claras e estáveis*, que podem servir como **pontos de ancoragem** e de focos de organização para a incorporação de *material novo e logicamente significativo*, os estudantes vêm-se presos numa teia de incompreensão e possuem poucas tarefas de aprendizagem, mas memorizadas, para fins de avaliação”. (Ausubel, 2003, p. 167, *grifo nosso*)

Segundo a concepção de Ausubel, a significância se encontra no sujeito cognoscente e não nas coisas. Partindo desse princípio qualquer proposta de ensino que tenha a pretensão de ser bem-sucedida deve levar consideração o *conhecimento prévio e as representações mentais dos alunos*. Para isso se torna **desafiador e necessário a busca por evidências** que estejam associadas às formas de linguagem, concepções e interpretação do mundo. A relevância do conhecimento prévio é bem conhecida e tem um papel de destaque nos estudos de Piaget, entre outros, atualmente são consideradas pedras angulares e necessidades legítimas que não devem ser negligenciadas dentro do processo de ensino-aprendizagem em diversas perspectivas teóricas, fundamentalmente na psicologia educacional cognitiva.

As buscas por evidências concretas e fidedignas nesse campo se torna ainda mais relevante, quando Ausubel (2003, p. 25-26) destaca três razões principais que explicam a escassez de conhecimento sólidos acerca da aprendizagem em sala de aula:

(1) A maioria dos estudos na área da aprendizagem são levadas a cabo por *professores e outros investigadores não profissionais na área da educação*; (2) as investigações mais rigorosas vem sendo conduzidas por *psicólogos experimentais que não estão ligados à área da educação*; (3) e principalmente *aos psicólogos educacionais não levam a cabo a investigação aplicada necessária* e contentam em extrapolar os resultados da aprendizagem por memorização dos colegas experimentais para a aprendizagem escolar. No primeiro se destaca a negligência conceptual e de um rigor científico-metodológico investigativo, muito associado a uma má formação tanto do ponto de vista de sua formação, em que o domínio conceitual deve ser algo propedêutico para o exercício docente de suas atribuições profissionais, como na necessidade de conduzir com diligência as investigações, sendo em grande parte realizadas com um caráter ingênuo psicoeducacional por neófitos pesquisadores.

Os psicólogos experimentais por sua vez têm sido mais rigorosos, porém com um foco distante das implicações mais “práticas” e emergenciais para a educação. Ao invés disso, as análises deveriam se concentrar na investigação de problemas mais gerais da aprendizagem e da instrução cognitivas, buscando uma aplicabilidade mais vasta para situações de aprendizagem e contribuindo de forma mais substancial para a psicologia da aprendizagem, em especial, aquelas voltadas para uma recepção significativa ou aquisição e retenção de conhecimentos (Ausubel, 2003, p. 25), que como consequência contribui para o desenvolvimento de competências. Sabe-se que a maioria dos pesquisadores na área do ensino não são propriamente psicólogos, sendo um problema contornável devido a obrigatoriedade de transitar na busca de subsídios e fundamentações dentro psicologia educacional, porém não é algo se garanta uma apropriação mínima no domínio desse campo da ciência, e ainda podem haver reflexos, lacunas e necessidades que absorvem desse campo. Sobretudo, o distanciamento entre psicólogos experimentais e os aplicados, como p.ex., os educacionais, seja um fator agravante.

Para isso, as implicações baseadas na psicologia educacional de Ausubel como condições de contorno se refere a sugestões que visem compreender, valorizar e enfrentar os erros dos alunos, seja na identificação prévia do *subsunçores* em potencial

(quando se planeja e constrói itens de um teste, quando se elabora determinada atividade entre os alunos, quando se seleciona e organiza determinados conteúdos) para a ancoragem dos novos conceitos envolvidos e reivindicados na situação-problema. Os *subsunçores* são, portanto, uma estrutura de conhecimento específica preexistente na cognição de quem aprende, e se caracteriza como um aspecto relevante capaz de ancorar de modo “não-literal” (substantivo, e não por retenção memorizada ou literal), “não-arbitrária” e “relacional” a nova informação.

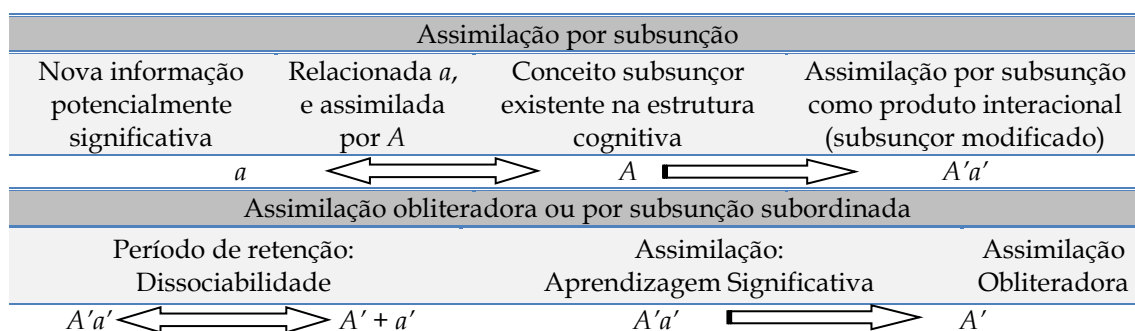
O processo de ancoragem da nova informação resulta em crescimento e modificação do conceito subsunçor, e esse processo de assimilação de conceitos se dá pela associação dessas ideias relevantes estabelecidas na estrutura cognitiva do aprendiz com o conteúdo potencialmente significativo, implícito na definição dos termos ou das “pistas” contextuais (atributos criteriais) (Moreira & Masini, 2006, p. 18), que no caso pode ser obtido pela análise do traço latente existente no enunciado do item, em que sugere um domínio sobre determinado conhecimento específico de acordo com a situação-problema. As respostas dos indivíduos sugerem evidências de subsunçores na estrutura cognitiva dos alunos, que podem ser subsunçores sustentado em equívocos ou concepções alternativas. Estes supostamente encontram-se organizados hierarquicamente na estrutura cognitiva dos alunos e representam abstrações da experiência do indivíduo (Ibid, p. 18).

Quando não se verifica a existência de subsunçores diante da resolução de uma situação-problema ou previamente, justifica-se uma aprendizagem mecânica baseada no processo de memorização de fórmulas, leis e conceitos, que na taxonomia de Bloom corresponderia ao nível mais baixo de abstração na dimensão processual do conhecimento, e na concepção de Ausubel é um processo de aprendizagem ou assimilação não-arbitrária, pois não busca se associar a outras formas de conhecimento em sua estrutura cognitiva. Uma vez internalizado esses subsunçores, eles se tornam ideias-âncoras com capacidade de serem associadas a determinados conhecimentos que reivindique essas ideias, por isso essas ideias são consideradas relevantes e conhecimento a ser associado potencialmente significativo, ou seja, possibilita fazer sentido para o sujeito. Por exemplo, entre os itens no teste, se pode observar que alguns

remetem a memorização de valores para a temperatura de determinados corpos em situações típicas, assim como noções de suas propriedades térmicas, esse tipo de conhecimento estando preexistentes se tornam “ancoradouros” de um conjunto de conhecimentos e de situações-problemas possíveis.

Na organização e sequencia dos conteúdos a serem abordados, considera-se relevante, de acordo com a concepção ausubeliana seguir um sentido linear e com lógica dedutiva ao serem exploradas noções conceituais específicas e intuitivas dos alunos, porém, para que seja mais adequada dentro dessa concepção é preciso levar em consideração os processos de *diferenciação progressiva*, *reconciliação integrativa*, *organização sequencial* e *consolidação*. As hipóteses de assimilação podem ser descritas como um processo de assimilação regida simbolicamente por um princípio de assimilação na qual se distingue em dois estágios: assimilação por subsunção e assimilação obliteradora, sendo a segunda também denominada de assimilação por *subsunção subordinada* (aprendizagem subsunção subordinada) (Braga, 2010).

Quadro 15: Processo inicial de assimilação por subsunção, seguida pelo processo de assimilação obliteradora.



Fonte: Braga (2010) adaptado a partir de Moreira (2001)

Os conceitos devem ser abordados inicialmente através dos conteúdos mais abrangentes e estruturadores, seguindo uma ordem hierárquica, inicialmente com os conceitos superordenados (muito gerais e inclusivos), seguido de conceitos subordinados (intermediários) e finalizando nos conceitos mais específicos (pouco inclusivos), essa perspectiva aparentemente contrapõe a «estratégia de indução dos objetos de aprendizagem» que é a mais comum, como geralmente se afirma (Gérard & Roegiers, 1998, pp. 65-66), considerada adequada na perspectiva do desenvolvimento de competências. Nesse sentido se admite uma progressão paulatinamente, agregando

aquisições conceituais, e capacidades específicas, construindo e desenvolvendo competências essenciais visando alcançar posteriormente competências mais complexas, envolvendo mais capacidades, mobilizando e associando uma quantidade e diversidade maior de recursos. Um exemplo da organização sequencial da Termodinâmica dentro da concepção ausubeliana foi adotada no curso de Física na UFRGS no final da década de 70, e sendo objeto de estudo posteriormente (Lima, 1981).

Nesta, vai-se do geral para o particular. Começa-se com uma discussão geral sobre Física Clássica, Moderna e Contemporânea, e situa-se a Termodinâmica e a Teoria Cinética nesse contexto. Fala-se em conceitos e **introduz os conceitos de temperatura, calor e entropia como os conceitos mais relevantes ao estudo da Termodinâmica e da Teoria Cinética**. A seguir, introduz-se a primeira e a segunda Lei da Termodinâmica. Nesta unidade são também introduzidos alguns conceitos e são dados alguns exemplos. A característica fundamental dessa unidade é o seu caráter geral e introdutório; os conceitos e leis fundamentais da Termodinâmica são apresentados de um ponto de vista macroscópico, geral e intuitivo, sem levar em conta a natureza atômica e molecular da matéria.

A segunda unidade é uma diferenciação da primeira: analisam-se os mesmos conceitos e leis do ponto de vista microscópico. Fazendo uso da Teoria Cinética aplicada ao modelo de um gás ideal, chega-se à interpretação microscópica dos conceitos e leis estudados na primeira unidade.

Na terceira unidade, conclui-se o estudo dos fenômenos térmicos dando ênfase a aplicações dos conceitos e leis estudados nas duas primeiras. (Moreira & Masini, 2006, pp. 64-65)

Esta abordagem sugere partir de conteúdos mais abrangentes para os mais específicos, porém não significa necessariamente que seja uma lógica totalmente dedutiva, pois apresenta a parte inicial se apresentam as leis e princípios que por sua vez se estruturando em um sentido indutivo na abordagem dos conceitos elementares.

Esse processo segue de maneira que atende a perspectiva de desenvolvimento de competências mais simples para mais complexas na termodinâmica, porém os conceitos são elementos que não se desvinculam das etapas de aprofundamento, assim como das concepções mais abrangentes e estruturantes que são as Leis e Princípios fundamentais.

Quadro 16: Comparação organizacional da abordagem da Termodinâmica pelo método convencional e ausubeliano.

	Organização Convencional	Organização segundo A teoria de Ausubel
Unidade I	Temperatura e 1ª Lei da Termodinâmica: Temperatura, equilíbrio térmico, calor, equivalente mecânico do calor, 1ª Lei da Termodinâmica, aplicações.	Conceitos e leis fundamentais da Termodinâmica: A Termodinâmica e a Teoria Cinética no contexto da Física, temperatura e calor, entropia, 1ª e 2ª Lei da Termodinâmica
Unidade II	Teoria Cinética dos Gases: Equação de estado de um gás ideal, modelo cinético de um gás, cálculo da pressão, interpretação cinética da temperatura, calores específicos de gases ideais, equipartição da energia, aplicações.	Teoria Cinética dos Gases: Equação de estudo de um gás ideal, modelo cinético de um gás, cálculo da pressão, interpretação mecânica dos conceitos de temperatura, calor e entropia, calores específicos de gases ideais, equipartição da energia.
Unidade III	Entropia e 2ª Lei da Termodinâmica: Processos reversíveis e irreversíveis, Ciclo de Carnot, máquinas térmicas, 2ª Lei da Termodinâmica, entropia, aplicações	Aplicações dos Conceitos e Leis Fundamentais da Termodinâmica: Aplicações dos conceitos e leis estudados nas unidades anteriores à solução de problemas e questões.

Fonte: Moreira & Masini (2006).

Também foi sugerida a necessidade de uma abordagem introdutória sobre o *Modelo Cinético-Molecular* explorando as propriedades gerais da matéria, servindo de auxílio na compreensão dos fenômenos térmicos na relação de observações indiretas com o nível microscópico (Braga, 2010).

É importante ressaltar que dentro dessa área do conhecimento, alguns fenômenos térmicos pertencem diferentes categorias explicativas, como é o caso dos processos de transmissão de calor entre corpos: condução, convecção e radiação. Comumente, professores (e livros didáticos) induzem aos estudantes que esses são processos de mesma natureza, apesar de constituírem uma natureza epistemológica diferente. A condução e convecção são fenômenos explicados no ramo da Termodinâmica e estatística, mais precisamente ao *Modelo Cinético-Molecular*, enquanto a radiação pertence mais ao campo da física moderna, ou mais precisamente ao eletromagnetismo clássico, ainda com *status* epistemológico, à mecânica quântica (Mattos, Drumond, & Paulo, 2004, p. 16). Dessa forma, um elo epistemológico se revela como um parâmetro integrador necessário para a compreensibilidade de determinados fenômenos térmicos, assim, na escolha da organização, estrutura e sequência na abordagem dos conteúdos.

Arrisca-se uma ilustração que tente aproximar desse processo de sequenciação e organização dos conteúdos. Nisso, poderia ser representada sendo “x” o conhecimento mais estruturado (Leis e Princípios), na qual preenche e delimita o ramo da física correspondente, sem desvincular os conceitos básicos “x” (no caso, os conceitos de temperatura, calor e entropia, sendo cada vez mais detalhados e especificados, de acordo com diversidade de situações que o envolvem). Nessa etapa encontram-se a necessidade de serem desenvolvidas competências científicas de base conceitual, em que geralmente são predominantes concepções prévias dos estudantes. Identificar as dificuldades conceituais nessa etapa e buscar mecanismos de superá-las é um dos recortes delimitados no estudo, sobretudo com o foco em contribuir na aquisição do conhecimento mais estruturado. A sequência segue na incorporação e associação de conhecimentos intermediários “x” (noções em torno dos conceitos e conhecimentos articuladores, como processos de transição que ocorrem) juntamente com específicos “x”, que, por fim, numa etapa de aplicação e integração do conhecimento, estando os conceitos básicos formando uma rede de saberes que estruturam e articulam uma compreensão mais alargada na Termodinâmica, momento em que se espera desenvolver competências mais complexas.

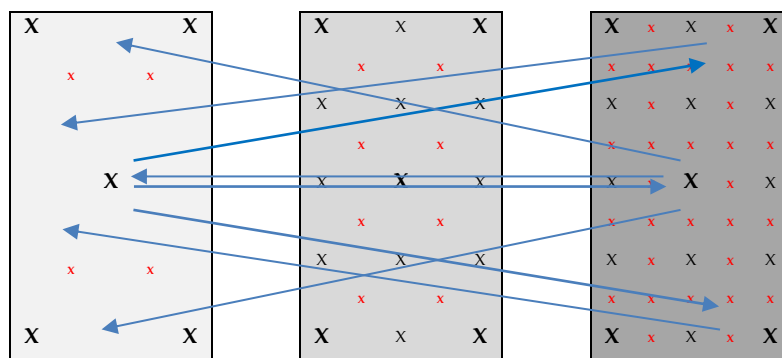


Figura 12: Etapas de incorporação dos conteúdos seguindo uma abordagem ausubeliana .

Obviamente que a estrutura cognitiva seja pessoal para cada aluno, com isso, não corresponda a estrutura organizacional dos conteúdos dentro dessa sequência, entretanto acredita-se que possa favorecer esse processo. No entanto, essa é apenas uma das possibilidades educativas propostas, de acordo com as condições de contorno fundamentadas na psicologia de David Ausubel, da aprendizagem significativa, que

supostamente demonstra ser adequada numa perspectiva de desenvolvimento de competências.

A representação visual em tons mais fortes serve para mostrar que o conhecimento vai sendo estruturado e se tornando mais coeso pela formação de uma rede de conceitos intermediários e mais básicos que são incorporados e desenvolvidos. As setas divergentes de ida sugerem o processo de *diferenciação progressiva*, e as setas de retorno a *reconciliação integrativa*, sendo a trajetória recursiva convergindo, porém não necessariamente pelo mesmo caminho de ida, e também não precisando chegar ao término da sequenciação de ensino. Essa abordagem segue possibilitando numa aprendizagem específica que proporciona uma percepção e compreensão mais geral correspondente as Leis e Princípios que regem todos os fenômenos. Dentro disso, se adequa a concepção de reconciliação integrativa proposta por Ausubel, na qual deve-se “descer e subir” nas estruturas conceituais hierárquicas (Moreira & Masini, 2006, p. 65) em cada etapa e nível de abordagem. Reforça-se em dizer que não deve ser confundida com uma forma sistemática de distribuição do conhecimento no texto com tempo, e muito menos como um modelo visual da estruturadas cognitivas, por ser inacessível diretamente e exclusivo de cada sujeito.

Uma «estratégia exclusivamente indutiva» busca induzir os objetos de aprendizagem (a regra, a fórmula, a definição, ...) depois da descoberta dos alunos das suas características, a partir de vários exemplos (Gérard & Roegiers, 1998). A «estratégia exclusivamente dedutiva» parte de uma apresentação geral (a definição, a regra, a fórmula, ...) para a sua aplicação em casos particulares. Os autores ressaltam: “esta estratégia pode ser importante num modelo experimental no qual uma hipótese emitida deve ser validada por um dispositivo de experimentação” (Ibid, 1998). A lógica dedutiva muitas vezes caracteriza um ensino mecanizado baseado na memorização, e no campo experimental se ajusta na modalidade verificativa das leis e princípios. Na sugestão apresentada, a concepção ausubeliana demonstra ser uma associação da estratégia dedutiva e indutiva.

Uma noção simples e esclarecedora sobre a noção do processo de *reconciliação integrativa* é representada através dos estudos sobre «mapas conceituais», onde seguem uma regra linear no sentido “de cima para baixo”, sendo interligados por

palavras ou ideias-link. Nas condições de favorecem o desenvolvimento de competências científicas, é imprescindível que haja uma *aprendizagem significativa* na aquisição dos conceitos básicos da Termologia. Dentro disso, acredita-se que o papel que o *Modelo Cinético Molecular* pode desempenhar no processo de aprendizagem implica sua importância na organização sequencial do conteúdo, pois através dele se explora a admissão da existência de átomos, na qual diversas modalidades experimentais, diferentes níveis de observação fenomenológica (micro, meso e macroscópico) (Braga & Kalhil, 2010), sendo tanto reais/concretas quanto virtuais/simuladas são uma alternativa promissora nessa fase da aprendizagem (Braga, 2010).

A concepção da existência de átomos é um passo inicial a ser explorado, e uma etapa de transição inicial importante para identificar e realizar diagnósticos sobre as concepções dos alunos e suas noções diante do conhecimento térmico. As ideias que contrapõem o conhecimento científico podem ser articuladas através de um modelo teórico simplificado que rege de modo geral o comportamento das partículas constituintes da matéria. A ilustração, demonstração e verificação na prática de um ensino experimental das propriedades gerais da matéria pode ser um mecanismo e recorte da complexidade eficiente, além do que, é torna-se cabível e supostamente favorável para ao nível de Física Geral no ensino médio, porém com um nível de maior aprofundamento no ensino superior, sobretudo, sendo relevantes nas abordagens dos livros texto e nas respectivas orientações metodológicas que facilitam o planejamento e a prática de ensino nas salas de aula.

Linn (1987, citado por Cachapuz et al., 2005) aponta como necessidade e desafio a associação na busca de inovações de ensino com o processo avaliativo. Um caminho para isso estaria em explorar o potencial das avaliações no sentido de desenvolver situações de aprendizagem mais *representativas* aos alunos (Alonso, Gil-Pérez E Martínez Torregrosa, 1996, citado por Cachapuz et al., 2005), e de forma semelhante para uma melhor articulação entre o currículo intencional e de ação. Nesse sentido, a teoria para a assimilação e aquisição significativa de conceitos de Ausubel (2003) fornece algumas contribuições a serem valorizadas nesse processo, destacando os

seguintes elementos para uma aprendizagem significativa: (1) *material* potencialmente significativo; (2) *mecanismo de aprendizagem* significativo; (3) *tarefas* logicamente significativa. O cenário pedagógico necessita de uma vigilância metodológica que antecede a prática de ensino, é preciso identificar e tomar a decisão sobre os meios, processos e recursos a serem utilizados sobre determinada abordagem, partindo de reconhecimento a natureza epistemológica de cada conteúdo e do conhecimento prévio dos alunos para assim tomar uma decisão didática apropriada ao invés de uma generalização metodológica. Recorrentemente esse equívoco ocorre pela ilusão de sustentar ou considerar concepções unicamente das psicologias da educação. Em resumo, existem quatro tarefas fundamentais do papel do professor na facilitação da aprendizagem significativa segundo Moreira (1999), que são:

Quadro 17: Recomendações para o ensino na concepção de Ausubel.

Papel do professor	Descrição da ação
1 - Identificar a estrutura conceitual e proposicional da matéria de ensino.	Identificar os conceitos e princípios unificadores, inclusivos, com maior poder explanatório e propriedades integradoras, e organizá-los hierarquicamente de modo que, progressivamente, abranjam os menos inclusivos até chegar aos exemplos e dados específicos.
2 - Identificar quais os <i>subsunçores</i> .	Identificar conceitos, proposições, ideias claras, precisas, estáveis, relevantes à aprendizagem do conteúdo a ser ensinado, que o aluno deveria ter em sua estrutura cognitiva para poder aprender significativamente este conteúdo.
3 - Diagnosticar aquilo que o aluno já sabe.	Determinar dentre os <i>subsunçores</i> especificamente relevantes (previamente identificados ao “mapear” e organizar a matéria de ensino), quais os que estão disponíveis na estrutura cognitiva do aluno.
4 - Ensinar utilizando recursos e princípios.	Ensinar de modo que facilitem a aquisição da estrutura conceitual da matéria de ensino de uma maneira significativa.

Fonte: Moreira (1999)

Os dois primeiros apresentam-se como atividades externas as serem realizadas pelo professor dentro do **domínio epistemológico**, certificando e estabelecendo *existência de materiais potencialmente significativos* de natureza *teórico conceptual e factual* da formação de conceitos. Esse processo foca em uma análise do conteúdo e da linguagem nos livros didáticos adotados e/ou quaisquer materiais de outra natureza que sirva de referência para a abordagem dos conteúdos no que diz respeito aos critérios de seleção, organização, estruturação e sequenciação.

Já nas implicações dentro do **domínio psicológico**⁴¹ na inter-relação com a didática, em comum acordo com as condições de Ausubel (em síntese) para uma aprendizagem significativa nessa perspectiva têm-se as duas últimas tarefas para o *papel do professor*⁴² nesse processo. Diagnosticar aquilo que o aluno já sabe deve envolver uma coordenação entre os materiais de natureza *teórico conceptual e factual* com as tarefas a serem realizadas pelos alunos, para que sejam lógicas e significativas dentro desse processo. Considerando os alunos protagonistas neste processo de construção do conhecimento (método construtivista), a didática da Física na intenção de atender *um currículo por competências* é articulada de forma favorável dentro de uma pedagogia integradora.

Apropriando-se de algumas dessas noções ausubelianas para um processo instrutivo de ensino, sugerem-se etapas para que o *desenvolvimento de competências científicas de base progrida* e alcance a aquisição de competências de aperfeiçoamento, ou seja, o domínio do corpo mais estruturado dentro de cada ramo da Física:

Quadro 18: Sugestões de etapas para um ensino ausubeliano favorável ao desenvolvimento de competências.

Etapas	Descrição da ação
Etapa 1: Determinar as competências disciplinares básicas que se pretende desenvolver	Definir os objetivos específicos (habilidades e os conteúdos específicos que envolvam conceitos básicos e introdutórios) com vistas a alcançar as competências mais complexas pretendidas.
Etapa 2: Selecionar as famílias de situações-problemas pertinentes a cada competência estabelecida a ser desenvolvida	As situações-problemas devem envolver um conjunto de habilidades que de forma conjunta caracterize de forma suficiente cada competência que se pretende adquirir.
Etapa 3: Identificar os supostos subsunçores entre os saberes que envolvem as famílias de situações-problemas	Interpretação e análise dos conteúdos dentro de um ramo e sub-ramos do conhecimento, nomeadamente relacionado aos problemas que se pretende ser proficiente em solucioná-los
Etapa 4: Verificar os subsunçores dentre os conhecimentos preexistentes nos alunos	Aplicar mecanismos que possam detectar a existência de subsunçores (testes, entrevistas, etc..).
Etapa 5: Diferenciar os subsunçores em patamares de dificuldades e abstração pelos alunos	Discriminar em um ambiente mais controlado os níveis de dificuldades para um perfil da classe através de critérios referenciados preliminarmente, e se possível para cada sujeito.

⁴¹ Considera-se que os domínios epistemológico, psicológico e didático devem estar inter-relacionados numa visão dialética, e não dissociados, independentes ou simplesmente complementar.

⁴² Como foi dito anteriormente, essa é uma questão fundamental a ser enfrentada para o EC, entretanto, serão consideradas essas implicações dentro do “problema do conteúdo”, iniciando a discussão defendendo um currículo baseado por competências.

Etapa 6: Definir um conjunto de tarefas potencialmente significativas, integradora e bem definidas dentro de cada família de situações-problema	Selecionar ou elaborar atividades diversificadas em níveis de abstração e dificuldade, tendo como subsídio o conhecimento potencialmente significativo que supostamente seja favorável a se associar aos subsunçores identificados preliminarmente.
Etapa 7: Propor uma progressão das tarefas potencialmente significativas com base na diversidade, diferentes níveis de dificuldades e abstração	Estabelecer uma sequência lógica das atividades utilizando os processos de reconciliação integrativa e de diferenciação progressiva, buscando assim garantir o desenvolvimento de competências de base
Etapa 8: Manter uma vigilância epistemológica no progresso da aprendizagem enfrentando o processo de obliteração subsunçiva	Avaliar continuamente a aquisição dos conceitos de base através de diagnósticos, em especial no âmbito intradisciplinar, e sendo ressignificadas as competências aperfeiçoadas como de base, e com isso sendo agregadas os novos subsunçores.

2.2.2 Implicações Ausubelianas para os Testes

A preocupação em medir a cognição ganhou destaque no início do século XX com os trabalhos sobre a “medida da inteligência” de Charles Spearman e o *fator g* (ou teoria g) em 1904, e de Binet em 1908, em que se acreditava que o teste media aquilo que pretendia medir, ou seja, media a inteligência! Na realidade, a análise fatorial desenvolvida por Spearman mostrava um fator predominante que explicava acima de 50%, podendo chegar a cerca de 80% de um fator geral de inteligência, na qual Thurstone já havia identificado provavelmente na década de 30 e buscado compreender sem sucesso. Esse problema, no entanto, começou a ser resolvido entre os psicometristas na década de 50, estando a Psicometria Clássica (Teoria Clássica dos Testes – TCT) já bem axiomizada (Pasquali & Primi, 2003). Esse avanço se deu em especial com sua *teoria do traço latente*⁴³ de Lazarsfeld (1959), e aos trabalhos de Lord (1952) e Rasch (1980) (Ibid, 2003).

O traço latente se refere a uma família de modelos matemáticos que relaciona *variáveis observáveis* (itens de um teste, por exemplo) e *traços hipotéticos não-observáveis ou aptidões*, estes responsáveis pelo aparecimento das variáveis observáveis ou, melhor, das respostas ou comportamentos emitidos pelo sujeito que são as variáveis observáveis. (Pasquali & Primi, 2003, p. 102)

O conceito de “traço latente” foi introduzido por Lazarsfeld em 1950 (Lazarsfeld, 1950). Ele considerou apropriado inicialmente relacionar esse “traço” à ansiedade como variável intermediária a influenciar nas respostas (R), de acordo com

⁴³ *Latent structure analysis.*

as situações (S), e deixou registrado sua perplexidade com o problema diante das sequencias S-R:

“[...] Mas porque os indicadores têm apenas uma relação de probabilidade com os traços, o problema crucial surge sobre como eles podem ser combinados se não todos seguem na mesma direção [...]”; “[...] É muito cedo mesmo para sugerir uma resposta aqui; mas é surpreendente a gravidade da questão que os estudiosos podem perder quando eles deixam a área de atuação de suas especialidades práticas de pesquisa [...]”. (Ibid, 1950, p. 482)

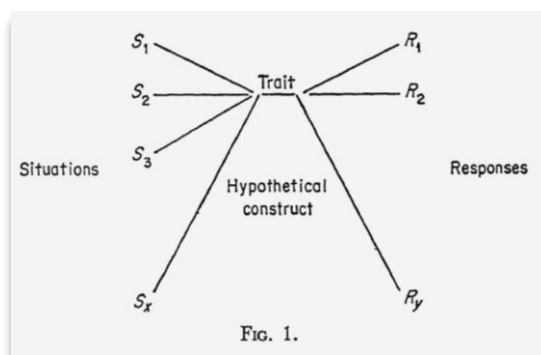


Figura 13: Esquema original da representação proposta de traço latente.

Fonte: Lazarsfeld (1950, p. 482)

O conceito de traço latente se apresenta como construto hipotético dentro de um *continuum*, na qual faz referência a um *continuum latente* distribuído em um espaço unidimensional, e que se projeta em características classificatórias.

Estamos lidando com *características latentes*, no sentido de que seus parâmetros devem, de alguma forma, ser derivados de observações manifestas. Os termos manifestos e latentes não têm conotação aqui além da distinção entre dados diretamente acessíveis ao investigador (manifesto) e parâmetros que, de alguma forma, devem ser inferidos a partir dos dados manifestos (latente). (Lazarsfeld, 1950, p. 490, *tradução livre*).

O conceito de “traço latente” acabou se constituindo como um parâmetro-chave aos modelos de ogiva normal da TRI para Psicometria Moderna, desenvolvidos por Frederic Lord (Lord, 1952).

A escassez de trabalhos na literatura dentro da psicologia educacional que adotam esses procedimentos e discute aquilo que está sendo medido como resultado desses processos dificultam uma melhor compreensão. Devido à preocupação com esse aspecto preliminar foram levantadas algumas definições encontradas na intenção de captar as noções características e contracenar com concepções exploradas dentro

das teorias de aprendizagem, que definem de em várias perspectivas a natureza do conhecimento prévio.

Quadro 19: Algumas ideias sobre o conceito de traço latente na Psicometria.

Referências	Ideias relacionadas ao conceito de traço latente na Psicometria
Moreira Junior (2010, p. 137)	<p>“Traços latentes são características do indivíduo que não podem ser observadas diretamente, isto é, não existe um aparelho capaz de medi-las diretamente, como, por exemplo, um termômetro que mede diretamente a temperatura. Portanto, essas características são mensuradas através de variáveis secundárias que sejam relacionadas com o traço latente em estudo.” (p. 137)</p>
Andrade, Tavares, & Valle (2000)	<p>“O que esta metodologia sugere são formas de representar a relação entre a probabilidade de um indivíduo dar uma certa resposta a um item e seus traços latentes, proficiências ou habilidades na área de conhecimento avaliada.” (p. 3).</p> <p>“[...] quantidade de traços latentes que está sendo medida [...]”; “[...] considerando modelos que avaliam apenas um traço latente ou habilidade [...]” (p. 3).</p> <p>“O modelo proposto pressupõe a unidimensionalidade do teste, isto é, a homogeneidade do conjunto de itens que supostamente devem estar medindo um único traço latente” (p. 16).</p> <p>“Este método baseia-se na existência de uma distribuição (latente) associada à habilidade dos indivíduos da população em estudo [...]” (p. 29).</p> <p>“De forma geral, podemos dizer que estamos trabalhando com variáveis latentes, e nessa situação sempre há a necessidade do estabelecimento da métrica” (p. 49).</p> <p>“Podemos citar, por exemplo, a questão da dimensionalidade do espaço de traços latentes envolvidos na avaliação. Todos os modelos que vêm sendo efetivamente utilizados pressupõem que o conhecimento que se deseja medir pode ser representado por uma única habilidade” (p. 134)</p> <p>“Um ponto importante dessa metodologia é que tanto os itens, através de seus parâmetros, quanto o traço latente associado são medidos em uma mesma métrica, permitindo com isso uma operacionalização dessa característica latente que está sendo medida, bem como a adequação e a contribuição de cada um dos itens aplicados nessa operacionalização” (p. 137).</p>
(Luiz Pasquali & Primi, 2003)	<p>“Estas teorias postulam que o comportamento humano é consequência de processos hipotéticos chamados de traços latentes”; “[...] A CCI define os parâmetros dos comportamentos, ditos itens (dificuldade, discriminação) em função do tamanho do traço latente, expresso como teta (q) [...]” (p. 99).</p>

Observa-se que muitos trabalhos que envolvem os modelos da TRI na Psicometria apresentam por vezes definições genéricas sobre o traço latente, outras ambíguas e confusas (fala-se em processos, proficiências, habilidades, distribuição associada a habilidade, medida de uma característica, etc.). Parece comum que os psicometristas foquem nos processos mentais, porém a natureza epistemológica do

conhecimento que está sendo medido é um fator de complexidade que pode estar influenciando esse processo cognitivo, a atribuição de sua importância está na validade do construto que deve explorar a natureza qualitativa do traço latente, pois ele representa mais dos desempenhos e interpretações diretas dos parâmetros dos itens.

O conceito de traço latente não é isento de ambiguidades e controvérsias entre os autores que trabalham com tal construto. A variedade de expressões utilizadas para representá-lo já indica tal dificuldade. Traço latente vem referido sob denominações como: variável hipotética, variável fonte, fator, construto, conceito, estrutura psíquica, traço cognitivo, processo cognitivo, processo mental, estrutura mental, habilidade, aptidão, traço de personalidade, processo elementar de informação, componente cognitivo, tendência e outros. A própria natureza ontológica de traço latente deixa dúvidas se ele é concebido como um rótulo, representando uma síntese hipotética de um conjunto de comportamentos reais, ou como uma realidade mental. (Pasquali, 1996, p. 74, *grifo nosso*)

Uma possível razão seria a concepção tradicional comportamentalista intrínseca e predominante entre os pesquisadores que desenvolvem e utilizam esses procedimentos, mas não avançam, na maioria das vezes, sobre as interpretações do que representa “o traço” diante da situação, compreendendo como fatores altamente explicativos diante de outros construtos, e componente representativo dos processos mentais que converge para expectativas de comportamentos: “*É a tradição na orientação da análise fatorial em Psicometria, onde os fatores são concebidos como variáveis-fonte responsáveis pela qualidade da execução das tarefas comportamentais*” (Pasquali, 1996, p. 75), ou seja, é uma concepção fatorista, criticado que Sternberg (1977).

Para os *fatoristas* os *traços latentes* representam um conjunto de processos cognitivos necessários para a execução de uma tarefa (de fato um barramento correlacionado de processos), falando de habilidades primárias, que seriam combinações de processos cognitivos elementares, isto é, de representações mentais de objetos e símbolos (Pasquali, 1996). De modo geral, há duas concepções: *elementarista (reducionista) e estrutural*.

Concepção	Traço latente	Enfoque	Característica	Ilustrativa	Referência	
Elemento	<ul style="list-style-type: none"> Processo elementar de informação Componente cognitivo 	Processamento da informação	Processo mais elementar possível	Elemento atômico da física nuclear	Newel & Simon (1972)	
		Psicologia cognitiva	Último elemento cognitivo de uma atividade cognitiva	Elemento da tabela periódica de química	Sternberg (1977)	
Estrutura	Fator	Psicometria	Holística	Aptidão	Fator g, conhecimento	Spearman, Piaget
				Emoção	Ansiedade, ...	Tipologias
				Motivação	Autoestima, ...	Tipologias
			Diferencial	Aptidões diferenciais	Habilidades	Thurstone, Cattell, Guilford ...
Emoções	Traços de personalidade	Idem (fatoristas)				
			Motivações	Necessidades básicas	Idem (fatoristas)	

Figura 14: Visão elementar e estrutural do traço latente.

Fonte: Adaptado de Pasquali (1996, pp. 74-75).

Testes conceituais exploram o conhecimento, e esse se relaciona a diversos conceitos. O termo *conceitual* e *conceptual* são comumente interpretados de forma equivalente quando se considera um conceito como uma explicação acerca de algo, algumas discussões sobre esse aspecto peculiar etimológico podem ser encontradas no trabalho de Dworkin (1986, p. 71). Considera-se, entretanto, que um determinado conceito possa remeter a diversas formas de ser compreendido ou concebido, ou seja, os conceitos térmicos envolvidos remeterem a diversas concepções entre os estudantes, sendo muitas dessas concepções consideradas equívocos. Entende-se que o teste dito “conceitual” não garante aferir propriamente o conceito em si, mas sim indícios de concepções associadas aos conceitos, podendo haver uma grande chance de ser concepções científicas quando a resposta é esperada, como concepções não-científicas quando as respostas são atraídas pelos distratores, nesse caso, tendo uma grande chance de ser uma concepção representadas com base em equívocos.

Diante disso, considera-se que:

- As *concepções alternativas* sejam conhecimentos implícitos preexistentes diante de conceitos ou que se constroem diante da necessidade de compreender uma situação, sendo carregados por crenças e formas alternativas para justificar, argumentar, explicar, resolver ou responder a um problema;
- Um *subsunçor*, de acordo com Ausubel, se refere a uma ideia-âncora favorável a se associar a um *conhecimento potencialmente significativo*, e que;
- Os *traços latentes* sejam medidas indiretas de características do indivíduo que não podem ser observadas diretamente.

Entende-se que entre essas características se tenha o conhecimento prévio, podendo ser CAs ou não. A incidência de respostas aos distratores em um teste podem ser considerados indícios de que essas características cognitivas não observadas diretamente sejam CAs, e as incidências diante das respostas corretas sejam evidências de um *subsunçor*.

As **observações** diante das respostas obviamente remetem a uma análise do estímulo que causa o item, podendo o enunciado ser um dos fatores externos a influenciar diretamente as respostas. Esse se torna um campo fértil para desenvolver muitas suposições, conjecturas e interpretações para tentar compreender as capacidades de compreensão e linguagem na aprendizagem, inclusive sendo algumas delas confrontadas dentro psicologia educacional e no campo da Linguística. Dentre os fatores interno nesse processo, o *inatismo* de Noam Chomsky busca explicações biológicas da genética advindas da hereditariedade para as construções cognitivas, enquanto Jean Piaget propõe uma *psicogênese dos conhecimentos* com significados epistemológicos advindos eminentemente de uma concepção construtivista e dialética na relação sujeito-objeto (Matui, 1995, pp. 43-44).

Os aspectos a serem *aprimorados* no item ou na questão de modo geral se referem a manter o item e buscar aperfeiçoar sua interpretação identificando e eliminando termos com problemas semânticos de interpretação, ou ainda por sua omissão. Esses problemas podem estar implícitos ou explícitos tanto no enunciado quanto nos distratores. Quanto ao enunciado podem haver problemas de indução ao erro de

forma explícita ou pela ausência de informações relevantes, esses podem ser considerados numa concepção ausubeliana parte integrante de um *conhecimento potencialmente significativo* a serem ancorados nas *ideias-âncoras* ou *subsunçores* supostamente preexistentes na estrutura cognitiva dos respondentes, na qual o teste visa evidenciar.

As informações da *situação-problema* devem caracterizar a identificação de um traço latente em potencial que possa sobressair diante de outros traços secundários que geralmente estão envolvidos de forma implícita. O traço latente em potencial em testes de aferição do conhecimento envolvem dimensões sobre o processo e sobre a categoria desse conhecimento, entretanto, também deve ser considerado dificuldades inerentes a natureza epistemológica do conhecimento envolvido dentro da situação, que podem se referir a aspectos relacionadas à categoria do conhecimento e de sua associação dentro do contexto, ou seja, relações de fatos, conceitos e processos do conhecimento científico que estão situados dentro do contexto. O fato é que o traço latente em potencial a ser aferido deve estar claro para garantir uma fiabilidade do instrumento, possuindo indícios de validade que garanta medir aquilo que se deseja medir.

Os aspectos a serem *reelaborados* no item ou na questão de modo geral se referem identificar o traço latente em destaque e buscar uma nova forma de aferi-lo situando em um contexto mais claro, específico e minimizando ao máximo vieses interpretativos. Dessa forma se torna necessário compreender e seguir um rigor sobre o processo da elaboração e proposição de situação-problema.

Imaginemos um elaborador de itens (ou professor) intencionado em avaliar a relação entre um suposto (!?) conhecimento “potencialmente significativo” (a, b e c – ou parte deles) representado de forma associada pelas figuras geométricas (unidade complexa de conhecimentos relacionados), e os respectivos subsunçores (A, B e C – ou parte deles) dos itens y_1 , y_2 e y_3 . O aluno por sua vez pode estar convicto de uma resposta correta caso esteja com os subsunçores disponíveis, ou predisposto a realizar essa associação com conhecimento potencialmente significativo.

Quadro 20: Concepção ausubeliana nas interpretações pedagógicas dos itens.

Unidade de Análise: Estrutura Cognitiva	Instrumento: Teste Conceitual
Fundamento: Aquisição e retenção de conceitos	Conhecimentos envolvidos: Calor, Temperatura, Processos Termodinâmicos e Propriedades Térmicas
<p>Processo cognitivo teoricamente hipotético para a ocorrência de uma aprendizagem significativa: Assimilação por subsunção.</p> <ul style="list-style-type: none"> • Consideração: O teste em si não é um processo de aprendizagem, mas de verificação de que ela supostamente tenha ocorrido anteriormente, apesar de que possa ocorrer eventualmente durante sua realização. 	<p>Conhecimento potencialmente significativo: Apresenta-se de forma explícita no enunciado da situação-problema, ou pode remeter propositalmente a esse conhecimento de forma implícita, caso o item esteja bem elaborado.</p> <ul style="list-style-type: none"> • Consideração: O conhecimento potencialmente significativo não significa que esteja em sua completude no enunciado, no caso, o enunciado da situação-problema pode remeter a uma associação de diversos fatos, conceitos, relações entre conceitos, processos, princípios e teorias conjuntamente (ou não) vistos anteriormente (função avaliativa verificativa), bem como a remeter para que ele ainda seja abordado (função avaliativa diagnóstica) ou reforçado (função avaliativa reguladora)
<p>Condição de ocorrência: Preexistência de subsunções na estrutura cognitiva, ou seja, ideia âncora favoravelmente capaz de ser associada a um conhecimento familiar ou potencialmente significativo.</p> <ul style="list-style-type: none"> • Consideração: Os subsunções não são necessariamente conhecimentos isolados de forma que tenham sido recebidos previamente de forma literal e arbitrária (aprendizagem mecânica), podendo estar previamente associados com outros conhecimentos científicos (aprendizagem significativa) ou estruturados por uma rede de conhecimentos relativamente estáveis como uma unidade complexa, porém, sendo considerados concepções ou ideias equivocadas diante dos conceitos (concepções alternativas). Dependendo da natureza desse conhecimento prévio, o aluno pode tomar a decisão por uma alternativa correta ou mais plausível, ou ser atraída por um distrator. De uma forma ou outra, o aluno responde conscientemente diante daquilo que seja mais familiar ao conhecimento ou crença preexistente em sua estrutura cognitiva 	<p>Identificação de um subsunção dentro de um suposto conhecimento “potencialmente significativo”: O subsunção a ser verificado é aquilo que se quer avaliar no item, portanto se caracteriza como aspecto central da proficiência necessária para responder corretamente o item, podendo assim configurar como traço latente do item. Para isso, acredita-se que possa supostamente ser evidenciado na resposta correta, estando literalmente expresso ou remeter a ele.</p> <ul style="list-style-type: none"> • Consideração: A verificação prévia de um subsunção é uma conjectura hipotética construída pela análise e interpretação de um conteúdo considerado relevante e necessário a se saber diante de uma situação-problema, entretanto somente se torna significativo aos sujeitos que tenham os subsunções disponíveis na estrutura cognitiva.

Convêm ressaltar que o propósito dos testes diagnósticos, nesse e na maioria dos casos, não possuem a intenção de que a aprendizagem significativa ocorra no momento da resolução dos itens. Eventualmente isso pode ocorrer e ser verificado de certa forma em um pós-teste, porém, sem que haja garantias. A intenção é que isso seja parte estratégica de um processo que favoreça uma assimilação subsunção, apresentando subsídios para que os desdobramentos que buscam esse tipo de aprendizagem ocorram de forma mais eficiente. Pressupor essas situações, entretanto, podem projetar condições que ajudam na elaboração e aperfeiçoamento dos itens, que por sua vez podem refletir em ações dentro de outros campos pedagógicos.

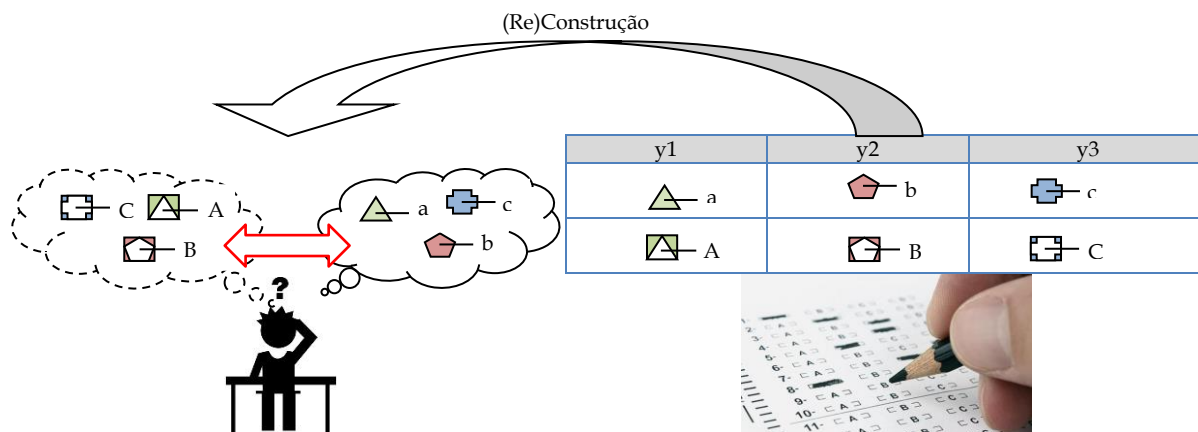


Figura 15: Suposição de assimilação subsunsa durante o processo de resolução de problemas.

Nisso, deve se considerar o problema com a *representação do constructo* no processo de interpretação do item, pois estes se encontram subjacentes à *extensão nomotética* (isto é, os correlatos dos resultados dos exames) (Embretson, 1994). Esse processo ocorre pela capacidade do item manipular estímulos em seu enunciado influenciando a extensão nomotética, e como consequência, a representação do constructo, sendo com isso relevante para a sua eficiência que sua (re)elaboração seja fundamentada em uma teoria cognitiva (Ibid., 1994), recorrendo as teorias da aprendizagem na elaboração dos constructos (Webb, 2006), e não sendo encarados meramente como um processo artístico.

[...] as manipulações dos estímulos dos itens devem influenciar a extensão nomotética do teste, bem como a representação da construção. Em terceiro lugar, o desenvolvimento de itens se torna um processo científico e não um processo artístico. Os itens são desenvolvidos para representar especificações explícitas que influenciam o processamento. O elaborador de um item torna-se um experimentador que projeta tarefas para refletir aspectos específicos de uma teoria na tarefa. Para algumas tarefas, as especificações tornaram-se suficientemente explícitas para que a inteligência artificial seja usada na construção de itens (por exemplo, Bejar, 1990; Homke & Habon, 1986). Em quarto lugar, **a teoria cognitiva pode melhorar a eficiência do desenvolvimento de testes. São esperados itens de melhor qualidade com dificuldades mais específicas. O desenvolvimento de itens por especificações da teoria deve produzir itens discriminantes melhores porque os estímulos de itens que influenciam processos irrelevantes podem ser eliminados.** Consequentemente, as tentativas empíricas devem resultar em menos desgaste de itens. Além disso, as tentativas empíricas precisam incluir apenas itens com os níveis de dificuldade desejados porque as especificações também produzem previsões de níveis de dificuldade do item. (Embretson, 1994, p. 108)

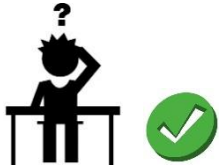
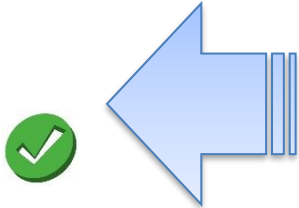

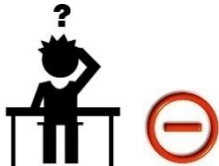
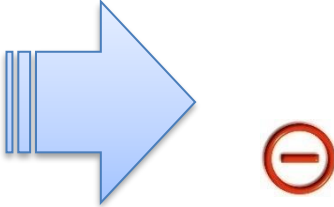

Os estímulos na leitura do texto é um processo interpretativo mediado com sua estrutura cognitiva, podendo influenciar as expectativas do sujeito ao determinar o

sentido da visão captado pelo respondente, e com isso, se projeta a responder. O método de recolha, sendo um teste ou questionário, poderá captar uma *visão ideal* quando se pretende obter opiniões dos participantes sobre o objeto (aquilo que ele gostaria que fosse), e *atual* quando se deseja uma opinião sobre aquilo que ele percebe que o objeto lhe transmite (aquilo que o objeto é) (De Ketele & Roegiers, 1993).

A natureza do instrumento e o seu conteúdo devem ser elaborados buscando estabelecer uma visão atual para os sujeitos, uma interpretação direta e clara do problema evitando causar introspecções distorcidas. Não há garantias de que isso possa ser eliminado, mas podem ser controladas e previstas algumas crenças e expectativas intuitivas de como ele possa interpretar e compreender as situações-problemas, influenciando em suas escolhas de resposta. Essa possibilidade está na capacidade de o aluno refletir sobre a situação-problema buscando se aproximar de situações análogas e familiares para realizar comparações e tomar um juízo de valor, nesse caso, o respondente acessa *subsunções* relativamente próximos, considerados ao mais plausíveis disponíveis. Quando isso ocorre, pode-se dizer que esse processo segue uma lógica relativamente consistente e consciente, características típicas de concepções alternativas, sendo mais comum de acontecer quando o item exige uma habilidade que o respondente não possui. Existe ainda a possibilidade do acerto nessas circunstâncias situação não ser totalmente ao acaso, pois o aluno não opta de forma aleatória, podendo haver tentativas de travessia em sua *zona de desenvolvimento proximal* (cf. Lev Vygotsky), e de forma pouco provável, uma aprendizagem durante a realização da prova, ainda que não haja a intenção para que isso ocorra.

Dessa forma, entende-se que o aluno com dificuldades de responder, possa buscar identificar novas formas de interpretar o problema. Nesse processo é esperado que faça novas varreduras no enunciado e nas alternativas, com intuito de identificar argumentos ou elementos que supostamente fariam sentido se existissem, podendo se aproveitar de variáveis parasitas, quando elas existem.

Quadro 21: Sentido atual e ideal no suposto processo interpretativo dos itens pelo respondente⁴⁴.

Sua opinião (crenças, concepções, <i>subsunçores</i> , ... – conhecimento prévio) diante daquilo que situação-problema é (ou aquilo que compreende que seja)!		
VISÃO ATUAL: A imagem/representação/compreensão que a situação-problema provoca no aluno, acessando <i>subsunçores</i> , caso estejam disponíveis.		
Como o aluno compreende a situação-problema!		
		
Sua opinião (crenças, concepções, <i>subsunçores</i> , ... – conhecimento prévio) diante daquilo que situação-problema deveria ser (ou aquilo que gostaria que fosse)!		
VISÃO IDEAL: A imagem/representação/compreensão que o aluno possui diante daquilo que ele gostaria que fosse a situação-problema (no caso dele possuir <i>subsunçores</i> diferentes daquele que se exige, com isso, o respondente deseja que o problema fosse outro, estando de acordo com o conhecimento disponível, muito provavelmente ele irá buscar situações análogas, próximas a situação proposta, com isso ele opta por aquela que seja mais ou menos plausível, caso isso não ocorra, ele arrisca um acerto totalmente ao acaso)		
Como o aluno compreende que deveria ser a situação-problema!		
		

Ocorrendo esse tipo de situação, o aluno tenderia a buscar por uma nova interpretação do problema, e assim, ao fazê-lo, poderia distorcer o sentido do enunciado a tomar uma decisão equivocada, talvez por aquela que parecer mais lógica diante do conhecimento que lhe predispõe, ou ainda, podendo justificar o erro ao acaso para respondentes proficientes que cometem equívocos por descuidos devido a “pegadinhas” no item. Apesar disso, deve-se considerar a possibilidade da utilidade na predisposição de situações análogas que possam exercer a função de facilitadores interpretativos, remetendo a saberes metodológicos articuladores e artifícios lógicos relativamente válidos, pois há um risco de chegar a uma conclusão certa através de equívocos.

⁴⁴ Créditos da foto: <<https://www.aprovaconcursos.com.br/noticias/wp-content/uploads/2013/04/Teste-de-multippla-escolha-size-598.jpg>>.

Quadro 22: Ideias centrais da psicologia educacional de Ausubel e algumas possíveis contribuições para as análises psicométricas.

Aspectos fundamentais	Psicologia educacional de David Ausubel - <i>Teoria da Assimilação e Retenção Significativas de Conceitos</i>	Possibilidades instrucionais ausubelianas no uso de testes conceituais para o processo de ensino-aprendizagem
Tipos de aprendizagem por recepção significativa	<p>Representacional (equivalência entre os símbolos arbitrários e os seus referentes correspondentes - objetos, exemplos, <i>conceitos</i>, passando a remeter o indivíduo ao mesmo significado - memorização nos primeiros anos de vida); Conceitos (significado de <i>conceitos</i>, objetos e acontecimentos representados por símbolos individuais e arbitrários, ou seja, abstrações dos atributos criteriais dos referentes ou regularidades em eventos ou objetos, seja como nomes e palavras e da sua relação com o próprio conceito); Proposicional (significados das ideias expressas por grupos de palavras (geralmente representando <i>conceitos</i>) combinadas em proposições ou sentenças).</p>	<ul style="list-style-type: none"> Definir a finalidade da avaliação diagnóstica com base na verificação de <i>subsunçores</i>; Analisar previamente o <i>conhecimento estruturante</i> do conteúdo de modo geral (ramo do conhecimento) identificando os <i>conceitos mais inclusivos</i> e <i>conhecimentos potencialmente significativos</i> envolvidos nas unidades temáticas subjacentes, e assim incorporá-los na elaboração de itens nos testes; Identificar os <i>conhecimentos potencialmente significativos</i> relacionados aos <i>conceitos mais inclusivos</i> dentro de cada unidade temática e da situação apresentada no problema, assim como dos possíveis subsunçores relacionados necessários para evidenciar uma apreensão do conhecimento de acordo com a tomada de decisão; Classificar previamente através de um quadro de tipicidade sobre as dificuldades conceituais (critérios referenciados com base em: concepções alternativas típicas, análise de gráficos; resolução literal de problemas, etc.) a serem exploradas entre os distratores no teste, relacionando o <i>subsunçor</i> com a habilidade envolvida a ser adquirida; Elaborar uma <i>matriz de habilidades cognitivas</i> de acordo com os itens que compõem um teste preestabelecido, destacando supostos subsunçores esperados e os conhecimentos potencialmente necessários;
Princípios na aquisição de conceitos (aprendizagem conceitual)	<p>Formação de conceitos é a aquisição espontânea de ideias genéricas por meio das experiências empírico-concreta em crianças de nível pré-escolar; Assimilação de conceitos por subsunção e obliteradora (ou por subsunção subordinada, relacionada ao esquecimento)</p>	<ul style="list-style-type: none"> Identificar os subsunçores preestabelecidos em um teste conceitual de acordo com os níveis de abstração e complexidade envolvidos e considerados dentro de uma <i>análise taxonômica de objetivos educacionais</i>; Identificar <i>famílias de situações-problemas</i> a serem desenvolvidas dentro de uma unidade temática de acordo com o objetivo educacional proposto, estando claro os <i>subsunçores</i> comuns dentro de um conjunto de habilidades cognitivas necessárias para se alcançar um patamar de competência científica que em envolva base conceitual;
Premissas	<p>Unidade de análise (intrapessoal) é a estrutura cognitiva do indivíduo; Fator isolado mais importante que influencia a aprendizagem é aquilo que o aluno já sabe, ou seja, o conhecimento prévio (conceitos subsunçores, representações, esquemas,</p>	<ul style="list-style-type: none"> Enfatizar as dificuldades conceituais encontradas nos resultados de desempenho do teste dando-lhes sentido, e buscando uma relação entre o conhecimento <i>subsunçor</i>, as dificuldades e os <i>conhecimentos potencialmente significativos</i> necessários; Apresentar sugestões para a reelaboração ou aprimoramento do item de modo a favorecer a verificação do conhecimento prévio <i>subsunçor</i> envolvido dentro do contexto da situação-problema;

	<p>modelos, construtos pessoais, concepções alternativas, invariantes operatórios, enfim, cognições já existentes na estrutura cognitiva do aprendiz – Moreira, 2011).</p>	<ul style="list-style-type: none"> • Selecionar uma família de situações-problemas estratégicas, significativas e diversificadas em níveis de dificuldades e abstração ao aluno relacionando o contexto de vida real, e a partir dos <i>subsunçores</i> e dificuldades encontradas, buscar promover o processo de <i>diferenciação progressiva</i>; • Sugerir um <i>conjunto de conhecimentos potencialmente significativos</i> a serem priorizados de acordo com cada <i>subsunçor</i>, assim como para tentar extirpar equívocos e possibilidades de maior inclusão de estabilidade de erros e resistências a “mudanças conceituais”; • Propor uma <i>organização hierárquica de conteúdos</i> mais inclusivos seguindo para os mais específicos tendo como parâmetro de relevância a necessidade de disponibilidade de subsunçores evidenciados, utilizando os conhecimentos potencialmente significativos dentro de unidades temáticas específicas por meio de tarefas e ideias novas não-familiares • Selecionar os conteúdos a serem abordados continuamente nos diversos ciclos de aprendizagem visando a elevação temporária do <i>limiar de disponibilidade</i> dos <i>subsunçores</i> evidenciados e considerados integradores entre diferentes campos do conhecimento, favorecendo a retenção, <i>minimizando</i> a competição de memórias alternativas e conflituosas, assim como o inevitável processo de <i>assimilação obliteradora</i>; • Elaborar <i>organizadores avançados</i> como “ponte cognitiva” entre aquilo que o aluno já sabe e deve aprender. Não ausência pode servir para influenciar ou alterar vários atributos inclusos na estrutura cognitiva com relação a ideais ancoradas em um nível subordinante (clareza, estabilidade, capacidade de discriminação, coesão e integração de subsunçores), e assim tornar disponível subsunçores relevantes e próximos, depois verificar na resolução de situações-problemas contidas em testes conceituais. • Elaborar <i>organizadores avançados</i> para exercer a função de facilitador, mediador e manipulador da estrutura cognitiva tendo como base os <i>subsunçores</i> evidenciados no teste conceitual, e assim potencializar a recepção significativa de uma <i>nova informação</i> de acordo com o contexto da situação e sequenciação da unidade temática, sugerindo favorecer um desenvolvimento de competências cada vez mais abstratas e complexas; • Elevar o nível de abstração, generalidade e inclusividade dos <i>organizadores avançados</i> diante dos novos materiais potencialmente significativos (para o sujeito) e lógicos (ter estrutura, organização, exemplos, linguagem adequada, etc.), tendo como suporte o
<p>Condições de ocorrência</p>	<p>A retenção significativa está na interação entre o subsunçor relevante e inclusivo (A) do indivíduo às novas ideias e conceitos, potencialmente significativos (a); O ensino se consuma quando o subsunçor do aluno é modificado pela associação com um novo conhecimento tornando-o mais inclusivo (A´a´); A aluno precisa de ter uma disposição para aprender, sem intenção de memorizar o material arbitrariamente e literalmente; O material a ser aprendido tem que ser potencialmente (natureza do material) e psicologicamente significativo (experiência de cada sujeito);</p>	
<p>Processos de aprendizagem significativa</p>	<p>Mecanismos cognitivos de aprendizagem significativa por descoberta, compreensão e incorporação de aspectos não-literais e de forma não-arbitrária (situação de aprendizagem significativa); Materiais potencialmente significativos: (1) <i>tarefas de aprendizagem logicamente significativas, ou seja, não aleatórias, sensíveis e plausíveis para se relacionarem, de forma não arbitrária e substancial aos subsunçores preexistentes</i>; (2) que de fato esteja <u>preexistente os subsunçores na estrutura cognitiva do aprendiz</u>.</p>	

<p>Formas organizacionais de <i>facilitação de transferência</i></p>	<p>Processo de diferenciação progressiva (as ideias mais gerais e inclusivas da disciplina e, depois, estas são progressivamente diferenciadas em termos de detalhes e especificidade); Reconciliação integrativa (explorar ideias relacionadas indicando semelhanças e diferenças significativas, para se reconciliarem as inconsistências reais ou aparentes e para se combinar ou integrar ideias semelhantes que sejam logicamente relacionais umas com as outras); Organização sequencial clara, explicativa e integradora do conteúdo substantivo (dependências sequenciais temáticas; hierarquia partindo de conteúdos mais inclusivos, gerais e abstratos para os mais específicos, relevantes e relacionáveis; “organizador” para cada unidade temática; disposição lógica das tarefas); Consolidação (através da confirmação, correção e clarificação, no decurso do retorno (<i>feedback</i>), e através da prática diferencial e da revisão, no decurso da exposição repetida, com retorno, ao material de aprendizagem).</p>	<p>conhecimento prévio <i>subsunçor</i> para sugerir <i>conhecimentos potencialmente significativos</i> a serem abordados;</p> <ul style="list-style-type: none"> • Elaborar <i>organizadores posteriores</i> após os ciclos de aprendizagem visando “subir e descer” estruturas hierárquicas cognitivas tendo como base a sequenciação organizacional dos conteúdos abordados, apontando semelhanças, diferenças e aspectos discrepantes, e assim promover um <i>processo de reconciliação integradora</i>; • Apresentar <i>organizadores explicativos</i> para alunos de baixo desempenho que não apresentam subsunçores disponíveis, ou para promover esses subsunçores diante de um conteúdo não-familiar, e assim tentar garantir uma progressão na aprendizagem; • Posicionar indivíduos alto e baixo desempenho numa escala de proficiência identificando tipicidade de grupos de respondentes e propriedades de subsunçores em diferentes níveis de proficiência, possibilitando realizar comparações e eventualmente verificar a existência de patamares de estabilização de dificuldades conceituais e de concepções alternativas • Propor intervenções específicas para o processo de ensino-aprendizagem de determinada unidade temática levando em consideração a tipicidade de grupos de desempenho (alto, mediano, baixo – tendo como parâmetro os valores bisseriais), de acordo com grupo de subsunçores evidenciados ou ausentes, assim como das dificuldades conceituais envolvidas; • Propor intervenções específicas individualizadas para alunos com baixo e alto desempenho, seja para ajudar na progressão daqueles com maiores dificuldades como para potencializar os de maiores proficiências (alunos <i>experts</i>); • Sugerir o desenvolvimento de uma sequenciação de unidades de ensino potencialmente significativas que esteja alinhada as necessidades diagnosticadas nas dificuldades conceituais dos alunos, levando em conta as diferentes composições de grupo sugeridas pelos bisseriais dos distratores;
<p>Etapas gerais e necessárias</p>	<p>Realizar uma aprendizagem significativa; Evidenciar a aprendizagem significativa e; Neutralizar o processo de assimilação obliteradora. Recomenda-se de modo geral: (1) <i>identificar no conteúdo escolar o hipotético subsunçor que irá se relacionar com o conhecimento potencialmente significativo</i>; (2) <i>verificar a preexistência do subsunçor no indivíduo</i>.</p>	<ul style="list-style-type: none"> • Estabelecer metas diferenciadas para que a maioria dos indivíduos possam alcançar um nível mais elevado na escala de proficiências de concepções térmicas em um pós-teste, levando em consideração intervenções de ensino que utilizem os mecanismos e materiais significativos com base nos subsunçores diagnosticados; • Desenvolver múltiplas alternativas de resolução (utilizando um arcabouço de saberes metodológicos possíveis para cada situação), diversidade de aplicações em diferentes níveis de dificuldade para o <i>conhecimento potencialmente significativos</i> dentro de unidades temáticas partido daqueles mais estruturados e inclusivos, e seguindo para situações de

<p>Implicações metodológicas</p>	<p>Uso de organizadores avançados⁴⁵ como “pontes cognitivas” entre aquilo que o aluno já sabe e deve aprender; Usar organizadores explicativos para promover subsunçores relevantes aproximados diante de um material totalmente não-familiar; Uso de organizadores comparativos para integrar novas ideias com conceitos basicamente similares existentes na estrutura cognitiva visando a discriminabilidade evitar confusão com aspectos similares; Uso de organizadores posteriores durante a após um ciclo de aprendizagem visando integrar, associar e comparar ideias e discrepâncias; Uso de organizadores expositivos para melhorar a relevância (disponibilidade), estabilidade e clareza das ideias ancoradas (não se preocupa com a discriminação); Uso de mapas conceituais</p> <p>atuação conceitual mais específicas e detalhadas visando favorecer o processo de <i>diferenciação progressiva</i>;</p> <ul style="list-style-type: none"> • Problematizar e instigar por meio de questionamentos, solicitar previsões, descrições de fenômenos observados, socializar as dificuldades, dúvidas e convicções (baseadas em subsunçores, concepções alternativas, entre outros) por meio do intercâmbio de ideias e discussões, oportunizando a argumentação científica entre os grupos de alto e baixo desempenho; • Consolidar e estruturar o conhecimento apreendido e verificado; • Propor a construção de mapas conceituais e compará-los as análises taxonômicas e mapas de construto de acordo com conceitos e ideias principais; • Propor questões-desafio dentro de um nível de proficiência mais elevado na escala para alunos <i>experts</i>, levando em consideração as evidências de subsunçores e de seu processo de desenvolvimento de competências diante das capacidades em solucionar famílias de situações-problema.
----------------------------------	---

Fonte: Elaborado com base em Ausubel (2003).

⁴⁵ Também denominado na literatura como Organizador Prévio.

2.3 CONTRIBUIÇÕES DA PSICOMETRIA

2.3.1 Contexto Histórico

“[...] um bom psicometrista, sobretudo prático, não necessita ser um exímio estatístico. Ele deve ser, sim, um exímio conhecedor da teoria psicológica.”
(Pasquali, 2011, p. 12)

Atualmente torna-se evidente a necessidade de buscar novas técnicas e métodos que melhor se adequem ao processo de avaliação no ensino. Algumas dessas técnicas estão organizadas na Psicometria. Etimologicamente o termo «Psicometria» representa a teoria e a técnica de medida dos processos mentais, especialmente aplicada na área da Psicologia e da Educação, e pode ser definido como a teoria da medida em ciências para explicar o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas e propor técnicas de medida dos processos mentais (Pasquali, 2011).

A Psicometria possui a finalidade de estudar fenômenos psicológicos, no entanto, a Psicometria vem sendo tradicionalmente dominada por pesquisadores com formação em estatística, fazendo com que fosse concebida erroneamente por muitos como um ramo da Estatística, configurando um equívoco de perspectiva, que vem sendo desde a década de 30 debatido pelo matemático e psicólogo Louis Leon Thurstone (1887-1955), pois os «números» nessa ciência são apenas uma representação dos fenômenos psicológicos (Ibid, 2011).

As contribuições educacionais que advêm da Psicometria para a Psicologia Educacional passam pela *necessidade* de uma concepção mais cognitivista nas interpretações pedagógicas dos resultados (além das expectativas comportamentais) e do exercício adequado de suas finalidades, bem como de uma compreensão minimamente suficiente pelos pesquisadores e educadores preocupados com o processo de avaliação e melhorias na qualidade do ensino, principalmente quando o próprio sistema adota esse paradigma de referência. Apesar disso, é comum perceber pesquisadores demonstrando aparentemente um ceticismo ingênuo diante aos estudos que tomam decisões ou constroem um juízo de valor com base em “números” e/ou medidas, alegando que a natureza dos fenômenos educacionais deva se

concentrar exclusivamente no paradigma investigativo qualitativo. Além disso, muitas críticas tem como base se apoiar em pressupostos epistemológicos mais gerais e reacionários da concepção positivista clássica, que por sua vez já foram em grande parte superadas pelo pós-positivismo. O posicionamento nesse discurso busca associar quase sempre às políticas educacionais mal sucedidas de responsabilização (*accountability*), argumentos esses muitas vezes sem uma fundamentação mais consistente e sem conhecimento de causa.

O rigor tanto na aplicação quanto na análise de um teste para avaliar conhecimento específico é por vezes bem diferente de um teste psicológico padronizado na qual se exige geralmente o acompanhamento de um psicólogo e possui um manual bem orientado. Supostamente, essa diferença ocorra pela distinção dos objetos de estudo. Entretanto, esse é um equívoco, pois, de modo geral, tanto um quanto outro explora a natureza humana diante daquilo que se pensa e sente, portanto ambas fazem parte da psicologia, assim como a fisiologia sensorial na qual se desenvolveu a *Psicofísica*⁴⁶ de Gustav Fechner (1801 - 1887) e do *behaviorismo* de Ivan Pavlov (1894 - 1936). Com o avanço, os desdobramentos nas perspectivas teóricas e de análise dentro das medidas e avaliações psicológicas que remontam o final do século XIX e início do século XX, acabaram por conduzir o desenvolvimento e consolidação da *Psicometria* como campo específico de estudo.

Evidentemente, as raízes históricas e epistemológicas da Psicometria mostram que a sua prática segue guiada por influências da concepção positivista baconiana do empirismo, lógica essa já refutada por David Hume (1711 - 1776) e por Karl Raimund Popper (1902 - 1994), sendo essa uma das questões preocupantes que vem sendo superada, felizmente, mais recentemente pela *Psicologia Cognitiva*. Outro ponto crítico na Psicometria é a concepção estatística sobre a psicológica, e justifica-se pelos precursores e desenvolvedores da Psicometria serem estatísticos de formação. Ao

⁴⁶ Se dividia em psicofísica *interna* e *externa*. Segundo Fechner (1860), a psicofísica interna se refere às funções neurais ou às relações sensoriais e as suas respectivas atividades neurais que a suportam, enquanto a psicofísica externa buscava a relação entre sensações e as correspondentes propriedades físicas e as variações dos próprios objetos.

contrário disso, se configura como um ramo da “*Psicologia dedicada à pesquisa de modelos matemáticos dos processos psicológicos, mas sempre a serviço deles*”, conforme concebido por Thurstone (Ibid, 2011).

O inglês Francis Galton (1822 – 1911), primo de Charles Darwin, é considerado o “Pai da Psicometria” e fundador da *Psicologia Diferencial*. As tendências dessa ciência partem por um lado da psicologia alemã voltada para a introspecção e *experiência subjetiva*, e ao empirismo inglês e norte-americano interessado no *comportamento*. Essa perspectiva histórica mais detalhada se encontra em diversos estudos (Ibid, 2011). Com o avanço nessa área de pesquisa surge a *Psicometria Moderna* com duas vertentes: a *Teoria da Resposta ao Item* (TRI) e a *Teoria Clássica dos Testes* (TCT), que são em grande parte complementares e sem rompimento, apesar das diferenças epistemológicas. Segundo Cohen, Swerdlik e Sturman (2014, p. 8), a Psicometria pode ser definida como a parte da psicologia preocupada com a medição psicológica ou ciência da mensuração psicológica, com isso a construção, análise, validação e uso de testes são seus principais objetos de estudo, cada etapa da elaboração ou de uso de testes possuem critérios específicos que precisam ser considerados. No caso da educação, esses testes estão a serviço de avaliar ou aferir conhecimentos e estes carregam em si crenças e questões epistemológicas que são geralmente interpretados dentro das teorias de aprendizagem, especificamente no campo da psicologia educacional. O *adjetivo* psicométrico se refere à mensuração de natureza psicológica, e o *substantivo* psicometrista ao profissional que usa, analisa e interpreta dados de testes psicológicos.

Para que se tenha uma visão mais ampla sobre o progresso da testagem e avaliação psicológica, Cohen, Swerdlik e Sturman (2014), apresenta uma cronologia abrangendo de 2200 a.C. até o presente.

Quadro 23: Contexto histórico da Psicometria.

Idade Antiga: de 4000 a.C. até 476 d.C.	<p>2200 a.C. – Sabe-se que um teste de proficiência foi conduzido na China. O imperador avaliava periodicamente os funcionários públicos.</p> <p>1115 a.C. – Na China, os exames do serviço público abertos e competitivos são comuns durante a Dinastia Chang. A proficiência é testada em áreas como aritmética, escrita, geografia, música, agricultura, equitação, ritos e cerimônias culturais.</p> <p>400 a.C. – Platão sugere que as pessoas devem trabalhar em empregos compatíveis com suas capacidades e seus talentos - um sentimento que será compartilhado muitas vezes através dos tempos por psicólogos, profissionais de recursos humanos e pais.</p> <p>175 a.C. – Cláudio Galeno (também conhecido como Galeno) cria experimentos para demonstrar que o cérebro, não o coração, é a sede do intelecto.</p>
---	--

<p>Idade Média: de 476 d.C. até 1453.</p>	<p>476 d.C. – A chamada Idade das Trevas (Idade Média) inicia, e a sociedade força a ciência a relegar a segundo plano (temporariamente) a questão da fé e da superstição.</p>
<p>Idade Moderna: de 1453 até 1789.</p>	<p>1484 – O interesse nas diferenças individuais concentra-se, sobretudo em questões como "<i>Quem está associado a Satanás?</i>" e "<i>A associação deles é voluntária ou involuntária?</i>". O martelo das bruxas é um manual diagnóstico de tipo primitivo, com sugestões para entrevistar e identificar pessoas suspeitas de terem se desviado do caminho justo.</p> <p>1550 – A Renascença testemunha um renascimento na filosofia, e o médico alemão Johann Weyer escreve que aqueles acusados de serem feiticeiros podem ter sido vítimas de transtornos mentais ou físicos. Para os fiéis, Weyer é visto como promotor da causa de Satanás.</p> <p>1600 – O pêndulo começa a se afastar de uma visão de mundo dominada pela religião para uma visão de natureza mais filosófica e científica.</p> <p>1700 – A causa da filosofia e da ciência é promovida com os escritos do filósofo francês René Descartes, do filósofo alemão Gottfried Leibniz e de um grupo de filósofos ingleses (John Locke, George Berkeley, Dave Nume e David Hartley) denominados coletivamente "os empiristas britânicos". Descartes, e.g., levanta questões fascinantes a respeito da relação entre a mente e o corpo. Essas questões seriam exploradas de uma forma menos filosófica e mais física pelo fisiologista Pierre Cabanis. Para fins humanitários, Cabanis realiza observação pessoal do estado de consciência de condenados à guilhotina da Revolução Francesa. Ele concluiu que a ligação entre a mente e o corpo é tão íntima que a guilhotina é provavelmente um modo de execução indolor.</p> <p>1734 – Christian von Wolff é autor de dois livros, <i>Psychologia Empírica</i> (Psicologia empírica) (1732) e <i>Psychologia Rationalis</i> (Psicologia Racional) (1734), os quais preveem a psicologia como uma ciência. Aluno de Gottfried Leibniz, von Wolff também aprofunda a ideia de Leibniz de que existem percepções abaixo do limiar da consciência, desse modo, antecipando a noção de Freud do inconsciente.</p> <p>1780 – Franz Mesmer "<i>mesmeriza</i>" não apenas pacientes parisienses, mas também alguns membros da comunidade médica europeia o uso do que denomina "<i>magnetismo animal</i>" para efetuar curas. O mesmerismo (ou <i>hipnose</i>, como conhecemos hoje) se tornaria um instrumento da avaliação psicológica; a entrevista hipnótica é uma das muitas técnicas alternativas para obtenção de informação.</p>
<p>Idade Contemporânea: de 1789 até aos dias atuais</p>	<p>1823 – O <i>Journal of Phrenology</i> (Jornal da Frenologia) é fundado para avançar o estudo da noção de Franz Joseph Gall de que a capacidade e os talentos especiais estão localizados em concentrações de fibras cerebrais que fazem pressão para fora. Experimentações extensivas eventualmente desacreditam a frenologia, e o jornal é encerrado no início do século XX. Em meados do século XX, a avaliação de "<i>protuberâncias</i>" em perfis de papel é preferível ao exame das protuberâncias na cabeça para se obter informações sobre capacidade e talentos.</p> <p>1829 – Em Análise dos fenômenos da mente humana, o filósofo inglês James Mill afirma que a estrutura da vida mental consiste em sensações e ideias. Mill antecipa uma abordagem à psicologia experimental denominada <i>estruturalismo</i>, cujo objetivo seria explorar os componentes da estrutura da mente.</p> <p>1848 – Em Vermont, uma descarga acidental de explosivos lança uma barra de ferro de 90 cm através do crânio do capataz de construção ferroviária Phineas Gage, destruindo quase toda a parte frontal do lado esquerdo de seu cérebro. Com intervenção médica, Gage sobrevive. Entretanto, antes visto como um trabalhador competente e capaz, após o acidente é descrito como imprevisível, irreverente e "<i>não mais o Gage</i>". Uma vez que seu intelecto parecia não ter sido afetado, o caso foi significativo para chamar a atenção para o papel do cérebro na personalidade e em sua avaliação.</p> <p>1859 – A publicação de <i>Sobre a origem das espécies por meio da seleção natural</i>, de Charles Darwin, aprofunda a noção, então radical, de que os seres humanos descendem dos macacos. A obra levanta questões sobre como os animais e os humanos se comparam em relação a variáveis como o estado de consciência. Darwin também escreve sobre seleção natural e a sobrevivência dos mais aptos da espécie. Essas ideias podem ter influenciado enormemente Freud, cuja teoria psicanalítica da personalidade enfatiza a importância dos impulsos sexuais e agressivos instintivos.</p> <p>1860 – O fisiologista Gustav Fechner publica <i>Elementos da psicofísica</i> ("<i>Elemente der psychophysik</i>"), no qual explora a forma como as pessoas respondem a estímulos como luz e som. O trabalho impulsiona experimentações nas áreas de percepção humana e animal.</p> <p>1869 – Sir Francis Galton, primo de Charles Darwin, publica <i>O gênio hereditário</i>, que foi notável tanto por (a) sua afirmação de que o gênio é hereditário como (b) seu uso pioneiro da técnica estatística que Karl Pearson mais tarde chamaria de <i>correlação</i>. Galton daria inúmeras e variadas contribuições para a mensuração com suas invenções e inovações.</p> <p>1879 – Wilhelm Max Wundt funda a primeira clínica de <i>psicologia experimental</i> em Leipzig, Alemanha; a psicologia é uma ciência autônoma, não apenas um ramo da filosofia. Estruturalista, Wundt revela forte embasamento em um instrumento de avaliação chamado de <i>introspecção</i> fielmente sua (no qual os indivíduos tentam descrever fielmente sua experiência consciente de um estímulo mediante verbalização). Os estruturalistas concentram a atenção nas capacidades sensoriais e no tempo de reação.</p> <p>1885 – Herman Ebbinghaus publica <i>A memória: uma contribuição à psicologia experimental</i>, no qual descreve o uso de sílabas sem sentido para pesquisar e avaliar a memória humana. Seus muitos <i>insights</i> perspicazes sobre as curvas de aprendizagem (e de esquecimento) comprovam que os processos mentais de ordem superior, como a memória, não apenas o tempo de reação ou a reação sensorial a estímulos podem ser avaliados de maneira eficiente.</p>

1890 – O psicólogo norte-americano James McKeen Cattell cria o termo *teste mental* em uma publicação. Ele fundaria diversas publicações, em especial *Science* e *Psychological Review*. Em 1921, formou a Corporação Psicológica com a meta de "*aplicações úteis da psicologia*". Também em 1890, Nova York se torna o primeiro Estado a assumir a responsabilidade por seus cidadãos mentalmente doentes. A legislação relacionada mudou o nome dos chamados "*asilo de lunáticos*" para hospitais estaduais - o lugar onde o doente mental indigente receberia avaliação e tratamento supervisionados por médicos.

1892 – O psiquiatra Emil Kraepelin, que estudou com Wundt, publica pesquisa que empregou um teste de associação de palavras. Também em 1892, a *American Psychological Association* (APA) é fundada com 31 membros, principalmente graças aos esforços de seu primeiro presidente, G. Stanley Hall. Para um relato fascinante, ver o artigo de Samuel Willis Fernberger, "*The American Psychological Association: 1892-1942*", na edição de janeiro de 1943 (volume 50) da *Psychological Review*.

1895 – Alfred Binet e Victor Henri publicam artigos requerendo a mensuração de capacidades cognitivas, como a memória, outras capacidades humanas, como a compreensão social. Curiosamente, Binet também se perguntava em voz alta sobre os possíveis usos dos borrões de tinta para estudar a personalidade.

1896 – Lightner Witmer estabelece a primeira clínica psicológica nos Estados Unidos, na Universidade da Pensilvânia. Subsequentemente, em 1907, funda uma revista chamada *Psychological Clinic*. Witmer escreveu "*Clinical Psychology*", o primeiro artigo dessa revista.

1904 – Charles Spearman, aluno de Wundt em Leipzig, começa a formar a base para o conceito de *confiabilidade* do teste. Ele também inicia a construção da estrutura matemática para a *análise fatorial*.

1905 – Alfred Binet e Theodore Simon publicam uma "*escala de medida da inteligência*" de 30 itens visando ajudar a identificar crianças com retardo mental das escolas de Paris. A noção de medida da inteligência encontraria uma plateia receptiva no mundo todo.

1910 – Como é sua caligrafia? Se você fosse estudante nessa época, teria sua caligrafia analisada por um dos primeiros testes padronizados - de autoria de E. L. Thorndike. Seu artigo intitulado "*Caligrafia*" (*Teachers College Record*, volume 11, item 2) fornece 16 exemplos de caligrafia organizados em ordem de mérito.

1912 – Nesse ano, o termo, agora familiar, "QI" (quociente de inteligência) passa a ser usado. William Stern cria uma fórmula pela qual a "*idade mental*" determinada pelo teste de Binet era o dividendo, o divisor era a *idade cronológica* do testado e o quociente, *multiplicado por 100*, era o QI. Embora o "QI" continue sendo usado no vocabulário mundial, as medidas de inteligência contemporâneas não são mais concebidas por essas relações. Também em 1912, outro marco relacionado ao QI (de várias classes): o livro de Henry Herbert Goddard, *A família Kallikak: estudo da hereditariedade da debilidade mental* é publicado.

1913 – O psiquiatra suíço Hermann Rorschach, filho de um professor de artes, publica ensaios sobre como a análise da produção artística dos pacientes pode fornecer *insights* sobre a personalidade. Em 1921, sua agora famosa monografia, *Psicodiagnóstico*, evoluiu para um teste que se viria a se tornar um ícone para os testes psicológicos aos olhos do público: o teste de borrões de tinta de Rorschach. Também em 1913, o agora famoso artigo de John Watson na *Psychological Review*, "*Psychology as the Behaviorist Views It*" (A psicologia vista por um *behaviorista*), se torna conhecido como o "manifesto behaviorista". Naturalmente, do ponto de vista do behaviorista, a observação comportamental é um instrumento fundamental da avaliação.

1914 – A Primeira Guerra Mundial constitui uma vantagem para o movimento dos testes, visto que o funcionamento intelectual e a aptidão emocional de milhares de recrutas devem ser rapidamente avaliados.

1916 – Após anos de pesquisa, Lewis M. Terman, trabalhando na Universidade de Stanford, publica a *Revisão Stanford da Escala de Inteligência de Binet-Simon*. Essa adaptação e revisão norte-americana do teste desenvolvido na França iria se tornar amplamente conhecida como o Stanford-Binet.

1920 – O *Army Mental Tests* (Testes Mentais do Exército), editado pelos majores Clarence S. Yoakum e Robert M. Yerkes (ambos psicólogos com carreiras destacadas) é publicado por Holt. Este volume editado fornece informações detalhadas sobre os testes *Alfa e Beta do Exército* desenvolvidos durante a "grande guerra" em um momento "nesta luta suprema [quando] se tornou claro [...] que a utilização adequada da força do homem, e mais particularmente da força da mente ou do cérebro, garantiria a vitória final" (p. vii).

1926 – O Conselho Educacional (College Board) patrocina o desenvolvimento do *Teste de Aptidão Acadêmica* (SAT) e o administra pela primeira vez.

Fonte: Cohen, Swerdlik, & Sturman (2014).

Historicamente, dentro do progresso da testagem e avaliação psicológica, se destacam (Pasquali, 1999, 2011):

- 1880: *A década de Sir Francis Galton* (1822 - 1911) – o primeiro a falar sobre traços mentais, após ter criado a correlação e regressão; precursor de estudos da psicologia associados à educação; pioneiro no uso de métodos estatísticos para o estudo das diferenças e heranças humanas da inteligência; referências para os princípios que

embasam os primeiros testes construídos; a grande contribuição para a Psicometria foi a elaboração de procedimentos estatísticos para análise de dados, obtidos por meio de instrumentos (questionários) que viabilizavam quantificar as avaliações psicológicas (Quaresma, 2014).

Seus trabalhos visavam a avaliação das aptidões humanas através da medida sensorial, salientando-se sua obra *Inquiries into Human Faculty*, de 1883. O trabalho de Galton terá enorme impacto tanto na orientação mais prática da Psicometria (Cattell e outros psicometristas americanos), quanto na teórica (Pearson e Spearman). (Pasquali, 2011, p. 15)

- 1890: *A década de James McKeen Cattell* (1860 - 1944) - o primeiro a utilizar o termo "testes mentais"; mediu "processos mentais simples" de forma independente; a pesquisa sobre as diferenças individuais desempenhou um papel importante ao introduzir e enfatizar a técnica experimental e a importância da metodologia na experimentação na América (Ibid, 2014);

Sob a influência de Galton, Cattell desenvolveu suas medidas das diferenças individuais e recolheu sua experiência no *Mental Tests and Measurements*, de 1890, inaugurando, inclusive, a terminologia de *mental test* (teste mental). (Pasquali, 2011, p. 15)

- 1900: *A década de Alfred Binet* (1857 - 1911) - conhecido na Europa como o principal representante da Psicometria; criou a escala métrica de inteligência desenvolvendo estudos que enfatizavam os processos mentais superiores, como o pensamento e a inteligência, diferentemente do que era abordado por Galton e Cattell (Ibid, 2014).

"Baseando-se nos trabalhos de Binet, na França, e no desenvolvimento de numerosos trabalhos americanos no domínio da medição, a Psicometria veio auxiliar a avaliação, pedagógica. Esse modelo preocupa-se sobretudo com duas questões: melhorar a fidelidade das medições dos desempenhos, decompondo a variância total em seus diversos componentes (entre elas, a variância erro), e tentar construir testes válidos com a ajuda de técnicas como a análise fatorial, que permite verificar o que um teste realmente mede". (De Ketele, 1986, citado por Ibid, 2014)

Nessa década predominaram os interesses da avaliação das aptidões humanas visando à predição na área acadêmica e na área da saúde. Embora Binet se destaque, outros expoentes aparecem neste período, salientando-se sobretudo Spearman na Inglaterra. Na verdade, no que se refere propriamente à teoria psicométrica, a década de 1900 deve ser considerada a *era de Spearman*, o qual lançou os fundamentos da teoria da Psicometria clássica com suas obras *The proof and measurement of association between two things* (1904a), *'General intelligence' objectively*

determined and measured (1904b), *Demonstration of formulae for true measurement of correlations* (1907) e *Correlations of sums and differences* (1913). (Pasquali, 2011, pp. 15-16)

- 1910 a 1930: *A era dos testes de inteligência;*

Vários foram os fatores que concorreram para o desenvolvimento dessa era, a saber, o teste de inteligência de Binet-Simon (1905), o artigo de Spearman sobre o fator G (1904b), a revisão do teste de Binet para os Estados Unidos (Terman, 1916) e o impacto da Primeira Guerra Mundial com a imposição da necessidade de seleção rápida, eficiente e universal de recrutas para o exército (os testes *Army Alpha* e *Beta*). (Pasquali, 2011, p. 16)

- 1930: *A década da análise fatorial;*

Já por volta de 1920, o entusiasmo com os testes de inteligência vinha caindo muito, sobretudo quando se mostrou que eram demasiadamente dependentes da cultura onde eram criados, não apoiando a ideia de um fator geral universal, como proposto por Spearman. Tais eventos fizeram com que os psicólogos estatísticos começassem a repensar as ideias de Spearman. De fato, Kelley quebrou com a tradição de Spearman em 1928. Esta tendência foi seguida, na Inglaterra, por Thomson (1939) e Burt (1941) e nos Estados Unidos da América, por Thurstone (1935, 1947). Este último autor é especialmente relevante nesta época, pois além de desenvolver a análise fatorial múltipla, atuou no desenvolvimento da escalonagem psicológica (1927, 1928, Thurstone & Chave, 1929), bem como por ter fundado, em 1936, a Sociedade Psicométrica Americana, juntamente com a revista *Psychometrika*, ambas dedicadas ao estudo e avanço da Psicometria. (Pasquali, 2011, p. 16)

- 1940 a 1980: *A era da sistematização.* Duas tendências opostas marcam esta época: os trabalhos de síntese e os de crítica.

Nas obras de síntese, temos Guilford (1936, *Psychometric Methods*, reeditada em 1954), tentando sistematizar os avanços em Psicometria até então conseguidos; Gulliksen (1950, *Theory of Mental Tests*), sistematizando a teoria clássica dos testes psicológicos e Torgerson (1958, *Theory and Methods of Scaling*), sistematizando a teoria sobre a medida escalar. Além disso, Thurstone (1947) e Hannan (1967) recolheram os avanços na área da análise fatorial; Cartell (1965; Cattell & Warburton, 1967) procurou sintetizar os dados da medida em personalidade e Guilford (1967) procurou sistematizar uma teoria sobre a inteligência. Por outro lado, Buros (1938) iniciou uma coletânea de todos os testes existentes no mercado, a qual vem sendo refeita periodicamente (mais ou menos a cada cinco anos), publicada no *Mental Measurement Yearbook*. Na mesma época, *A American*

Psychological Association - APA (1954, 1974, 1985) introduziu as normas de elaboração e uso dos testes.

No lado da crítica, temos Stevens (1946) questionando o uso das escalas de medida que deu/dá muita polêmica na área (Lord, 1953; Gaito, 1980; Miehell, 1986; Townsend & Ashby, 1984) e, sobretudo, surge a primeira grande crítica à teoria clássica dos testes na obra de Lord e Novick (1968 - *Statistical Theory of Mental Tests Scores*), que iniciou o desenvolvimento de uma teoria alternativa, a teoria do traço latente, que vai desembocar na teoria moderna da Psicometria, a Teoria de Resposta ao Item (TRI), mais tarde sintetizada por Lord (1980). Outra tendência de crítica para superar as dificuldades da Psicometria clássica foi iniciada pela Psicologia Cognitiva de Sternberg (1977, 1982, 1985; Sternberg & Detterman, 1979; Sternberg & Weil, 1980) com seu modelo, procedimentos e pesquisas sobre os componentes cognitivos, na área da inteligência. (Pasquali, 2011, pp. 16-17)

- 1980 até a atualidade: *A era da Psicometria Moderna* (TRI).

Chamar a era atual de era da TRI talvez seja inadequado, porque (1) esta teoria, embora esteja sendo o modelo no dito primeiro mundo, ainda não resolveu todos seus problemas fundamentais para se tomar o modelo moderno definitivo de Psicometria e (2) ela não veio para substituir toda a Psicometria clássica, mas apenas partes dela. De qualquer forma é o que há de mais novo no campo. (Pasquali, 2011, p. 17)

Pasquali destaca ainda que os campo promissores na atualidade na qual a Psicometria avança são: *Sistematização da Psicometria Clássica; Pesquisa na TRI; Pesquisa em uma série de áreas paralelas da Psicometria (testes com referência a critério; testes sob medida "Computer adaptive testing"; banco de itens; equiparação dos escores; validade dos testes; vieses dos testes; construção de itens; impacto dos trabalhos da Psicologia Cognitiva)*.

2.3.2 Teoria Clássica dos Testes

"Tornar-se-á possível associar às análises profundas de Edgar Morin (1977-1991) e de Jean-Louis Le Moigne (1990) a instrumentos conceptuais quantitativos de precisão, que não lhes amputarão o sentido." (Mugur-Schächter, 2007, p. 150, *In: Edgar Morin; Jean-Louis Le Moigne, 2007*)

Dentro da *Psicometria Moderna*, podem-se destacar a *Teoria da Resposta ao Item* (TRI) e a *Teoria Clássica dos Testes* (TCT). Apesar de toda uma discussão sobre qual deve ser utilizada neste processo, onde todos os elementos apontam para a TRI, existem técnicas da TCT que não devem ser descartadas como, por exemplo, o teste de confiabilidade do item. Dentro do ensino da Física existem diversos trabalhos desenvolvidos utilizando a TCT (Engelhardt, 1997; Tarekegn, 2009), apesar de ser um

campo de interesse durante o MCA desde a década de 70 e 80 (Treagust, 1988). Pesquisadores na educação científica e tecnológica interessados em explorar os testes conceituais vem avançando ao explorar as contribuições os modelos psicométricos da TRI, apesar de ser um processo em expansão e de familiaridade com esses processos de análise, sendo bem mais difundido e desenvolvidos dentro da psicologia experimental.

Como seria de esperar, os pesquisadores em educação científica e tecnológica tornaram-se mais sofisticados no projeto de testes conceituais de múltipla escolha desde a FCI. Mais atenção é agora dada às questões de confiabilidade e validade. Assim como os pesquisadores da educação científica se tornaram mais sofisticados, o mesmo ocorre dentro do campo da área da Psicometria. *A Teoria de Resposta ao Item (TRI) ainda não fez avanços significativos no léxico ou na “caixa de ferramentas” da maioria dos pesquisadores em educação científica e tecnológica.* (Engelhardt, 2009, p. 35, grifo nosso, tradução livre)

É comum observar trabalhos que reservam um espaço para destacar uma comparação entre a utilização da TRI a TCT como procedimentos de análise. Pode-se dizer que a principal diferença é que a TRI apresenta uma análise de acordo com *cada item*, lhe atribuindo diversas características (como, por exemplo, a discriminação, dificuldade, acerto aleatório - “*guessing*”, erro aleatório - “*upper*”, e parâmetros de assimetria na curva), enquanto a TCT apresenta uma análise do *teste como um todo*, em que todas as análises dos itens contemplam uma única perspectiva mais geral.

O fato de diversas avaliações educacionais de desempenho em larga escala, tanto nacionais como internacionais, utilizarem os modelos e princípios da TRI, o torna atualmente um método consagrado e com credibilidade. Dentre as avaliações nacionais no contexto brasileiro podem-se destacar o *Exame Nacional do Ensino Médio (ENEM)*, *Sistema Nacional de Avaliação da Educação Básica (SAEB)*, *Sistema Mineiro de Avaliação da Educação Pública (SIMAVE)*, *Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP)*, além de aproximadamente 30 sistemas de avaliação em nível estadual. Em outros países também se aplicam a TRI: na Holanda tem o *Institute for Educational Measurement (CITO)*; nos Estados Unidos da América, onde tem o *National Assessment of Educational Progress (NAEP)*, *Educational Testing Service (ETS)*, *Graduate Management Admission Test (GMAT)*; além de ser utilizado no *Programme for International Student Assessment (PISA)*, entre outros.

2.3.2.1 Fundamentos da análise clássica

A Teoria Clássica dos Testes (TCT) baseia-se em parâmetros descritivos, e segue contemplando a proposta metodológica de item comentado da *plataforma Devolutivas Pedagógicas*⁴⁷, que são os seguintes parâmetros:

- O «índice de dificuldade» do item (ID) corresponde a porcentagem de acertos;
- O percentual de acerto superior (de maior habilidade) e inferior (de menor habilidade) a 27% (Borgatto & Andrade, 2012), conforme as conclusões e recomendações da obra clássica de Kelley (1939, p. 24), na qual considera-se todos os respondentes com o mesmo escore na linha corte dos 27% superior e inferior;
- O «índice de discriminação» (DISCR) que corresponde à diferença entre o percentual de acerto superior e inferior, de acordo com o trabalho clássico de Findley (1956, p. 177);
- O «coeficiente bisserial» (r_{biss}) entre a resposta numa dada categoria do item e a pontuação total no teste.

Também foram acrescentadas outras informações, como: *índice de fiabilidade (IF)*; *fator de concentração de análise*; *médias do grupo de acerto*; incluindo os respectivos índices clássicos corrigidos. De modo geral, a análise dos itens depende do modelo teórico a partir do qual foi construído o teste (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, p. 54), isso justifica a inclusão de alguns procedimentos de análise que são tradicionais em testes conceituais na categoria de inventário de conceitos, como o ganho percentuais normatizados e fator de concentração de análise. Contudo, a avaliação da qualidade de itens dentro da teoria clássica, deve conter basicamente a (Ibid, 2005, p. 54): *dificuldade*; *discriminação*; *análise dos distratores*; *fiabilidade*; *validade e dimensionalidade*. Diante disso, serão comentados alguns desses procedimentos de análise, e outros de forma mais detalhada na etapa metodológica e dos resultados, devido à necessidade na tomada de decisão específicas.

⁴⁷ Desenvolvida pelo INEP no Brasil, e disponibilizada em 2015. Esta plataforma possui a intenção de *feedback* ao apresentar os resultados do SAEB de todas as escolas dos Estados e municípios nas disciplinas de matemática e português, comentando os itens de forma pedagógica e possibilitando comparações, para que essas informações possam ser utilizadas na escola através da troca de experiências entre gestores e na tomada de decisão do professor em suas aulas.

2.3.2.1.1 Índice de dificuldade

O «índice de dificuldade» (ID_i) de um item é um parâmetro correspondente ao parâmetro “b” da TRI, sendo um conceito por vezes controverso pelo fato de que o índice de dificuldade na realidade mede o percentual de acerto em um determinado item. Dessa forma, considera-se que quanto maior a porcentagem de acerto maior será o “grau de dificuldade do item”, e obviamente mais fácil é o item, dessa forma seria conveniente ser chamado de *índice de facilidade* (Rabelo, 2013, p. 133), termo esse adotado por alguns psicometristas (Erthal, 1987), e recomendado por outros por ser do ponto de vista puramente semântico mais adequado (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, p. 61). O ID do item é expressa como:

$$ID_i = \frac{\text{total de acertos}}{\text{total de respondentes}} = \frac{A}{n}$$

Sendo:

- A o total de indivíduos que acertaram ou total de respostas certas de um determinado item;
- n o tamanho da amostra ou total de respostas válidas (todos os que erraram e acertaram).

Existe também a *média das pontuações do teste*, que corresponde a soma dos índices de dificuldade dos itens ((Ibid., 2005, p. 61):

$$\bar{X} = \sum_{i=1}^n ID_i$$

Para uma melhor discriminação quanto ao ID , Cerdá (1984) propõe uma tabela como “modo de apresentação dos itens em um teste heterogêneo e sua interpretação” (Erthal, 1987), de forma que pode ser útil na interpretação das características clássicas dos itens.

Quadro 24: Classificação para o Índice de Dificuldade.

Tipologia do item	Número aproximado	Limite aproximado dos itens de dificuldade
Muito fáceis	10%	de 0,75 a 0,95
Fáceis	20%	de 0,55 a 0,74
Normais	40%	de 0,45 a 0,54
Difíceis	20%	de 0,25 a 0,44
Muito difíceis	10%	de 0,05 a 0,24

Fonte: Cerdá (1984) e Erthal (1987).

Existem outras sugestões a classificar as dificuldades dos itens, levando em consideração as assimetrias. Por exemplo, quando o item é mais fácil apresenta uma

forte assimetria negativa (80% de acerto), quando é difícil a assimetria é positiva (80% de erro), e quando é totalmente simétrico possui um índice de dificuldade de 0,5 (50%).

Para verificar a relação entre a dificuldade dos itens e a variância do teste, quando o item é dicotômico (ou se acerta ou erra), tem-se a variância dos escores X dada por:

$S_x^2 = p \cdot q$, onde p é a proporção de acerto e q a proporção de erros. Assim, naturalmente pode-se dizer que: $q = 1 - p$.

$$\text{Dessa forma: } S_x^2 = p \cdot (1 - p) = p - p^2.$$

Para calcular o máximo da função é necessário calcular a primeira e a segunda derivada da função, com isso, tem-se: $\delta'_p = 1 - 2p$ e $\delta''_p = -2$. Igualando a primeira derivada igual a zero (0):

$$0 = 1 - 2p \rightarrow p = \frac{1}{2} = 0,5$$

Ignorando a segunda derivada, já que deu negativa, pode-se dizer então que a variância de um item dicotômico é máxima quando seu índice de dificuldade é 0,5 (50%), ou seja, a variância de um teste formado por itens dicotômicos estará mais eficiente quando seus itens possuírem um índice de dificuldade igual a 0,5 (50%). Assim, se favorece estabelecer máximas discriminações entre os respondentes de um teste no que diz respeito a diretamente a dificuldade de seus itens (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, pp. 57-58).

Isso significa dizer que **itens mais fáceis** permitem discriminar bem entre os **sujeitos com posição mais desfavorecida** no traço latente que mede o teste. **Itens mais difíceis** vão discriminar melhor entre **aquelas mais capacitadas** na variável apresentada para a medida. Ou seja, itens difíceis e fáceis ajudam a discriminar os extremos da distribuição. Apesar disso, considera-se que **se obtêm os melhores resultados nas avaliações de desempenho quando a maioria dos itens possuem uma dificuldade média**, para isso torna-se imprescindível incluir itens fáceis e difíceis para melhor discriminar os extremos da distribuição, porém, convêm que no teste **não se**

comece com itens demasiadamente difíceis, pois itens mais fáceis no início das provas ajudam a manter um bom nível de motivação entre aqueles que devem responde-las, evitando assim um desânimo pessoal com a ideia de que seja incapaz de responder o restante (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, pp. 60-61).

No estudo feito por Lord (1952, p. 188), esse é considerado um critério comum na qual se afirma que **a dificuldade do item ideal está a meio caminho entre o nível de chance de acerto ao acaso** ($\bar{p}_i = \frac{1}{k}$, onde k é o número de alternativas) e **1**, ou seja:

$$ID_{ideal} = \frac{\frac{1}{k} + 1}{2}$$

Apesar disso, Lord buscou discutir a dificuldade dos itens em termos de:

- A proporção real das respostas corretas ao item (\bar{p}_i);
- uma estimativa da proporção de examinados que realmente conhecem a resposta ao item: $p_i = \frac{k \cdot \bar{p}_i - 1}{k - 1}$
- h_i , o desvio relativo da curva normal padronizada acima da qual p_i .

Ao analisar o quadro elaborado por Lord (1952, p. 189) quanto ao parâmetro de dificuldade dos itens para 5, 4, 3 e 2 alternativas, é possível identificar para coeficiente de correlação bisserial os percentuais que são mais fiáveis (mais elevados), que correspondem aos percentuais ideais de acerto. Diante de um teste difícil ou mal-elaborado, esses são percentuais elevados a serem alcançados.

Tabela 1: Percentuais ideais para o índice de dificuldade, segundo Frederic Lord, 1952.

Formato	Dificuldade Ideal (%)
Cinco respostas de múltipla escolha	70
Quatro respostas de múltipla escolha	74
Três respostas de múltipla escolha	77
Verdadeiro-Falso (duas respostas de múltipla escolha)	85

Erthal (1987) sugere ainda duas formas mais sofisticadas para calcular o ID , em que a técnica de Davis *corrige o acerto ao acaso* ($k - 1$), e leva em conta o tempo de

realização do teste (*NNR*). Evidentemente pode acontecer de o respondente parar o teste antes do tempo limite de entrega, supostamente por outros motivos.

Quadro 25: Técnicas do Índice de Dificuldade.

Técnicas do <i>ID</i>	Relação do <i>ID</i>	Descrição dos termos
Técnica de Davis	$ID = \frac{A - \frac{E - o}{k - 1}}{n - NNR}$	Sendo: <i>A</i> = acertos; <i>o</i> = omissão; <i>E</i> = erros; <i>k</i> = número de alternativas; <i>n</i> = número de indivíduos; <i>NNR</i> = número de indivíduos que não responderam ao item e que pararam o teste.
Técnica Abreviada dos 27%	$ID = \frac{A_{27\%}^- + A_{27\%}^+}{n^- + n^+}$	Sendo: $A_{27\%}^-$ a composição do grupo correspondente aos 27% dos desempenhos inferiores; $A_{27\%}^+$ grupo dos 27% com desempenho superior; e o <i>n</i> sendo o número de respondente de acordo com cada grupo, porém deve se considerar todos os indivíduos (independentemente de ter assinalado ou omitido), entretanto é preciso subtrair dos indivíduos que não atingiram o item, ou seja, que desistiram antes de chegar ao item.

Fonte: Adaptado de Erthal (1987).

No teste TCE aplicado em 2001 (Yeo & Zadnik, 2001), foi utilizada a técnica abreviada. O índice de dificuldade corrigido segundo a técnica de Davis exclui os casos omissos, entretanto, é possível também obter um índice semelhante em termos de percentual de acerto e erro (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, p. 59):

$$ID_{corr} = p - \frac{q}{k - 1}$$

Sendo:

- ID_{corr} : índice de dificuldade corrigido;
- *p*: proporção de acertos (se refere ao índice de dificuldade sem ter sido corrigido);
- *q*: proporção de falhas ou erros;
- *k*: Número de alternativas que tem um item.

2.3.2.1.2 Índices percentuais de discriminação

Quanto aos *Índices Percentuais de Discriminação (IPD)* representam indicadores de **consistência interna dos itens com relação ao teste** como um todo, podendo ser a diferença entre o grupo de desempenho superior e inferior dentro dos 27% (*DISCR*), o «coeficiente de correlação biserial» (r_{biss}), e o «coeficiente de correlação biserial pontual», por pontos ou ponto-biserial (r_{pb}), sendo esses dois últimos variando entre -1 a +1. O caso do valor obtido ser +1 corresponde a uma perfeita correspondência entre os que obtiveram escores totais ou baixos em um teste e o número de acertos ou

erros em determinado item. Considera-se ainda a possibilidade de cálculo através do índice de discriminação do «método abreviado dos 27%» (Erthal, 1987, p. 86):

$$IPD = \frac{A_{27\%}^- + A_{27\%}^+}{\frac{n^- + n^+}{2}}$$

Ou seja, há uma relação entre os índices de dificuldades e de discriminação pelo método abreviado:

$$IPD = 2 \cdot ID$$

Para todos os casos, se admite aceitável como linha de corte valores de discriminação maiores que 0,25.

Considera-se que o grupo intermediário corresponda a estimados 46% ($A_{46\%}^\pm$) e sugere-se que (Rabelo, 2013, p. 135) os percentuais de acerto dentro de cada grupo siga a seguinte ordem:

$$A_{27\%}^- < A_{46\%}^\pm < A_{27\%}^+$$

Para determinar os 27% de maior (superiores) e menor (inferiores) habilidades ou desempenho no teste é importante considerar todos os indivíduos com a mesma nota, portanto não corresponde precisamente em 27% da amostra para cima ou para baixo dentro do rol. Após determinado esses parâmetros, a diferença entre eles se configura como o índice de discriminação (*DISCR*) na qual mede a capacidade do item de diferenciar os participantes com maior e menor desempenho, para isso leva em conta o percentual dos grupos de maior e pior desempenho dentro dos 27%. O índice de discriminação (*DISCR*) da TCT também possui correspondência a um parâmetro da TRI, no caso o parâmetro “a”. Considera-se que um item de boa qualidade, a porcentagem de acerto seja maior para o grupo de melhor desempenho, que de “quanto maior for a diferença entre as porcentagens de acertos dos dois grupos (com melhor e com pior desempenho), maior será a discriminação do item” (Borgatto & Andrade, 2012; Rabelo, 2013).

$$DISCR = ACIM - ABAIX = A_{27\%}^+ - A_{27\%}^-, \text{ onde } -1 \leq DISCR \leq 1.$$

Admitindo a magnitude do índice de discriminação, tem-se a diferença entre o grupo superior (ACIM) e o inferior (ABAIX), em outras palavras, do percentual de acertos entre os indivíduos nos 27% com maior desempenho ($A_{27\%}^+$), menos os o percentual de acertos entre os indivíduos nos 27% com menor desempenho $A_{27\%}^-$. Considera-se uma boa discriminação quando a diferença fica maior ou igual a 30%.

Quadro 26: Classificação para o Índice de Discriminação.

Valores	Tipologia do item	Classificação de Ebel (1965)
$DISCR \geq 0,4$	Adequado (<i>favorável a reaplicação</i>)	Discrimina muito bem
$0,3 \leq DISCR < 0,4$	Bom (<i>sujeito a aprimoramento</i>)	Discrimina bem
$0,2 \leq DISCR < 0,3$	Marginal (<i>sujeito a reelaboração</i>)	Discrimina pouco
$DISCR < 0,2$	Inadequado (<i>sujeito a rejeição</i>)	$0,1 \leq DISCR < 0,2$: Item limite. Se deve melhorar $DISCR < 0,1$: Item sem utilidade para discriminar

Fonte: Adaptado de Rabelo (2013, p. 136) e Ebel (1965, citado por Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, p. 62).

Considerando os IPD **baseados na correlação**, também são chamados de *índice de homogeneidade*, tem-se em destaque o: *coeficiente bisserial* e *ponto-bisserial* (ou item-teste). Nessa perspectiva, o «coeficiente bisserial» (r_{biss}) é uma estimativa do coeficiente de Pearson, estando sujeito às mesmas limitações. Em outras palavras, pode-se dizer que corresponde a uma «medida de associação» entre o «desempenho no item» e o «desempenho na prova» (score total), dessa forma, estima a correlação entre a variável de desempenho no teste e uma variável latente (não observável) com distribuição normal que, por hipótese, representa a habilidade que determina o acerto ou erro do item (Borgatto & Andrade, 2012). Sua expressão é dada por:

$$r_{biss} = \frac{M^+ - M^-}{\sigma} \cdot \frac{p \cdot (1 - p)}{h(p)}$$

Sendo o M^+ a média da medida de desempenho para os alunos que acertaram o item, M^- a média da medida de desempenho no teste para os alunos que erraram o item, σ o desvio-padrão da medida de desempenho no teste para todos os alunos, p o percentual de respostas e $h(p)$ o valor da densidade da distribuição normal com média 0 e variância 1 no ponto em que a área da curva à esquerda deste ponto é igual a p . (Borgatto & Andrade, 2012, p. 149)

A média da medida dos respondentes M^+ é a média total dos escores relativa ao percentual de respostas de uma dada alternativa assinalada, e pode surgir na literatura como \bar{X}_p , e $q = 1 - p$, sendo q o percentual de respostas às outras

alternativas, quando o p se refere ao garabarito o q se refere aos distratores da questão. O coeficiente bisserial também pode se relacionar com o **coeficiente de correlação ponto-bisserial** (r_{pb}), que se refere à correlação linear de Pearson entre as respostas dos itens e seus respectivos escores no teste. As relações são dadas por (Garcia, Abad, & Tello, 2000, p. 44):

$$r_{pb} = \frac{\bar{X}_p - \bar{X}_q}{\sigma_x} \cdot \sqrt{p \cdot q} = \frac{\bar{X}_p - \bar{X}}{\sigma_x} \cdot \sqrt{\frac{p}{q}}$$

A combinação dos coeficientes pode ser expressa como:





$$r_{biss} = \frac{r_{pb} \cdot \sqrt{p \cdot q}}{\sigma_x}$$

Borgatto & Andrade (2012) advertem que o uso do escore (número total e acertos de cada indivíduo) somente é válido quando os indivíduos são submetidos à mesma prova, ou provas paralelas equivalentes. No caso desse estudo, o teste é o mesmo para todos indivíduos. Convém lembrar, entretanto, que o teste foi traduzido para o português do Brasil e Portugal.

A correlação ponto-bisserial é considerado um *índice de homogeneidade do item* e informa o quanto determinado item mede diante da prova como um todo, e em que grau contribui para a homogeneidade ou consistência interna do teste. Caso a finalidade do teste seja medir um construto unitário se deve eliminar itens com valores próximo a zero, pois baixos valores nos índices indicam que o item esteja medindo algo diferente da prova em seu conjunto (Abad, Garrido, Olea, & Ponsoda, 2006). Em casos de teste com poucos itens, torna-se apropriado sob critério de avaliar a sua consistência realizar uma correção no produto item-teste subtraindo uma unidade do escore bruto antes de correlacionar com o item (Ibid., 2006), ou seja, considerando $r_{pb} = r_{i,x}$, onde i indica os valores para o item e x seus respectivos escores, sendo então a correlação item-teste entre eles, com a correção passaria a ser $r_{i,x-i}$, este é um procedimento padrão realizado e visualizado nos *outputs* do programa BILOG-MG, na qual foram realizadas algumas análises neste trabalho.

Para a interpretação pedagógica dos indicadores r_{pb} e r_{biss} , tem-se:

Quadro 27: Interpretação a ser feita na discriminação dos sujeitos.

Valores	Interpretação		
$-1 < r_{pb} < 1$	Se o valor estiver negativo ou próximo de zero, os indivíduos de bom desempenho de <i>modo geral</i> no teste estão errando o item, contrariando aquilo que se espera. Recomenda-se nesse caso rever o item ou eliminá-lo.		
$-1 < r_{biss} < 1$	É uma transformação do r_{pb} . O seu valor revela como determinado grupo de indivíduos que assinalaram uma alternativa se distribui entre aqueles de baixo e alto desempenho, de modo geral, ou seja, possibilita inferir sobre um perfil de determinado grupo distinguindo proporções de escolhas sobre de indivíduos sobre a alternativa.		
Perfis de composição de grupos (de baixo e alto desempenho) para determinada alternativa do item, de acordo com os valores do r_{biss} .			
Com um valor <i>negativo</i> e próximo de -1 : perfil com predomínio de sujeitos de baixo desempenho [↑].	Com valor <i>igual ou próximo a zero</i> : perfil de composição misto, mais ou menos equilibrado.	Um valor <i>positivo</i> e próximo de 1 : perfil com predomínio de sujeitos de alto desempenho [↓].	
			
 Determinada Alternativa			

⇒ A identificação de **perfis de composição de grupos** pode auxiliar o professor na tomada decisão sobre as atividades a serem desenvolvidas pelos alunos diferentemente, buscando identificar e valorizar as potencialidades de cada sujeito, além de propor diferentes níveis de dificuldades, rompendo com um ensino linear, considerando que todos devem aprender e progredir da mesma forma.

Em um teste dicotômico o coeficiente de correlação ponto-bisserial (ou item-teste) corrigido pode ser obtido **eliminando o efeito do item** dentro do escore, na qual seria a correlação entre a *resposta do item i* (1 ou 0) com o *escore menos a resposta do item* ($X - i$):

$$DISCR_{\text{corrigido}} = r_{i,x-i}$$

Também pode ser considerado a correção nos índices de discriminação corrigindo a correlação entre o item i e o escore total X , quando **não se elimina o item** para o cálculo da correlação da pontuação do teste

$$DISCR_{corrigido} = \frac{r_{ix} \cdot \sigma_x - \sigma_i}{\sqrt{\sigma_x^2 + \sigma_i^2 - 2 \cdot r_{ix} \cdot \sigma_x \cdot \sigma_i}}$$

O *Índice de Fiabilidade* (IF) de um item pode ser utilizado para medir a característica da variável que se mede no teste total. Para ser medido, tem-se como:

$$IF = \sigma_i \cdot DISCR_i$$

Sendo:

- $\sigma_i = \sqrt{p_i \cdot q_i}$: desvio-padrão típico das pontuações do item i ;
- $DISCR_i$: índice de discriminação do item i . Recomenda-se utilizar a **correlação item-teste corrigida** ($DISCR_{corrigido}$), eliminando a influencia do item com a pontuação total do teste.

O *Índice de Validez do Item* (IV) também pode ser obtido considerando a correlação das pontuações de um item com um critério externo.

$$IV = r_{i,y}$$

Sendo:

$r_{i,y}$: correlação entre as pontuações de um item e as pontuações em um critério externo.

De acordo com alguns autores, entre eles Laveault e Grégoire (1997), propõe-se como índice de validade do item a **correlação item-teste**, ponderada pelo **desvio-padrão do item** (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, p. 73), semelhante ao *índice de fiabilidade*.

O índice de consistência mais conhecido é o coeficiente de alfa de Cronbach, que é igual a:

$$\alpha = \left(\frac{n}{n-1} \right) \cdot \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_t^2} \right)$$

Sendo:

- n o número de itens;
- σ_i a variância do item;
- σ_t a variância total do teste.

Kuder-Richardson 20 (KR20) desenvolveram uma adaptação para os testes dicotômicos, sendo similar ao coeficiente alfa, sendo p_i a porcentagem de alunos que responderam corretamente o item, e r_{tt} o coeficiente do teste-total:

$$KR20 = r_{tt} = \left(\frac{n}{n-1}\right) \cdot \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_t^2}\right)$$

O método das duas metades se refere ao coeficiente de Spearman-Brown, na qual r_{hh} é o coeficiente de correlação entre as duas metades das respostas, ou seja, se tem como procedimento dividir os escores em duas colunas e realizar a correlação entre elas:

$$r_{tt} = \frac{2 \cdot r_{hh}}{r_{hh}}$$

Apesar disso, se recomenda o uso do KR20 para os testes conceituais dicotômicos.

A maioria dos testes de escolha múltipla são marcados dicotomicamente, e é por isso que o KR-20 tende a ser o mais utilizado na literatura do ensino de física. Tanto o coeficiente alfa quanto o KR-20 produzem resultados semelhantes, mas ambos os métodos produzem coeficientes de confiabilidade inferiores aos produzidos pelo método das metades divididas ou de Spearman-Brown. (Engelhardt, 2009, p. 23, *tradução livre*)

Dependendo das finalidades do teste, torna-se importante identificar os valores mais apropriados para os coeficiente de correlação. Apesar disso, sabe-se que o objetivo da maioria dos testes conceituais de escolha múltipla é identificar áreas de dificuldade e avaliar o ensino, portanto, os valores de 0,70 ou acima são aceitáveis (Engelhardt, 2009, p. 24).

Quadro 28: Classificação para os coeficientes de correlação.

Valores	Classificação
De 0,95 a 0,99	Muito alto , raramente encontrado
De 0,90 a 0,95	Alto , suficiente para medir indivíduos
De 0,80 a 0,90	Relativamente alto , possível para a medição de indivíduos
De 0,70 a 0,80	Suficiente para medidas de grupo, não indivíduos
Abaixo de 0,70	Baixo , útil apenas para médias e pesquisas de grupo

Além disso, também se torna interessante determinar não somente a consistência dos itens em relação ao teste, mas também se os escores são próximos da pontuação ou escore verdadeiro. Para isso se tem o erro padrão de medida (*“Standard*

Error of Measurement” ou “SEM”) na qual depende do coeficiente de correlação total do teste r_{tt} e do desvio-padrão dos escores do teste σ_t :

$$SEM = \sigma_t \sqrt{1 - r_{tt}}$$

Deve-se interpretar que quando o SEM é grande, a incerteza no escore verdadeiro do indivíduo também é grande, e quando o SEM é pequeno, a incerteza é pequena, isso significa dizer que seja mais certo se esta da pontuação. A classificação desses valores se dá pelo intervalo de confiança estabelecido, e diante disso considera-se que a distribuição de pontuação para um único indivíduo seja normal.

Quadro 29: Níveis de confiança para o escore verdadeiro.

Níveis de confiança	$\pm SEM$
68%	1,0
90%	1,64
95%	1,96

2.3.2.1.3 Zonas de “guessing” do teste e dos itens

O TCE possui 26 itens, sendo cinco itens (y9, y11, y16, y18 e y26) com cinco alternativas e o restante com quatro alternativas, com isso a probabilidade de *acerto ao acaso* (“chute no teste”, que seria o “*guessing rate*” ou taxa de adivinhação do teste) são de 7 itens ($21 \cdot 0,25 + 5 \cdot 0,20 = 6,25 \approx 7$ **itens**) de modo geral, isso corresponde no caso a responder cerca de 26,92% ($\approx 27\%$) do teste, ou seja, todos alunos que adquirirem esse percentual de escore bruto podem ser considerados dentro de um grupo com grande probabilidade de não dominarem o conhecimento que o teste exige, e admite-se que seja um percentual de acerto válido para a média das amostras de qualquer grupo normativo analisado separadamente, bem como para as análises da amostra de modo geral. Para diferenciar, mantêm-se a denominação de “*guessing rate*” quando se referir a probabilidade de “chute” (acerto ao acaso do teste) diante do **escore bruto**, ou seja, considerando todos os itens do teste, diferentemente do “*guessing*” para a **alternativa de cada item**, que seria o endosso probabilístico minimamente esperado de um item com relação ao seu respondente, que corresponde a proporção simples de assinalar uma entre todas as alternativas disponíveis, em especial, a resposta correta.

O “guessing” de um respondente com relação a um item, naturalmente se expande a média de todos os respondentes para o mesmo item. Para visualizar melhor essas diferenças, vejamos um exemplo de um teste com 10 itens e 11 respondentes, diante disso tem-se:

Quadro 30: Determinando o valor do guessing do teste.

Total de itens	n	10
Total de alternativas por item	k	4
%Guessing do item para um respondente - %G(i)	$\frac{1}{k} \cdot 100\%$	25%
Cálculo do escore do guessing do teste - G(t)	$G(t) = G(i) \cdot n$	2,50
Escore guessing do teste (verdadeiro)	Arredondar o G(t)	3
%Guessing do teste (verdadeiro)	$\frac{G(t)}{n} \cdot 100\%$	30%

Pode-se perceber que o “guessing do item” é a mesma para o conjunto de respondentes do item, pois ao tirar a média das probabilidades, multiplica-se e divide-se por n, ou seja, não se altera os 25%, isso significa que a taxa de adivinhação do item é a mesma independentemente da quantidade de respondentes. Assim, como o número total de acertos (A) corresponde ao mesmo número de sujeitos que responderam a determinada alternativa, é possível associar o “guessing do item” (percentual como probabilidade estimada) com o quantitativo percentual desses sujeitos. Por exemplo, se por um lado o “guessing de um item” (y_3 , por exemplo) com quatro alternativas ($k = 4$) for invariavelmente de 25%, e 3 dos 11 respondentes acertarem, ou seja, a taxa média de acertos for esse valor ($ID_i = \frac{A}{n} = \frac{3}{11} = 0,273$, ou seja, 27,3%), pode-se dizer que no caso de “chute” ou “acerto aleatório”, para cada uma das quatro alternativas existe 25% de probabilidade (inclusive para a alternativa correta!), e se somente 27,3% evidentemente acertaram, observa-se que esse percentual de acerto está equivalente ao limite percentual de sujeitos que espera acertar ao acaso (pela divisão dos números inteiros, estimam-se 3 respondentes que seria: $A = ID_i \cdot n = 0,25 \cdot 11 = 2,75$, ou seja, 3 respondentes). Diante disso, torna-se um equívoco assumir diretamente que esse percentual de sujeitos de fato dominam o conhecimento avaliado para o item y_3 , pois encontram-se na zona de “guessing do item”. Vale ressaltar que dentro desse grupo de respondentes, não se excluí a possibilidade da existência de respondentes de alto desempenho no perfil de composição desse grupo, apesar de que seja menos provável

que tenham “chutado” o item, salvo caso o item possua baixa qualidade ou que ocorra respostas pouco cuidadosas, como nos casos de “erro ao acaso”.

Extrapolando essa perspectiva, pode-se observar que **respondentes dentro da zona de “guessing do item”, não necessariamente estarão na zona de “guessing do teste”,** pois o primeiro depende do número de alternativas e se refere a realidade de acertos de um determinado item, enquanto o segundo do percentual de itens do teste como um todo. Uma analogia por ser feita com um lançamento de dados, a probabilidade de dar cara ou coroa em cada lance será sempre de 50% ($p = 1/2$), ainda que se lance 1000 vezes, a probabilidade será a metade, ou seja, em média serão 500 caras e 500 coroas, da mesma forma serve para a probabilidade de acerto de cada item, mas para um conjunto de itens, teria que ser uma análise combinatória de caras e coroas, ilustrando os acertos e erros na sequência de respostas dentro do conjunto de itens que formam o teste.

Tabela 2: Tabela-exemplo explicativa do guessing do item e do teste.

Zona de “guessing do item” y3 (27,3% da amostra):
ID₃ = 0,273 (≈ 25%)

N	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	Escore	Zonas
1	1	1	1	1	1	1	1	1	1	1	10	Zona de representatividade dos respondentes - escores com proporção de acertos acima do “guessing rate” (64% da amostra): Escore > 3
2	0	1	1	1	1	1	1	1	1	1	9	
3	0	0	1	1	1	1	1	1	1	1	8	
4	0	0	0	1	1	1	1	1	1	1	7	
5	0	0	0	0	1	1	1	1	1	1	6	
6	0	0	0	0	0	1	1	1	1	1	5	
7	0	0	0	0	0	0	1	1	1	1	4	
8	0	0	0	0	0	0	0	1	1	1	3	Zona de “guessing do teste” (36% da amostra): Escore ≤ 3
9	0	0	0	0	0	0	0	0	1	1	2	
10	0	0	0	0	0	0	0	0	0	1	1	
11	0	0	0	0	0	0	0	0	0	0	0	
ID emp	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	⇒ % empírica de certos	
% emp	9,1%	18,2%	27,3%	36,4%	45,5%	54,5%	63,6%	72,7%	81,8%	90,9%	⇒ % de respondentes	
% esp	≈ 10%		≈ 20%	≈ 40%		≈ 20%		≈ 10%			⇒ % esperado de respond.	
ID esp	até 24%		até 44%	até 54%		até 74%		até 95%			% esperado de acertos	
Classif.	Muito difícil		Difícil	Normal		Fácil		Muito Fácil			Tipologia das dificuldades	
Zonas	Zona empírica de guessing para os itens (27% da amostra): ID ≤ 25% *aprox.			Zona empírica de representatividade do teste - itens com proporção de acertos acima do guessing do item (70% dos itens do teste): percentual de itens com ID > 25% *aprox.								

Tem-se acima como exemplo uma matriz-resposta dicotômica, na qual “1” representa o acerto e “0” o erro. As proporções de acerto empírico estão apresentadas linearmente acumulativas na sequência dos itens, apenas para tentar ilustrar da melhor forma um panorama da situação. Pode-se observar que os três primeiros respondentes (N=1, 2 e 3) estão na zona de representatividade dos respondentes, ou seja, com escores mais elevados, e portanto, com mais evidências de que dominam o conteúdo do item y3 (porém, sem haja garantias), entretanto, eles estão na zona empírica de *guessing* do item, pois o percentual de acerto para o item foi muito baixo, praticamente no limiar da faixa de acerto ao acaso, e com isso, é um risco afirmar que dominam especificamente o conteúdo nesse item, ainda que sejam alunos com mais elevada performance.

O escore do “*guessing* do teste”, obviamente não pode ser um número decimal, não faria sentido, pois não se considera “meio acerto”, mas sempre números inteiros, portanto é preciso arredondar para mais o escore do “*guessing rate*”, que no caso representara a porcentagem de acerto ou proporção de respondentes.

Quanto às porcentagens de acerto para um item, recomenda-se que o **ideal seja a metade entre o valor máximo (ou seja, 100%) e a porcentagem de puro acerto ao acaso (“*guessing*”)** (Thompson & Levitov, 1985), que no caso, varia conforme a quantidade de alternativas por item. Seguindo esse critério, os itens com **5 alternativas** (y9, y11, y16, y18 e y26) devem ter **valores maiores e iguais a 60,0%**, e para o restante de 4 alternativas, **62,5%** (p.ex.: *Itens com 4 alternativas*: $\frac{25\%+100\%}{2} = 62,5\%$), ou seja, é desejável que todos os itens sejam fáceis para os respondentes, isso significaria que eles dominam supostamente o conhecimento que se exige.

Em termos **mínimos de classificação**, para identificá-los, se estabelece como critério uma porcentagem que esteja **acima do “*guessing*” esperado**, com isso, os itens com **5 alternativas fica sendo 20%** dentro do total de respostas para cada alternativa, e o **restante dos itens em torno de 25%**. Foi observado que esses mesmos critérios foram adotados pelos autores do TCE em 2012 (Chu, Treagust, & Yeo, 2012, p. 1520), entretanto, focando apenas no percentual de respostas para distratores específicos e não para o percentual de respostas acertadas. No caso, valores menores ou igual ao

“guessing” do item são indicativos de inconsistência no item ou que os alunos ficaram na margem do “acerto aleatório”, ou seja, a probabilidade de *acertar ao acaso* está coincidindo com o percentual de respondentes na alternativa, e dessa forma não foi considerado razoável assumir que sejam representativamente fortes evidências de concepção entre os alunos. A zona de representatividade do teste está relacionada a critérios mínimos referenciados preliminarmente, com isso, o alcance percentual de acertos dos itens e os totais de itens favoráveis que atendem a esses critérios são metas estabelecidas pelo ensino, e também dependem da etapa de aprendizagem na qual ele é aplicado ou administrado. A zona de representatividade dos respondentes descrita na tabela explicativa será comentada mais adiante em detalhes através de outros gráficos.

2.3.2.1.4 Função de densidade de distribuição normal

A medida do $h(p)$, sendo o valor da densidade da distribuição normal padrão no ponto em que a área da curva à esquerda desse ponto é igual a p , se refere a proporção de acertos do item (Rabelo, 2013). A função densidade de probabilidade da distribuição normal é dada por:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Sendo:

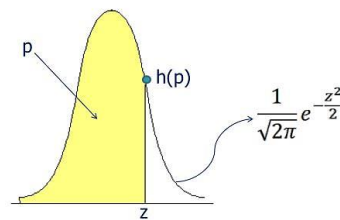
- μ a média;
- σ o desvio-padrão populacional.

O ponto p tem como abscissa o valor normal padronizado na distribuição Z representado por:

$$Z = \frac{x - \mu}{\sigma}$$

Considerando $\mu = 0$ e $\sigma = 1$, no ponto p se tem:

$$h(p) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2}$$



O processo de transpor a os escores brutos para escores padronizados (Z) é um artifício na qual “força” a distribuição, que na maioria das vezes possui uma assimetria, para que se tenha uma distribuição normal. Esse procedimento foi considerado para psicometristas do início do século XX como uma forma de obtenção do escore verdadeiro, argumento atualmente ultrapassado, mas que, entretanto, é um procedimento utilizado comumente, principalmente na construção de escalas psicométricas, como no caso do SAEB e ENEM, e em especial, na comparação de escores com testes diferentes que medem o mesmo «traço latente».

2.3.2.2 Análise Gráfica do Item (AGI)

A AGI é uma técnica desenvolvida originalmente por Theo A. Van Batenburg e Jacob A. Laros a partir da análise de testes aplicados no SAEB em 1997 (Batenburg & Laros, 2002), sendo aplicado em outro momento na Turquia (Yurdugül & Batenburg, 2006). Essa técnica demonstra ser útil em análises complementares da TCT e TRI sobre a qualidade dos itens sem a necessidade de softwares mais sofisticados, além de contribuir em análises que visam a eliminação de itens, na construção de testes e no processo de melhoria dos itens, pois possibilita detectar facilmente itens com baixa qualidade além de fornecer informações essenciais e de fácil interpretação para as características do itens como o índice de dificuldade, discriminação, e “acerto ao acaso” (também conhecido como “guessing”) (Ibid., 2002). Rodrigues (2006, p. 51) afirma que o principal pressuposto dessa análise é que: *“a proporção da alternativa correta deve aumentar com um aumento do escore total, e a proporção de alternativas falsas deve decrescer com um aumento do escore total”* (Batenburg & Laros, 2002), e acrescenta que sejam válidos mais dois pressupostos para a AGI (Rodrigues, 2006, p. 51): (1) *um aluno que dá uma resposta certa em um item de múltipla escolha sabe mais que um aluno que dá a resposta errada; e (2) um aluno que tem mais itens certos sabe mais que um aluno com menos itens certos.* Essas afirmações seguem de acordo com a TCT.

Para interpretar a AGI, foi adaptado o exemplo de um bom item teórico (e não empírico) com cinco alternativas apresentada pelos autores Batenburg & Laros (2002) que propuseram essa técnica. Os autores fazem uma ressalva na análise para quatro aspectos que visam caracterizar a qualidade ou eliminação do item: (1) *analisar o número de violações contra a suposição de que os aumentos das alternativas corretas devam seguir crescentemente com a pontuação total*; (2) *analisar o número de violações contra a suposição de que a proporção dos distratores (falsas alternativas) deva diminuir o aumento da pontuação global*; (3) *analisar o número de intersecções entre a alternativa correta e as falsas após o início do intervalo de discriminação*; (4) *analisar a ausência de discriminação ou de uma baixa potência de inclinação*.

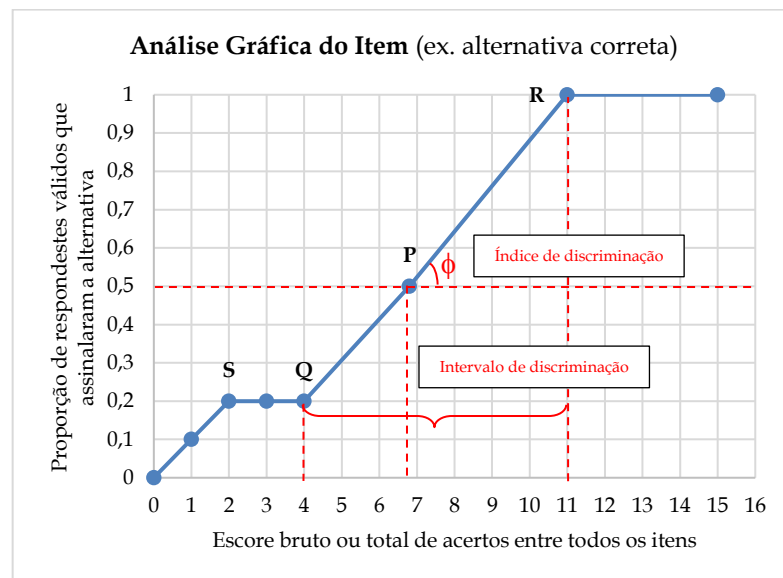


Figura 16: Exemplo de item teórico considerando apenas a alternativa correta.

Fonte: Adaptado de Batenburg & Laros (2002).

Com isso seguem as considerações principais para realizar a análise:

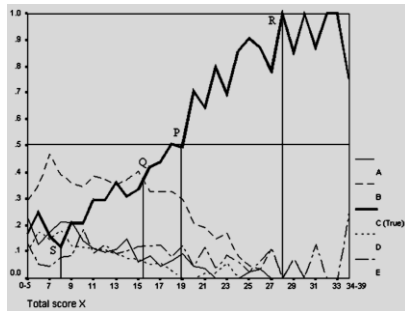
- Para um bom item, espera-se que a proporção de acerto ao item tende a aumentar de 0 para 1 conforme vai aumentando o escore total, enquanto a proporção dos distratores (alternativas “falsas”) do item tende a diminuir;
- A soma das proporções em cada nível de escore para todas as alternativas assinaladas tende a 1 (ou 100%), quando não ocorre é devido a eventual existência de casos omissos;

- Em certo nível de escore, se espera que as alternativas certas e as falsas fiquem nos valores da chance de acerto ao acaso (p. ex., 0,25 para item com 4 alternativas e 0,2 para itens com 5 alternativas, como no caso do teste TCE);
- A partir em que os escores dos distratores assumam a perspectiva de acerto ao acaso, se espera que a marcação da alternativa correta aumente conforme o escore total específico, ou seja, a proporção de marcação da alternativa correta aumenta, e a proporção de marcação das alternativas falsas decresce.
- O ponto S corresponde ao início em que as proporções das respostas corretas começam a aumentar, nesse ponto a proporção das respostas corretas é mínima, ou seja, é mais próxima das abscissas. É a partir desse ponto que se tem como referência na discriminação do item, que vai de S ao ponto R, que corresponde ao pico alcançado;
- O ponto Q corresponde a interseção ou cruzamento da “linha” da alternativa correta com a “linha” do distrator com proporção mais elevada, e a partir desse ponto, na discriminação entre os grupos de maior e menor desempenho a proporção da alternativa correta se torna cada vez maior;
- O ponto P corresponde a proporção de 0,5 (50%) no item e passando por ele se estabelece uma “linha sólida” que divide igualmente o total de respondentes. Dessa forma representa a mediana do total das respostas, com isso, a partir desse ponto se tem um predomínio proporcional de respondentes em determinado nível de escore. Considera-se de modo geral que **a abscissa nesse ponto seja o nível de desempenho que corresponde a dificuldade do item na escala**, no exemplo teórico corresponde em torno de 6,8 valores. Obviamente que em uma situação empírica corresponderia a um valor discreto, pois os escores brutos são valores inteiros em um teste dicotômico;
- O ponto R corresponde ao “pico” de proporção na assinaladas na sequência dos níveis de escore bruto, que nem sempre alcançam o valor máximo e podem ocorrer diversos picos (R_1, R_2, \dots), e conseqüentemente novos pontos Q (Q_1, Q_2, \dots) e P (P_1, P_2, \dots) cruzando a “linha sólida”. Quando isso ocorre, não é um bom sinal para a qualidade do item, pois surgem diversas inclinações e índices secundários de discriminação no mesmo item que vai de S a R_1 , S a R_2 , etc., assim como pontos extras para a determinação da dificuldade do item no nível de desempenho. Isso resulta numa

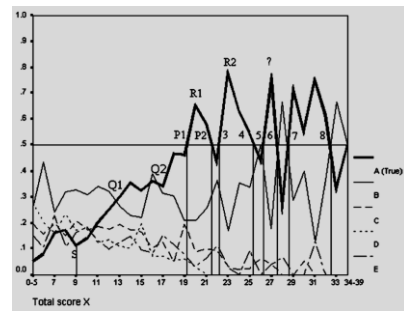
indeterminação do intervalo de discriminação, do posicionamento da dificuldade do item na escala e do poder de discriminação, e será considerado um item ruim;

- Quando o ponto R alcança uma proporção máxima (100%) no item é o momento em que em todos os respondentes assinalaram a alternativa (podendo ser a correta ou não), sendo no caso a alternativa correta, é desse ponto que se determina tanto o intervalo de discriminação do item (eixo X) que é o comprimento linear horizontal de Q a R, quanto o índice de dificuldade, que se refere a tangente do ângulo de inclinação (ϕ) da reta entre S e R com o eixo horizontal ou a linha de corte de 0,5 (50%) em que P se encontra;
- No exemplo os pontos S e Q (entre os escore 2 e 4) a proporção fica igual a todos os respondentes de cada uma das cinco alternativas, pois a proporção de acerto se torna igual a todos com 0,2, dessa forma o ângulo de inclinação é nulo, ou seja, não ocorre discriminação entre os grupos (o mesmo ocorre com o intervalo de escore de 11 a 15). Deve-se considerar a reta de inclinação nesse caso partindo de Q;
- Nos intervalos que não ocorrem discriminação ($\phi = 0$) os dados empíricos desse grupo serão extremamente baixos;
- Quanto maior o ângulo de inclinação maior será o poder de discriminação do item, e quanto maior o intervalo de discriminação (distância de Q a R) menor será o ângulo entre S e R, conseqüentemente justifica a diminuição do poder de discriminação;
- Quando um distrator tiver uma proporção superior ao da alternativa correta em um intervalo de desempenho, ou seja, até o ponto Q, convém analisar o conteúdo do distrator para compreender o fato dele ser mais atrativo para esse determinado grupo de sujeitos (essa é uma situação que eventualmente ocorre em itens empíricos);
- Quando a proporção de um determinado distrator sobressai em relação a alternativa correta e segue com padrões elevados na progressão dos níveis de desempenho, estes devam ser analisados para averiguar a causa de sua força atrativa. Também podem ser eliminados esses distratores numa reformulação do teste, tendo em vista que eles diminuem o poder de discriminação e não contribuem para medir aquilo que o teste tem como finalidade medir ou verificar.

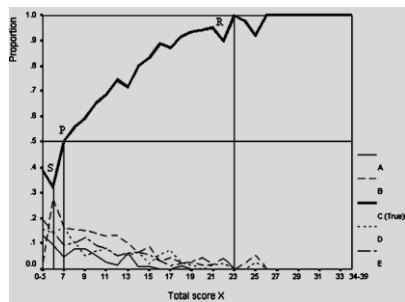
Item de boa qualidade



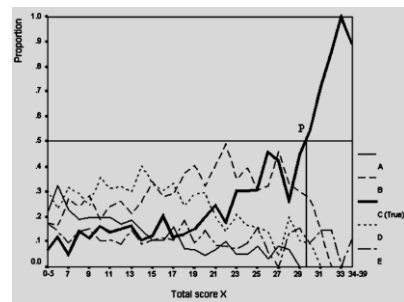
Item de qualidade baixa



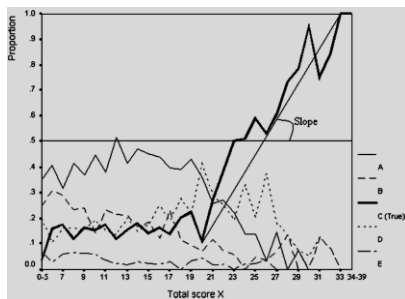
Item fácil



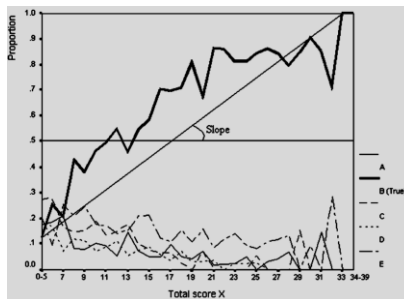
Item difícil



Item com alta discriminação



Item com baixa discriminação



Item com distrator com forte poder atrativo

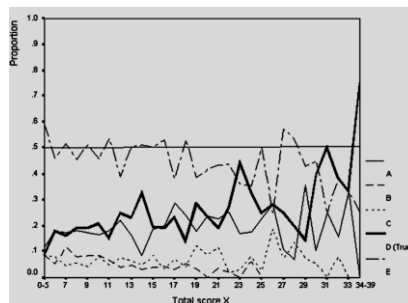


Figura 17: Exemplos tipológicos na AGI de acordo com diferentes combinações de características psicométricas.

Fonte: Batenburg & Laros (2002)

Batenburg & Laros (2002) e Rodrigues (2006) apresentam exemplos de algumas AGI no que se refere à qualidade, poder de discriminação, efeito atrativo de distratores e do acerto ao acaso. Os exemplos apresentados por Batenburg & Laros (2002) demonstram serem úteis como referência na interpretação e para as análises clássicas, reforçando a interpretação dos diferentes dos diferentes grupos de respondentes (Figura 17).

Analisando a classificação dos autores a partir dos gráficos, pode-se dizer que a qualidade de um item está relacionada com a discriminação e a força dos distratores, e de modo geral é quando eles demonstram manter um crescimento percentual mais consistente para cada nível subsequente na pontuação, dando pico cada vez mais elevados na porcentagem de acerto. Um item de baixa qualidade seria uma disputa mais equilibrada da alternativa correta com um ou mais distratores, e uma qualidade mediana seria um meio termo entre as duas situações.

Para um **item fácil**, no caso, visualiza-se um crescimento vertiginoso desde o início do gráfico, alcançando o mais brevemente possível um percentual de 100% de acerto ou com uma curva apresentando picos percentuais de resposta elevados desde o início do gráfico, quanto os distratores ficam relativamente próximos ao eixo das abscissas. Um **item difícil** mostra que apenas indivíduos com alta performance (elevados escores brutos) conseguem se destacar, ou seja, a elevação dos percentuais ocorrem próximo a zona do grupo superior de desempenho, porém observa-se a força dos distratores por quase todos os níveis de pontuação, inclusive os níveis mais elevados. Um item com dificuldade mediana ilustra uma curva com crescimento próximo a diagonal vindo do canto inferior esquerdo para o superior direito.

Um item com **alta discriminação** apresenta um certo equilíbrio no percentual entre os respondentes de baixa performance, porém demonstra um crescimento abrupto (com picos percentuais elevados) quando alcança a zona de respondentes de elevado desempenho, com isso o ângulo que representado o poder discriminatório apresenta-se com maior atividade. Um item com **baixa discriminação** por sua vez apresenta um ângulo relativamente baixo de inclinação, para isso, geralmente

apresenta uma curva que explora a diagonal do gráfico iniciando o crescimento mais à esquerda indo até o ponto superior mais à direita.

Um item com **distrator com forte** poder atrativo apresenta características de acordo com aquele considerado de péssima qualidade, geralmente os percentuais dos distratores são elevados para alunos na zona de maior desempenho ou do grupo superior da amostra (27% acima).

A AGI é uma aliada para as análises clássicas dos testes (TCT), sobretudo na interpretação de seus parâmetros, em que fornecem informações sobre a qualidade psicométrica dos itens, tanto das respostas certas quanto para os distratores, sendo, portanto, informações relevantes do ponto de vista pedagógico na interpretação dos conteúdos e análises dos perfis de composição dos grupos nos diversos níveis de desempenho, além de auxiliar uma melhor apropriação das análises na TRI.

2.3.2.3 *Ganho normalizado e tamanho do efeito*

Através desse instrumento, Richard R. Hake mostrou que até graduados em nível universitário não conseguiram compreender algumas das ideias mais fundamentais da física (Hake, 1998), utilizando o famoso teste conceitual *Mechanics Baseline Test – MBT* (Halloun & Hestenes, 1985b), e o posterior e mais conhecido como “FCI” (Hestenes & Wells, 1992), que traduzindo seria o *Inventário Conceitual de Forças*. A contribuição neste trabalho envolve a proposição da *normalização de ganhos* $\langle g \rangle$ (“*normalized again*”) entre pré-testes e pós-testes se tornando uma referência para as análises clássicas dos testes conceituais assim como objeto de análise e críticas por outros pesquisadores, porém recorrentemente utilizado e com indícios de validade. Hake definiu o ganho normalizado médio: “*como uma medida aproximada da eficácia de um curso na promoção da compreensão conceitual*” (Hake, 1998, *tradução livre*), sendo expresso como:

$$\langle g \rangle \equiv \frac{\langle G \rangle}{\langle G_{\text{máx}} \rangle} = \frac{\langle \text{pos} \rangle - \langle \text{pre} \rangle}{100 - \langle \text{pre} \rangle}$$

Equação 1: Fator de Hake (ganho médio normalizado).

Sendo:

$\langle g \rangle$ o ganho normalizado médio;

$\langle G_{m\acute{a}x} \rangle$ o máximo ganho possível médio;

$\langle pos \rangle$ a média do pós-teste, e;

$\langle pre \rangle$ a média do pré-teste.

Torna-se necessário discriminar os impactos de ganho percentual normalizado entre os diferentes grupos de respondentes, buscando assim verificar as necessidades de mudança e/ou melhoria no cenário pedagógico. Diante disso, entende-se ainda que esse processo não impede a possibilidade de verificar esses impactos por grupos de respondentes dentro de cada item, ao invés do teste como um todo.

Esses impactos classificam-se como:

Alto- g : $\langle g \rangle \geq 0,70$

Médio- g : $0,70 > \langle g \rangle \geq 0,30$

Baixo- g : $\langle g \rangle < 0,30$

O gráfico sugerido por Hake para esse tipo de análise é o ganho normalizado ($\langle g \rangle$) *versus* porcentagem do valor médio no pré-teste ($\% \langle pre \rangle$), na qual sugere-se que seja significativo um impacto com ganho de $\langle g \rangle \geq 0,60$ (Hake, 1998, p. 16) que seria:

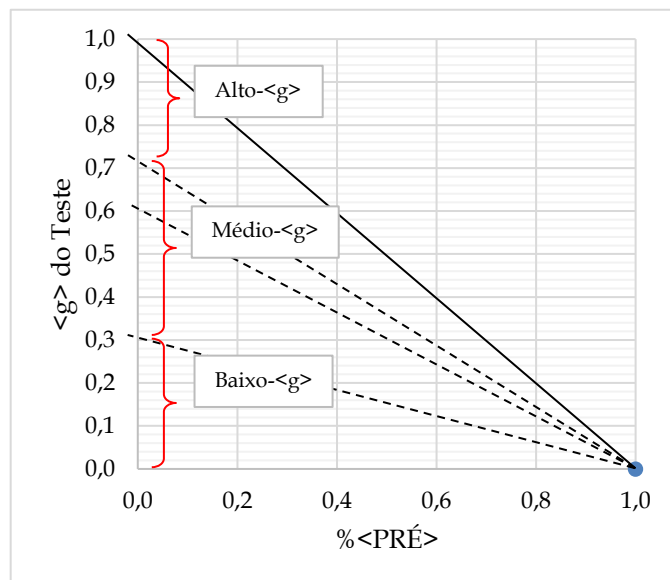


Figura 18: Gráfico $\langle g \rangle$ *versus* $\% \langle pré \rangle$.

Hake propôs o uso de ganhos normalizados por ser evidentemente diferenciado entre os convencionais métodos de ensino, ele considera que esse efeito permite "uma

análise consistente sobre diversas populações estudantis com estados de conhecimento inicialmente variados", acreditando que fosse independente dos resultados da população ou pré-teste, permitindo que os instrutores comparassem a aprendizagem de seus alunos com os de outros alunos em diferentes tipos de instituições. Estudos posteriores que não verificaram maiores diferenças entre *ganhos médios* e *média dos ganhos* dos escores (Lei Bao, 2006), e algumas críticas relacionadas de que o fator de Hake *não contabiliza as perdas de respondentes do pré-teste e pós teste* (Miller et al., 2010); demonstra ser fiável apenas em instituições com baixa evasão (5-10%), pois quando fica entorno de 50%, acaba não considerando uma grande fração de estudantes que fizeram o pré-teste e abandonaram; de *não ser independente da preparação dos respondentes* (Coletta & Phillips, 2005); e mais recentemente, uma meta-análise *não encontrou uma correlação significativa entre os ganhos normalizados do pré-teste e os ganhos normalizados em resultados na literatura* (Von Korff et al., 2016).

Apesar disso, alguns estudos vem utilizando o fator Hake obtendo resultados relativamente consistentes na aquisição do conhecimento térmico (Baser, 2006; Crossley, 2010). Apesar disso, o ganho normalizado pode ser calculado de diversas formas, podendo variar em torno de até 10% com relação ao fator de Hake. Uma das formas diretas é calcular o ganho através da *média de ganhos* ao invés do *ganho médio normalizado*.

$$g = \left\langle \frac{pós - pré}{100 - pré} \right\rangle$$

Equação 2: Média dos ganhos.

Comparando, se tem:

Quadro 31: Comparação entre o Fator de Hake e a Média dos Ganhos.

Fator de Hake: $\langle g \rangle = \frac{\langle pós \rangle - \langle pré \rangle}{100 - \langle pré \rangle}$	Média dos ganhos: $g = \left\langle \frac{pós - pré}{100 - pré} \right\rangle$
<ul style="list-style-type: none"> • É a definição "padrão" de ganho normalizado, segundo Hake; • É a única coisa que você pode calcular se você não tem pontuação individual de alunos, mas tem pontuação média da classe. • Ainda está bem definido e significa algo razoável, mesmo que alguns 	<ul style="list-style-type: none"> • Provavelmente é muito usado, possivelmente ainda mais comumente usado do que a definição oficial. • Isso força você a usar dados correspondentes. • Você pode relacionar significativamente os ganhos individuais dos alunos com os ganhos médios da classe. Se você traçar um histograma de

alunos obtenham 100% no pré-teste ou tenham pontuações que diminuam ligeiramente do pré-teste para o pós-teste. Para a outra medida, você deve abandonar esses alunos ou usar mudanças normalizadas.

ganhos de alunos individuais, a média desse histograma será o ganho médio da turma.

- Para uma mudança normalizada, Marx & Cummings (2007) argumentam que o cálculo de g é melhor do que $\langle g \rangle$, porque torna os ganhos mais importantes do que as perdas, e também porque captura com mais precisão a propagação.

Fonte: McKagan, Sayre, & Madsen (2017, *tradução livre*)

Marx & Cummings (2007) sugerem como alternativa um efeito denominado “mudança normalizada” $\langle c \rangle$, que seria o equivalente “ganho normalizado médio de alunos”. Para isso remove alunos que pontuam 0 ou 100% tanto no pré-teste como no pós-teste, e usa um cálculo diferente para estudantes com ganhos negativos:

Mudança normalizada	Condição quanto aos escores	Justificativa
$\langle c \rangle$ $\left\{ \begin{array}{l} = \left\langle \frac{pós - pré}{100 - pré} \right\rangle \\ \\ Retirar \\ \\ 0 \\ \\ = \left\langle \frac{pós - pré}{pré} \right\rangle \end{array} \right.$	$pós > pré$	Considera-se o ganho máximo possível
	$pós = pré = 100\% \text{ ou } 0$	As pontuações extremas devem ser removidas dos conjuntos de dados porque o desempenho do aluno está além do escopo do instrumento de medição.
	$pós = pré$	Simplesmente não há ganhos
	$pós < pré$	Considera-se a perda máxima possível

Equação 3: Mudança normalizada.

Fonte: Adaptado de Marx & Cummings (2007, p. 88, *tradução livre*)

O argumento para a mudança normalizada é de que mesmo em cenários de ensino efetivos, alguns alunos não melhoram suas pontuações. Diante disso é preciso calcular a média das mudanças normalizadas lidando com as pontuações de cada indivíduo adequadamente, caso contrário, isso irá refletir na precisão da avaliação e perda no diagnóstico (Marx & Cummings, 2007). Uma preocupação foi levantada devido pelo fato de que as pontuações que geram a média de valores de mudança normalizados positivos e a de mudança normalizadas negativas são feitas duas proporções diferentes. Diante disso, foi proposta uma relação única que produz os

valores gerados da relação de mudança normalizada a uma decisão arbitrária:

$$c = g \left(\theta \left(\frac{100}{\text{pré}} - 2 \right) + 1 \right)$$

Equação 4: Relação geral da mudança normalizada.

, onde θ é dado por:

$$\theta = \frac{1}{\alpha^{\beta(\text{pós-pré})} + 1}$$

Equação 5: Parâmetro responsável pela mudança [normalizada].

No caso, α e β são parâmetros arbitrários que podem ser variados para obter o grau de precisão desejado. Sugere-se que θ deva ser pensado como uma mudança (Marx & Cummings, 2007, p. 91).

Vem sendo defendido na literatura que os cálculos devam incluir *apenas dados para os alunos que fizeram o pré e o pós-teste*, comumente designados como **dados "correspondentes"**, para evitar inflar o ganho ao incluir apenas as pré-pontuações de estudantes que caíram ou pararam de frequentar a aula (McKagan et al., 2017). O estudo de meta-análise de Von Korff et al. (2016) mostra no entanto, que a maioria dos estudos que relatam ganhos normalizados não mencionam se eles usam dados correspondentes, então não devemos assumir que eles fazem.

Nas pesquisas em ciências sociais diferentemente da Física, é mais comum a utilização do termo "tamanho do efeito" ("*effect size*") ao invés de "ganhos". Nesse âmbito, Jacob Cohen propõe o "tamanho do efeito" sendo obtido "*ao padronizar o tamanho do efeito bruto conforme expresso na unidade de medida da variável dependente dividindo-o pelo desvio padrão (comum) das medidas em suas respectivas populações, o último também na unidade de medida original*" (Cohen, 1977).

Para duas amostras independentes:

$$d = \frac{m_A - m_B}{\sigma}$$

Figura 19: Tamanho do Efeito de Cohen.

Para o caso direcional (de uma só cauda):

$$d = \frac{|m_A - m_B|}{\sigma}$$

Figura 20: Tamanho do Efeito de Cohen para uma cauda.

Sendo:

d o índice de tamanho de efeito em termos de médias padronizadas de unidade;
 m_A e m_B são as médias de cada grupo da população expressos em unidades brutas (medição original);
 σ sendo o desvio padrão de qualquer população (uma vez que são assumidos como iguais) (Cohen, 1977, p. 20, tradução livre).

A classificação para os efeitos, são relativas, entretanto sugere-se:

Quadro 32: Classificação para o tamanho do efeito do "d de Cohen".

Em termos de efeito	Em termos de correlação
Baixo: $d \approx 0,20$	O r de Pearson (equivalente ao ponto-bisserial) corresponde a 0,10 e r^2 , portanto, a 0,01, ou seja, a associação da população representa 1% da variância da variável dependente Y nas populações A e B combinadas.
Médio: $d \approx 0,50$	O r de Pearson (equivalente ao ponto-bisserial) corresponde a 0,243 e r^2 , portanto, a 0,059, ou seja, a associação da população representa 5,9% da variância da variável dependente Y contabilizada na adesão das populações A e B.
Alto: $d \approx 0,80$	O r de Pearson (equivalente ao ponto-bisserial) corresponde a 0,371 e r^2 , portanto, a 0,138, ou seja, a associação da população representa 13,8% da variância da variável dependente Y contabilizada na adesão das populações A e B.

Fonte: Elaborado a partir de Cohen (1977, pp. 24-27, tradução livre)

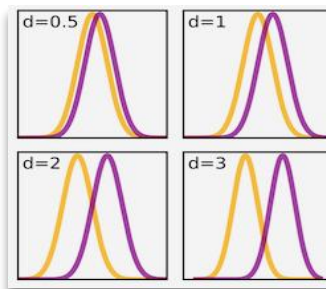


Figura 21: Comparativo das diferenças entre ganhos normalizados de um pré-teste e pós-teste segundo o "d de Cohen".

Fonte: Skbkekak (2012)

Cohen (1977) lembra que os cientistas comportamentais que trabalham com coeficientes de correlação (como, por exemplo, psicólogos educacionais) não consideram normalmente um r de 0,371 como grande. Nas linguagens dos testes conceituais, o "d de Cohen" pode ser expresso como:

$$d = \frac{\langle pos \rangle - \langle pre \rangle}{\sigma}$$

Equação 6: "d de Cohen" para pré-teste e pós-teste com grupos correspondentes.

Cohen argumentou que o desvio padrão de qualquer grupo poderia ser usado quando as variâncias dos dois grupos são homogêneas (Becker, 2000), e com isso deve-se utilizar o valor absoluto no numerador. Uma das vantagens do tamanho do efeito de Cohen é porque:

“O desvio padrão inclui quantos alunos você possui, e com isso, ao usar o tamanho do efeito você pode *comparar a eficácia do ensino entre as classes de diferentes tamanhos de forma mais justa*. O tamanho do efeito é uma medida popular entre pesquisadores e estatísticos educacionais por esse motivo” (Madsen et al., 2016, *tradução livre, grifo nosso*)

Pode-se dizer que, ao contrário do *ganho normalizado*, o *tamanho do efeito* não tem limite superior, embora os tamanhos do efeito sejam geralmente inferiores a 2. Para os testes conceituais, o desvio-padrão “ σ ” é a combinação dos escores do pré-teste e pós-teste, sendo $\langle pos \rangle$ e $\langle pre \rangle$ as médias da classe no pré-teste e pós-teste. Vale ressaltar que o *tamanho do efeito* não é igual a *significância estatística p-values*, pois a significância se refere ao quanto é provável que um resultado seja devido ao acaso e o tamanho do efeito indica a importância do resultado. E sabe-se que, o *tamanho do efeito* **é calculado apenas para alunos correspondentes que tomaram o pré-teste e o pós-teste** (Madsen et al., 2016).

"O significado estatístico não é equivalente ao significado científico, humano ou econômico. Os valores *p-values* menores não implicam necessariamente a presença de efeitos maiores ou mais importantes, e **valores *p-values* maiores não implicam falta de importância ou mesmo falta de efeito**. Qualquer efeito, não importa quão minúsculo, possa produzir um pequeno valor de *p-values* se o tamanho da amostra ou a **precisão da medição for alta o suficiente**, e os efeitos maiores podem produzir valores *p* não impressionantes se o tamanho da amostra for pequeno ou se as **medidas forem imprecisas**. Da mesma forma, efeitos idênticos estimados terão diferentes valores *p-values* se a **precisão das estimativas for diferente**” (Wasserstein & Lazar, 2016, p. 132, *tradução livre, grifo nosso*)

Na comparação com do *Fator de Hake* ou “ $\langle g \rangle$ de Hake” com o “*d* de Cohen”, pode se dizer ainda que:

A maior diferença matemática entre o *ganho normalizado* e o *tamanho do efeito* é que o ganho normalizado não explica o tamanho da classe ou a variação nos alunos dentro da classe, mas o tamanho do efeito faz. Ao explicar a variação nos escores dos indivíduos, o tamanho do efeito é uma medida muito mais sensível do que o ganho normalizado. A diferença é mais pronunciada em classes muito pequenas ou diversas. Como o erro geralmente diminuiu com o aumento do tamanho da amostra, as classes pequenas são muito mais vulneráveis no ganho normalizado do que o tamanho efetivo: o mesmo ensino ano após ano pode fazer balanços selvagens no ganho normalizado, mas pequenas mudanças no tamanho do efeito. (McKagan et al., 2017, *tradução livre*)

Também é possível adaptar o fator de Cohen para grupos diferentes, como na meta-análise, considerando grupo controle e experimental para uma única observação (Becker, 2000). Na meta-análise, os teste-t e testes-F são inadequados pelo fato de estarem em função do tamanho da amostra. Ao contrário disso, as estimativas do tamanho do efeito não são influenciadas pelos tamanhos da amostra, com isso, o “d de Cohen” é o mais comum utilizado nesse tipo de análise. Para compreender como fazer isso, tem-se o desvio-padrão é expresso por:

$$\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{N}}$$

, sendo x o escore bruto, \bar{x} a média e N o número de casos.

Considerando o desvio-padrão de grupos distintos, pode-se ser adequado utilizar o desvio-padrão combinado, que corresponde a raiz quadrada da média da soma dos quadrados dos desvios-padrões de cada amostra (Cohen, 1988, p.44), lembrando que quando os dois desvios-padrão são semelhantes o quadrado médio da raiz não diferirá muito da média simples das duas variâncias. Na prática o desvio-padrão combinado é comumente utilizado (Rosnow & Rosenthal, 1996), e pode ser expresso como:

$$\sigma_{comb} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

Quando se aplica a uma turma que não recebeu o tratamento (podendo ser um grupo de pré-teste – *grupo controle*) com outra que recebeu (podendo ser pós-teste – *grupo experimental*), sendo grupos não são homogêneos, os sinais positivos devem

indicar uma direção de melhoria, e sinais negativos uma direção de deterioração ou oposta a direção prevista. Como o tamanho da amostra pode eventualmente mudar para cada grupo, considera-se o tamanho da amostra do “pré-teste” sendo $n_{pré}$ e $n_{pós}$ para o pós-teste. O desvio-padrão combinado reuni informações de ambos os grupos e será dado por (Yeo & Zadnik, 2001):

$$\sigma_{comb} = \sqrt{\frac{(n_{pré} - 1) \cdot \sigma_{pré}^2 + (n_{pós} - 1) \cdot \sigma_{pós}^2}{n_{pré} + n_{pós} - 2}}$$

Equação 7: Desvio-padrão relacionado a dois grupos não correspondentes.

Dentro das propriedades do desvio-padrão úteis nesse tipo de caso, tem-se o desvio-padrão da soma algébrica de dois termos (p. ex.: os escores brutos de dois grupos), sendo:

$$\sigma_{pré,pós} = \sqrt{\sigma_{pré}^2 + \sigma_{pós}^2 + 2 \cdot \sigma_{pré} \cdot \sigma_{pós} \cdot \rho(pré, pós)}$$

E para determinar uma razão de desvios-padrões, sendo $\rho(pré, pós)$ o coeficiente de correlação entre os dois grupos, o resultado pode ser obtido como:

$$\sigma \approx f \cdot \sqrt{\left(\frac{\sigma_{pré}}{pré}\right)^2 + \left(\frac{\sigma_{pós}}{pós}\right)^2 - 2 \cdot \frac{\sigma_{pré} \cdot \sigma_{pós}}{pré \cdot pós} \cdot \rho(pré, pós)}$$

$$\rho(pré, pós) = \frac{COV(pré, pós)}{\sigma_{pré} \cdot \sigma_{pós}}$$

Existem ainda outras formas de tamanhos do efeito para dois grupos independentes, seja com sua relação com o teste-*t* e os graus de liberdade, o “*g*” de Hedges, Δ de Glass, entre outros (Ver em Becker, 2000; Cohen, 1988; Espírito-Santo & Daniel, 2015; Rosenthal & Rosnow, 1991; Rosnow & Rosenthal, 1996). De modo geral, recomenda-se algumas diretrizes para a escolha do tamanho (Espírito-Santo & Daniel, 2015).

Quadro 33: Diretrizes para apresentar e interpretar os Tamanhos Do Efeito (TDE).

1. Escolher o TDE mais adequado ao tamanho da amostra de cada grupo.
2. Indicar os dados descritivos básicos (media e desvio- padrão).
3. Calcular o TDE, independentemente de os resultados serem ou não significativos.
4. Designar o tipo de TDE usado, indicar as referências apropriadas e/ou apresentar a equação.
5. Calcular o Intervalo de confiança a 95% para o TDE.
6. Classificar o TDE de acordo com os critérios de Cohen (1988) e de Rosenthal (1996).

-
7. Interpretar o TDE em linguagem comum (uma das seguintes hipóteses):
 - a. Converter em percentis.
 - b. Calcular a percentagem de sobreposição.
 - c. Calcular a probabilidade de superioridade.
 8. Interpretar o TDE no contexto da área de investigação:
 - a. Rever a literatura, optando por estudos com planeamento e métodos de cálculo do TDE semelhantes.
 - b. Calcular os TDEs dos estudos revistos quando ausentes.
 9. Interpretar o TDE no contexto de variáveis da Realidade.
-

Fonte: Espírito-Santo & Daniel (2015).

2.3.2.4 Fator de concentração de análise «C»

Uma aplicação interessante para verificar tendências de modelos no padrão de respostas de cada item em um teste conceitual, bem como nas mudanças em termos de médias entre um pré-teste e pós-teste, pode ser obtido pelo Fator de Concentração da análise, proposta por Bao & Redish (2001). Os autores destacam uma coerência na relação dessa análise com as teorias dos modelos mentais, na qual destacam uma concordância com os seguintes princípios: *A memória é associativa; As respostas cognitivas são produtivas; As respostas cognitivas são dependentes do contexto (incluindo o contexto do estado de espírito do aluno).*

No âmbito de um entendimento sobre a aprendizagem de disciplinas complexas, como a física no nível da faculdade, eles se apoiam em pesquisadores da neurociência, ciências cognitivas e pesquisadores educacionais, na qual destacam os seguintes princípios: *Padrões de associações (redes neurais); Primitivas/Facetas; Esquemas; Modelos mentais; Modelos físicos.*

Padrões de associações (redes neurais) – é a estrutura de ligação fundamental representada pelas conexões de neurônios e modelos de redes neurais. Uma associação entre elementos da memória (declarativa ou processual) depende do contexto e, como todos os fatores determinantes que uma ativação não pode ser especificada, devem ser tratados de forma probabilística;

Primitivas/Facetas – Uma primitiva é uma regra, muitas vezes indivisível para o usuário, que quando aplicada em um contexto físico, produz uma faceta de uma declaração sobre como um sistema físico particular se comporta.

O conhecimento *declarativo*, as *primitivas* e as *facetatas* estão ligadas em **padrões associativos dependentes do contexto**. Quando um padrão particular (contendo

poucos ou muitos elementos) é robusto e ocorre com uma alta probabilidade em contextos particulares, referimo-nos ao padrão de associação como um *esquema*. Nós chamamos esquemas que são modelos mentais particularmente robustos e coerentes. Se um modelo mental é baseado em um conjunto de ideias sobre objetos físicos e suas propriedades, nós o chamamos de *modelo físico*. (Bao & Redish, 2001, p. 5, *tradução livre, grifo nosso*)

Considerando como exemplo um *teste de escolha múltipla com resposta única* (“*multiple-choice single-response*” - MCSR) com cinco alternativas (a, b, c, d, e) aplicado a 100 estudantes, diante pode-se admitir 3 padrões de respostas:

Tabela 3: Exemplos de concentração de respostas sem pico (tipo I), com dois picos (tipo II) e com 1 pico (tipo III).

Tipo	A	B	C	D	E
I	20	20	20	20	20
II	50	10	30	5	5
III	100	0	0	0	0

A descrição para os tipos de padrões de respostas pode-se:

Tipo I - **Sem pico**: as respostas são distribuídas de forma um tanto uniforme entre três ou mais escolhas.

Tipo II - **Dois picos**: a maioria das respostas está concentrada em duas opções, geralmente uma correta e uma incorreta.

Tipo III - **Um pico**: a maioria das respostas está concentrada em uma escolha (não necessariamente correta).

Considerando conveniente construir uma medida simples que forneça informações sobre a distribuição das respostas, os autores propuseram o **fator de concentração (C)** como uma função da resposta do aluno que leva um valor no intervalo de $[0,1]$, sendo $C = 1$ para um padrão de resposta perfeitamente correlacionada (tipo III), e $C = 0$ para um padrão de resposta aleatória (tipo I).

Quadro 34: Demonstração do Fator de Concentração (C).

Sendo: $m = n^{\circ}$ de diferentes respostas (as vezes chamado de k); $N = n^{\circ}$ de respondentes

Resposta de um único respondente:

$$\vec{R}_k = (y_{k1}, \dots, y_{ki}, \dots, y_{km}),$$

Onde $k = 1, \dots, N$ respondentes diferentes.

$y_{ki} = 1$ se a resposta $i^{\text{enésima}}$ for selecionada ou $y_{ki} = 0$ caso não seja selecionada.

Resposta para todos os respondentes:

$\vec{R} = \sum_{k=1}^N \vec{R}_k = (n_1, n_2, \dots, n_m)$, fornece a informação da concentração.

n_i é o número de respondentes que selecionaram a resposta/alternativa i .

$N = \sum_i^m n_i$ sendo o número total de respostas, ou seja, considerando as alternativas de i a m .

$S = \sum n_i$, onde a alternativa i é a correta, ou seja, representa o escore para o item ou a soma dos acertos do item.

Para o padrão tipo III (máximo), um pico:

$$|\vec{R}| = N$$

Para o padrão tipo I (mínimo), sem pico:

$$|\vec{R}| = \sqrt{\left(\frac{N}{m}\right)^2 \times m} = \frac{N}{\sqrt{m}}$$

Pode-se considerar que todas as concentrações estão no seguinte intervalo ente os padrões mínimo e máximo:

$$\frac{N}{\sqrt{m}} \leq |\vec{R}| \leq N$$

A medida da concentração pode ser construída subtraindo o comprimento mínimo e renormalizando. Com isso define-se r como comprimento de escala de \vec{R} :

$$r = \frac{\sqrt{\sum_{i=1}^m n_i^2}}{N}$$

Onde:

$$\frac{1}{\sqrt{m}} \leq r \leq 1$$

Onde, para $r = 1$ (tipo III, máximo, com 1 pico), tem-se: $\sqrt{\sum_{i=1}^m n_i^2} = N$

E para $r = \frac{1}{\sqrt{m}}$ (tipo I, mínimo, sem pico), tem-se: $\sqrt{\sum_{i=1}^m n_i^2} = \frac{N}{\sqrt{m}}$

Obtém-se o fator de concentração (C) subtraindo o comprimento mínimo de r e renormalizando:

$$C = \frac{\sqrt{m}}{\sqrt{m} - 1} \times \left(r - \frac{1}{\sqrt{m}} \right) = \frac{\sqrt{m}}{\sqrt{m} - 1} \times \left(\frac{\sqrt{\sum_{i=1}^m n_i^2}}{N} - \frac{1}{\sqrt{m}} \right)$$

Comentário 1: Esta relação matemática do fator de concentração no trabalho de Lei Bao e Redish se refere a (7) e inclui equivocadamente dentro da raiz quadrada o denominador, contrariando a relação do comprimento de escala determinado na equação (5).

Comentário 2: Bao e Redish afirmam que para a concentração mínima $C_{\min}(S)$ entre os respondentes que acertaram, tem-se: $\sum_{i=1}^m n_i^2 = (m - 1) \cdot \left[\frac{N-S}{(m-1)} \right]^2 + S^2$, na qual $N - S$ representa a quantidade de respostas distribuída entre os distratores $m - 1$. E para a concentração máxima $C_{\max}(S)$ entre os respondentes que acertaram, tem-se:

$\sum_{i=1}^m n_i^2 = (N - S)^2 + S^2$. Essas seriam responsáveis pelas curvas mínimas e máximas, respectivamente sendo as parábolas da função do gráfico S-C *plot* nessas condições.

Na verificação da concentração mínima, se admitiria:

$$\frac{\sqrt{\sum_{i=1}^m n_i^2}}{N} = \frac{1}{\sqrt{m}}, \text{ com isso, } C \text{ seria zero, com distribuição uniforme entre as alternativas.}$$

Para a concentração máxima:

$$\frac{\sqrt{\sum_{i=1}^m n_i^2}}{N} = 1, \text{ com isso, } C \text{ seria 1, com um pico concentrado em uma das alternativas.}$$

Fonte: Elaborado a partir de Bao & Redish (2001).

Também se sabe que há diferentes critérios, porém muito próximos, para determinar a qualidade ou classificação do item quanto aos níveis de acerto. Apesar disso, os autores propuseram as seguintes classificações no padrão de respostas:

Tabela 4: Esquema de codificação de três níveis para pontuação e fator de concentração.

Escore	Nível	Concentração (C)	Nível
0,0~0,4	Baixo (B)	0,0~0,2	Baixo (B)
0,4~0,7	Médio (M)	0,2~0,5	Médio (M)
0,7~1,0	Alto (A)	0,5~1,0	Alto (A)

Fonte: Elaborado a partir de Bao & Redish (2001).

Combinando o escores com o fator de concentração (C), tem-se:

Tabela 5: Padrões típicos de resposta ao usar o sistema de codificação de três níveis.

Discrepância	Picos	Tipo	Implicações dos padrões
Nula	1	AA	Um modelo correto
Alta	1	BA	Um modelo incorreto
Média	2	BM	Dois possíveis modelos incorretos
Média	2	MM	Dois modelos populares (correto e incorreto)
Sem leitura	0	BB	Situação quase aleatória

Fonte: Elaborado a partir de Bao & Redish (2001).

De acordo com o exemplo sugerido por Bao e Redish ($N = 100; m = 5$), considera-se uma questão de múltipla-escolha com resposta única. Para um C_{min} se tem $(N - S)$ respostas a serem distribuídas nas alternativas restantes, sendo com isso mais próximas de uma distribuição uniforme:

$$C_{mín} = \frac{\sqrt{5}}{\sqrt{5} - 1} \times \left(\frac{\sqrt{4 \left(\frac{N-S}{4} \right) + S^2}}{N} - \frac{1}{\sqrt{5}} \right)$$

E para o $C_{máx}$ tem-se todas as respostas concentradas em uma das opções:

$$C_{máx} = \frac{\sqrt{5}}{\sqrt{5} - 1} \times \left(\frac{\sqrt{(N-S)^2 + S^2}}{N} - \frac{1}{\sqrt{5}} \right)$$

O gráfico *S-C plot* sugerido por Bao e Redish possibilita uma análise de 6 regiões de padrões de resposta, conforme a combinação entre os três níveis de escore e tipos de concentração. Eles destacam que:

Os padrões de resposta não só fornecem uma medida do desempenho dos alunos, mas também indicam se a questão desencadeia um "equivoco" comum. Além disso, o padrão da mudança de pré para pós-instrução informa como o "estado" de uma classe evolui com a instrução. (Bao & Redish, 2001, p. 6, *tradução livre*)

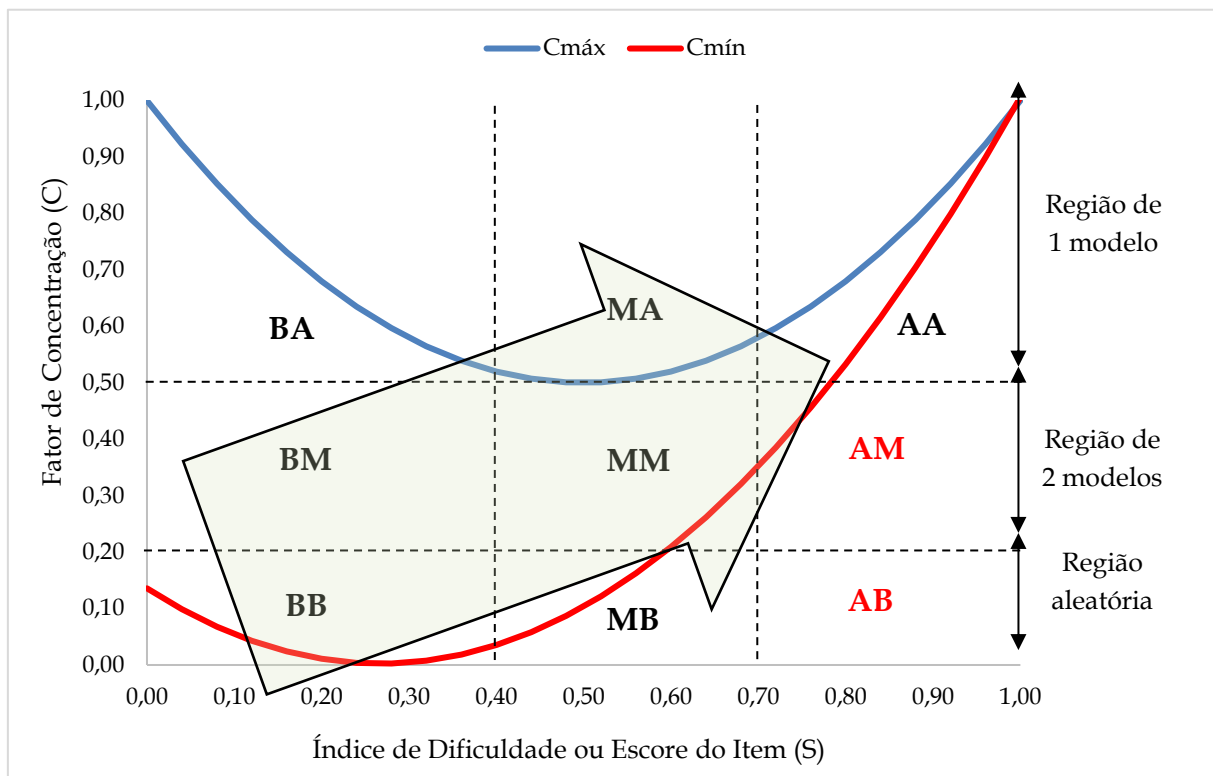


Figura 22: *S-C plot*.

As linhas de corte preestabelecidos para as regiões seguem critérios dos autores, porém, pode-se levar em consideração as delimitações de corte para as classificações

dos diferentes níveis de dificuldade dos itens, assim como o critério de “*guessing*” do item, de acordo com a quantidade de alternativas que ele possui.

Pode-se perceber que o gráfico *S-C plot* indica da esquerda para a direita o índice de porcentagem na alternativa correta, e de baixo para cima, se essa alternativa correta está concentrada ou mais dispersa entre os distratores, dessa forma, o canto superior a direita indica uma situação ideal, na qual os respondentes demonstram na sua grande maioria domínio sobre o conhecimento envolvido ou predominância de um modelo mental representativo, enquanto o canto inferior à esquerda uma baixa porcentagem de acertos e baixa concentração. É natural que à medida em que a porcentagem de acerto se eleva também aumenta o *fator de concentração*. Desse modo, algumas regiões do gráfico acabam se tornando pouco prováveis de existirem incidências de respostas, como no caso da AB e AM, sendo AB uma zona definitivamente não considerada por não conter a zona intermediária entre as curvas mínima e máxima. A região BA é uma situação típica, de predominância de um único modelo mental incorreto, apesar disso, entende-se que um teste bem projetado deve concentrar as regiões de 1 e 2 modelos, com uma tendência favorável aos modelos corretos, no caso, deslocando itens de um pré-teste para um pós-teste conforme indica a seta no gráfico.

Para pesquisas que envolvem testes conceituais ou inventários de conceitos, considera-se que o fator de concentração poderá ser útil em dois aspectos:

- Uma análise de concentração pode ajudar a confirmar a presença (e o nível) de modelos corretos e errados detectados através da pesquisa, ou seja, itens que correspondam aos perfis de proficiência no que diz respeito ao domínio e aos equívocos dos alunos, ajudando a identificar possíveis problemas nos itens, ou resistências a mudanças conceituais específicas diante das concepções e conceitos envolvidos;
- O contraste dos itens entre grupos que não receberam a instrução e os que receberam a instrução poderá sinalizar evidências sobre a eficácia dos eventuais tratamentos ou modelos de ensino desenvolvidos nos alunos, bem como uma caracterização mais

detalhada para explicar e diferenciar as perspectivas e interpretações mais gerais para cada contexto.

2.3.2.5 *Desvio na concentração das análises «Γ»*

Se refere à *concentração das respostas incorretas*. Uma função análoga a anterior pode ser obtida ao remover o *deslocamento absoluto* S^2 criado pela pontuação (Bao & Redish, 2001, p. 9), e com isso, ao se desconsiderar a alternativa correta, se diminui obviamente uma das alternativas ($m - 1$) na função, que resulta da seguinte forma:

$$\Gamma = \frac{\sqrt{m-1}}{\sqrt{m-1}-1} \times \left(\frac{\sqrt{\sum_{i=1}^m n_i^2 - S^2}}{(N-S)} - \frac{1}{\sqrt{m-1}} \right)$$

A remoção do deslocamento absoluto de acerto torna o *desvio na concentração* Γ independente do *percentual de acerto* S , com isso, a construção um gráfico S - Γ possibilita explorar os detalhes das respostas incorretas. Uma vez que se tem duas variáveis independentes como os eixos, não há restrição na área de traçado (Ibid., 2001).

Considera-se que os Γ 's elevados indicam distrações fortes, pode-se inferir que os *itens de baixo desempenho* são dominados por situações em que as respostas dos alunos possuem *fortes modelos alternativos*. Sabe-se que o termo «distrator» geralmente é associado a ideia de distração ou aquilo que serve para distrair causando falta de atenção, assim pode ser considerado por vezes uma *resposta plausível*, mas incorreta, apresentada como alternativa em testes com respostas de escolha múltipla. No caso, esses modelos mentais alternativos apresentados entre as alternativas incorretas em testes conceituais geralmente estão associados a concepções alternativas relativamente consistentes, mas também podendo ser considerados de fato absurdos, dependendo do contexto. Nas análises de desempenho de um pós-teste, ou seja, após as instruções do ensino, pode-se dizer que os alunos que dão respostas incorretas ainda são fortemente afetados por determinadas forças distratoras desses itens.

2.3.2.5.1 Análise dos distratores dos itens

Para verificar diferenças significativas entre os distratores, é sugerida a realização da prova do qui-quadrado χ^2 (necessitando a verificação na tabela do teste), na qual considera-se apenas os distratores, excluindo assim a alternativa correta (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, p. 61):

$$\chi^2 = \sum_{i=1}^k \frac{(FT - FO)^2}{FT}$$

Sendo:

FT : as frequências teóricas;

FO : as frequências observadas;

k : o número de alternativas, sendo o número de graus de liberdade com $(k - 1)$.

Uma correção sobre os efeitos do “azar” ou acertos aleatórios em testes formados por itens com opções de múltiplas escolhas consiste em estabelecer uma eliminação do número total de acertos aleatórios (A_a) dentro do total de acertos (A) considerados. Sabe-se que, quando não se conhece a resposta correta de um item, todas as alternativas são equiprováveis, ou seja, todas as alternativas possuem a mesma probabilidade de acerto ao acaso (Abad, Garrido, Olea, & Ponsoda, 2006, pp. 18-20):

$$P(A_a) = \frac{1}{k}$$

, sendo k o número de alternativas do item.

Da mesma forma, a *probabilidade de errar* do item será:

$$P(E) = 1 - \frac{1}{k} = \frac{k - 1}{k}$$

Considerando que R_a seja o *número de respostas aleatórias* (o número de itens que foi respondido sem saber a solução). Das R_a , alguns serão acertos aleatórios (A_a) e outras são erros (E). Estimando o acerto aleatório (A_a) é possível descontá-los do número total de acertos que a pessoa realmente teve, para isso, tem-se que o *número total de erros* (E) pode ser estabelecido como o produto do valor R_a pela probabilidade de cometer um erro $P(E)$:

$$E = R_a \cdot P(E) = R_a \cdot \frac{k - 1}{k}$$

Com isso, ao conhecer os valores de k e E , torna-se possível determinar o

número de respostas aleatórias R_a :

$$R_a = E \cdot \frac{k}{k-1}$$

Usando o mesmo procedimento, é possível estimar o número de acertos aleatórios (A_a) em função do número de erros (E), bastando multiplicar o número de respostas aleatórias R_a pela probabilidade de acertos aleatórios $P(A_a)$:

$$A_a = R_a \cdot P(A_a) = E \cdot \frac{k}{k-1} \cdot \frac{1}{k} = E \cdot \frac{1}{k-1}$$

Dessa forma, para cada acerto aleatório o erro deve ser ponderado pelo fator $(\frac{1}{k-1})$, ou seja, em testes com 2 alternativas desconta-se 1 ponto por cada erro, para 3 alternativas desconta-se 0,5 ponto por erro, **com 4 alternativas seria 0,33 pontos**, com **5 alternativas 0,2 pontos**, e assim sucessivamente.

Tomando emprestado um exemplo dado que ilustra bem esses aspectos (Abad, Garrido, Olea, & Ponsoda, 2006, p. 20), tem-se: Um teste de conhecimento da língua inglesa consiste de 140 itens ($n = 140$) com 5 opções de resposta cada ($k = 5$). Abaixo estão o número de sucessos (A), erros (E) e omissões (O) que receberam 3 pessoas:

Tabela 6: Três exemplos de valores de acertos (A), erros (E) e omissões (O).

Respondentes	A	E	O
1	112	28	0
2	110	12	18
3	109	0	31

Observando o número total de acertos brutos obtidos (A) entre todos os n itens, podendo também ser chamado de *score do sujeito* (j) no teste (X_j), pode-se dizer em ordem decrescente que os respondentes que “sabem mais” são supostamente aqueles com maior desempenho, obviamente, que no caso seria: 1 ($A_1 = X_1 = 112$), depois 2 ($A_2 = X_2 = 110$), em seguida o 3 ($A_3 = X_3 = 109$). Entretanto, ao corrigir os acertos ou os escores dos sujeitos ($X_{corrigido}$) retirando o número de acertos aleatórios estimados (A_a), tem-se:

$$X_{corrigido} = A - A_a$$

Assim, são obtidos os seguintes escores dos sujeitos:

- $X_{\text{corrigido}_1} = 112 - \frac{28}{4} = 105;$
- $X_{\text{corrigido}_2} = 110 - \frac{12}{4} = 107;$
- $X_{\text{corrigido}_3} = 109 - \frac{0}{4} = 109.$

Pode-se perceber que a ordem dos respondentes que supostamente dominam mais o inglês se altera sensivelmente, no caso, o que “sabe mais” inglês seria em ordem decrescente o respondente: 3 ($X_{\text{corrigido}_3} = 109$), seguido de 2 ($X_{\text{corrigido}_2} = 107$), e por fim o respondente 1 ($X_{\text{corrigido}_1} = 105$). Observa-se que a alteração do sujeito 3 se dá pelo fato de que ele não cometeu nenhum erro, ao contrário do sujeito 1, com isso, deve-se retirar as pontuações estimadas que estão contidas nos erros empíricos observados.

2.3.3 Teoria da Resposta ao Item

“Tornar-se-á possível associar às análises profundas de Edgar Morin (1977-1991) e de Jean-Lois Le Moigne (1990) a instrumentos conceituais quantitativos de precisão, que não lhes amputarão o sentido.” (Mugur-Schächter, 2007, p. 150, *In*: Morin & Moigne, 2007)

Segundo Andrade *et al* (2000), a TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar certa resposta a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente. Essa relação é sempre expressa de tal forma que quanto maior a habilidade, maior a probabilidade de acerto no item.

Existem diversos modelos presentes na literatura, que podem ser subdivididos em: natureza do item – dicotômicos ou politômicos; número de população; quantidade de traços latentes medidos (Andrade et al., 2000).

Dos modelos existentes, a análise da compreensão conceptual dos alunos será baseada no *Modelo Dicotômico*, onde assumem na resposta dos alunos apenas duas possibilidades, de estar certo ou errado.

Esse modelo pode ser com análise de um, dois, três, quatro ou cinco parâmetros logísticos. O modelo de cinco parâmetros acrescenta o parâmetro ou fator de assimetria da sigmoide (φ), na qual surgem discussões sobre vantagens e desvantagens em relação ao modelo de 4PL (Cumberland et al., 2015; Gottschalk &

Dunn, 2005), porém não se apresenta de forma difundida na literatura no campo da avaliação educacional, apesar de ser um modelo bem conhecido no campo médico, com aplicações em diversos estudos (Ricketts & Head, 1999).

Quadro 35: Funções de 1, 2, 3, 4 e 5PL da TRI.

Parâmetros logísticos	Função de Probabilidade de Acerto do Item
1PL	$P(x = 1 \theta, b) = \frac{1}{1 + \exp(-D(\theta + b))}$
2PL	$P(x = 1 \theta, a, b) = \frac{1}{1 + \exp(-Da(\theta + b))}$
3PL	$P(x = 1 \theta, a, b, g) = g + \frac{(1 - g)}{1 + \exp(-Da(\theta + b))}$
4PL	$P(x = 1 \theta, a, b, g, u) = g + \frac{(u - g)}{1 + \exp(-Da(\theta + b))}$
5PL	$P(x = 1 \theta, a, b, g, u, \varphi) = g + \frac{(u - g)}{[1 + \exp(-Da(\theta + b))]^\varphi}$

No modelo Rasch especifica-se com $D = 1$, sendo P a probabilidade de acerto ao item como função da habilidade do sujeito θ , parâmetro de dificuldade do item b , na assíntota inferior sendo o acerto aleatório ou “guessing”(g), na assíntota superior o erro aleatório ou “upper”, e o fator de assimetria da curva (φ), sendo o D conhecido como o fator de normalização, e assume valor igual a 1,702 para a função logística⁴⁸.

Para a realização dos cálculos estatísticos, utilizam-se diversos softwares de estimação e calibração. Dentre os diversos existentes podemos destacar o BILOG-MG e PARSALE, SAS, Mplus, entre outros, porém vários desses são todos pagos e muito caros, recorre-se na utilização de softwares livres e fiáveis como o *IRT Command Language - ICL*⁴⁹, desenvolvido por Bradley A. Hanson no ano de 2002, e um dos ambientes estatísticos mais utilizados nesse âmbito denomina-se R. Nos trabalhos de trabalhos de Jurich & Goodman (2009) e Mead *et al* (2007) apresentam a comparação do software *ICL* com o *BILOG-MG* e *PARSALE* na recuperação dos parâmetros e traços latentes em estudos, em ambos os resultados foram semelhantes. Segundo Mead *et al* (2007) a principal desvantagem do *ICL* é o fato de não apresentar interface gráfica, já o

⁴⁸ O conhecimento estatístico necessário nessa análise pedagógica será aprofundado e enfatizado no desenvolvimento da tese.

⁴⁹ Disponível em: <<http://www.b-a-h.com/software/irt/icl/>> acessado em 05/09/2012>.

R pode ser mediado com interface mais amigável através do *R stúdio*.

2.3.3.1 Fundamentos da TRI

A *Teoria da Resposta ao Item* (TRI) surgiu na década de 60 do século passado, visando atender discrepâncias nos resultados de testes de inteligência, que variavam conforme o instrumento de medida utilizado (Rabelo, 2013). A solução dada adquiriu uma independência do instrumento de medida em relação ao objeto que se deseja medir estava em utilizar modelos e algoritmos complexos, na qual se tornou viável e mais propício a difundir a partir da década de 80 com o desenvolvimento de *softwares*.

Frederic Lord foi o responsável direto pela introdução a TRI em 1952 utilizando um modelo teórico, possibilitando a estimação de parâmetros de itens, aplicados a respostas dicotômicas, binárias, do tipo *certo ou errado*⁵⁰. Porém em meados de 1969/1970, Samejima generalizou a teoria incluindo modelos politômicos (respostas graduais, p.ex., Escala *Likert*) e atendendo variáveis contínuas.

Muitas limitações advindas da *Teoria Clássica dos Testes* (TCT) foram sanadas pela TRI, nomeadamente quanto a discriminação dos itens, fidedignidade dos testes e comparabilidade dos desempenhos a partir de testes diferentes.

No Brasil, o ENEM de 1998 a 2008 era baseado na TCT, e algumas pessoas tentavam realizar inferências comparativas de modo desprovido de fundamento, o que não poderia ser feito, pois cada avaliação somente servia para comparações pontuais do mesmo teste, e não para testes diferentes e anos anteriores. O SAEB até 1993 utilizou a TCT, mas a partir de 1995 introduziu a TRI para a construção de instrumentos, atribuição de escores e a análise, sendo complementares entre si. Uma comparação simples e direta da TCT e TRI pode ser observada:

⁵⁰ Na *matriz de respostas* é atribuído o número “1” (certo) para o item assinalado corretamente e “0” (errado) quando não corresponder à chave-de-correção. A matriz de respostas é base de dados principal a ser manipulada para a realização das análises nos programas computacionais estatísticos.

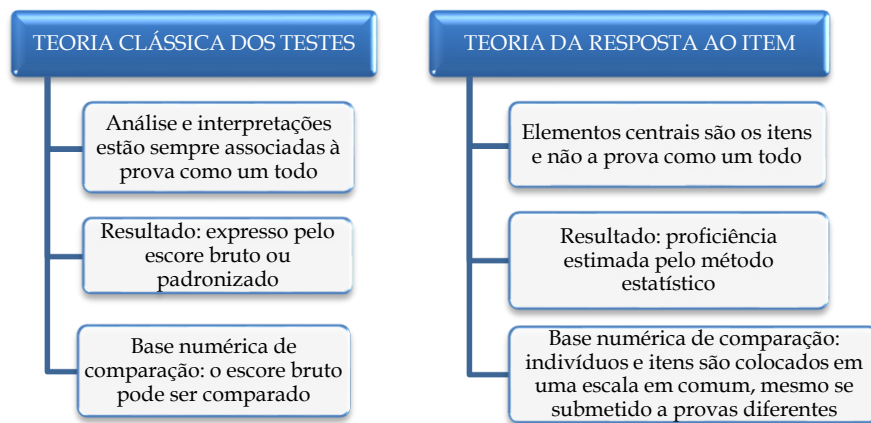


Figura 23: Comparação entre TCT e TRI.

Fonte: Rabelo (2013).

No sistema educacional brasileiro, a TRI foi sendo utilizada gradativamente nos exames, depois do SAEB foi o ENCCEJA, PROVA BRASIL, ENEM e ANA. Em 2009 o ENEM adotou a TRI, com a perspectiva de análise em uma escala de proficiência e assim possibilitou inferências comparativas através das *séries históricas de desempenho*. Atualmente o SAEB passa por um processo de reformulação sendo denominado de *Sistema Nacional de Avaliação da Educação Básica (SINAEB)*, prevista anteriormente sua implementação (Rabelo, 2013, p. 67), e já apresentada oficialmente como proposta de adequação ao cumprimento do *Plano Nacional de Educação* (Santos, Neto, & Junqueira, 2017), que mantém incluída a recente implementação da *Avaliação Nacional da Educação Infantil (ANEI)*, de 2016.

Segundo as notas técnicas do INEP (Brasil, 2012b, p. 1), a inclusão da TRI no ENEM possui duas finalidades principais: (1) *permitir a comparabilidade dos resultados entre os anos* e (2) *permitir a aplicação do Exame várias vezes ao ano*. Entre alguns países e exames que utilizam a TRI como mecanismo de análise tem os (a): Estados Unidos, França, Holanda, Coreia do Sul, China, e entre os *exames conhecidos*⁵¹ e semelhantes ao ENEM, tem: os internacionais PISA e TOEFL; os norte-americanos NAEP, ETS, GMAT e SAT; na Holanda o CITO, entre outros, como foi dito anteriormente.

⁵¹ *Programme for International Student Assessment; Test of English as a Foreign Language; The National Assessment of Educational Progress; Graduate Management Admission Test; Scholastic Aptitude Test ou Scholastic Assessment Test.*

Apesar de comumente se ouvir falar em “medir o conhecimento”, dentro da Psicometria utiliza-se o termo “**medir a proficiência**”, que em termos teóricos pode ser de $-\infty$ a $+\infty$. Devido o foco maior da TCT estar no *teste em questão* e não nos itens, ela possui limites quanto à realização de comparações entre desempenhos de alunos submetidos a diferentes provas. A TRI por sua vez, complementa as análises superando essas dificuldades tendo como foco o *item em questão*. Enquanto na TCT a proficiência é medida pela soma dos acertos (escore), na TRI, para se obter a “nota do respondente” é preciso *estimar* (encontrar) o valor do parâmetro “teta” (θ), que representa a nota do participante, levando em consideração o vetor de respostas dele e os parâmetros dos itens que compõem a prova, sendo adotado o modelo 3PL. Segundo a nota técnica do INEP:

“Dentro do contexto da TRI, a medida de proficiência de um aluno não depende dos itens apresentados a ele e os parâmetros de discriminação e de dificuldade do item não dependem do grupo de respondentes. Em outras palavras, um item mede determinado conhecimento, independentemente de quem o está respondendo, e a proficiência de um aluno não depende dos itens que estão sendo apresentados a ele”. (Brasil, 2012b, p. 2)

2.3.3.2 Vantagens e desvantagens

Uma das limitações da TRI está em identificar uma única habilidade por item, na qual se tem a compreensão de que o aluno na realidade utiliza outras habilidades para resolver problemas. No entanto, estudos mostram que suas vantagens superam suas limitações. Entre algumas das vantagens e desvantagens, tem-se:

Quadro 36: Vantagens e desvantagens entre a TCT e a TRI.

Vantagens	Desvantagens
<ul style="list-style-type: none"> • A estatísticas de itens são independentes da amostra a partir do qual eles foram estimados. • A pontuações examinadas são independentes da dificuldade do teste. • A análise de itens acomoda itens do teste correspondentes ao nível de conhecimento examinando. • A análise do teste não exige testes paralelos rigorosos para avaliar a confiabilidade. • As estatísticas de itens e capacidade nos exames são relatadas na mesma escala. 	<ul style="list-style-type: none"> • Os pressupostos subjacentes ao uso de modelos da TRI são mais rigorosos do que as exigidas da TCT. • Os modelos da TRI também tendem a serem mais complexos e os modelos-saídas mais difíceis de entender, principalmente com o público não-técnicamente orientado. • Os modelos da TRI exigem grandes amostras para obter estimativas de parâmetros precisos e estáveis, embora os modelos de Rasch⁵² de medição são úteis com pequenas amostras de moderados.

⁵² O modelo de Rasch mede apenas o parâmetro dificuldade do item (b), considerando o parâmetro discriminação constante e igual a 1 ($a = 1$), e o parâmetro de acerto aleatório nulo ($c = 0$).

Um dos aspectos interessantes na TRI é a sua capacidade em separar as características dos itens com os traços latentes (habilidades) dos alunos. Pasquali (2009) apresenta os axiomas que a fundamentam a TRI:

- O desempenho do sujeito numa tarefa (item do teste) se explica em função de um conjunto de fatores ou traços latentes (aptidões, habilidades etc.). O desempenho é o efeito e os traços latentes são a causa;
- Sujeitos com aptidão maior terão maior probabilidade de responder corretamente ao item e vice-versa.

Sabe-se que o modelo logístico da TRI parte do princípio de que quanto maior a *proficiência do respondente*, maior a sua probabilidade de acerto e o traço latente acumulativo, falando de modo mais apropriado seria dizer que ele representa a proficiência mínima que um respondente deve possuir para que sua probabilidade de acerto seja alta, dessa forma ele poderia ser chamado de “proficiência do item”. O seu *parâmetro de dificuldade* é medido na mesma escala da *proficiência*, fato este que permite a comparabilidade entre os resultados de diferentes testes e a construção/interpretação de escalas de proficiência, como a escala nacional de proficiência do SAEB construída pelo INEP/MEC, tendo a Matemática e Língua Portuguesa como as componentes disciplinares avaliadas dentro do currículo.

2.3.4 Escalas de Proficiência: Comentários e Fundamentos

A OCDE tem o Brasil como parceiro-chave e parte integrante da comunidade global no compartilhamento de experiências, intercâmbio das melhores práticas no desenvolvimento da educação e nas **políticas educacionais baseadas em competências** (OCDE, 2015, p. 26). A exemplo disso, tem-se como meta alcançar a média de desempenho no PISA ao dos países da OCDE até 2021, entretanto, a elevação de investimentos não tem demonstrado correspondência nas melhorias dos indicadores de desempenho do PISA que permanecem inalterados desde 2006, e comparativamente, tanto os indicadores de desempenho quanto os investimentos ainda encontram abaixo da média dos países da OCDE. Semelhantemente é possível constatar essa estagnação nos indicadores de desempenho do *Sistema de Avaliação da*

Educação Básica (SAEB) para os alunos do *ensino médio*⁵³, sendo esses indicadores baseados na concepção de *proficiências relacionadas às competências necessárias* a serem desenvolvidas pelos alunos, precisamente para as disciplinas de matemática/álgebra e língua portuguesa/leitura, estando ausente os conhecimentos em ciências naturais (Física, Química, Biologia, Geologia). Os exames PISA, SAEB, TOEFL (entre outros) utilizam **escalas de proficiência** como instrumento avaliativo de desempenho e competência, apresentando resultados diagnósticos que caracterizam «evidências» da qualidade da educação, como também problemas relacionados à equidade no sistema educacional, entretanto, geralmente lhe são atribuídas funções certificadoras, classificatórias e de *ranking*.

O uso de **escalas de proficiência** como instrumento avaliativo se tornou comum na década de 90 como parte de um movimento geral no sentido de dar mais transparência aos sistemas educacionais (North, 2000, p. 9). No Brasil, por exemplo, tem-se a escala do SAEB construída em 1995, e no contexto internacional a escala do PISA em 2000, entre outros exames para avaliação de proficiência. Em especial aqueles voltados para linguística, como o *British National Language Standards* na Inglaterra em 1992, *Finnish Scale of Language Proficiency* na Finlândia em 1993, *Eurocentres Scale of Language Proficiency* em 1993 e o teste *Association of Language Testers in Europe* em 1994 na Europa (Ibid., 2000, p. 9), sendo o *Test of English as a Foreign Language* (TOEFL) o mais conhecido, elaborado em 1964.

Muitas destas escalas representam o que Bachman (1990: 325-330) descreveu como a **abordagem "da vida real" para avaliação em que eles tentam dar uma imagem do que um aprendiz em um nível especial de realização pode fazer no mundo real**. Para outras escalas concebidas por Bachman, ele descreve como a abordagem "capacidade-interativa" focando sobre os aspectos de um desempenho teste de articulação (por exemplo Milanovic et al 1992/6; Fulcher 1993; Upshur e Turner, 1995; Brindley 1998). (North, 2000, p. 9, *tradução livre, grifo nosso*)

Entre alguns termos comuns e definições relacionadas as **escalas de proficiência**, tem-se:

"[...] *band scores, band scales, profile bands, proficiency levels, proficiency scales, proficiency ratings*" (Alderson 1991a: 71) ou "*guidelines, standards, levels, yardsticks, stages, scales, or grades*" (DeJong 1992: 43). O que todos eles têm em comum é que tentam fornecer

⁵³ Ensino médio e ensino secundário serão considerados termos equivalentes no texto.

"**uma série ascendente de níveis de competência linguística**" (Page em North et al. 1992: 7) ou "**uma hierarquia caracterizações de desempenho global de soluções integradas**" (ACTFL 1986), "**a sequência hierárquica de faixas de desempenho**" (Galloway, 1987: 27) ou "**perfis característicos dos tipos e níveis de desempenho que pode ser esperado de alunos em representante diferentes fases**" (Apare 1978: 6). (North, 2000, p. 9, *tradução livre, grifo nosso*)

Entre os atrativos e contribuições que advêm do uso de escala de proficiências, pode-se destacar:

Quadro 37: Potencialidades no uso de Escalas de Proficiência.

Potencialidades no uso de Escalas de Proficiência
<ul style="list-style-type: none">• Fornecer um "estereótipo" de modo que o aluno possa comparar a sua autoimagem e avaliar aproximadamente sua posição.
<ul style="list-style-type: none">• Estabelecer um quadro de referência que pode descrever a realização em um sistema complexo em termos significativos para todos os diferentes parceiros ou usuários desse sistema de uma forma que dezenas de itens de teste não pode (Trim, 1978; Brindley, 1986, 1991; Richterich & Schneider 1992).
<ul style="list-style-type: none">• Estabelecer metas de aprendizes e descrições de proficiência em níveis nominais a fim de proporcionar alvos para os alunos, para permitir que os resultados alcançados sejam mensurados em relação aos resultados esperados e para fornecer à sociedade um meio pragmático de colocar estudantes em ambientes de aprendizagem ou de trabalho adequados referindo-se ao perfil de um indivíduo nas subescalas do sistema (Clark, 1985).
<ul style="list-style-type: none">• Fornecer coerência interna nas ligações dentro de um sistema no intervalo de um pré-curso ou nas verificações de entrada, planejamento de currículo, organização de materiais, avaliação de progresso e de saída e de certificação (North, 1991).
<ul style="list-style-type: none">• Fornecer evidência de progresso (desde que as etapas são pequenas o suficiente) e, assim, ajudar a aumentar a motivação (Liskin-Gasparro 1984a; Page, 1992; North, 1992a).
<ul style="list-style-type: none">• Aumentar a confiabilidade das classificações subjetivamente julgadas, especialmente das habilidades linguísticas produtivas, além fornecer um padrão e significado comum para tais julgamentos (Alderson 1991a)
<ul style="list-style-type: none">• Relatar os resultados das avaliações dos professores, dos testes marcados, testes avaliados e de autoavaliação em termos do mesmo instrumento e evitar a sugestão espúria de precisão dada por uma escala marcada (por exemplo, 1 a 1 000) (Alderson, 1991a; Griffin, 1989).
<ul style="list-style-type: none">• Fornecer estágios de desempenho e notas que refletem o currículo da sala de aula, mas que podem ser traduzidos em uma declaração de proficiência e grau em um quadro comum (Trim, 1978; Ingram & Wylie, 1989; Hargreaves, 1992).
<ul style="list-style-type: none">• Permitir a comparação entre sistemas ou populações usando uma métrica comum ou padrão (Trim, 1978; Lowe, 1983; Liskin-Gasparro, 1984b; Bachman & Savignon, 1986, Carroll B.J. & West, 1989).

Fonte: North (2000, p. 12, *tradução livre*)

John Clark define **escala de proficiência** apresentando sua principal fraqueza: "*Descrições dos resultados esperados, ou gravuras impressionistas de que a proficiência poderia parecer como se movesse através de pontos hipotéticos ou níveis em um desenvolvimento contínuo* (Clark 1985: 348)", e Brindley ressalta: "*as descrições generalizadas dos níveis que figuram nas escalas de classificação representam uma simplificação excessiva e possivelmente enganosa do processo de aprendizagem de línguas* (Brindley 1998: 22)" (North, 2000, *tradução livre*).

A escolha das **escalas de proficiência** como procedimentos de análise no estudo busca alinhamento com a concepção de *Educação Baseada em Evidências* (EBE). Uma das justificativas está no fato da EBE estar relacionada com a qualidade da informação que são utilizadas para a «tomada de decisões» desde a prática nas salas de aulas quanto de providências a serem tomadas no sistema educacional. Essas evidências se configuram tanto em «constatações científicas» empíricas que trazem implicações diretas sobre a área da educação, como resultados de estudos da psicologia e neurociência, assim como na identificação das melhores práticas pedagógicas que são reforçadas através de diferentes pesquisas e contextos, tanto qualitativas, como quantitativas e/ou mistas. A concepção de EBE não é um modismo ou inovação, pois já vem sendo utilizada há algumas décadas e proporcionado grandes avanços na área da medicina, seja na prática clínica, cirúrgica, como na produção de medicamentos (Araujo e Oliveira, 2013). Com o avanço tecnológico, dos novos processos metodológicos na pesquisa e da produção de computadores modernos e mais eficientes tornou-se possível o desenvolvimento de processos de análises mais robustos (Ibid., 2013), e esse conceito vem apresentando reflexos na educação desde a década de 80, através de estudos de larga escala. Entre alguns modelos estatísticos que se tornou viável para análises estão aqueles relacionados a *Teoria da Resposta ao Item* (TRI), entre outros dentro da Psicometria que são amplamente reconhecidos e utilizados em todo o mundo, e em diversos segmentos.

As «evidências» na área da educação devem ser caracterizadas pela consistência e reforço observado em diversas circunstâncias e contextos, na qual muitas pesquisas sobre temas específicos vêm sendo abundantemente publicadas, e dentro disso, essas informações trazem argumentos e condições que devem sustentar as «tomadas de decisões» sobre os investimentos públicos na área da educação. Alguns países vem adotando mudanças nas práticas pedagógicas e no modelo de sistema educacional utilizando parâmetros de evidências, como no caso dos resultados **diagnósticos** de desempenho de estudantes em matemática, leitura e conhecimento científico abordados no PISA, sendo um exame considerado «equivalente» (não exatamente igual, mas com itens correspondentes calibrados com características psicométricas

equivalentes) ministrado em diferentes países que através de rigorosos pressupostos e critérios estatísticos possibilitam a comparação entre diferentes populações, contextos e países, além de contribuir no processo de «engenharia de construção de itens» e alimentar repositórios de banco de dados de itens.

Recentemente foi publicado na revista *The Guardian* (Weale, 2017), um estudo de neurocientistas e psicólogos de Harvard e Oxford na qual afirmam ser considerado "mito" acreditar no sucesso da aprendizagem por meio de um ensino baseado em «estilos de aprendizagem» individual, ou até mesmo classificar um estilo como sendo superior a outro, pois o segredo do sucesso da aprendizagem está no cérebro. Com isso ressaltam que as melhores práticas educacionais devam se «basear em evidências».

Ensinar as crianças de acordo com seu 'estilo de aprendizagem' individual não obtém melhores resultados e deve ser abandonado pelas escolas a favor da prática baseada em evidências, de acordo com cientistas líderes". (Weale, 2017, *tradução livre*)

Nesse sentido, as **construções de escalas de proficiência** juntamente com as análises advindas de testes psicométricos sobre os conhecimentos científicos podem emergir informações *baseadas em evidências*, e com isso servir de **diagnósticos** e orientação para alternativas e práticas educacionais levando em conta diferentes níveis de dificuldades conceituais específicas e relacionadas. Essas evidências podem ser úteis para substanciar parâmetros significativos dos alunos na contextualização das abordagens dos conteúdos, em critérios prioritários dentro de metodologias de ensino, com também na elaboração de propostas de «situações-problema» na Física (entre outras disciplinas) de modo favoreçam uma aprendizagem por competências. Metodologicamente as análises psicométricas fornecem indicadores em uma «linguagem de números», em que na *interpretação pedagógica da escala de proficiência* e no *item comentado* ocorre uma transposição para uma «linguagem pedagógica» sobre o *que fazer*⁵⁴ e não o *como fazer*.

⁵⁴ Essa perspectiva metodológica segue o modelo do Projeto Devolutivas Pedagógicas do INEP, apresentado em 2015. Ver: <<http://devolutivas.inep.gov.br/proficiencia>>.

O interesse de construir escalas de proficiência a partir de testes padronizados na área da educação faz parte de uma visão que acredita em contribuições de boas práticas no uso de testes padronizados. Uma das potencialidades em defender o uso de testes se apoia na necessidade de avaliações diagnósticas que visam atender as necessidades de aprendizagem, sendo essa função avaliativa muito negligenciada por escolas e uma grande lacuna na realidade da prática de ensino, argumentos comuns, e muitas vezes pertinentes está relacionado ao “trabalho extra” que se concebe pelos professores e a “falta de tempo” diante da sobrecarga de compromissos que eles são incumbidos, e excedente de função não remunerada por vezes defendida. Esse é um problema que extrapola para dimensões de gestão e política educacional, envolvendo questões relacionadas a políticas meritocráticas e de valorização do professor, entre outras.

A utilização de escalas de proficiência como instrumento construído a partir de testes padronizados é uma possibilidade educativa prática para aferir e comparar desempenhos, enriquecendo a compreensão do professor sobre as capacidades e necessidades de indivíduos em seu progresso, possibilitando assim estabelecer metas, intervenções diferenciadas e comuns que atendam às dificuldades de aprendizagem específicas, buscando potencializar e aprofundar alunos com bom desempenho e maior facilidade de aprendizagem, mas fundamentalmente contribuir diante daqueles com maiores dificuldades. Os modelos estatísticos psicométricos como procedimentos de análise para avaliar traços latentes relacionados ao conhecimento são consideradas as mais eficientes atualmente nesse sentido, e podem de maneira rica contribuir na tomada de decisão no campo pedagógico, seja dando maior enfoque em temáticas e em modelos de ensino mais eficientes, com maior consciência, consistência e fundamentação para isso.

Acredita-se ainda que seja um diferencial e contributo para estudos que envolvem *testes conceituais de física* no contexto brasileiro, considerando que não foi encontrado nenhum trabalho até então seguindo essa perspectiva na comunidade de pesquisadores do EC. Convém lembrar que as escalas proficiência mais conhecidas no Brasil são as do *Sistema de Avaliação da Educação Básica (SAEB)* e do *Exame Nacional do*

Ensino Médio (ENEM) que determinam padrões de conteúdo e desempenho, pois exercem a função de indicadores para avaliar a qualidade do sistema educacional, além de serem indexados na elaboração do *Índice de Desenvolvimento da Educação Básica* (IDEB). Apesar disso, o SAEB não avalia o conhecimento científico, portanto, não se conhece a capacidade ou desempenho de estudantes nessa área do conhecimento, ressaltando a necessidade de instrumentos que possam contribuir nesse aspecto.

A importância de compreender essas escalas contribui nas interpretações dos resultados de avaliação externa em nível nacional e estadual, assim como nos processos de reformulação e análise de suas respectivas matrizes de referência nesses exames, que por sua vez serão reorientadas e alcançarão maior destaque após a conclusão e vigência da *Base Nacional Comum Curricular* (BNCC) prevista para o segundo semestre de 2017. Em decorrência a isso, as mudanças nos LDs são um efeito imediato dentro do PNLD, assim como na definição das competências docentes no *Exame Nacional de Avaliação do Magistério da Educação Básica* (ENAMEB) e *Prova Docente*, e para os 19 sistemas de avaliação de desempenho escolar desenvolvidos pelas Secretarias de Estado de Educação pelo país (como o SADEAM, SISPAE, SEAPE, SARESP, SAERJ, entre outros), que se fortalecem e se expandem seguindo por indução o modelo do SAEB (Machado, Alavarse, & Arcas, 2015). Pode-se dizer que se representam mecanismos de avaliação e interpretação que justificam diversas políticas públicas educacionais nacionais e regionais, em especialmente quanto a cursos de formação docente e elaboração de materiais didáticos.

Entre outras, admite-se o uso da «escala de proficiência» como um mecanismo para verificar e estabelecer níveis ou patamares de competência científica diante de conceitos/conhecimentos, podendo evidenciar parâmetros de estabilização nos níveis de dificuldades para diferentes populações, assim como em subsídios para a proposição de situações-problema com diferentes dificuldades. O processo de interpretação da escala segue de modo semelhante a do SAEB, ou seja, *escalamento comportamental* do tipo II, segundo Carrol (1993, p. 303) citado por (Primi, 2004), sendo à definição da proficiência, associada aos pontos da escala, com base na análise, por

especialistas, dos conteúdos e processos necessários à resolução de itens típicos desses pontos, chamados «itens-âncora».

Carrol (1993, p.303) classificou quatro maneiras pelas quais essas informações podem ser consolidadas em afirmações comportamentais: (a) **Tipo I**: *apresentando ilustrações de itens ou tarefas com diferentes níveis de dificuldade, exemplificando os vários níveis da escala de dificuldade/habilidade com situações concretas*; (b) **Tipo II**: *apresentando descrições verbais dos comportamentos típicos esperados ao longo dos diferentes níveis de dificuldade, essas afirmações são fundamentadas na análise das competências necessárias à resolução de certos itens, representantes dos diferentes níveis de dificuldade, os quais são denominados itens-âncoras* (Beaton & Allen, 1992); (c) **Tipo III**: *vinculando informações paramétricas que descrevem como os diferentes níveis de dificuldade das tarefas presentes no teste se relacionam com outros atributos ou universo de tarefas* (esse procedimento se assemelha à referência ao critério e à referência ao conteúdo, apresentadas anteriormente); (d) **Tipo IV**: *descrevendo o tipo ou nível de processamento cognitivo necessário para acertar itens de diferentes níveis de dificuldade*. (Primi, 2004, p. 56, grifo nosso)

2.3.4.1 Matriz de referência nos exames nacionais: fundamentações epistemológicas

“A lógica das competências perpassa de forma contraditória, mas complementar, o mundo do trabalho: no núcleo do trabalho formalizado, a apropriação pelo capital do saber, do *saber-fazer* e do *saber-ser* dos trabalhadores, implica uma busca constante de ampliação e atualização do seu portfólio de competências e uma **renúncia permanente aos seus interesses de classe em favor dos interesses empresariais**”. (Deluiz, 2001, grifo nosso)

Os interesses nem sempre estão explícitos, porém com um olhar mais cuidadoso é possível encontrar as raízes epistemológicas que regem os interesses e impõem suas forças ideológicas. Essas raízes estão supostamente relacionadas com as bases que formam um conjunto de diretrizes, como no caso das matrizes de referências utilizadas nos exames avaliativos de larga escala no Brasil e em outros contextos.

Pode-se dizer que o termo “matriz de referência” é uma forma sistemática de apresentar critérios referenciados organizados sob a forma de uma lista de descritores em termos de habilidades, organizados em tópicos ou construtos que representam competências que se pretende desenvolver dentro de um sistema curricular, além disso, é um recurso de auxílio da elaboração de itens em testes de larga escala, verificando e interpretando o desempenho, bem como outras atividades finalidades pedagógicas. Entretanto, convém destacar que:

“[...] a **matriz de referência** não pode ser confundida com a **matriz curricular**, pois essa é muito mais ampla e norteia as estratégias de ensino nas escolas, enquanto aquela é utilizada para subsidiar a elaboração de um teste específico, contemplando apenas as habilidades consideradas fundamentais para a construção da avaliação”. (Rabelo, 2013, p. 14, *grifo nosso*)

A matriz de referência, ao exercer o papel de construir os itens, determina as capacidades e competências que devem ser desenvolvidas e avaliadas nos itens. Espera-se que cada item possa avaliar um traço latente diante múltiplas respostas, por exemplo, podendo uma competência ser representada por um conjunto de itens e ser considerado um construto a ser medido e avaliado. Por esses motivos que em exames usando testes de larga escala, como o SAEB, por exemplo, as matrizes de referência acabam por fundamentar e nortear os princípios básicos da *Teoria da Resposta ao Item*, possibilitando a interpretação qualitativa das escalas de proficiência construídas após a aplicação dos testes, além alimentar um *Banco Nacional de Itens* (BNI) (Rabelo, 2013). As matrizes também servem de referência nas devolutivas ou *feedback* com os gestores, professores e alunos, orientando na tomada de decisões e possibilitando uma visão ampla sobre o desempenho de diferentes grupos de indivíduos avaliados (Ibid., 2013).

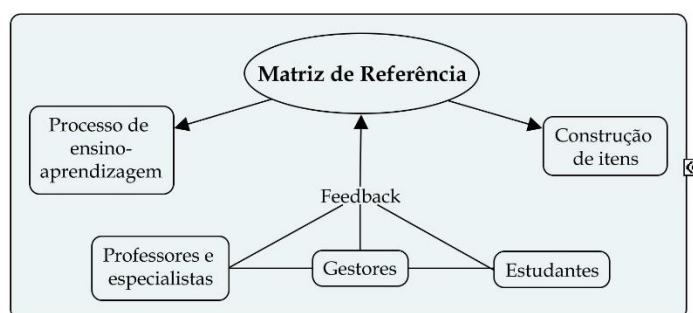


Figura 24: Relações da Influência e Utilização da Matriz de Referência.

Fonte: Adaptado de Rabelo (2013).

Concebe-se como uma visão construída coletivamente e democraticamente visando estabelecer critérios sobre a qualificação formativa, determinando conteúdos, objetivos e processos pedagógicos que podem estar referenciados epistemologicamente dentro de diversas concepções filosóficas. Essas concepções fazem parte da *matriz epistemológica* que fundamenta a *matriz de referência*. As concepções filosóficas inerentes a *matriz epistemológica*, se se estabelece de forma alinhada com a *matriz curricular*, representada pelos documentos oficiais curriculares

nacionais, estaduais e municipais, sendo um processo legitimamente democrático por envolver professores de todos os níveis da educação básica, especialistas curriculares e em livros didáticos.

Nesse processo, ao recorrer a taxonomia de Bloom revisada (Anderson et al., 2001), delimita-se explorar o domínio cognitivo em uma perspectiva comportamentalista quanto as ações do pensamento. O SAEB segue esse processo (Brasil, 2015), assim como o ENEM e ENCCEJA, que compartilham da mesma *matriz de referência* (Rabelo, 2013), com isso se configuram dentro de uma *matriz epistemológica condutivista ou behaviorista*, enquanto as modalidades de ensino técnico e profissionalizante, mais voltadas para o mercado de trabalho, dentro da *concepção funcionalista* (Roque, 2004) e estruturalista. Essas concepções situam-se dentro do **paradigma positivista** das ciências sociais (Ibid., 2004). De modo geral, as matrizes epistemológicas podem ser (Deluiz, 2001; Rabelo, 2013, p. 88):

- **Condutivista/Behaviorista:** fundamentada na psicologia de Skinner e na pedagogia dos objetivos de Bloom;
- **Fundamentada:** na Teoria dos Sistemas Sociais (a lógica de construção de competências é dedutiva, partindo-se das funções mais gerais para as mais específicas);
- **Construtivista/Behaviorista:** construção do conhecimento considerada como processo individual, subjetivo. de desenvolvimento de estruturas cognitivas, em uma perspectiva naturalista da aprendizagem, sem enfatizar o papel do contexto social;
- **Crítico-Emancipatória:** considera o conceito de competência como multidimensional, envolvendo facetas que vão do individual ao sociocultural, situacional (contextual-organizacional) e processual não podendo ser confundida com mero desempenho.

2.3.4.2 Fundamentos da taxonomia de objetivos educacionais

Em 1948, durante a Convenção da *American Psychological Association* (APA) foi discutido a tarefa de classificar metas e objetivos educacionais por um grupo de educadores liderados por Benjamin Bloom. Em novembro de 1995, um grupo de

psicólogos cognitivos, teóricos curriculares e especialistas em testes e avaliação reuniram-se em Siracusa e Nova Iorque, para discutirem a conveniência de revisar a Taxonomia de Objetivos Educacionais publicados em 1956, e depois de uma discussão considerável, o grupo, presidido conjuntamente por David Krathwohl, um dos autores do *Handbook* (edição original), e Lorin Anderson, um dos editores do volume de 1995, denominado *Bloom's Taxonomy: A Forty Year Retrospective*, decidiram prosseguir para uma nova edição (Anderson, 1999), na qual foi concluída e publicada em 2001 (Anderson et al., 2001). Essa taxonomia é adotada pela OCDE visando a classificação dos conhecimentos em uma **concepção behaviorista**, conjuntamente com a taxonomia *Structure of the Observed Learning Outcome* (SOLO) de John Biggs de 1982 (Biggs & Collis, 1982) na **concepção construtivista** (elaborado com inspiração nos trabalhos de Piaget), sendo adotadas também no PISA, pois consideram que: “Os níveis devem abraçar não apenas uma nova compreensão de conhecimentos, habilidades e competências, mas também novas formas em que a aprendizagem pode ser representada usando resultados de aprendizagem” (Keevy & Chakroun, 2015).

Em uma perspectiva de progressão dentro do domínio das competências, entende-se que seja a aplicação de conhecimentos e habilidades em contexto, ou seja, é mais sobre a especialização, em comparação com os *níveis de compreensão* (na taxonomia SOLO) ou a *classificação dos comportamentos de pensamento* (como na taxonomia da Bloom), foi adotada o modelo Dreyfus (Dreyfuss & Dreyfuss, 1986) para o processo de *aquisição de competências* em que se concentra na aquisição de habilidades através da hierarquia, desde novatos até especialistas (Keevy & Chakroun, 2015), mais especificamente como *novato, iniciante avançado, competente, proficiente e especialista* (Dreyfuss & Dreyfuss, 1986, p. 21), estando o especialista numa área entre irracional e racional, na qual ele denomina de “*arational*” (Ibid., 1986, p. 36), que remete ao entendimento de “a-razional”, fora da competência das regras da razão.

Quadro 38: Modelo Dreyfus de competências.

Nível de Progressão	Descritores
Novato	- Aderência rígida às regras ou planos ensinados - Pouca percepção situacional - Sem julgamento discricionário - Atos sem referência ao contexto.
Iniciante avançado	- Diretrizes para ação com base em atributos ou aspectos (os aspectos são características globais de situações reconhecidas apenas como experiência alguma experiência anterior) - A percepção situacional ainda é limitada - Todos os atributos e aspectos são tratados separadamente e tem igual importância.
Competente	- Liderança com a multidão - Agora vê ações pelo menos em parte em termos de metas de longo prazo - Planejamento consciente e deliberado - Procedimentos padronizados e rotineiros - Analítico.
Proficiente	- Considera as situações de forma holística e não em termos de aspectos - Considera o que é mais importante em uma situação - Percebe os desvios do padrão normal - A tomada de decisões menos laboriosa - Usa máximas para orientação, cujos significados variam de acordo com a situação - Racional.
Especialista	- Já não depende de regras, diretrizes ou máximas - Sensação intuitiva de situações baseadas em um entendimento tácito profundo - abordagens analíticas usadas apenas em situações novas ou quando ocorrem problemas - Visão do que é possível.

Fonte: Keevy & Chakroun (2015, pp. 68-69, tradução livre)

As contribuições da *Taxonomia Revisada de Bloom* (Anderson et al., 2001) se restringem na elaboração de uma matriz de competência que seja correspondente a escala de proficiência das concepções térmicas elaboradas no estudo. Nesse processo poderá ser fundamental na intenção de associar as *concepções térmicas* com as *habilidades e competências* correspondentes a cada item, e como consequência, na interpretação da escala, dos itens comentados e no desempenho dos alunos, e com isso proporcionar argumentos e supor diagnósticos instrutivos dentro da Termodinâmica.

No trabalho original publicado em 1956 de Bloom, Engelhart, Furst, Hill, & Krathwohl (1988, pp. 55-165), na versão traduzida para o português brasileiro, tem-se os seguintes *níveis e aptidões*, que na versão revisada serão úteis na interpretação dos itens e na elaboração dos descritores das habilidades e competências dentro da matriz de referência do teste.

Dentro disso se entende que a aquisição dos conceitos básicos seja um requisito necessário para o alcance e desenvolvimento de determinadas competências científicas associadas a um conjunto de situações expostas, porém, deve-se considerar que o conteúdo envolvido no teste se encontra dentro de um *continuum* de complexidade entre a compreensibilidade do conhecimento mais simples e concreto para o mais abstrato, de acordo como se admite na taxonomia de Bloom revisada (Anderson et al.,

2001), porém algumas das limitações existentes estão em considerar a interdependência ou subordinação no processo de progressão e acumulação entre os níveis hierárquicos (Marzano & Kendall, 2007), ignorando que as dimensões do processo cognitivo estejam acessíveis de forma independentes.

O processo de seleção, escolha e organização de um **conteúdo** é uma questão complexa e necessária que deve ser tratada *a priori*, pois se configura em um campo de levar em consideração tanto as questões de natureza epistemológica do conteúdo, como da forma como esse conteúdo é compreendido e de suas implicações para o processo de como ensinar e aprender, sendo dessa forma um aspecto que não pode ser negligenciado quando se busca evidências para contribuir numa otimização do processo educativo. Considera-se ainda relevante seguir para uma eficiência no ensino, na qual encontra-se em outra problemática de igual ou maior complexidade. Sobretudo o estudo possui um olhar mais pragmático e operacional, estando preocupado na busca de oferecer subsídios relevantes que contemplem uma vasta aplicabilidade para o processo de aprendizagem dentro de um recorte temático, estando restrito em um campo da didática que envolve dificuldades conceituais de alunos no final da educação básica, numa unidade temática da Física que envolve a introdução dos conceitos térmicos, que por sua vez se configura como uma delimitação entre as mais diversas implicações relacionadas com as dificuldades de aprendizagem.

Os objetivos educacionais de forma mais específica ao conteúdo que se pretende avaliar estão incluídos tanto no teste, quanto aos itens dos testes, e estes são elaborados a partir das *Matrizes de Referência do Sistema de Avaliação da Educação Básica*. Nessa dimensão, os papéis desses objetivos são definidos por *descritores* que associam determinados *conteúdos curriculares* (tópicos e temas) e *operações mentais* que indicam as *habilidades* a serem aferidas. A elaboração dos descritores é fundamentada numa concepção pedagógica sobre competências, nomeadamente, uma *pedagogia de competências*, uma *pedagogia de integração* ou uma *Abordagem Por Competências* (APC).

O paradigma mais antigo que perdurou por séculos foi o da *Pedagogia na Transmissão de Conteúdos* (PTC), que antecedeu a mudança da cultura curricular para a da *Pedagogia Por Objetivos* (PPO). Esta por sua vez predominou (e que de certa forma

ainda predomina) nas matrizes curriculares da maioria dos países a partir da segunda metade do século XX. O paradigma PPO foi difundido pelo estudo do grupo de pesquisadores norte-americanos liderados por Benjamin S. Bloom, na qual estabeleceram três **domínios** ou áreas fundamentais para o conhecimento: o *cognitivo*, o *psicomotor* e o *afetivo*. O produto desse estudo ficou mundialmente conhecido como a *Taxonomia dos Objetivos Educacionais de Bloom* (1956), ou simplesmente “Taxonomia de Bloom”. Na sequência houve uma difusão e aceitação bastante significativa em diversos países devido à carência ou praticamente ausência de materiais instrucionais voltados para o planejamento curricular e do ensino de forma sistematizada na qual se caracteriza o conhecimento científico. Também se tornou uma ferramenta importante para discutir a seleção de conteúdos, atividades pedagógicas e finalidades didáticas específicas.

Os *Estudos Taxonômicos de Objetivos* (ETO) educacionais de modo geral tiveram grande influência em diversos sistemas educacionais do mundo, a partir de Bloom, mas posteriormente verificaram limitações e distorções na interpretação do fenômeno educacional tendo em vista que se centrava em um paradigma positivista, paradigma esse que preconizou e justificou em parte um ensino “formal” e mecanicista, regido predominantemente pelo princípio da transmissão de conteúdos, que posteriormente buscou superar essas limitações adotando-se uma *Abordagem Por Competências* (APC). A pedagogia por competências surgiu no final do século XX (Brasil, 2015), e mais recentemente tem tido destaque na elaboração de metas curriculares comuns nos sistemas de ensino, assim como na configuração de competências gerais como base estruturadora e integradora em diversas universidades do hemisfério norte. Perrenoud (2013) nos dá uma noção do paradigma das competências no âmbito da educação:

O cerne dessa abordagem reside na mudança da relação da escola com o saber, rompendo a lógica extensiva, enciclopedista e inerte da relação com o saber em favor de uma lógica integradora, mobilizadora e atuante da relação dos cidadãos com esse mesmo saber. (Perrenoud, 2013, *citado por* Brasil, 2015)

Os ETOs apresentam os saberes/competências em diferentes níveis de conhecimento distribuídos nos domínios cognitivo, psicomotor e afetivo. Os trabalhos taxonômicos iniciais de Bloom, Guilford e De Corte, Gerlach e Sullivan, De Block,

French, entre outros, se concentram apenas nas categorias do processo de forma hierárquica e com interdependência ou subordinação entre níveis subsequentes mais estruturados, inclusivos e de maior complexidade, apresentando-se assim de forma *unidimensional e simplista* quanto à natureza do pensamento e sua relação com a aprendizagem (Marzano & Kendall, 2007). Cabe uma ressalva para o recente trabalho de revisão realizado por David Krathwohl e Lorin Anderson em 2001 para a *taxonomia de Bloom (Taxonomia Revisada de Bloom)*, onde avançam na especificidade dos domínios de conhecimento (factual, conceptual, processual, metacognitivo). Marzano & Kendall (2007) reconhecem dificuldades de utilizar este tipo de taxonomia como prelúdio na avaliação do conhecimento apreendido, sendo mais fiável defini-la como um instrumento instrucional de auxílio a uma compreensão mais ampla e profunda. Outra crítica para a linha de pensamento taxonômico de Bloom está a incoerente interdependência entre níveis de conhecimento. Limitações semelhantes são apontadas por Biggs & Collis (1982).

Apesar de serem reconhecidas às limitações de análise na perspectiva advinda dos trabalhos de Bloom, é notória a sua influência e contribuição tanto para o ensino, quanto nas pesquisas educacionais no século XX, onde sentimos seus reflexos ainda hoje. Algumas limitações também são apontadas na taxonomia SOLO,

“Investigadores e revisores apontaram as lacunas da [*Taxonomia*] SOLO causadas pela sua estrutura e os critérios de categorização (Chan et al., 2001, Chick, 1998). A ambiguidade conceitual da estrutura da [*Taxonomia*] SOLO torna a categorização instável, juntamente com a confiabilidade inter-avaliadores e o acidente de confundir ‘extensões abstratas’ (resposta muito boa) como ‘pré-estrutural’ (resposta muito pobre).” (Chan, Tsui, Chan, & Hong, 2002, p. 512, *tradução livre, grifo nosso*).

Um dos principais fatores para uni a taxonomia de Bloom com a SOLO está na potencialidade de articulação com um *pensamento construtivista*, associando níveis de dificuldade a uma *perspectiva integradora* ou de estrutura organizacional de ações ao invés de um único fator mensurável. Outro aspecto taxonômico aliado ao pensamento construtivo está a associação do cruzamento com o *campo crítico* (King, 2009, *citado por* Keevy & Chakroun, 2015), *competências de base* (Lokhoff et al., 2010, *citado por* Keevy & Chakroun, 2015), *competências centrais e habilidades transferíveis*.

Diante dos exames nacionais, um dos principais papéis exercido atualmente pelas taxonomias de objetivos educacionais se concentram na elaboração da *matriz de referência*, sendo este um instrumento essencial tanto no processo de construção de itens, elaboração de teses, quanto na avaliação de desempenhos.

Quadro 39: Descrição de alguns estudos taxonômicos envolvendo objetivos educacionais dentro do domínio cognitivo.

Estudos Taxonômicos de Objetivos Educacionais			
Domínio do Processo – Cognitivo			
N	Autor(es)	Dimensões Taxonômicas	
		Categorias do Processo (ação ou critérios)	Domínios do Conhecimento (níveis do objeto ou conteúdo)
1	Anderson et al., (2001). <i>*Taxonomia Revisada de Bloom.</i>	1. Lembrando; 2. Entendendo; 3. Aplicando; 4. Analisando; 5. Avaliando; 6. Criando.	1.1 Factual; 2.1 Conceptual; 3.1 Processual; 4.1 Metacognitivo; 5.1 Metacognitivo; 6.1 Metacognitivo.
2	Bloom, Englehart, Furst, Hill, & Krathwohl, (1956).	1. Conhecimento; 2. Compreensão; 3. Aplicação; 4. Análise; 5. Síntese; 6. Avaliação.	1.1 Conhecimento de específicos; 1.2 Conhecimento de maneiras e meios de tratar com específicos; 2.1 Translação; 2.2 Interpretação; Extrapolação; 4.1 Análise de elementos; 4.2 Análise de relações; 4.3 Análise dos princípios organizacionais; 5.1 Produção de uma comunicação singular; 5.2 Produção de um plano ou indicação de um conjunto; 5.3 Derivação de um conjunto de relações abstratas; 6.1 Julgamentos em termos de evidência interna; Julgamentos em termos de critérios externos.
3	Guilford (1967). <i>*Influência de Bloom.</i> <i>**Sua taxonomia não é diretamente sobre os objetivos da educação (objetivos operacionais).</i>	Estrutura do Modelo do Intelecto • <i>Operações:</i> 1. Cognição; 2. Memória; 3. Produção-divergente; 4. Produção-convergente; 5. Avaliação.	• <i>Conteúdos:</i> 1. Figurativo; 2. Simbólico; 3. Semântico; 4. Comportamental. • <i>Produtos:</i> 1. Unidades; 2. Classes; 3. Relações; 4. Sistemas; 5. Transformações.
4	De Corte (1973). <i>*Adaptado do modelo de Guilford.</i> <i>**Ênfase no Domínio Cognitivo.</i>	• <i>Operações Receptivo-Reprodutivo:</i> 1. Apercepção de informações; 2. Reconhecimento de informações; 3. Reprodução de informação. • <i>Operações Produtivas:</i> 1. Produção interpretativa de informações; 2. Produção convergente de informações; 3. Produção avaliativa de informações; 4. Produção divergente de informações.	• <i>Conteúdos:</i> 1. Figurativo; 2. Simbólico; 3. Semântico; 4. Comportamental. • <i>Produtos:</i> 1. Unidades; 2. Classes; 3. Relações; 4. Sistemas; 5. Transformações.
5	De Block (1975). <i>*Modelo influenciado por Guilford.</i>	• <i>Processos de aprendizagem:</i> 1. Saber; 2. Compreender; 3. Aplicar; 4. Integrar • <i>Cultura subjetiva:</i> 1. Transferência dentro de um ramo; 2. Transferência mais geral: áreas científicas, técnica, social, moral, estético; Transferência geral.	• <i>Conteúdos culturais dos objetivos:</i> 1. Fatos; 2. Conceitos; 3. Relações; 4. Estruturas; 5. Métodos; 6. Atitudes.
N	Autor(es)	Categoria do Processo (ação ou critérios)	
6	McGuire (1963).	<i>Taxonomia dos Processos Intelectuais</i> 1. Recordar; 2. Reconhecer significados (Compreensão); 3. Selecionar uma generalização relevante; 4. Interpretar dados simples; 5. Aplicar princípios (em situações familiares); 6. Aplicar princípios (em situações não familiares); 7. Avaliar (incluindo a análise) de uma situação total; 8. Síntese de dados com todos os significados.	
7	Gerlach & Sullivan, (1967). <i>*Ênfase behaviorista.</i>	1. Identificar; 2. Nomear; 3. Descrever; 4. Construir; 5. Ordenar; 6. Demonstrar.	
8	NLSMA ⁵⁵	1. Conhecimento dos fatos; 2. Cálculo; 3. Compreensão; 4. Aplicação; 5. Análise.	

⁵⁵ *National Longitudinal Study of Mathematical Abilities*, citado por Choppin & Postlethwaite (1979, p. 110).

Domínio Cognitivo Integrado a Outros Domínios e Categorias do Processo				
9	D' Hainaut, (1970). <i>*Tentativa de síntese geral sobre da metodologia de definições de objetivos (baseado em critérios operacionais).</i>	<ul style="list-style-type: none"> • Atividade intelectual: 1. A reprodução ou de repetição; 2. Conceituação; 3. Aplicação das regras ou algoritmos; 4. Mobilização e associação de respostas (divergência); 5. Resolução de Problemas. • Comportamentos observáveis: 1. Estado, descrever, nome...; 2. Reconhecer; 3. O membro de uma classe, identificado como...; 3. Calcular, utilizar...; 4. Imaginar, encontrar uma variedade de...; 5. Localizar, identificar, produzir (mas inventando a solução). 	<ul style="list-style-type: none"> • Contexto – Objetos: 1. Elementos; 2. Classes; 3. Relações; 4. Operações e operadores; 5. Estruturas. • O grau de integração: 1. Execução imediata; 2. Retenção de curto ou longo prazo; 3. Transferência escolar; 4. Transferência operacional; 5. Transferência integral. 	
10	Gagné (1965) & Merrill (1971). <i>*Baseado em comportamentos observáveis.</i>	<ul style="list-style-type: none"> • Afetivo: 1. Emocional (respostas involuntárias); 2. Emocional (aprendizagem de sinais). • Psicomotor: 1. Psicomotor (nível de consciência um pouco baixo; 2. Comportamento topográfico; 3. Cadeias; 4. Habilidades complexas (comportamento habilidoso). • Cognitivo/Psicomotor: 1. Memória (comportamento consciente); 2. Apelação (nomeação); 3. Memória serial (associação verbal); 4. Memória discreta (discriminação múltipla). • Cognitivo: 1. Cognição complexa (intensa consciência da transferência de comportamento); 2. Classificação (aprendizagem de conceitos); 3. Analisar (aprendizagem de princípios); 4. Resolução de problema (síntese). • O princípio da redução – Merrill: Um comportamento aprendido em um determinado nível é adiado para um nível inferior logo que as condições mudarem o suficiente. 	<ul style="list-style-type: none"> • Condições de aprendizagem – Gagné: 1. Aprendizagem de sinais (resposta condicionada de Pavlov); 2. Aprendizagem de conexões estímulo-resposta (aprendizagem E-R); 3. Aprendizagem de cadeias de transmissão (encadeamento de conexões E-R); 4. Aprendizagem de cadeias verbais (associação verbal); 5. Aprender com a discriminação múltipla; 6. Aprendizagem de conceitos; 7. Aprendizagem de princípios; 8. Resolução de problemas. 	
11	Tuckman (1972).	<ul style="list-style-type: none"> • Domínios: 1. Percepção; 2. Cognitivo; 3. Afetivo; 4. Psicomotor. • Processo: 1. Aquisição; 2. Aplicação; 3. Avaliação; 4. Comunicação. 	<ul style="list-style-type: none"> • Objetos: 1. Coisas; 2. Ideias; 3. Pessoas. 	
12	Scriven (1967). <i>*Voltado para a avaliação (Objetivos operacionais).</i>	<ul style="list-style-type: none"> • Descrição conceitual dos objetivos da educação: 1. Conhecimento; 2. Compreensão; 3. Motivação; 4. Capacidades não-cognitivas; 5. Variáveis não-educacionais. • Descrição fenomenológica de variáveis críticas: 1. Conhecimento; 2. Compreensão; 3. Atitudes; 4. Capacidades não-cognitivas. 	<ul style="list-style-type: none"> • Nível conceitual do evento (desenvolver o pensamento crítico). • Nível operacional do evento (objetivo definido em termos de formas de avaliar). 	
13	Biggs & Collins (1982).	<ul style="list-style-type: none"> • Níveis de complexidade: 1. Pré-estrutural; 2. Uni-estrutural; 3. Multi-estrutural; 4. Relacional; 5. Extensões abstratas. 	<ul style="list-style-type: none"> • Tipos de dados na estrutura das respostas (sinais ou “deixas” ↔ respostas): 1. irrelevante ou inapropriado; 2. relacionados e hipotéticos, não determinados; 3. relacionados e determinados em exposição. 	
N	Autor (es)	Categorias do Processo (ação ou critérios)	Domínios do Conhecimento (níveis do objeto ou conteúdo)	Níveis de Processamento
13	Marzano & Kendall (2006).	(Transcende todos os domínios e categorias do processo).	1. Procedimentos psicomotores; 2. Procedimentos mentais; 3. Informações.	1. Auto-sistema; 2. Sistema metacognitivo; 3. Sistema Cognitivo.
Normas para as Taxonomias de Objetivos Educacionais				
14	Tyler (1969). <i>*Sua taxonomia não é diretamente sobre os objetivos da educação (objetivos operacionais).</i>	Princípios para a fundamentação curricular (Objetivos Gerais Empíricos): 1. Análise da sociedade; 2. O aluno; 3. Conteúdo; 4. Relevância para a filosofia da educação; 5. Consistência com a nossa teoria de aprendizagem.		
15	Ebel (1973). <i>*Sua taxonomia não é diretamente sobre os objetivos da educação (objetivos operacionais).</i>	Princípios para a determinação de objetivos educacionais: 1. Objetivos específicos; 2. Objetivos Comportamentais; 3. Testes com Critérios Referenciados; 4. Estatísticas para Testes com Critérios Referenciados.		
16	Gronlund (1970). <i>*Sua taxonomia não é diretamente sobre os objetivos da educação (objetivos operacionais).</i>	Declarando os Objetivos de Instrução (em sala de aula): 1. Definir objetivos gerais; 2. Definir objetivos específicos.		
17	Walbesse (1963). <i>*Baseado em Gagné</i> <i>**Sua taxonomia não é diretamente sobre os objetivos da educação (objetivos operacionais).</i>	Objetivos operacionais: 1. Nomeando; 2. Identificando; 3. Reconhecendo; 4. Distinguindo; 5. Descrevendo; 6. Ordenando.		

18	Vandavelde & Vanderelst (1975).	Três níveis de definição (de objetivos): 1. Descrição verbal; 2. Definição pela lista de objetivos educacionais; 3. Definição para o comportamento correspondente em cada categoria.
19	Dressel & Mayhew (1954). <i>*Objetivos operacionais.</i>	Objetivos gerais educacionais: 1. Formar um cidadão; 2. Desenvolver a compreensão do método científico e da vontade de aplicar os conhecimentos e competências adquiridos na resolução de problemas sociais; 3. Comunicar; 4. Sendo adaptado emocionalmente e socialmente; 5. Desenvolver o pensamento crítico.

Fonte: Elaborado a partir de Anderson et al. (2001); Biggs & Collis (1982); Bloom, Engelhart, Furst, Hill, & Krathwohl (1988); Landsheere & Landsheere (1978); Marzano (2006); e Choppin & Postlethwaite (1979). Algumas referências se encontram diretamente no quadro. A tradução é livre.

Tendo a taxonomia SOLO e Revisada de Bloom como referência nos trabalhos da UNESCO (2015), e a Revisada de Bloom na fundamentação teórico-metodológica nos exames do Brasil (BRASIL, 2015), de acordo com a plataforma Devolutivas Pedagógicas do INEP, tem-se uma breve descrição delas.

John Biggs em seu site⁵⁶ apresenta uma definição em termos práticos da *taxonomia SOLO*⁵⁷:

“[...] um meio de classificar os resultados da aprendizagem, em termos de sua **complexidade**, o que nos permite avaliar o trabalho dos alunos em termos da sua qualidade e não como uma pequena porção disto e do que eles têm direito. No primeiro, pegar apenas um ou alguns aspectos da tarefa (uniestrutural), em seguida, vários aspectos, mas eles não estão relacionados (multiestrutural), então aprendermos a integrá-los em um todo (relacional) e, finalmente, somos capazes de generalizar o todo para aplicações ainda não desenvolvidas (abstrações estendidas)”.

Biggs declara que essa taxonomia foi primeiramente descrita por Kevin Collis, e posteriormente em seu trabalho denominado “*Avaliação da Qualidade da Aprendizagem: A taxonomia SOLO (1982)*” (“*Evaluating the Quality of Learning: The SOLO Taxonomy*”). Ressalta ainda que possa ser usada não apenas na avaliação, mas na concepção do currículo em termos de resultados de aprendizagem pretendidos, o que é útil para a aplicação do *alinhamento construtivo*⁵⁸. Considera também que possa explicar por que aqueles que usam argumentos de **baixa complexidade** em disputas políticas geralmente ganham – no curto prazo.

⁵⁶ Ver: <http://www.johnbiggs.com.au/academic/solo-taxonomy/>.

⁵⁷ *Structure of the Observed Learning Outcome (SOLO)*, traduzindo do inglês significa *Estrutura dos Resultados de Aprendizagem Observados*.

⁵⁸ John Biggs e Catherine Tang. *Teaching for Quality Learning at University, The Society for Research into Higher Education, Third Edition, 2007.*

Retornando a perspectiva dos itens na configuração dos exames, pode-se afirmar, portanto, que tanto em seu processo de elaboração e concepção textualizada quanto nas respectivas resoluções ou opções de resposta estão envolvidos em processos de pensamento previamente realizados ou que se exija um processamento mental para resolvê-lo. Nesse sentido, as *taxonomias de objetivos educacionais* podem e são utilizadas como instrumentos para identificar distinções entre os **diversos níveis cognitivos de abstração do objeto**, no caso em questão, *objetos de aprendizagem*.

Primeiramente, a *classificação dos objetos de aprendizagem* (às vezes chamados de objetos do conhecimento) na perspectiva de D' Hainaut (1983, p. 242 e 257) citado por Gérard e Roegiers (1993, p. 48 e 49):

1. **Fatos particulares:** *elementos ou acontecimentos singulares que não têm nenhum carácter de generalidade (data de um acontecimento, a fórmula química de uma substância, o nome de uma pessoa que tenha um lugar marcante na História, o título de um livro, o tempo de gestação de um animal, um meio de locomoção, a capital de um país...);*
2. **Classes:** *conjunto de elementos que possuem pelo menos uma propriedade comum (noções de cor, de quadrado, de triângulo, de adição, de frase, de verbo, de substantivo, de adjetivo, de força, de respiração, etc.). Os conceitos são representações mentais provenientes das classes;*
3. **Relações:** *proposições que contêm "variáveis", que dizer, termos gerais que podem assumir valores particulares. Por exemplo, a área do retângulo é igual ao produto do comprimento vezes a largura; é uma relação, porque o termo retângulo pode ser substituído por um retângulo específico e as palavras "comprimento" e "largura" por valores particulares relativos ao retângulo designado. Entre outras, fazem parte desta categoria as leis, as fórmulas, as regras, etc.*
4. **Estruturas e sistemas:** *conjunto de relações. Uma estrutura contém, simultaneamente, elementos e relações entre estes elementos: uma teoria, um organograma, uma demonstração, um processo de decisão, um algoritmo, etc.*

Gérard e Roegiers (1993, p. 48 e 49) ressaltam que o *mesmo objeto* pode ser uma **entidade particular** ou uma **estrutura**, conforme o uso que dele se faz. Ao exemplificar diz que uma cadeira é uma *entidade particular* para aquele que nela senta, mas para o marceneiro que a construiu é uma *estrutura*.

Por sua vez, o estudo taxonômico diante desses objetos de aprendizagem é um dos processos utilizados na perspectiva da *Abordagem por Competências* (APC), pode-se afirmar que apesar de ser uma fundamentação que predominou no paradigma da *Pedagogia Por Objetivos* (PPO) dentro da corrente condutivista/behaviorista tem sua utilidade restrita em uma perspectiva estruturadora e organizacional dos conteúdos, em especial, na construção das matrizes de referência ou de competências.

A Taxonomia Revisada de Bloom, contribui para uma compreensão mais ampla de sua utilização na elaboração dos descritores na matriz de referência, pertinente no âmbito da discussão do texto. Algumas ideias dessa plataforma serão utilizadas para um breve esclarecimento.

Em termos gerais, esta revisão redefiniu o domínio cognitivo como a interseção da *Dimensão dos Processos Cognitivos* com a *Dimensão do Conhecimento*, proporcionando uma análise taxonômica bidimensional. Cada dimensão é representada em *etapas hierárquicas*, porém as distinções entre categorias nem sempre esclarecedoras. Por exemplo, todo conhecimento procedimental não é necessariamente mais abstrato do que todo conhecimento conceitual; e um objetivo que envolve a análise ou avaliação pode exigir habilidades de pensamento que não são menos complexas do que uma que envolve a criação ou criatividade. É compreendido geralmente, contudo, que as habilidades de pensamento da ordem inferior são subordinadas e fornecem a fundação para habilidades de pensamento da ordem mais elevada. Estes aspectos são alvo de críticas por não ser capaz de explicar a independência dos níveis de processamento mental, conforme dito anteriormente por Marzano & Kendall (2007). Entretanto, as análises taxonômicas ainda são consideradas de grande utilidade para análise e organização de currículo, definição de tarefas com diversos níveis de abstração,

avaliação de livros didáticos, na elaboração de descritores nas matrizes de referência, entre outros.

É importante compreender que na elaboração de um descritor, a declaração de um *objetivo de aprendizagem* deve conter um **verbo** (uma ação) que representa a capacidade ou habilidade, geralmente se refere a [ações associadas] com o processo cognitivo pretendido. O **objeto** (geralmente um substantivo) que se refere ao conteúdo na qual a capacidade de realiza, geralmente descreve o conhecimento que os alunos esperam adquirir ou construir (Anderson et al., 2001, pp. 4-5). Um aspecto relevante a ser considerado seria que os objetivos de aprendizagem não atividades de aprendizado, talvez possa ser útil pensar em preceder cada objetivo com algo como: "*Os alunos serão capazes de...*".

Quadro 40: Descrição detalhada das dimensões da Taxonomia de Bloom (Revisada)

DOMÍNIOS		Dimensão dos Processos Cognitivos - Habilidades do Pensamento -		Dimensão dos Conhecimentos - Graus do conhecimento -			
		Verbos (simples) Ilustrativos para Declarar os Objetivos (Como?)		Tipologia dos Conhecimentos (O quê?)			
Habilidades Cognitivo (Anderson et al 2001)	Habilidades do Pensamento de Ordem Inferior	1. Recordar	1.1 Reconhecendo (identificando)	Concreto	1. Factual	1.1 Conhecimento da terminologia	
			1.2 Recordando (recuperando)			1.2 Conhecimento de detalhes específicos e elementos	
		2. Compreender	2.1 Interpretando (esclarecendo, parafraseando, representando, traduzindo)			2. Conceptual	2.1 Conhecimento de classificações e categorias
			2.2 Exemplificando (ilustrando, instanciando)				2.2 Conhecimento dos princípios e generalizações
			2.3 Classificando (categorizando, subsumindo)				2.3 Conhecimento das teorias, modelos e estruturas
			2.4 Resumindo (abstraindo, generalizando)				
			2.5 Inferindo (concluindo, extrapolando, interpolando, prevendo)				
			2.6 Comparando (contrastando, mapeando, correspondendo)				
			2.7 Explicando (construindo modelos)				
		3. Aplicar	3.1 Executando (realizando)		3. Processual	3.1 Conhecimento de habilidades e algoritmos específicos de um assunto	
			3.2 Implementando (usando)			3.2 Conhecimento de técnicas específicas de assunto e métodos	
		4. Analisar	4.1 Diferenciando (discriminando, distinguindo, convergindo, selecionando)			3.3 Conhecimento dos critérios para determinar quando utilizar processos adequados	
	4.2 Organizando (encontrando coerência, integrando, delineando, analisando sintaticamente, estruturando)						
	4.3 Atribuindo (desconstruindo)						
	5. Avaliar	5.1 Verificando (coordenando, detectando, monitorando, testando)	4. Metacognitivo	4.1 O conhecimento estratégico			
		5.2 Criticando (julgando)		4.2 Conhecimento sobre tarefas cognitivas, incluindo apropriado contexto e condicional conhecimento			
	6. Criar	6.1 Gerando (admitindo hipóteses)		4.3 O autoconhecimento			
		6.2 Planejando (projetando)					
		6.3 Produzindo (construindo)					
	Habilidades do Pensamento de Ordem Superior			Abstrato			

Fonte: Traduzido e Adaptado do site <<http://www.celt.iastate.edu/teaching/effective-teaching-practices/revise-blooms-taxonomy>>.

Quadro 35: Sugestões de Ideias Associadas a Taxonomia de Bloom (Revisada) nas Dimensões do Conhecimento e dos Processos Cognitivos.

SISTEMA TAXONÓMICO BIDIMENSIONAL		A DIMENSÃO DO CONHECIMENTO			
		Factual	Conceptual	Processual	Metacognitivo
DIMENSÃO DO PROCESSO COGNITIVO	Recordar (Recuperar o conhecimento relevante da memória de longo prazo).	Recordar + Factual <i>Listar as cores primárias e secundárias.</i>	Recordar + Conceptual <i>Reconhecer sintomas de exaustão</i>	Recordar + Processual <i>Lembrar como executar os procedimentos de ressuscitação cardiopulmonar</i>	Recordar + Metacognitivo <i>Identificar estratégias para reter informações</i>
	Compreender (Construir o significado a partir de mensagens instrucionais, incluindo comunicação oral, escrita e gráfica).	Compreender + Factual <i>Resumir características de um novo produto.</i>	Compreender + Conceptual <i>Classificar os adesivos por toxicidade</i>	Compreender + Processual <i>Esclarecer as instruções de montagem</i>	Compreender + Metacognitivo <i>Prever a resposta de alguém ao choque da cultura</i>
	Aplicar (Realizar ou usar um procedimento em uma situação dada).	Aplicar + Factual <i>Responder às perguntas mais frequentes.</i>	Aplicar + Conceptual <i>Fornecer conselho aos principiantes</i>	Aplicar + Processual <i>Realizar testes do pH de amostras de água</i>	Aplicar + Metacognitivo <i>Usar as técnicas que combinam com seus pontos fortes</i>
	Analisar (Decompor o material em partes constituintes e determine como as partes se relacionam entre si e a uma estrutura ou finalidade geral).	Analisar + Factual <i>Selecionar uma lista de atividades de modo mais completo.</i>	Analisar + Conceptual <i>Diferenciar uma baixa e elevada cultura</i>	Analisar + Processual <i>Integrar a conformidade de acordo com os regulamentos</i>	Analisar + Metacognitivo <i>Desconstruir os seus preconceitos</i>
	Avaliar (Fazer julgamentos baseados em critérios e padrões).	Avaliar + Factual <i>Verificar a consistências entre fontes</i>	Avaliar + Conceptual <i>Determinar a relevância dos resultados</i>	Avaliar + Processual <i>Julgar a eficiência das técnicas de amostragem.</i>	Avaliar + Metacognitivo <i>Refletir sobre o progresso</i>
	Criar (Juntar elementos para formar um todo coerente ou funcional; reorganizar elementos em um novo padrão ou estrutura).	Criar + Factual <i>Gerar um registro de atividades diárias</i>	Criar + Conceptual <i>Reunir uma equipe de especialistas</i>	Criar + Processual <i>Planejar um eficiente projeto de fluxo de trabalho</i>	Criar + Metacognitivo <i>Criar um portfólio de aprendizagem</i>

Fonte: Traduzido e Adaptado do site: <http://www.celt.iastate.edu/teaching/effective-teaching-practices/revised-blooms-taxonomy>.

Em outubro de 2009, Andrew Churches, codiretor da área de Estudos de Informática da *Kristin School* de Auckland, Nova Zelândia, sendo um entusiasta das TIC⁵⁹ como um poderoso mecanismo na transformação educacional publicou em seu *blog*⁶⁰ uma nova atualização da Taxonomia de Bloom, na qual denomina “*Taxonomia de Bloom para a Era Digital*”. Com o avanço das TIC cada vez mais intenso no campo educacional, considera-se um campo de investigação promissor este recurso para o desenvolvimento mais sistematizado de uma *pedagogia digital*⁶¹, que aparentemente se encontra em processo preliminar de análise, ainda numa perspectiva unidimensional. Entretanto, o avanço nesse campo de estudo poderá trazer contribuições na elaboração de materiais de educação *online* e na modalidade à distância com novas perspectivas curriculares, em especial, no desenvolvimento de *objetos virtuais de aprendizagem* (ou objetos educacionais digitais), na definição de mecanismos de interação e comunicação, e na organização e estruturação dos módulos cursos *online*.

2.3.4.3 Fundamentos para a construção de uma escala de proficiências

Para a construção de uma escala de proficiências, basicamente, deve-se partir de uma *matriz de respostas* dos testes, submetê-la às análises estatísticas que irão resultar em um agrupamento de itens que podem ser separados por intervalos predefinidos e flexíveis. Os critérios utilizados para agrupar itens se baseiam no conceito de *item âncora*, em que:

“[...] para um item ser âncora em um determinado *nível âncora da escala*, ele precisa ser respondido corretamente por uma grande proporção de indivíduos (pelo menos 65%) com este nível de habilidade e por uma proporção menor de indivíduos (no máximo 50%) com o nível de habilidade imediatamente anterior. Além disso, a diferença entre a proporção de indivíduos com esses níveis de habilidade que acertam a esse item deve ser de pelo menos 30%. Assim, para um item ser âncora ele deve ser um item “típico” daquele nível, ou seja, bastante acertado por indivíduos com aquele nível de habilidade e pouco acertado por indivíduos com um nível de habilidade imediatamente inferior”. (Andrade, Tavares, & Valle, 2000, p. 110)

⁵⁹ *Tecnologias de Informação e Comunicação*.

⁶⁰ Ver: <http://eduteka.icesi.edu.co/articulos/TaxonomiaBloomDigital>.

⁶¹ O conceito *Pedagoware* apresentado por Souza e Mol (2013) com essa perspectiva taxonômica parece um caminho promissor de investigação.

Dessa forma, a definição de «item âncora» deve obedecer a determinadas condições estatísticas, p. ex.: considere dois *níveis âncora*⁶² consecutivos Y e Z com $Y < Z$. Dizemos que um determinado item é âncora para o nível Z se, e somente se, as 3 condições abaixo forem satisfeitas simultaneamente:

- $P(X = 1|\theta = Z) \geq 0,65$.
- $P(X = 1|\theta = Y) < 0,50$.
- $P(X = 1|\theta = Z) - P(U = 1|\theta = Y) \geq 0,30$.

Rabelo (2013), afirma que o conceito de item-âncora adotada no SAEB é baseado na metodologia de Beaton & Allen (1992). Os critérios acima descritos são considerados muito rígidos, diante disso tem-se adotado uma modificação:

- $P(X = 1|\theta = Z) \geq 0,65$.
- $P(X = 1|\theta = Y) < 0,50$.
- Ajuste da CCI.

Os *níveis âncora* representam o *nível de proficiência* na qual os respondentes demonstraram terem desenvolvido em termos de habilidades, portanto, representam um *patamar de competência*. Dentre alguns problemas que eventualmente surgem no processo de construção de uma *escala de proficiência* estão à quantidade insuficiente de itens para caracterizar um *patamar de competência*, em especial, os níveis extremos da escala que são geralmente representados pelos itens mais fáceis e difíceis. Vale lembrar que no nível superior encontram-se uma quantidade limitada de respondentes (os de elevada proficiência), portanto, a interpretação pedagógica nesse nível fica baseada por um percentual muito baixo de alunos que dominam os conhecimentos descritos por esse nível, e este é um aspecto de limitação a ser considerado.

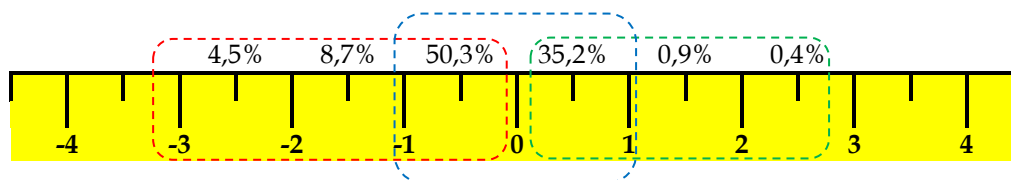


Figura 25: “Régua” ilustrativa do percentual de respondentes em uma escala de proficiência.

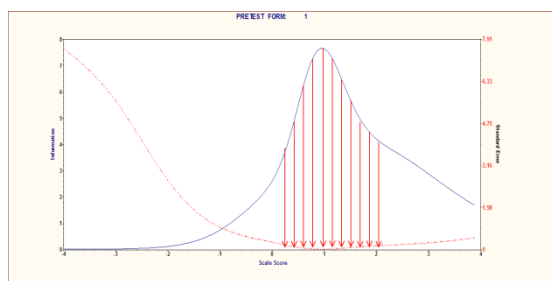
⁶² Formado por um conjunto de itens que enquadram dentro de mesmo nível de proficiência.

Observando a figura – *meramente ilustrativa*, pode-se notar que nas extremidades o percentual é mais baixo, e no meio se concentra a maioria dos respondentes, e é nela que se concentram as maiores informações do teste.

Nesse exemplo, tem-se que mais de 60% dos respondentes encontram-se abaixo da média, sinalizando que o teste foi relativamente difícil para a maioria. Para facilitar a compreensão sobre o processo de construção de uma escala de proficiência diante das análises e critérios envolvidos, as ilustrações abaixo representam algumas dessas etapas. Esses tipos de análises psicométricas da TRI ocorrem nos seguintes exames: SAEB, PROVA BRASIL, ANA, ENCCEJA, ENEM, e está previsto de ocorrer também na PROVA DOCENTE, que segue em elaboração.

Os grupos de itens que estão destacados na tabela (figura à esquerda) são indicadores de probabilidade de acerto dentro de um determinado nível âncora, e correspondem aos itens-âncora do respectivo nível por obedecerem aos critérios supracitados. O próximo passo seria agrupar os itens-âncora em um nível de proficiência na escala para serem interpretados pedagogicamente, e seguindo esse processo a escala de proficiência vai sendo construída. Por último tem-se a opção de definir valores (através de transposição linear) para os níveis de proficiência da escala e determinar um valor para o nível de referência, na qual representa o valor médio de proficiência da amostra (semelhante às escalas do SAEB). Depois de construída, as análises posteriores possibilitam o posicionamento tanto de itens quanto de respondentes dentro da mesma escala, assim como estudos comparativos entre contextos históricos e sociais diferentes.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
001	0,889	2,42197	0,2708	0,47809	0,52608	0,58083	0,64011	0,69152	0,74408	0,79718	0,82226	0,84111	0,85954	0,87628													
002	0,898	1,6279	0,1728	0,74411	0,37277	0,38232	0,39556	0,42578	0,46852	0,52628	0,73357	0,81431	0,85347	0,86981													
003	1,3145	0,7971	0,5015	0,30243	0,30683	0,32011	0,32893	0,33803	0,35075	0,36500	0,38089	0,38738	0,39236	0,39429													
004	0,842	1,1415	0,2868	0,29049	0,29559	0,29817	0,29944	0,3096	0,32622	0,34954	0,35711	0,36455	0,37187														
005	0,7408	1,4211	0,2032	0,20789	0,21384	0,22009	0,22582	0,23034	0,23445	0,23824	0,24141	0,24408	0,24625	0,24793													
006	1,4111	1,2584	0,1929	0,19739	0,19901	0,19936	0,19945	0,20011	0,20077	0,20170	0,20289	0,20414	0,20536	0,20658													
008	0,4715	2,0514	0,1442	0,14749	0,14971	0,15193	0,15415	0,15637	0,15859	0,16081	0,16303	0,16525	0,16747	0,16969													
009	1,6528	1,9376	0,1704	0,17049	0,17049	0,18024	0,18024	0,18024	0,18024	0,18024	0,18024	0,18024	0,18024	0,18024													
010	1,4091	0,5520	0,0881	0,08123	0,08628	0,09021	0,09387	0,09722	0,10027	0,10302	0,10557	0,10792	0,11007	0,11202													
011	0,4088	2,6276	0,1811	0,18254	0,18487	0,18621	0,18754	0,18887	0,19021	0,19154	0,19287	0,19421	0,19554	0,19687													
012	0,2908	0,4680	0,2813	0,38136	0,48137	0,43774	0,49683	0,40771	0,49054	0,44208	0,49439	0,43399	0,49048	0,43036													
013	1,8990	0,1887	0,1859	0,18812	0,19079	0,19347	0,19614	0,19881	0,20148	0,20415	0,20682	0,20949	0,21216	0,21483													
014	1,4994	0,7391	0,1335	0,1335	0,13350	0,13351	0,13352	0,13353	0,13354	0,13355	0,13356	0,13357	0,13358	0,13359													
015	1,4414	2,4232	0,1400	0,14004	0,14002	0,14985	0,14985	0,14985	0,14985	0,14985	0,14985	0,14985	0,14985	0,14985													
016	1,2154	1,8146	0,1199	0,12100	0,12215	0,12331	0,12446	0,12561	0,12676	0,12791	0,12906	0,13021	0,13136	0,13251													
017	0,8912	2,4627	0,1202	0,12768	0,12920	0,13182	0,13446	0,13668	0,13890	0,14112	0,14334	0,14556	0,14778	0,15000													
018	0,9615	1,2940	0,1078	0,10936	0,11086	0,11234	0,11382	0,11530	0,11678	0,11826	0,11974	0,12122	0,12270	0,12418													
019	1,9319	2,1780	0,1006	0,10064	0,10061	0,10059	0,10056	0,10053	0,10050	0,10047	0,10044	0,10041	0,10038	0,10035													
020	0,4208	1,4742	0,1900	0,14238	0,16623	0,19272	0,19724	0,24480	0,19070	0,46488	0,19378	0,29017	0,49038	0,49017													
021	1,0256	0,4680	0,1950	0,17956	0,18173	0,19726	0,19263	0,19647	0,19920	0,19920	0,19920	0,19920	0,19920	0,19920													
022	1,3660	0,0076	0,1763	0,17718	0,19738	0,18945	0,20141	0,14232	0,38954	0,18118	0,14017	0,18667	0,19648	0,19301													
024	3,2785	1,9728	0,1518	0,15118	0,15158	0,15181	0,15216	0,15242	0,15276	0,15310	0,15344	0,15378	0,15412	0,15446													
025	1,3099	1,2744	0,1219	0,1215	0,12179	0,12203	0,12228	0,12252	0,12276	0,12300	0,12324	0,12348	0,12372	0,12396													
026	1,3030	1,2216	0,2012	0,20143	0,20600	0,20809	0,20564	0,21218	0,21448	0,21678	0,21908	0,22138	0,22368	0,22598													



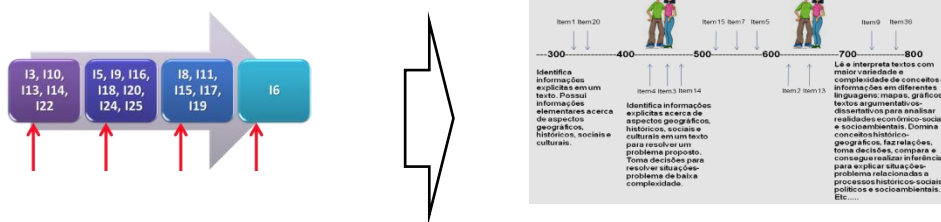


Figura 26: Ilustração do processo de construção e interpretação de uma escala de proficiência.

Para que se tenha uma boa qualidade do teste em um exame é necessário um bom conhecimento dos parâmetros de seus itens. Isto é conseguido através de pré-testagens de itens em amostras apropriadas de alunos (respondentes com as mesmas características na qual se pretende aplicar o exame) na intenção de estimar os parâmetros e calibrar os itens em uma mesma escala de proficiência, e com isso, os itens são posicionados em uma escala de acordo com o nível de proficiência que se é exigido. Esse processo de validação é o que define a seleção de itens a serem utilizados no exame, itens “ruins” serão descartados ou reformulados.

Este processo também é útil para compor um *banco de itens*⁶³ na escala de proficiência desejada e a partir dele pode-se construir um ou mais testes com graus de dificuldade equivalentes para atender os objetivos de uma ou mais avaliações posteriores. Com relação aos exames em si, uma boa vantagem é o fato de que as *proficiências dos alunos* submetidos a esses *diferentes testes* são medidas na *mesma escala* e, portanto, são **comparáveis entre si**. De forma equivalente, as medidas que se obtêm da proficiência de um aluno submetido a dois testes construídos com os itens desse banco serão iguais (Brasil, 2012b), com tanto que se tenham os mesmos itens-âncoras nos dois testes, pois esse é o viés que possibilita os estudos comparativos. Em outras palavras, a *TRI* permite que **indivíduos** e **itens** possam ser localizados na mesma escala, e esta é uma das grandes vantagens sobre a TCT.

⁶³ Também denominado “*Banco Nacional de Itens*” (BNI) administrado pelo INEP. Nele são realizados chamadas públicas com o intuito de convocar interessados em elaborar e revisar itens para a composição de provas de avaliações desenvolvidas por este órgão. Fazem parte da BNI os exames: *PROVINHA BRASIL*; Sistema Nacional de Avaliação da Educação Básica - SAEB; ENEM; ENCCEJA; PISA; ENADE; *Revalidação dos Diplomas Médicos - Revalida*; *Prova Nacional de Concurso para o Ingresso na Carreira Docente*; *Certificado de Proficiência na Língua Brasileira de Sinais - PROLIBRAS*; *Certificado de Proficiência em Língua Portuguesa - CELPE-BRAS*.

Tendo em vista que o acerto casual é considerado no exame tanto do SAEB quanto do ENEM, é importante lembrar que a medida de proficiência da TRI leva em conta não só o número de acertos, mas também o *padrão de respostas do aluno*. Diante disso, *dois alunos com o mesmo score podem receber da TRI diferentes valores de proficiência*. Receberá maior proficiência aquele aluno que apresentar respostas aos itens de forma mais coerente com o construto que está sendo medido.

Ainda assim podem surgir problemas dentro do processo seletivo classificatório dos candidatos (ENEM) que obtém o mesmo nível de proficiência em valores absolutos. Surge então a necessidade de ajustes para um maior refinamento e precisão, e não é possível obter esse valor diretamente, mediante a uma fórmula simples, é preciso utilizar um método matemático mais sofisticado para calcular e definir a nota. Entre diversos métodos para o cálculo da nota, o SAEB e o ENEM utilizam o método denominado *Expected a Posteriori* (EAP).

“O método EAP tem como princípio usar uma função de probabilidade *a priori* para calcular a nota. Essa função tem a informação da nota média e do desvio padrão dos participantes concluintes do ensino médio da rede pública que realizaram o ENEM em 2009”. (Brasil, 2012, p. 31)

O método EAP pode ser representado pelas seguintes funções matemáticas (Brasil, 2012a, p. 32):

$$E(\theta|X) = \frac{\int_R \theta \cdot P(X|\theta) \cdot f(\theta) \cdot d\theta}{\int_R P(X|\theta) \cdot f(\theta) \cdot d\theta}$$

Equação 8: Método EAP.

Legenda:

- $f(\theta)$ é a função de probabilidade *a priori*;
- $P(X|\theta)$ é uma função matemática associada ao padrão de respostas dos participantes e aos parâmetros dos itens, valores conhecidos;
- θ representa a proficiência do participante que será calculada.

Para resolver a expressão acima, o INEP (Ibid., 2012, p. 32) adota o *método da quadratura Gaussiana* sugerida por Stroud & Secrest (1966). A fórmula da quadratura Gaussiana é (Stroud & Secrest, 1966, p. 1):

$$\int_a^b w(x) \cdot f(x) \cdot dx \approx \sum_{i=1}^n A_i \cdot f(x_i)$$

Equação 9: Método da Quadratura Gaussiana.

Sendo:

- $w(x)$ é uma função para os momentos (integrais monomiais);
- $f(x)$ é um função polinomial;
- x_i são os pontos (ou nós) da fórmula;
- A_i são os chamados coeficientes (ou pesos)

Fazendo a analogia, tem-se (Andrade et al., 2000, p. 106):

$$\hat{\theta} = \frac{\sum_{k=1}^q X_k \cdot P(X_k) \cdot f(X_k)}{\sum_{k=1}^q P(X_k) \cdot f(X_k)}$$

Sendo:

- Os pontos médios de cada retângulo, $\hat{\theta}_k$, $k = 1, \dots, q$, são denominados de nós (ou pontos de quadratura);
- $\hat{\theta}$ representa a estimativa de θ , sendo, portanto, a nota dos participantes do ENEM.

2.3.4.4 Accountability: alguns contrapontos

“Eu acredito em automóveis padronizados. Eu não acredito em pessoas padronizadas. A padronização é um grande perigo que ameaça a cultura Americana”. (Albert Einstein)

É preciso considerar os riscos existentes no exercício demasiado de uma concepção educativa que explora o conceito de padronização, seja de conteúdos como de desempenho educacional, assim posicionar uma função adequada para o uso de testes avaliativos da psicologia voltados para cognição dentro do processo de aprendizagem, em princípio diante dos conhecimentos, mas também considerando relevante o papel dos processos mentais envolvidos. Esse estudo declara defender um enfoque moderador dentro do processo educativo, em que o uso de testes padronizados deve exercer uma função mais voltada para a orientação (diagnóstica, de previsão e predição) e regulação (formativa, podendo ainda ser diagnóstica), estando a serviço tanto do professor na melhoria de seu ensino, mas fundamentalmente para o aluno na otimização de sua aprendizagem e necessidades.

A avaliar é natural em nosso cotidiano, que deve passar de modo geral em três etapas: *constatar*, *refletir* e *agir*. Na área da educação é um processo complexo e muitas vezes conduzida e concebida de forma equivocada nas escolas, onde um aluno malsucedido em seu desempenho é considerado um problema e condicionando uma necessidade de acompanhamento extra através de aulas de reforço com professores “particulares”, predizendo uma condenação o seu sucesso caso não seja feito. Essa

postura pode ser entendida como uma forma de abdicação de sua função propedêutica em ajudar alunos com maiores dificuldades na aprendizagem, com o argumento de que não acompanham o andamento de aprendizagem imposto pelos professores. Essa imposição pode ainda estar relacionada a um princípio funcional da escola, quase sempre preocupada no cumprimento de ementas a qualquer preço, isso condiciona alguns professores a privilegiarem um ensino conteudista, que sua função principal é ensinar única e exclusivamente conteúdos. Diante de um ensino conteudista, em que se concentra em um modelo de ensino de transmissão de conhecimentos, vale a pena ressaltar que atualmente, as diversas mídias de tecnologia da informação disponíveis já conseguem exercer esse papel de forma eficiente, seja em plataformas de vídeos explicativos entre outros.

Muitos especialistas em educação consideram a preocupação das escolas, professores, e com o consentimento dos pais, em ensinar conteúdos que estejam relacionados aos critérios de avaliação externa, seja em decorrência dos exames nacionais classificarem as escolas em boas e ruins como na função seletiva em condicionar o ingresso de alunos em novas etapas e níveis de escolarização, exercendo assim um papel deformador na área da educação. O uso desses testes como instrumento avaliativo na educação é um campo controverso entre especialistas de avaliação, na qual as críticas passam desde uma função a privilegiar uma hegemônica psicologização para os fenômenos de fracasso escolar (medicalização e diagnósticos sobre transtornos de aprendizagem, etc.), como seu uso demasiado sobrecarregando os alunos, a imposição de padrões de conteúdo e desempenho tendo efeitos reducionistas para o currículo (“currículo mínimo”) e conseqüentemente para a formação dos sujeitos, a rigidez de critérios avaliativos certificatórios ao exercer funções seletivas e classificatórias, assim como das limitações de avaliar o sistema como um todo com base em um modelo de APC.

Críticas relacionadas ao uso de testes padronizados são enfatizados pela associação a interesses macroeconômicos nacionais e internacionais voltados a atender as necessidades do mercado de trabalho, se configurando assim numa tentativa de padronização de comportamento e da aprendizagem, na qual escolas são

responsabilizadas (“*accountability*”) pelo desempenho dos alunos. Esse é um modelo que vem ganhando força no Brasil e recebendo críticas devido a supostas influências do modelo americano, também muito criticado como mostra o documentário “*Standardized Lies Money & Civil Rights How Testing is Ruining Public Education*”. Acredita-se, entretanto, que este modelo se torna um risco quando malsucedido.

A *responsabilidade* e o termo *accountability* são considerados muitas vezes sinônimos. Do ponto de vista do indivíduo, a *responsabilidade* é um processo relacionado aos deveres e capacidades profissionais em atingir suas metas e tarefas que lhe são atribuídas para desempenhar sua função, seguindo regras e procedimentos, ou seja, está ligada ao um sentido propedêutico das competências funcionais. O termo *accountability* se refere às capacidades que se tem de apropriar das situações em que está envolvido como resultado de suas próprias escolhas e ações. Essa perspectiva é geralmente atribuída como política de gestão no sentido de “prestação de contas”, porém, em uma perspectiva individual de trabalho podem se estar associadas. Individualmente é uma virtude pessoal, relacionada uma capacidade predisposta de se apropriar da responsabilidade e procurar gerar respostas com resultados positivos. Nesse sentido, uma pessoa pode ter responsabilidade sobre uma atividade ou projeto e não apresentar *accountability*, não ser uma pessoa *accountable*, o inverso é óbvio, sendo *accountable* evidentemente a pessoa é responsável, porém em um patamar mais elevado.

Alves & De Ketele (2011) apresentam fortes críticas aos modelos matemáticos na qual se inclui a TRI dentro de políticas educacionais *accountability* para a avaliação de sistemas educacionais. Consideram excelentes para a avaliação dos recursos, mas «inadequados para a avaliação de competências», pois acarretam pressupostos que infringem a aceção do termo, como a unidimensionalidade e reducionismo na abordagem e avaliação de conteúdos, além da exclusão de itens considerados essenciais na perspectiva das competências.

De Ketele e Gerard (2005) mostram que este modelo (TRI, Teoria da Resposta ao Item Rasch), assenta no pressuposto da unidimensionalidade do objeto a avaliar (que pode ser o caso de um certo número de recursos a avaliar), enquanto, por

definição, **uma competência é necessariamente multidimensional**. A análise que fizemos das provas do PASSEC, como parte de um estudo coordenado pelo CIEP (2009), mostra que as provas não avaliam nenhuma competência, quando muito, e no melhor dos casos, avaliarão “as suas aplicações contextualizadas”. O projeto HARMOS, na Suíça, está, atualmente, em grande dificuldade, porque o recurso à TRI levou à rejeição de muitos itens, incluindo os mais importantes (**competências tão importantes quanto a produção escrita são inevitavelmente postas de lado**). A Bélgica francófona vive um conflito semelhante: os conceptores das provas nacionais atuais, querem, de forma a serem coerentes, introduzir uma parte da prova que avalie as competências; mas não conseguem, porque as autoridades ministeriais, em última análise, que querem proteger-se (e compreendemo-las), refugiam-se atrás das estatísticas que recomendam a TRI. Não é de admirar que os professores que analisam atentamente as provas que são obrigados a passar aos seus alunos, apenas ensinem e avaliem os recursos e não as competências impostas por decreto. (Alves & De Ketele, 2011, p. 19, *grifo nosso*)

Apesar das críticas, se entende que os autores reconhecem a relação dos itens com a perspectiva das competências, mas que o rigor dos critérios estatísticos põe em risco eliminar itens que representam competências essenciais, além de acusarem os interesses políticos perversivos por esses mecanismos, da incoerência unidimensional concebida, e das distorções que tendem a causar no cenário pedagógico. Algumas dessas limitações são críticas comuns no contexto das avaliações externas no Brasil, entretanto, o processo de calibração e ajuste no enunciado dos itens durante as análises preliminares podem assegurar a consistência daqueles considerados mais relevantes, que a perversão nos interesses político-ideológicos são causas externas e de natureza distintas aos modelos psicométricos, que por sua vez, são considerados robustos por especialistas na área da avaliação psicológica, em especial, os modelos unidimensionais, entretanto, ainda pode ser contornado utilizando modelos multidimensionais da TRI como alternativa (Oliveira, 2014, 2015).

O uso de testes padronizados na avaliação e gestão de sistemas é geralmente configurada dentro das políticas *accountability* (Vaughan, 2002), sendo um termo polissêmico e multidimensional (Afonso, 2009) tem sido objeto de análise no contexto brasileiro e considerado em construção (Pinho & Sacramento, 2009), assim como no contexto português, em grande parte está relacionado ao desenvolvimento de processos de prestação de contas e de responsabilização (Afonso, 2011) nos diversos níveis e dimensões dos sistemas de gestão administrativa. Nessa perspectiva Afonso

(2011) apresenta algumas críticas e polêmicas quanto a acepção do termo diante de modelos de avaliação dentro de políticas educacionais baseadas em testes estandardizados. Entre os exemplos frequentemente criticados e citados como política malsucedida está a reforma educativa de George W. Bush, conhecida como *No Child Left Behind Act* - NCLB (“Nenhuma criança deixada para trás”), em que o autor ressalta (Ibid., 2011, p. 89): “Para aquela organização, este modelo de *accountability* não apenas mostrou grandes insuficiências como criou efeitos secundários indesejáveis (cf. FairTest, 2007)”.

Uma noção de *accountability* criticada remete ao termo *high-stakes testing* (“*testagem de alto-risco*” ou “*testes de alto impacto*”), em que os examinadores exercem em última instância uma função certificadora e determinante na avaliação a partir de testes aplicados aos alunos (entre outras situações), podendo remeter reprovação em um nível de escolaridade, perda de bolsas de estudo, licença para o exercício profissional, entre outros. O exercício dessa função para os testes é atualmente alvo de fortes resistências entre especialistas na área da educação, setores sociais, e prestigiadas associações científicas⁶⁴ nos Estados Unidos, ainda porque constituem como **único fator para a tomada de decisões cruciais** sobre as escolas, os professores e estudantes (Ibid., 2011, p. 90). Na análise de uma década do programa NCLB (chamada de “*década perdida*”), destacam-se três aspectos negativos:

- 1) A NCLB tem prejudicado seriamente a qualidade e a equidade da educação, com efeitos de estreitamento e limitação que afligem os pobres de modo mais implacável.
- 2) A NCLB fracassou significativamente em aumentar a performance acadêmica média ou em diminuir expressivamente as diferenças de desempenho.
- 3) As chamadas “reformas” da lei, como as isenções do governo Obama e a revalidação, pela Comissão de Educação do Senado, da *Elementary and Secondary Education Act* (ESEA - Lei do Ensino Fundamental e Médio), não afetam muitos dos erros fundamentais da NCLB e, em certos casos, agravaram-nos, pois vão dar continuidade a uma “*década perdida para as escolas estadunidenses*”. (Guisbond, Neill, & Schaeffer, 2012, p. 407)

⁶⁴ American Evaluation Association (AEA). Ver: <www.eval.org/hst3.htm>.

Esses contrastes servem de parâmetros moderadores para uma política de EBE, em que, o mal-uso na implementação de um sistema de gestão de responsabilização e *accountability* na qual testes exercem funções exclusivas certificadoras podem acarretar uma distorção educacional, e ainda ser perversiva para as práticas educacionais, causar falhas nas políticas internas de gestão administrativa, induzir atividades docentes e principalmente causar injustiças para os estudantes.

Quadro 41: O uso de testes e o efeito da concorrência em resultados educacionais.

Autor (ano)	Diferença em desvio padrão	Efeito significativo	Metodologia	País	Variável de responsabilidade
Koretz & Barron (1998)	0,15	sim	Análise de Tendência	Kentucky - EUA	Prêmio/sanções
Ladd (1999)	0,14	sim	Estimação de Painel	Dallas - EUA	Prêmio financeiro
Klein <i>et al.</i> (2000)	0,08	sim	Análise de Tendência	Texas - EUA	Prêmio financeiro
Greene (2001)	0,28	misto	Médias de teste em grupos comparáveis	Florida - EUA	Ameaça de sanção
Carnoy & Loeb (2002)	5,25*	sim	Modelo recursivo	EUA	Prêmio/sanções
Shaw <i>et al.</i> (2003)	-0,004*	misto	Modelo multinível	Inglaterra	Inspeção/Ameaça de sanção
Rosenthal (2004)	-0,004*	sim	Estimação de Painel	Reino Unido	Inspeção/Ameaça de sanção
Hanushek Raymond (2004)	0,08	sim	Estimação de Painel	EUA	Prêmio/sanções
Jacob (2005)	0,29	sim	Estimação de Painel	Chicago - EUA	Ameaça de sanção
Figlio & Rouse (2005)	0,06	misto	Diferença em diferença	Flórida - EUA	Ameaça de sanção
West & Peterson (2006)	0,04	sim	Análise de regressão com grupos comparáveis	Flórida - EUA	Ameaça de sanção
Rouse <i>et al.</i> (2007)	0,09	sim	Regressão descontínua	Flórida - EUA	Ameaça de sanção
Reback (2007)	0,008	sim	Efeitos fixos por escola-ano	Texas - EUA	Ameaça de sanção
Springer (2007)	0,15	sim	Efeitos fixos de escola, classe e ano	EUA	Ameaça de sanção
Rockoff & Tumer (2008)	0,075	sim	Regressão descontínua	Nova Iorque - EUA	Ameaça de sanção
Winters, Greene & Trivitt (2008)	0,08	sim	Regressão descontínua	Flórida - EUA	Ameaça de sanção
Chalcrabarti (2008)	0,355	sim	Diferença em diferença	Flórida - EUA	Ameaça de sanção
Ahn & Vigdor (2009)	0,01	não	Regressão descontínua	Carolina do Norte - EUA	Ameaça de sanção
Chiang (2009)	0,115	sim	Diferença em diferença	Flórida - EUA	Ameaça de sanção
Gill <i>et al.</i> (2009)	1,97*	misto	Regressão descontínua	Dois Estados dos EUA	Ameaça de sanção
Luginbuhl, Webbink & de Wolf (2009)	0,02	misto	Efeitos fixos de escola / projeto experimental	Holanda	Inspeção / Ameaça de sanção
Dee & Jacob (2011)	0,13	misto	Diferença em diferença	EUA	Prêmio/sanções
Lee e Reeves (2012)	0,02	misto	Diferença em diferença	EUA	Prêmio/sanções
Burges, Wilson & Worth (2013)	0,08	sim	Diferença em diferença	País de Gales e Inglaterra	Abolição de informação de desempenho
Sims (2013)	-0,125	sim	Regressão descontínua	Califórnia - EUA	Ameaça de sanção
Siegler (2013)	0,13	sim	Regressão descontínua	Flórida - EUA	Ameaça de sanção
Custard (2014)	2,13*	sim	Regressão descontínua - Regressao Quantitativa	Geórgia - EUA	Ameaça de sanção

Fonte: Christophe et al. (2015, p. 99)

Estudos preliminares mostram que os modelos de *accountability* que se concentraram “**apenas nos resultados em testes padronizados**, eles não conseguiram controlar de modo adequado outros componentes associados às políticas de responsabilização das escolas, como foi o caso da Carolina do Norte, Texas e Kentucky, nos EUA” (Christophe et al., 2015, p. 97), ou seja, a exploração excessiva desses mecanismos pode se tornar um risco associado. Apesar disso, 27 estudos relacionados aos sistemas estaduais dos EUA e da Europa **reforçam o fator positivo de políticas de responsabilização** (Quadro 41), na qual encontra-se incluído o modelo do NCLB. Em 2015, o presidente Barak Obama assinou uma nova lei que substituiu o NCLB, chamado *Every Student Succeeds Act* (ESSA), em que se refere a uma lei de sucesso para todos os estudantes, com as mudanças publicadas recentemente (Congress of the United States of America, 2017). Esta busca superar alguns problemas e críticas do NCLB, como aspectos críticos considerados punitivos, descentralização do controle federativo para os estados, porém **foi mantida a política do uso de testes padronizados** nas escolas em leitura e matemática, tendo as metas de desempenho sendo estabelecidas para cada estado, assim como um maior controle para evitar um demasiado uso de testes.

Apesar disso, o Brasil adota esse tipo de política ao utilizar os indicadores do PISA, SAEB, entre muitos outros que compõem o IDEB, para a tomada de decisões e estabelecimento de metas, como por exemplo a meta 7 do *Plano Nacional de Educação 2014-2024* (Brasil, 2014).

Tabela 7: Meta 7 - Fomentar a qualidade da educação básica em todas as etapas e modalidades, com melhoria do fluxo escolar e da aprendizagem de modo a atingir as seguintes médias nacionais para o IDEB.

IDEB	2015	2017	2019	2021
Anos iniciais do ensino fundamental	5,2	5,5	5,7	6,0
Anos finais do ensino fundamental	4,7	5,0	5,2	5,5
Ensino médio	4,3	4,7	5,0	5,2

Fonte: Brasil (2014, p. 61).

Os valores do PISA mostram que a média do Brasil encontra-se abaixo da média dos países da OCDE, e com isso, sinalizam a necessidade de investimentos na formação docente, infraestrutura e reestruturação no sistema educacional.

Tabela 8: Estratégia 7.11 – Melhorar o desempenho dos alunos da educação básica nas avaliações da aprendizagem no Programa Internacional de Avaliação de Estudantes (PISA), tomado como instrumento externo de referência, internacionalmente reconhecido, de acordo com as seguintes projeções.

PISA	2015	2018	2021
Média dos resultados em matemática, leitura e ciências	438	455	473

Fonte: Brasil (2014, p. 63).

Mesmo adotando uma política de responsabilização pelos estados federativos com suas próprias metas, seja premiando escolas, gestores, professores e alunos pelo mérito de desempenho, ainda são **escassos os estudos sistemáticos na área da educação tendo como base as evidências** (Christophe et al., 2015). Desde 2013 o Brasil não vem alcançando as metas (médias totais) do IDEB, com exceção para os anos iniciais do ensino fundamental, apesar disso os estados mais ricos continuam com médias superiores aos mais pobres. O exemplo de boas práticas na política de responsabilização na educação da cidade de Sobral, no estado do Ceará, tem se tornado **modelo de gestão a sendo replicado em todo o país**, modelo incluso na base do PNE atual, e mostra que mesmo sendo um contexto adverso com uma renda capita igual ou inferior a R\$ 255,00 mensais, esse tipo de política é um caminho promissor a ser seguido, porém é necessário avançar em estudos na área da educação que adotem essa concepção (“Sobral vira modelo nacional de gestão - Educação - Estadão,” 2014).

A evidência proveniente de estudos realizados em **países da América Latina** ou de renda média ou baixa é **quase inexistente**, porque os sistemas de responsabilização, em geral, têm sido implementados em resposta ao aumento de gastos com educação, **sem que tenha havido melhorias significativas e observáveis na aprendizagem dos alunos** (Hanushek, 1994). Nos países de renda média ou baixa, o gasto com educação ainda é baixo. Uma exceção é o Chile, onde recentemente se implementou uma política de responsabilização que ameaça fechar as escolas com baixo desempenho, caso não o melhorem em um período de quatro anos. **Ainda**

não existe um estudo sistemático dos efeitos dessa política sobre os resultados de aprendizagem. (Christophe et al., 2015, p. 98, *grifo nosso*)

Em Portugal, os exames nacionais é uma política reinserida em 1996, e possui algumas peculiaridades, uma delas é não adotar a TRI em seus exames nacionais.

Quadro 42: Contexto históricos dos exames nacionais em Portugal.

Contexto históricos dos exames nacionais em Portugal	
1926 a 1974 (período ditatorial)	Estabeleceu-se exames nacionais obrigatórios, estando a progressão e certificação dos alunos dependentes deste processo, sendo posteriormente abolidos após o regime ditatorial (Afonso, 2011).
1974 a 1995	Foi adotado um sistema de avaliação de aprendizagens essencialmente interna e baseada na escola (Fernandes, 2007, p. 589, citado por Afonso, 2011, p. 92)
1996	Os exames nacionais foram reintroduzidos no final do ensino secundário, porém, as avaliações internas prevaleceram e prevalecem até os dias atuais com um peso decisório em termos de progressão e certificação dos estudantes
2001	Divulgação pública (no <i>mass media</i> ⁶⁵) dos resultados dos exames externos estandardizados, sendo constituído como base para a organização de <i>rankings</i> das escolas.
2005	Os exames sendo aplicados no final da escolaridade obrigatória.

Fonte: Elaborado a partir de Afonso (2011)

Afonso (2011) defende uma *accountability* voltada para uma educação democrática, em que:

“[...] pode e deve ser conceptualizada como um sistema denso do ponto político, axiológico e epistemológico, bem como deve ser assumida como um sistema complexo em termo de procedimentos, dimensões e práticas em que a avaliação, a prestação de contas e a responsabilização estarão, sempre que possível, integradas (ou serão integráveis).” (Afonso, 2011, p. 97)

Apesar disso, apresenta algumas críticas no contexto português para o modelo de *accountability* com predominância em exames nacionais ou testes estandardizados, na qual a política de *rankings* de escolas tem pressionado e reconfigurado as funções cotidianas do professor, possuindo um fator de peso em termos de progressão e certificação dos estudantes, além de ser considerado por grande parte dos professores que os critérios utilizados na elaboração dos *rankings* não possuem credibilidade e confiança. Afirma que o processo de avaliação dos professores e das escolas seja uma forma “parcelar” de *accountability* ao adotar como **único fundamento o desempenho**

⁶⁵ Meios de comunicação de grande massa, como televisão, rádio, filmes, entre outros, na incluem-se ainda as mídias digitais.

dos estudantes, sendo assim um modelo desarticulado e incongruente, que na realidade não configura um modelo legítimo de *accountability*, mas sim um modelo com orientações políticas e normativas em processo de evolução que caminha a curto e médio prazo para esse modelo (Ibid., 2011, p. 94).

CAPÍTULO 3 - METODOLOGIA

3.1 FASE PRÉ-DESCRITIVA OU EXPLORATÓRIA - Indícios de Credibilidade Investigativa, Avaliação do Instrumento e o Teste-Piloto

Os métodos científicos consagrados são o *Positivismo* de Auguste Comte; a *Fenomenologia* de Edmund Husserl; o *Estruturalismo* de Ferdinand de Saussure e Claude Lévi-Strauss; e o *Materialismo Histórico-Dialético* que parte de contribuições de Sócrates & Platão, e culmina com Karl Marx & Engels, mais propriamente, também denominada de «diamate». E quanto as principais *teorias da aprendizagem*, essas podem ser classificadas em: *Associacionistas* (Pavlov, Watson, Guthrie, Hull, Thorndike e Skinner) e *Mediacionais* (Kofka, Köhler, Whertheimer, Maslow, Rogers, Piaget, Vygotsky).

O “método” nas ações de um pesquisador é uma das bases legítimas para se “fazer” ciência ou se construir conhecimento científico, ou seja, não há ciência sem método, epistemologicamente falando, sendo a recíproca não necessariamente verdadeira. Esse é um processo muito discutido dentro do campo gnosiológico da filosofia, e um componente essencial na busca de uma verdade concreta ou como aliado aos fundamentos de um *critério da verdade* de um conhecimento (Bazarian, 1994).

Legitimar nas bases da ciência não significa considerá-lo com *status* único e rígido, sendo imperativamente válido por determinada forma, apesar de necessariamente propedêutico na conduta de um pesquisador pós-aspirante. Diante disso, surgem novas exigências e maiores competências nas suas ações ligadas ao rigor científico, não somente nos momentos de elaboração/definição de estratégias e desenhos metodológicos, mas fundamentalmente forma de conduzir os procedimentos na coleta dos dados, onde a consciência, conhecimento e a diligência

no *modus operandi* na pesquisa de campo. O processo de recolha de informações deve ser tratado em seu devido e legítimo valor epistemológico na *metodologia científica*, fato negligenciado em muitos livros nesse segmento (De Ketele & Roegiers, 1993).

3.1.1 Delimitações Sobre os Aspectos de Credibilidade

A intenção de destacar e descrever aspectos de credibilidade está em apresentar indícios de coerência interna no estudo, assim como pontos fracos, delimitações e compromisso com a clareza e objetividade. Podem ser destacados três aspectos fundamentais para a qualidade na investigação: *pertinência*, *validade* e *fiabilidade*. Em linhas gerais, podem ser definidas como (De Ketele & Roegiers, 1993):

- A **pertinência** se refere a uma característica mais ou menos apropriada que se inscreve na linha do objetivo visado.
- A **validade** se refere ao grau de adequação entre o que se pretende fazer (avaliar ou recolher informação) e o que se faz realmente.
- A **fiabilidade** se refere a qualidade que consiste em encontrar os mesmos resultados, quer em pessoas diferentes quer de uma mesma pessoa em momentos diferentes, quer por uma nova pessoa em relação a um protocolo definido, etc.

O controle da influência de fatores intrapsicológicos e subjetivos do pesquisador dentro do processo investigativo também pode ser considerado um fator de credibilidade, e é denominado como *neutralidade*, aspecto esse, considerado fundamental na maioria dos tipos de pesquisa, sendo necessário e desafiador em abordagens qualitativas.

O estudo incide sobre diferentes processos de análise, adotando o método hipotético-dedutivo na abordagem quantitativa, com um enfoque empírico-analítico (busca de evidências empíricas e suas análises), e na abordagem qualitativa um enfoque fenomenológico-interpretativo ou fenomenológico-hermenêutico (busca de uma compreensão sobre as influências que envolvem o fenômeno de escolha dos respondentes). Em parte, se admite um pressuposto na abordagem da pesquisa sendo de natureza gnosiológica (mais precisamente, ao se referir a uma concepção de objeto

e de sujeito e a sua relação no processo do conhecimento (empírico-analítica, centrado na objetividade) (Barbosa & Miki, 2007):

Os dados são obtidos por observação controlada, com instrumentos previamente testados, enunciados de uma única interpretação e uma codificação numérica. Os dados são analisados sem a subjetividade do pesquisador pelo raciocínio lógico-dedutivo (Barbosa & Miki, 2007, p. 53).

Por outro lado, há uma investigação sobre os processos do conhecimento envolvidos previamente, estando preconizados no teste, quanto as análises sobre os mecanismos de acesso ao conhecimento disponível do respondente, e sobre as interpretações pedagógicas das respostas com base no conteúdo, nos resultados e dos parâmetros psicométricos, delimitando nesse âmbito uma ênfase interpretativo-analítica e comparativa entre as análises. Diante disso, foi elaborado um quadro para organizar e sistematizar as ações tomadas dentro de cada critério de credibilidade.

Quadro 43: Critérios de credibilidade da investigação diante das principais análises e processos realizados de acordo com os enfoques-metodológicos.

		Paradigmas/ Enfoque metodológico	
		Hipotético-dedutivo	Fenomenológico-interpretativo
		Critérios de credibilidade e análises/processos	
Veracidade	Validade interna	<ul style="list-style-type: none"> Validade da chave-de-correção por meios da estatística descritiva das opiniões de pares (“juízes”); Validade de construto do TCE por meio da análise de multidimensionalidade, de acordo com o modelo adotado Identificação do tipo de <i>design</i>. 	<ul style="list-style-type: none"> Confirmação da chave-de-correção e validade dada ao instrumento TCE de acordo com a creditação dada ao site https://www.physport.org/ pela AATP; Validade de conteúdo com base na análise taxonômica dos itens do TCE; Validade de construto do TCE por meio da análise interpretativa da natureza tipológica das concepções incidentes nos itens, assim como seu quantitativo e disponibilização no teste; Validade interna com base na análise interpretativa a partir de um <i>check-list</i> de potenciais riscos associados ao <i>design</i>.
Aplicabilidade	Validade externa/Generabilidade	<ul style="list-style-type: none"> Representatividade da amostra entorno de 20% da população estabelecida em cada contexto; Identificação do tipo de <i>design</i>; Consideração das respostas de todos os sujeitos da amostra na base de dados; Identificação da estimação de parâmetros relativamente estáveis entre os contextos; Análise comparativa de desempenho entre diferentes grupos por meio dos escores brutos da TCT e das habilidades estimadas na TRI; Análise de DIF por meio de técnicas de Mantel-Haenszel, regressão logística e de DIF uniforme diante das dificuldades entre diferentes grupos de desempenho e de gênero entre os contextos. 	<ul style="list-style-type: none"> Validade externa e ecológica com base na análise interpretativa a partir de um <i>check-list</i> de potenciais riscos associados ao <i>design</i>; Interpretação pedagógica dos parâmetros estimados relativamente estáveis entre os contextos com implicações particularizadas de grupos específicos de desempenho dentro de cada contexto; Interpretação pedagógica de DIF diante das dificuldades entre diferentes grupos específicos de desempenho e de gênero entre os contextos; Interpretação da análise comparativa de desempenho entre diferentes grupos específicos por meio dos escores brutos da TCT e das habilidades estimadas na TRI.
			Credibilidade
			Transferibilidade

Consistência	Fiabilidade	<ul style="list-style-type: none"> • Escolha do modelo logístico mais ajustado aos dados, de acordo com os pressupostos da TRI; • Eliminação de itens com baixa consistência; • Aceitação da análise de acordo com critérios mínimo de consistência do valor do alfa de Cronbach; • Seleção de itens-âncoras para a construção da escala de proficiência segundo critérios estatísticos; • Análise de consistência das informações na escala de proficiências através da matriz item-sujeito, de acordo com o modelo da TRI; • Interpretação pedagógica com base em índices/parâmetros de discriminação e análise da CII. • Verificação do comportamento dos itens em diferentes modelos logísticos. 	Confiança	<ul style="list-style-type: none"> • Verificação da chave-de-correção do teste TCE com base em análises interpretativa e crítica das respostas de 5 especialistas no ensino da Física, 4 deles sendo doutores; • Validade da adaptação transcultural do teste TCE através da análise e julgamento de 4 pesquisadores e doutores na área do ensino, 3 deles com experiência na área da avaliação, e 3 sendo do campo do ensino da Física.
Neutralidade	Objetividade	<ul style="list-style-type: none"> • Interpretação com base em critérios estatísticos e dados obtidos; • Inexistência de interferência no ambiente de sala de aula a eventualmente enfatizar uma abordagem da unidade temática ou de influenciar os escores na aplicação do teste; • Interpretação subjetiva dos parâmetros psicométricos comprometida com uma coerência horizontal (evidências empíricas <i>versus</i> fundamentação teórica). 	Confirmabilidade	<ul style="list-style-type: none"> • Comparação da chave-de-correção com outros estudos; • Comparação de resultados com outros estudos quanto as dificuldades dos alunos no teste TCE.

Fonte: Elaborado a partir de Amado (2013, p. 358).

Enfim, é comum o surgimento de dúvidas quanto aos procedimentos mais eficientes e pertinentes ao estudo. Para ilustrar, a relação que deve existir entre os diferentes tipos de instrumento para a recolha de informações e o tipo de variável que se pretende “medir” (ou quantificar), concretizado na sua definição operacional, pode ser:

Quadro 44: Tipo de instrumento em função do tipo de variável.

Tipo de Instrumento	Questionários	Escala, Inventário	Grelha de observação; entrevista	Testes
Objetivo do instrumento (tipo de dados a obter)				
Classificação/ordenação (variáveis nominais/ordinais)				
Background pessoal (idade, sexo, nível de instrução)	X			
Classe social (indicadores)	X		X	
Tipo de organização (escola)	X		X	
Preferências (políticas, disciplinares, etc.)	X		X	
Traços quantificáveis que originam dados de natureza contínua (variáveis intervalares)				
Atitudes, percepções, opiniões, grau de empenho		X	X	
Valores		X	X	
Aproveitamento			X	X
Aptidões			X	X
Inteligência, criatividade			X	X

Fonte: Coutinho (2011, p. 108, adaptado de Black, 1999, p. 191).

3.1.1.1 Pertinência

A *pertinência nos critérios avaliativos* é declarada pretendendo atender os objetivos visados com o máximo juízo de valor. A pertinência é a característica mais

ou menos apropriada que se inscreve na linha do objetivo analisado (De Ketele et al, 1989, citado por De Ketele & Roegiers, 1993). Em outras palavras, está relacionada com a *adequação do objeto em relação ao objetivo visado*. Segundo De Ketele & Roegiers (1993, p. 74): “um critério é sempre um objetivo de mais fraco nível em relação ao objetivo global. Exprime-se em termos qualitativos” (1993, p. 74), se busca responder a seguinte indagação: *Os critérios permitem verificar o que afirmo querer verificar?* Para isso se torna necessário primeiramente situar de forma apropriada os critérios dentro da natureza avaliativa, ou seja, compreender a que tipo de avaliação ela corresponde.

Compartilha-se da definição de que “*avaliar significa confrontar um conjunto de informações com um conjunto de critérios com o fim de tomar uma decisão*” (De Ketele, 1980, citado por De Ketele & Roegiers, 1993, p. 46). As avaliações de modo geral estão divididas em três finalidades (Ibid., 1993): orientação, regulação e certificação, e elas podem exercer as seguintes especificidades:

- **Avaliação de orientação:** *Avaliação preditiva* – consiste em prever as hipóteses de êxito de uma pessoa num domínio de formação, mantendo-se as coisas iguais; *Avaliação de previsão* – consiste em antecipar a realidade, emitindo hipóteses sobre o futuro (para além das possibilidades de êxito), ou seja, avalia-se a oportunidade de tomar uma determinada decisão em função da evolução estimada deste ou daquele parâmetro; *Avaliação diagnóstica* – para identificar forças e fraquezas de um sistema ou de uma pessoa, podendo estar a serviço da orientação (por ex.: antes da aprendizagem) ou da regulação (p. ex.: durante a aprendizagem);
- **Avaliação de regulação:** *Avaliação formativa* – se aplica a uma pessoa em aprendizagem, ou seja, diretamente a serviço do indivíduo visando determinar decisões apropriadas às dificuldades de um estudante, a fim de o fazer progredir ou muito simplesmente de melhorar a sua progressão. Tem como objetivo localizar o mais precisamente possível a origem das suas dificuldades a fim de resolver, por exemplo através de uma análise dos seus erros. Ou ainda ter a intenção de modificar a situação de aprendizagem ou o ritmo desta progressão, para introduzir melhorias ou corretivos apropriados (Gérard Scallon 1988, p. 155, citado por De Ketele & Roegiers, 1993, p. 53);

Avaliação Formadora – consiste se apropriar das dificuldades na aprendizagem e rever as estratégias de ensino, estando, portanto, a serviço do professor

- **Avaliação de certificação:** *Avaliação de seleção* – realizada ao final de um processo visando selecionar um candidato (De Ketele & Roegiers, 1993, p. 54). Pode-se ainda selecionar recursos (p. ex.: livro didático mais apropriado, aparatos físicos, aplicativos, etc.) ou determinados procedimentos, atividades ou métodos, que no caso, pode estar a serviço da aprendizagem e/ou do ensino; *Avaliação de Classificação* – diz respeito principalmente aos desempenhos das pessoas. Prolonga uma avaliação de certificação, ordenando as pessoas umas em relação às outras (pode eventualmente exercer a função de certificação); *Avaliação aditiva (ou somativa)* – está associada ao estabelecimento de um balanço que descreve a soma dos conhecimentos adquiridos. O termo “aditivo” refere-se principalmente aos meios postos em prática para avaliar e não ao objetivo visado pela avaliação (a tomada de decisão), não sendo exclusivamente uma avaliação de certificação (no caso, pode eventualmente ser aplicada no processo formativo), entretanto está mais ao serviço da função (põe-se a questão de saber se determinada pessoa corresponde ao perfil esperado).

Dentro dos processos avaliativos, os critérios de avaliação a enunciar, em síntese, podem ser organizados em **nível da ação** quando o objeto a ser avaliado se refere ao funcionamento ou qualidade de um recurso ou de um sistema, ou em **nível da pessoa** (De Ketele & Roegiers, 1993) quando o objeto a ser avaliado se refere ao desempenho de sujeitos individualmente ou de grupos.

Quadro 45: Tipos de avaliação e os respectivos critérios avaliativos relacionados.

	Avaliação de Orientação	Avaliação de Regulação	Avaliação de Certificação
Nível da ação	<ul style="list-style-type: none"> • Critérios de orientação (da ação); 	<ul style="list-style-type: none"> • Critérios de eficácia; • Critérios de qualidade do funcionamento; • Critérios de satisfação. 	<ul style="list-style-type: none"> • Critérios de êxito (da ação)
Nível da pessoa	<ul style="list-style-type: none"> • Critérios de correção (S, SFC); • Critérios de apreciação (SFP); • Critérios comportamentais (SS). 	<ul style="list-style-type: none"> • Critérios de correção (S, SFC); • Critérios de apreciação (SFP). • Critérios comportamentais (SS). 	<p>Critérios de seleção; Critérios de êxito; Critérios de deliberação.</p> <ul style="list-style-type: none"> • Critérios de correção (S, SFC); • Critérios de apreciação (SFP); • Critérios comportamentais (SS).

Legenda: S – Saber; SFC – Saber-Fazer Cognitivo; SFP – Saber-Fazer Prático; SS – Saber-Ser. Fonte: De Ketele & Roegiers (1993, p. 73).

Em uma perspectiva ampla, a investigação se concentra em uma perspectiva diagnóstica. Entretanto, as diversas tomadas de decisão durante o percurso exercem diversas funções, em especial dentro do processo de recolha de informações. Em detalhes, considera-se que a pesquisa tenha uma ênfase nas seguintes funções:

- *Função descritiva* – restringe-se na descrição analítica do conteúdo e construtos do instrumento, do mapeamento das competências e conteúdos enfatizados em cada contexto, e não a um método de observação participante;
- *Função heurística* – em fase exploratória realiza-se um teste-piloto com uma pequena amostra visando identificar a eficiência do instrumento e a elaboração de hipóteses sobre o desempenho e comportamento dos itens no teste TCE, para uma verificação posterior em uma amostra maior;
- *Função preditiva* – faz-se um diagnóstico diante das dificuldades conceituais dos alunos representado pelos distratores, e com isso se busca avaliar hipoteticamente as causas de concepções equivocadas, além de prever êxito futuro de indivíduos (ou não);
- *Função prospectiva* – a construção de uma escala de proficiências visa estudos longitudinais na intenção de tendências de desempenhos e dificuldades diante das concepções térmicas, além da análise de DIF tendo como aspecto normativo as dificuldades relacionadas aos grupos buscando implicações necessárias sobre intervenções futuras;
- *Função certificativa*⁶⁶ – pelo processo de escolha na resposta ser discriminatório, sendo medido no instrumento em uma unidade de escala dicotômica; declara-se êxito de aprendizagem (ou não) em um pós-teste com base em escores e probabilidades de acerto; se exerce uma função de seleção e ordenação de indivíduos e grupos de desempenho ao comparar suas habilidades, e posicionando-os em uma escala de proficiência;

⁶⁶ É importante ressaltar que, efetivamente no campo de avaliação pedagógica, os testes conceituais não são projetados para dar nota e portanto exercer a função certificatória, mas apenas quando esta a serviço de sondagens para o pesquisador ou professor.

- *Função formativa* – sugestões delimitadas quanto à sequenciação de ensino, organização de conteúdos, temáticas complementares, identificação de situações-problemas, e métodos de ensino, tendo como base interpretações das evidências e fundamentações teóricas;
- *Função reguladora* – sugestões delimitadas visando reorientar a aprendizagem de acordo com as dificuldades de grupos de desempenho, com evidências para que sejam destacados determinados conteúdos e atividades no sistema ou paradigma curricular (conteúdos padrões e de desempenho).

Com isso, os critérios operacionais em relação aos objetivos enunciados se referem a descrição dos critérios avaliativos de acordo as modalidades e perspectivas avaliativas. Porém, existem alguns critérios que não devem ser considerados por não fazerem parte do estudo. Por exemplo, os critérios de eficácia se referem à avaliação da qualidade de um produto, podendo ser externo quando se quer verificar os efeitos procurados, ou interno quando se quer verificar os objetivos visados – esses critérios podem ser individuais ou coletivos, nesse sentido somente é pertinente quanto a uma comparação dos dois sistemas educacionais diante ao paradigma das competências em cada contexto, e se correspondem às necessidades dos alunos (eficácia do sistema quanto ao alinhamento curricular).

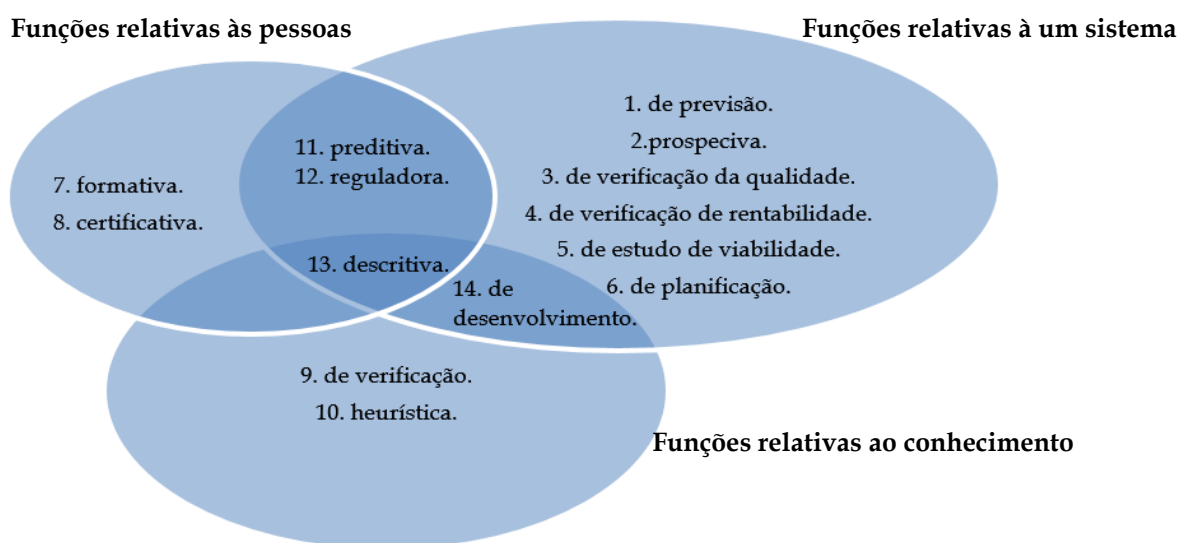


Figura 27: Relações das funções mais enfatizadas relativas às pessoas, a um sistema e ao conhecimento.

Fonte: De Ketele & Roegiers (1993, p. 136).

Os critérios de qualidade do funcionamento estão relacionados com a eficácia, e se referem a termos de motivação ou de autonomia, ou ainda de organização de equipes, estresse, etc., sendo esse um aspecto não pertinente ao estudo. Dentro da mesma ideia, os critérios de satisfação avaliam o desenvolvimento geral de uma formação, porém é necessária uma avaliação em termos de efeitos reais, portanto, também não um critério para a pesquisa. Já os critérios de apreciação se relacionam com o *saber-fazer prático* (SFP), e foge do escopo do estudo, pois não são consideradas observações e atividades práticas dos alunos na avaliação, ou seja, em situações-didáticas em salas de aula.

3.1.1.1.1 Critérios e normas referenciados na pesquisa

De um modo geral, foram organizados dois quadros que buscam sintetizar os critérios adotados de forma correspondente com a perspectiva avaliativa.

Quadro 46: Critérios em nível da ação dentro da pesquisa.

Nível da ação - ênfase na estratégia (Relacionado aos objetivos do estudo quanto ao instrumento, <i>design</i> investigativo e documentos oficiais)		
Avaliação de orientação	Critérios de orientação	<ul style="list-style-type: none"> • Texto explicativo aos professores para a aplicação do teste; • Texto de explicativo aos alunos para realizar a prova; • <i>Checklist</i> quanto os riscos de validade interna e externa no <i>design</i>.
Avaliação de regulação <i>*alinhamento curricular quanto ao conteúdo e o teste TCE</i>	Critérios de eficácia.	<ul style="list-style-type: none"> • Aspectos comuns nos documentos oficiais sobre o currículo no Brasil e em Portugal, com relação aos conteúdos e competências enfatizadas em termodinâmica; • Correspondência entre os conteúdos e competências sendo contempladas e enfatizadas nos programas curriculares e exames nacionais enfatizadas em termodinâmica; • Correspondência entre os conteúdos e competências entre o teste e os documentos oficiais em cada contexto, destacando as evidências das dificuldades e necessidades

Avaliação de certificação⁶⁷	Critérios de êxito	<ul style="list-style-type: none"> • Auxílio de juízes no processo de tradução, confirmação da chave-de-correção do teste e análises de riscos do <i>design</i> (critérios deliberativos). • Interpretação das situações-problemas, no que diz respeito ao conteúdo, habilidades, competências científicas e proposições de melhoria fundamentada na Didática da Ciência; • Definição das habilidades e categorias do conhecimento envolvido de acordo com a Taxonomia de Revisada de Bloom revisada; • Interpretação das situações-problemas e das evidências empíricas com o auxílio da psicologia educacional de Ausubel para da relação do traço latente com conhecimento prévio, bem como interpretação do subsunçor e conhecimentos potencialmente significativos; • Interpretação dos conceitos e implícitas concepções contidas no teste que supostamente possam influenciar a resposta correta ou errada no teste.
---	--------------------	--

A avaliação de desempenhos em um teste compõe a base das análises do estudo. Esse tipo de avaliação de medidas tipicamente pode ser feito por *norma* e/ou por *critério*, em palavras mais gerais seria focar na comparação entre os desempenhos de indivíduos (ou diferentes grupos – *aqueles que dominam ou não, sendo os grupos caracterizados por uma norma, tipo: idade, sexo, nível de instrução, região geográfica, etc.*) ou avaliando aquilo que ele domina diante de um quadro de critérios previamente referenciados (*critérios do tipo: completar a tarefa sem erro em 20s; completar com 2 erros/equívocos no máximo*). Se o indivíduo é comparado com relação as pontuações médias de um grupo, seria um *teste referenciado por norma*, ou seja, os escores referenciados pela norma são interpretados em relação aos escores alcançados por outros. As interpretações associadas aos resultados dos testes com base em critérios externos são chamadas *interpretações referenciadas por critérios*, como é o caso de avaliar um quantitativo de respostas referente a um determinado subconstruto no teste. No caso, o significado é associado aos escores e/ou percentuais de representatividade desses critérios externo no total de respostas, associando características significativas do atributo medido com pontuação de teste específica. As interpretações referenciadas por norma e as referências de critério atribuem significado às pontuações de teste, mas elas ocorrem de maneiras diferentes.

⁶⁷ Não se refere ao TCE como um instrumento a exercer a função avaliativa de certificação, mas sim das análises prévias relacionadas a sua coerência interna e rigor científico.

Uma das vantagens em avaliar por uma escala proficiência é que ela nos fornece ao mesmo tempo uma comparação por desempenho informando e da interpretação qualitativa das proficiências alcançadas, com base em itens consistentes de informação, dentro disso pode-se facilmente explorar interpretações com base em norma ou critérios referenciados (relacionados a um inventário de concepções, matriz de habilidades em anexo no teste), ou conjuntamente, tornando mais a rica a interpretação por explorar diferentes perspectivas. Um obstáculo ou inconveniente na escala de proficiências diante esses aspectos é o fato que nem sempre ela poderá ser formada por todos os itens do teste. Com isso, o estudo explora as análises comparativas das diferentes (e aspectos comuns) concepções entre alunos brasileiros e portugueses pelos dois enfoques, para assim ampliar uma compreensão sobre os conhecimentos preexistentes e avaliar de modo mais eficiente e substancial essas diferenças em propósitos específicos declarados previamente.

As diferenças entre os resultados referenciados pela norma e as referências por critérios são importantes para avaliar a adequação de um teste para um propósito específico. Em testes com referência à norma, o carácter do grupo norma é uma preocupação primordial ao fazer inferências sobre os resultados dos exames. Nos testes de referência por critérios, é fundamental que o domínio do critério seja apropriado e claramente definido. (Sireci, Wainer, & Braun, 1998, p. 37, *tradução livre*)

Quadro 47: Critérios e normas referenciadas em nível da pessoa (grupos normativos) e as implicações operacionais fundamentais das análises realizadas.

Nível da pessoa - ênfase nos grupos normativos (Relacionado aos objetivos do estudo quanto às diferenças e similaridades das concepções térmicas por critério e normas)
Grupos normativos considerados na avaliação comparativa de desempenhos:
<ol style="list-style-type: none"> 1. Alunos pré-instrução (grupos de 10º ano - Portugal, e 1º ano - Brasil); 2. Alunos pós-instrução (grupos de 11º e 12º ano - Portugal; 2º e 3º ano - Brasil); 3. Alunos por turno (Brasil); 4. Alunos agregados em escolas; 5. Alunos por Livro Didático adotado.
Parâmetros avaliativos relacionados a um <i>Teste Referenciado por Critérios (TRC)</i>
<p>Avaliação sobre os parâmetros do perfil de proficiência térmica dos respondentes, conforme critérios referenciados (diferenciando grupos por meio de informações advindas de uma quantidade mínima de incidências que represente uma determinada categoria preestabelecida):</p> <ul style="list-style-type: none"> • Correspondência representativa na matriz de referência (habilidades e competências) do teste para os 26 itens: <ul style="list-style-type: none"> ⇒ Espera-se que os resultados no que se refere as evidências de modelos mentais corretos (quantidade e percentual de itens acertados) devam atender a uma quantidade⁶⁸ de habilidades no quadro da matriz de referência que represente

⁶⁸ Não há garantias que a quantidade de habilidades que caracteriza uma determinada competência científica identificada esteja contemplada no teste, nesse sentido, deve entender que as evidências visem uma contribuição para o desenvolvimento dessa competência ou de atendimento eficiente as características do instrumento.

satisfatoriamente (cerca de 50% no mínimo daquelas que consideram-se representar efetivamente o construto) a competência correspondente – «Maior precisão para avaliar as evidências do conhecimento prévio (pré-teste) ou domínio de conhecimento do respondente (pós-teste)»);

- Correspondência representativa nos grupos categóricos de **concepções térmicas** dentro do inventário:
 - ⇒ Espera-se que os resultados no que se refere as evidências de **modelos mentais incorretos** (quantidade de distratores de um determinado grupo de concepções térmicas nos diferentes itens) possam representar satisfatoriamente um grupo de concepções térmicas equivocadas (cerca de 50% no mínimo daquelas que consideram-se representar efetivamente o construto, ou seja, da correspondência entre o quadro de tipicidade de concepções e dos respectivos distratores nas alternativas dos itens) – «Maior precisão para avaliar a natureza dos equívocos de um respondente e/ou para um perfil de grupos de respondentes». O ideal é que não haja picos entre os distratores e o percentual de incidência seja homogêneo e baixo, pois assim, suponha-se que os equívocos sejam mínimos, e consequentemente, não hajam grupos de respondentes caracterizados por um perfil de concepções térmicas equivocadas.

Parâmetros avaliativos relacionados a um Teste Referenciado por Normas (TRN)

Avaliação sobre os parâmetros do perfil de proficiência térmica dos respondentes, conforme **normas referenciadas** (diferenciando grupos por meio de informações advindas do escore e/ou habilidades θ):

- Posicionamento diferenciado nas faixas de **valores dos índices e parâmetros** psicométricos dos itens:
 - ⇒ Espera-se que os resultados no que se refere aos valores dos parâmetros (de acordo com os critérios preestabelecidos em tabela) possam identificar diferentes grupos de respondentes, em especial, que sejam predominantemente representados dentro das indicações para uma boa qualidade do item (elevada discriminação); uma distribuição relativamente homogênea entre itens fáceis, medianos e difíceis do teste; baixo acerto aleatório (fora da zona de “guessing” do item e do teste), e baixo erro casual;
- Posicionamento diferenciado de **itens e das habilidades** na escala de proficiências térmicas:
 - ⇒ Espera-se que os resultados no que se refere as habilidades θ possam representar de forma relativamente homogênea os diferentes grupos de respondentes por toda a extensão da escala, bem como, para os itens-âncoras na escala – «Maior precisão da escala para diferenciar as competências ou patamares de proficiência em diferentes grupos». A descrição das capacidades cognitivas dos itens na escala de proficiência também possibilite diferenciar grupos de acordo com os critérios referenciados na matriz de referência e no quadro de tipicidade das concepções térmicas, ou seja, também fornece subsídios de uma avaliação comparativa baseada em critérios referenciados (grupos de respondentes em patamares de proficiência que representam determinado domínio (matriz) ou dificuldade (inventário)).

Ações específicas e as respectivas implicações operacionais das análises

As análises podem se basear no **enfoque interpretativo-descritivo** (análise textual dos itens e no percentual de acerto e erro dos itens, perspectiva predominante na TCT) e no **enfoque hipotético-dedutivo** ou da **análise fatorial** (estimativas desenvolvidas pelos modelos da TRI).

De modo geral, todas as ações se distribuem quanto aos aspectos de identificar, comparar e diferenciar as **habilidades/competências** e **concepções térmicas** de respondentes em cada contexto de acordo com as faixas classificatórias dos índices e parâmetros de discriminação e dificuldades dos itens estimados para a TCT e TRI (aproximação ou dentro de um mesmo intervalo ou classe). Para efetivar as ações, tem-se algumas implicações operacionais predefinidas para o processo das análises:

Implicações operacionais das análises no **enfoque interpretativo-descritivo**:

- i. A identificação pode se dar através dos níveis de corte para o índice de dificuldade com base em critérios ideais e pertinentes aos dados (de acordo com cada item – p. ex.: eliminação de respondentes na zona de “guessing”, além do nível de complexidade do conhecimento envolvido). No caso, classificar de acordo com os critérios e procedimentos estatísticos definidos na literatura quanto aos melhores índices de discriminação e adequação aos níveis de dificuldades para TCT e TRI. Vale ressaltar que ao considerar os valores dos *índices corrigidos* torna-se possível elevar a consistência das informações e um aprofundamento interpretativo nas análises;
- ii. Na TCT, é possível obter uma identificação e comparação para as evidências que caracterizam os *modelos mentais corretos e incorretos* através do *Fator de Concentração das Análises* (destacando itens em zonas de desempenho inferior e superior) e do *Desvio na Concentração das Análises* (destacando itens em zonas de desempenho inferior e superior) de Bao & Redish (2001). Enquanto as evidências dos modelos corretos remetem a matriz de referência, os modelos errados remetem ao quadro de tipicidade dos equívocos;
- iii. Comparar os escores médios dos grupos com o parâmetro de dificuldade do item;
- iv. Comparar as concepções alternativas distribuídas nos itens entre os grupos através da incidência percentual de acertos e pela força atrativa na AGI pelos grupos de desempenho superior (ACIM), destacando assim a relevância de equívocos devido a sua atração por respondentes mais proficientes;
- v. Uma comparação entre alunos mais proficientes e menos proficientes pode ser feita através da observação dos escores por nível de escolaridade, na qual se espera que hajam maiores ganhos percentuais com o avançar dos níveis. Para isso, o *Fator de Hake* (ganhos percentuais brutos e normatizados) sinalizam esses ganhos, remetendo de forma indireta sobre a qualidade do processo educacional envolvido;
- vi. Interpretar as evidências dos escores com base na psicologia educacional de Ausubel, visando trazer reflexões, sugestões, implicações e recomendações em caráter diagnóstico. Para isso, busca-se justificar a identificação de subsunçores com base nas evidências empíricas, interpretações do conhecimento envolvido no enunciado do item e das concepções térmicas envolvidas e destacadas, tendo em vista procedimentos que possam auxiliar os professores e favorecer o desenvolvimento de uma aprendizagem significativa.

Implicações operacionais das análises no enfoque hipotético-dedutivo ou da análise fatorial:

- i. Para a identificação das habilidades e competências na TRI se deve respeitar os pressupostos de *monotonicidade*; *independência local* e *unidimensionalidade*, além da realização do processo de ajuste nos modelos;
 - ii. Comparar as diferenças significativas através do teste paramétrico *t-student* entre escores padronizados médios dos grupos normativos de Portugal e Brasil;
 - iii. Identificando as probabilidades nas habilidades de acordo com a CCI para cada contexto (quanto mais um grupo ou indivíduo domina o conteúdo, maior são as habilidades θ estimadas na CCI);
 - iv. Comparar as habilidades e dificuldades entre os contextos de acordo com a CCI na análise de DIF multigrupo (inclui a comparação das CCI's de cada grupo), com isso, é possível verificar a estimação das probabilidades de acerto diferenciados para respondentes com a mesma habilidade;
 - v. Identificar e comparar as habilidades dos grupos dentro dos níveis da escala de proficiências técnicas (quanto mais um grupo (ou indivíduos) domina o conteúdo, maior são as habilidades θ estimadas em probabilidade de acerto na escala de proficiências). Para isso, após identificar os itens-âncora que constituem a escala de proficiência por meio de critérios específicos, deve-se interpretar diretamente a proficiência na escala levando em consideração o caráter acumulativo de todos os itens-âncoras posicionados em níveis inferiores, podendo recorrer ao auxílio das análises gráficas das CCI e CII. Recomenda-se complementar as interpretações de probabilidade das habilidades na TRI com os resultados na TCT.
-

Concebe-se dessa forma como *critérios mínimos avaliativos* sobre as análises das diferentes concepções envolvidas a busca por uma tipicidade entre os contextos, mais em caráter qualitativo em sentido interpretativo e de correspondência em diferentes perspectivas de análise, visando dar suporte ao um julgamento mais consistente. Enquanto o *desempenho dos indivíduos* envolve parâmetros normativos de referência para a comparação entre as diferentes concepções, sendo mais em caráter quantitativo visando delimitar e especificar a amostra, possibilitando uma interpretação mais refinada e coerente.

3.1.1.2 Validade

3.1.1.2.1 Considerações de Validade em Testes

A *Psicometria* como parte da psicologia preocupada com a medição psicológica, ou, a ciência da mensuração psicológica (Cohen et al., 2014, p. 8), considera a construção, análise, **validação e uso de testes são seus principais objetos de estudo**, nisto, cada etapa da elaboração ou de uso de testes possuem critérios específicos que precisam ser considerados. No caso da educação esses testes estão a serviço de avaliar ou aferir conhecimentos e estes carregam em si crenças e questões epistemológicas que são geralmente interpretados dentro das teorias de aprendizagem.

Entre as primeiras definições sobre o conceito de validação em um teste realizadas por psicometristas do século XX, se tem Lindquist (1942) afirmando que:

A validade de um teste pode ser definida como a precisão com que mede o que se pretende medir, ou como o grau em que ele se aproxima da infalibilidade na medição do que ele pretende medir. O grau de validade de um teste, portanto, depende da magnitude dos "erros" (devido a qualquer e todas as causas que acabamos de considerar) que estão presentes nas medidas obtidas a partir dele. (Lindquist, 1942, p. 213, *tradução livre*)

Enquanto Gulliksen (1950):

A confiabilidade foi considerada como a correlação de um determinado teste com uma forma paralela. Correspondentemente, a validade de um teste é a correlação do teste com algum critério. Nesse sentido, um teste tem muitas "validades" diferentes. (Gulliksen, 1950, p. 88, *tradução livre*)

As duas opiniões de validade foram desenvolvidas e passaram a ser conhecidas como **validade de conteúdo** e **validade de critério**.

O conceito de validade segue dentro de um processo reformulação e aprimoramento na contemporaneidade, partindo da visão de Cronbach e Meehl sobre a *validade de construto* como um dos três tipos de *validade interna* (validade de conteúdo, validade de critério e validade de construto), inicialmente expandido por Loevinger em três componentes de validade de construto (componentes substantivo, estrutural e externo), e ampliado ainda mais por Messick para incluir seis componentes (conteúdo, substantivo, estrutural, generalização, externo e consequencial) (Wren & Barbera, 2013).

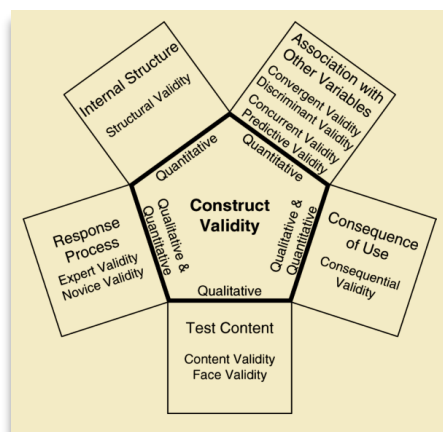


Figura 28: Perspectivas quanto a validade de construto.

Fonte: Wren & Barbera (2013)

Atualmente, é bem conhecida uma diversidade de tipos de validade, e recomenda-se utilizar o termo "evidências" ou "indícios" de validade, tendo em vista

que é inconcebível admitir que uma investigação seja plenamente válida em todos os processos.

Quadro 48: Fontes de evidências de validade recomendadas no uso de testes psicológicos.

Fonte	Definição
<i>Evidências baseadas no conteúdo.</i>	Levanta dados sobre a representatividade dos itens do teste, investigando se esses consistem em amostras abrangentes do domínio que se pretende avaliar com o teste.
<i>Evidências baseadas no processo de resposta.</i>	Levanta dados sobre os processos mentais envolvidos na realização das tarefas propostas pelo teste.
<i>Evidências baseadas na estrutura interna.</i>	Levanta dados sobre a estrutura das correlações entre itens avaliando o mesmo construto e também sobre as correlações entre subtestes avaliando construtos similares.
<i>Evidências baseadas nas relações com variáveis externas.</i>	Levanta dados sobre os padrões de correlações entre os escores do teste e outras variáveis, medindo o mesmo construto ou construtos relacionados (convergência), e com variáveis medindo construtos diferentes (divergência). Também traz dados sobre a capacidade preditiva do teste de outros fatos de interesse direto (critérios externos) que têm importância por si só e associam-se ao propósito direto do uso do teste (por exemplo, sucesso no trabalho).
<i>Evidências baseadas nas consequências da testagem.</i>	Examina as consequências sociais intencionais e não intencionais do uso do teste para verificar se sua utilização está surtindo os efeitos desejados de acordo com o propósito para o qual foi criado.

Fonte: Primi et al. (2009)

Considera-se que algumas das evidências recomendadas para testes psicológicos por instituições internacionais⁶⁹ (AERA, APA, & NCME 1999; Primi, Muniz, & Nunes, 2009) possam servir de referência aos testes educacionais (Quadro 48)⁷⁰. Essas recomendações continuam sendo reforçadas na última edição do *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014).

3.1.1.2.2 Tipos de Validade

O processo de validação de um instrumento se insere no que se denomina de **validade interna**. A validade interna está diretamente ligada ao instrumento de medida, aos seus itens, e dá-nos indicadores de precisão das inferências que fazemos, com base neles.

⁶⁹ AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION.

⁷⁰ Um quadro semelhante e mais completo em Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno (2005, pp. 49-52).

3.1.1.2.2.1 Validade de construto

É o sentido de validade mais ampla e atual, englobando as restantes. Na prática é usada para descrever uma escala ou medida que se correlaciona com medidas de outras variáveis tal como é suposto com base nos conhecimentos existentes (ou como a teoria prevê que as variáveis se relacionem); a validade de construto acompanha todo o processo de construção de um instrumento e não se expressa por um só indicador, mas por vários. Neste sentido, e segundo Shepard (1993), “a validade de construto não pode produzir conclusões definitivas, nem pode alguma vez ser terminada” (p. 430). Atualmente, considera-se que a questão mais relevante é “o teste é adequado para os propósitos específicos que o investigador tem em mente?” Deste modo, aferir a validade de uma medida consiste em encontrar um conjunto de argumentos que fundamentem a sua utilização, tendo em conta os objetivos que se pretendem alcançar, podendo recorrer-se a muitas formas de validação tradicional, agora enquadradas no conceito unitário de validade de construto.

O fato é que a validade de construto varia de acordo com os *objetivos da pesquisa*. As principais técnicas são: *validade convergente, validade discriminante, análise fatorial exploratória e confirmatória*.

Dois aspectos da validade de construto:

- Rede *nomotética* (ou *nomológica*) investigar a rede de relações entre os escores no teste e outras variáveis como critérios externos (Cronbach & Meehl, 1955; Trochim, 2007, pp. 83–84).

Também se pode dizer que a *validade nomológica* ou teórica (teste de hipótese) consiste na capacidade de um instrumento de medida de comprovar uma hipótese que se baseia em uma teoria. Em certos casos, a hipótese pode prever que dois grupos distintos de indivíduos deveriam teoricamente ter resultados diferentes em uma determinada escala (Bandeira, n.d.).

- Representação do *constructo* (Embretson, 1994): identificação dos processos cognitivos básicos subjacentes à resolução ou resposta aos itens do teste.

A coerência interna depende da *validade dos instrumentos de coleta de dados* (ou de medida), isso significa dizer que os objetivos do instrumento devam estar alinhados a contribuir na decisão que se pretende tomar. Para isso ele deve possuir a capacidade de medir aquilo que ele pretende medir (e não outro construto), ou seja, possuir *validade de construto*.

A *validade de construção* também é chamada de *validade teórica* e diz respeito à generalização dos resultados para o conceito ou teoria subjacente ao experimento, deve superar a riscos ao *design* do estudo quanto à(ao) (Prado, 2009):

- *Explicação pré-operacional inadequada* devido à falta de compreensão teórica (ou seja, uma explicitação inadequada dos constructos);
- *Viés mono-operação* devido à simplificação de variáveis independentes necessárias a elaboração do construto, ou seja, falta de riqueza necessária para chegar a conclusões;
- *Viés mono-método* está na necessidade de métrica (relativa à métodos) adequada para medir a variável dependente, não levando em conta toda a sua complexidade (ou seja, um problema é estudado sempre pelo mesmo método);
- *Confusão entre construtos e seus níveis* ocorre quando não se identifica a relação entre variáveis devido a ignorância de fatores no grau de conhecimento de indivíduos;
- *Interação de tratamentos* devido à dificuldade de identificar a influência na medida por um determinado tratamento ou outro;
- *Interação entre um teste e um tratamento* ocorre pelo fato do indivíduo saber que está sendo testado influenciar nos resultados;
- *Generalização restrita entre construto* é um efeito colateral não intencional que ocorre no processo a influenciar nas conclusões, isentando comentários sobre seus efeitos ocasionados e de suas implicações.

3.1.1.2.2.2 Validade de conclusão estatística

Devem ser revisados os riscos (ou ameaças) à *validade da conclusão estatística*, geralmente quando se extraem inferências inexatas dos dados devido à potência estatística inadequada ou suposições estatísticas. A falta de rigor estatístico é outra

causa que está muitas vezes ligada à inobservância ou imprudência de não respeitar os critérios e pressupostos, ou falta de competência de utilizar as estratégias mais adequadas.

3.1.1.2.2.3 Validade de traço

A validade *de traço* (ou validade concomitante) se refere à capacidade de um instrumento de medida correlacionar com um outro instrumento de medida que avalia um conceito teoricamente semelhante e que já foi validado anteriormente. Esse é um aspecto que pode ser verificado em outros estudos, tornando mais consistente as informações e interpretações da escala e dos itens (Bandeira, n.d.).

3.1.1.2.2.4 Validade de conteúdo

O instrumento também deve conter as variáveis ou conjunto de categorias previamente definidas e associadas aos eixos temáticos do estudo (não podendo abranger conteúdos que não pertencem ao seu construto), ou seja, possuir *validade do conteúdo*. A validade de conteúdo existe quando os conteúdos dos itens cobrem a totalidade do construto que pretendem medir. É obtido através da opinião de peritos/especialistas no assunto quanto à sua adequação e exaustividade. Entretanto, os procedimentos para avaliar a *validade de conteúdo* de um teste envolvem: *Consulta a uma Comissão de Especialistas* e a *Consulta a um grupo da população-alvo familiarizado com o fenômeno* (Prado, 2009).

3.1.1.2.2.5 Validade de critério

Se o instrumento for *capaz de prever o resultado* a ser obtido em um desempenho ou comportamento específico dos indivíduos, presente ou futuro, podemos dizer que possui *validade de critério*. Deve-se, então, ser avaliado através do cálculo da correlação entre os escores do teste e uma *outra medida*, presente ou futura, de *desempenho dos sujeitos*, que chamamos de *critério*. Ou seja, a *validade de critério* ou de *predição* se refere

a capacidade de prever o futuro comportamento de indivíduos, ou seja, apresenta correlação entre testes diferentes. No estudo, esse tipo de consistência se apoia na estimação do alfa de Cronbach, na precisão da escala através de itens-âncoras (sendo necessário uma distribuição entre itens fáceis, medianos e difíceis), para assim, estimar as habilidades ou probabilidades de acerto na escala de proficiência em situações futuras. Índícios desse tipo de validade são explorados no teste-piloto.

3.1.1.2.2.6 Validade concorrente

Está relacionada com o desempenho atual e diz-nos da precisão com que mede o construto (p. ex.: *será que um teste interno de matemática do 12º ano avalia tão bem um aluno como um teste nacional?*).

3.1.1.2.2.7 Validade preditiva

Está relacionada com o desempenho futuro e nos indica em que medida o resultado de um teste permite prever desempenhos futuros dos sujeitos noutras situações;

3.1.1.2.3 Validade do Instrumento: Tradução e Adaptação

A adaptação de um teste envolve várias decisões que primeiro determinam se o teste pode medir o mesmo conceito em uma cultura e idioma diferentes, em seguida, escolha os tradutores e as modificações a serem feitas na preparação do teste que deve ser usado em outro idioma, até que, no final do processo, modifique o teste e verifique sua equivalência na versão adaptada. (Hambleton, Merenda, & Spielberger, 2005, p. 4, *tradução livre*)

Um processo eventualmente necessário é o da *tradução*⁷¹ de questionários/testes. O processo de tradução representa o início de uma análise mais

⁷¹ Testes traduzidos não são testes adaptados, segundo (Hambleton, 2005, citado por Borsa, Damásio, & Bandeira, 2012): “os termos adaptação e tradução são distintos, e tem-se preferido o uso do primeiro, uma vez que compreende todos os processos concernentes à adequação cultural do instrumento, para além da mera tradução. A

aprofundada, tendo em vista a necessidade de **validação**, em que um teste somente é válido quando ele está adaptado e não traduzido. Nesse sentido, o trabalho de Guillemín, Bombardier, & Beaton (1993, p. 1422) é uma referência clássica para isso, e sugerem os seguintes aspectos:

Quadro 49: Diretrizes de Guillemín, Bombardier, & Beaton (1993) para a adaptação transcultural de medidas.

Tradução:

- Produza várias traduções;
- Use tradutores qualificados.

Retradução:

- Produza tantas retraduições como traduções;
- Use retraduições adequados.

Revisão por uma comissão de especialistas

- Constituir um comitê para comparar versões originais e finais;
- A composição do comitê deve ser multidisciplinar;
- Use técnicas estruturadas para resolver discrepâncias;
- Modificar instruções ou formatar, modificar/rejeitar itens inapropriados, gerar novos itens;
- Garantir que a tradução seja totalmente compreensível;
- Verificar a equivalência transcultural das versões de origem e final.

Pré-testagem (estudo piloto)

- Verifique a equivalência nas versões de origem e final usando uma técnica de pré-teste;
- Ou use uma técnica de sonda;
- Ou envie a fonte e as versões finais para leigos bilíngues;
- Imigrantes: escolha o idioma de administração ou use uma medida de formato duplo.

Ponderação das pontuações

- Considere adaptar os pesos das pontuações ao contexto cultural

Fonte: Guillemín, Bombardier, & Beaton (1993, p. 1422, *tradução livre*).

Recomenda-se que tradução deva ser feita por um bilíngue que seja nativo da língua a ser traduzida, sendo a primeira versão em língua materna, e se possível realizar no mínimo duas traduções por dois tradutores. A retradução, seria um *feedback* da língua traduzida para a língua original, realizando o mesmo processo da tradução, porém sendo os tradutores possuindo sendo nativos da língua original. A comissão de especialistas deve ser composta por profissionais que possuem *domínio sobre o tema* do teste e da *escala ou unidade de medida* que o teste foi elaborado ou se propõe (dicotômico, politômico – Likert, tipo-Likert, diferencial semântico, etc.), sendo importante duas tarefas:

- Comparação entre as versões retraduzidas *lado a lado* com a versão original, juntamente com os tradutores. Com isso devem ser feitas análises semânticas sobre as

adaptação é um processo árduo e trabalhoso, que envolve vários indivíduos, nível na qual este estudo ainda não alcançou. Na fase inicial do processo investigativo tomou-se cuidado com o clássico erro conhecido como "*Traduttore, Traditore*" (Tradutor, traidor).

palavras que possam eventualmente mudar o sentido das frases para que assim sejam corrigidas, e, posteriormente, serem corrigidas na língua que se deseja a tradução;

- Discussão sobre os aspectos relacionados à adaptação transcultural de frases, palavras, exemplos, expressões verbais e situações da escala original, que necessitam ser adaptadas para o contexto, ou seja, frases e expressões verbais cotidianas do contexto que se pretende devem ser escolhidas para que se formem as frases, diante disso se obtêm a versão preliminar do teste.

Se concebe a pré-testagem como estudo-piloto, sendo aplicada em uma amostra reduzida de sujeitos dentro da normativa pretendida (nível mínimo adequado com relação a idade e escolaridade, gênero, situação socioeconômica, região geográfica, etc.) e relacionadas à população-alvo, tendo como finalidade testar se a escala possui uma fácil compreensão, se a escolha das palavras são familiares e de uso cotidiano, se as instruções estão claras, se as alternativas de resposta são fáceis de serem discriminadas. Recomenda-se que seja aplicada para cada pessoa procurando identificar termos ou frases que ela considera que não sejam familiares ou de difícil compreensão, buscando superar encontrando termos substitutos, e de melhor aceitação. Após a modificação, segue-se sendo aplicada a outra pessoa da amostra normativa até que não haja mais problemas de compreensão. A intenção é obter uma versão que seja fácil para qualquer pessoa, independentemente de seu nível socioeconômico e de escolaridade. Este procedimento foi denominado *Técnica de Sondagem* (“*Probe Technique*”).

Uma outra maneira de se proceder, no estudo piloto, é introduzir uma avaliação escrita das questões, de forma que o respondente classifique cada pergunta da escala em termos do seu grau de compreensão, em uma escala de alternativas (ex. de 1 a 4, onde 1 é muito confusa e 4 muito clara). Calcula-se, então o grau de compreensão da média do grupo, eliminando-se os itens que obtiveram baixa cotação.

Pode-se ainda aplicar a versão brasileira e a versão original em um mesmo grupo de pessoas bilíngues e verificar se as respostas destes sujeitos apresentam uma correlação elevada, o que deveria ocorrer se as duas formas são equivalentes. Este último procedimento, entretanto, não garante que a formulação das questões seja de fácil compreensão para todos e um procedimento específico para isto deve ser feito, tal como descrito acima, perguntando-se aos sujeitos o seu grau de compreensão. (Bandeira, n.d.)

Apesar de não haver um consenso na literatura sobre quais são as etapas rigorosamente necessárias, existem aspectos essenciais concernentes a este processo a serem verificados (Borsa, Damásio, & Bandeira, 2012):

Quadro 50: Diretrizes de Borsa, Damásio, & Bandeira (2012) para adaptação transcultural de medidas.

Etapas	Condições ou tarefas serem desenvolvidas para o processo de adaptação
Tradução do instrumento para o novo idioma:	<ul style="list-style-type: none"> • De modo geral evitar a <i>tradução literal</i> dos itens (Hambleton, 1994, 2005), porque, muitas vezes, resulta em frases incompreensíveis ou, pelo menos, não coerentes com a fluência do idioma-alvo; • No mínimo <i>dois tradutores bilíngues independentes</i> devem ser convocados para adaptar os itens ao novo idioma (Beaton et al., 2000; Gudmundsson, 2009; Hambleton, 2005; ITC, 2010), preferencialmente <i>fluentes no idioma de origem do instrumento e nativos no idioma-alvo</i>; • No mínimo que um dos tradutores possa ter uma <i>compreensão do construto</i> a ser avaliado e que tenham <i>habilidade e familiaridade com a escrita de artigos científicos</i> (Cassepp-Borges et al., 2010; Hambleton, 1994, 2005; ITC, 2010).
Síntese das Versões Traduzidas	<ul style="list-style-type: none"> • <i>Equivalência semântica</i> – objetiva avaliar se as palavras apresentam o mesmo significado, se o item apresenta mais de um significado e se existem erros gramaticais na tradução; • <i>Equivalência idiomática</i> – refere-se a avaliar se os itens de difícil tradução do instrumento original foram adaptados por uma expressão equivalente que não tenha mudado o significado cultural do item; • <i>Equivalência experiencial</i> – refere-se a observar se determinado item de um instrumento é aplicável na nova cultura e, em caso negativo, substituir por algum item equivalente; • <i>Equivalência conceitual</i> – busca avaliar se determinado termo ou expressão, mesmo que traduzido adequadamente, avalia o mesmo aspecto em diferentes culturas.
Avaliação da Síntese por Experts	<p>Recomenda-se que seja submetido ao <i>auxílio de um comitê de experts na área da avaliação psicológica</i>, ou, se possível, com conhecimento específico acerca do construto avaliado pelo instrumento, que avaliará aspectos ainda não contemplados, tais como a estrutura, o layout, as instruções do instrumento e a abrangência e adequação das expressões contidas nos itens, tais como:</p> <ul style="list-style-type: none"> • Se os <i>termos ou as expressões podem ser generalizados para diferentes contextos e populações</i> (isto é, diferentes regiões de um mesmo país); • Se as <i>expressões são adequadas</i> para aquele público a que o instrumento se destina; • Se os <i>aspectos da diagramação</i> do instrumento estão adequados, uma vez que estes são tão imprescindíveis quanto os aspectos linguísticos dos itens (populações específicas como crianças, jovens, idosos, etc.); • Se há <i>clareza do rapport</i>⁷², ou seja, uma adequação do tipo e do tamanho da fonte utilizada, a disposição das informações no instrumento, entre outros aspectos;
Avaliação pelo Público-Alvo	<ul style="list-style-type: none"> • Ser <i>avaliado por crianças com diferentes idades</i> (dentro da faixa etária a que o instrumento se destina); • Ser <i>avaliado por residentes em diferentes localidades e regiões</i> (já que, uma vez validado, o instrumento poderá ser aplicado em diferentes populações, de diferentes regiões do país). • Solicitar que o respondente <i>forneça sinônimos que melhor exemplifiquem o vocabulário do grupo</i> a quem o instrumento se destina; • Solicitado que as questões sejam lidas em voz alta pelo respondente e que o

⁷² É um conceito originário da psicologia que remete à técnica de criar uma ligação de empatia com outra pessoa (Marques, 2016).

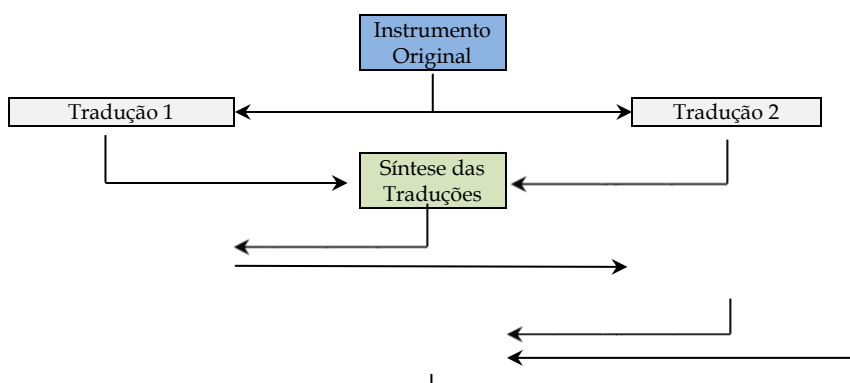
	<p>mesmo realize uma breve explicação sobre o significado de cada um dos itens;</p> <ul style="list-style-type: none"> • Realizar uma aplicação para que os respondentes preencham o instrumento e, posteriormente, realizem uma discussão acerca da compreensão de cada item, propondo modificações, se caso for; • Repetir o procedimento uma ou mais vezes, dependendo da necessidade e da complexidade do instrumento a ser adaptado.
Tradução Reversa (Back-translation)	<p>O objetivo da tradução reversa não deve ser a obtenção de uma equivalência literal entre as versões traduzidas e a versão original. Ao invés disso, o processo da tradução reversa deve ser utilizado como uma ferramenta para identificar palavras que não ficaram claras no idioma-alvo, buscando encontrar inconsistências ou erros conceituais na versão final, quando comparada à versão original (Beaton et al., 2000). É importante dizer que a retrotradução não pressupõe que o item necessita se manter literalmente igual ao original, mas, sim, manter uma equivalência conceitual.</p> <ul style="list-style-type: none"> • Traduzir a versão sintetizada e revisada do instrumento (instrumento “pronto”) para o idioma de origem; • Submeter a dois tradutores que não aqueles que realizaram a primeira tradução; • O processo de tradução reversa pode incorrer na ênfase dos aspectos gramaticais em detrimento dos aspectos contextuais. • Disponibilizar acesso ao autor à versão retrotraduzida (<i>back-translated</i>) do instrumento, pois assim pode se afirmar se os itens têm, em sua essência, a mesma ideia conceitual que os itens originais.
Estudo-Piloto	<ul style="list-style-type: none"> • A aplicação prévia do instrumento em uma pequena amostra que reflita as características da amostra/população-alvo (Gudmundsson, 2009); • Avaliar a adequação dos itens em relação ao seu significado e à sua dificuldade de compreensão, bem como as instruções para a administração do teste; • Realizar um segundo estudo-piloto (ou quantos forem necessários), para avaliar se o instrumento está, finalmente, pronto para ser utilizado. • O procedimento do(s) estudo(s)-piloto devem ser realizadas com o auxílio do comitê de experts, e nunca apenas pelo pesquisador que foi a campo.

Fonte: Adaptado de Borsa, Damásio, & Bandeira (2012).

A validação nesse caso deve passar por um processo de *adaptação transcultural* (Quadro 50) de instrumentos psicológicos e exige o cumprimento de diretrizes aos pesquisadores sobre os diferentes passos desse processo (Borsa, Damásio, & Bandeira, 2012).

Em geral, a literatura aponta que a adaptação de um instrumento deve ser constituída por cinco etapas essenciais: (1) tradução do instrumento do idioma de origem para o idioma-alvo, (2) realização da síntese das versões traduzidas, (3) análise da versão sintetizada por juízes *experts*, (4) tradução reversa para o idioma de origem (*back translation*), e (5) estudo-piloto. (Hambleton, 2005; Sireci et al, 2006, citado por Borsa, Damásio, & Bandeira, 2012, p. 424)

Um modelo esquemático sobre esses processos, pode ser:



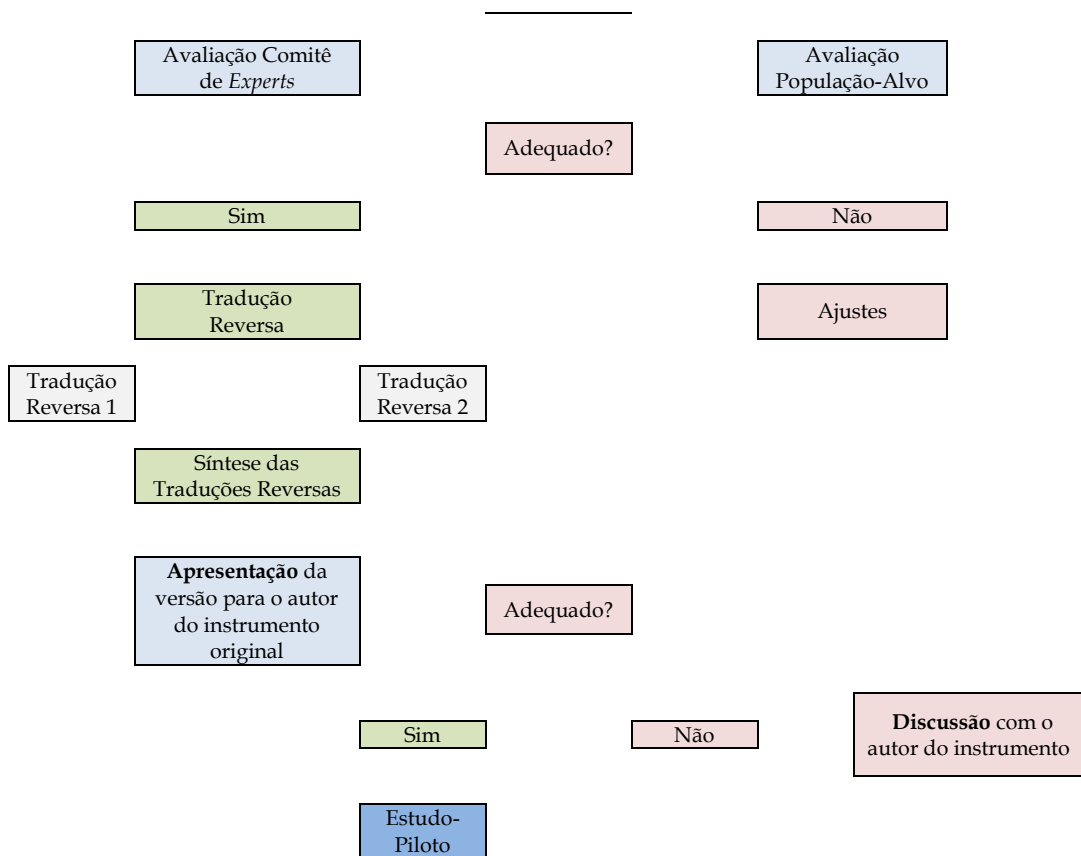


Figura 29: Procedimentos para adaptação transcultural de instrumentos psicológicos.

Fonte: Borsa, Damásio, & Bandeira (2012).

É importante ressaltar que a *International Test Commission*⁷³ (ITC) vem desde 1992 trabalhando com o objetivo de propor diretrizes para a tradução e a adaptação de instrumentos psicológicos entre culturas (Ibid., 2012, p. 423), estando disponível um documento atual dessa instituição que rege normas internacionais (Cheung, Iliescu, & Bartram, 2014b), e uma versão em português do Brasil (Cheung et al., 2014a). Ressalta-se que se os *juízes* incumbidos da tarefa de traduzir não garantirem recomendações específicas e criteriosas, se assumem riscos e equívocos interpretativos dos escores.

3.1.1.2.4 Implicações na Pesquisa

3.1.1.2.4.1 Validade de construto

A *validade de construto* consiste na capacidade de relacionar adequadamente a variável operacional (itens da escala) com o construto ou conceito teórico (definição do

⁷³ Ver: <http://www.intestcom.org/files/guideline_quality_control.pdf>.

conceito, com base na teoria). Dessa forma, torna-se importante definir (**qualitativa**) rigorosamente cada item e seus conglomerados (categorias/construtos e subcategorias/subconstrutos) para que hajam correspondências interpretativas consistentes com as “medidas” (**quantitativa**).

3.1.1.2.4.2 Qualitativa

Considera-se que o instrumento de medida tem **validade de construto** quando está realmente medindo o conceito teórico que pretendemos medir, e não sendo verificado isso, se tem uma *validade aparente*, ou seja, o teste parece medir algo, porém, menos válido. Para se obter indícios de validade nesse sentido, o teste foi analisado identificando a sua finalidade de medida (aquilo que ele pretende medir) com base nas finalidades de medida de cada item, sendo incluído os conceitos básicos e processos termodinâmicos envolvidos, juntamente com uma análise taxonômica visando a elaboração de uma *matriz de referência* para o teste. Essa matriz se configura como os *critérios referenciados do teste* a partir da elaboração de descritores sobre as habilidades, conhecimentos e situação contextual envolvida para cada item, estando organizado *lado a lado* com o *inventário de concepções* ligado aos itens.

Esse processo qualitativo busca contribuir na validação prévia dos subconstrutos preexistentes como base nos conceitos envolvidos, concepções, e situações predefinidas, verificando assim os objetivos, as habilidades e competências para cada item.

O quadro de tipicidade ou *inventário de concepções* que acompanha o teste TCE se configura como critérios referenciados que auxiliam na avaliação. O inventário é composto por um *conjunto de concepções alternativas* específicas, típicas e relacionadas a um conhecimento introdutório da Física Térmica. Diante disso, foi realizada uma análise inicial dos subconstrutos dentro do inventário para certificar as correspondências de pertinência dos itens e aproximações das concepções apresentadas, podendo nesse processo serem excluídas ou deslocadas algumas categorias e subcategorias para ajustar os subconstrutos e não ameaçar a sua validade

e fiabilidade nas conclusões, e assim serem formuladas contribuições de cada um grupo de concepções.

Essas contribuições são parciais, por serem consideradas subgrupos de construtos (subconstrutos), ou seja, entendidos como componentes a serem explicadas por um *fator predominante*⁷⁴ podendo ser denominada de modo geral como um *perfil das concepções térmicas básicas* para os diferentes grupos normativos analisados, referente a todas as concepções envolvidas. Busca-se, contudo, apresentar as diferenças existentes do perfil de concepções térmicas de cada grupo e de aspectos comuns, sinalizando assim possíveis necessidades gerais e intercambiáveis.

3.1.1.2.4.3 Quantitativa

Para a **validade de construto** foi realizada uma análise fatorial confirmatória de acordo com o modelo de melhor ajuste, bem como uma análise gráfica do *Scree Plot* para identificar os autovalores da matriz de correlação tetracórica, e com isso verificar a existência (ou não) de uma dimensionalidade predominante entre os construtos. Esse processo também é denominado de *validade fatorial*, ou seja, um método utilizado para avaliar a validade de construto de um instrumento de medida verificando se consiste em identificar sua estrutura dimensional, ou seja, identificar quantas dimensões do construto a escala está avaliando com os seus itens. A linha de corte no *Scree Plot* é igual a 1, os traços ou construtos acima são considerados dimensões os mais explicativos. Seguindo os pressupostos clássicos da TRI, o modelo deve respeitar o critério da unidimensionalidade, ou seja, conceber que haja um único construto predominante explicativo. As interpretações das análises partem desse critério estatístico, de acordo com o modelo que melhor se ajusta a essas condições.

3.1.1.2.5 Validade de conteúdo

Para a *validade do conteúdo* foi realizada uma análise interpretativa item-a-item, bem como as classificações específicas dos distratores dentro do quadro de tipicidade.

⁷⁴ Com base no enfoque da análise fatorial (*a posteriori*), porém, existe também a perspectiva de um perfil das concepções térmicas através do estabelecimento de construtos *a priori*.

Espera-se que verificação do conteúdo possa cobrir de forma representativa a parte introdutória da Termodinâmica dentro dos programas. Um aspecto complementar, seria verificar se os conhecimentos no teste são relevantes dentre aqueles que correspondem (ou foram correspondidos) e são exigidos em exames nacionais.

3.1.1.2.5.1 Validade nomológica (teste de hipóteses)

Para a **validade nomológica**, espera-se que indivíduos de baixa habilidade errem itens difíceis, bem como, que os itens tenham um mesmo comportamento funcional para os diferentes gêneros. Caso isso não se evidencie, pode-se afirmar que não atenda a essas hipóteses.

3.1.1.2.5.2 Validade do TCE

Quanto ao teste TCE, o site *PhysPort*⁷⁵ (*Supporting Physics Teaching With Research-Based Resources*), ligado a Associação Americana de Professores de Física (AAPT), apresenta argumentos sobre alguns indícios de validade:

“As questões de escolha múltipla sobre o TCE foram desenvolvidas com base em um inventário de concepções alternativas identificadas sobre Termodinâmica na literatura de pesquisa. O TCE foi então testado e revisto. As perguntas também foram revisadas por especialistas e consideradas válidas, e foram testadas em entrevistas com os alunos e foram interpretadas como pretendidas. Foram realizadas análises estatísticas apropriadas de confiabilidade, discriminação e dificuldade e valores razoáveis. Uma análise fatorial foi realizada e quatro grupos conceituais de questões identificadas: 1) transferência de calor e mudanças de temperatura, 2) ebulição, 3) condutividade térmica e equilíbrio, e 4) congelamento e fusão. O TCE tem sido usado com mais de 1000 estudantes de ensino médio e iniciantes universitários em nove instituições na Austrália e Coreia. Existem duas publicações revisadas por pares que apresentam dados de TCE”. (Physport, n.d., tradução livre)

No caso do TCE, o site da *PhysPort* apresenta validação “prata” aceita pela *American Association of Physics Teachers* para o teste e detalha suas especificações.

Quadro 51: Validade do TCE Creditada pela AAPT.

Avaliação sobre a validade do teste TCE
Desenvolvido por: Shelley Yeo e Marjan Zadnik (2001).

⁷⁵ Ver: <<https://www.physport.org/assessments/assessment.cfm?I=17&A=TCE>>.

Objetivo: Avaliar a compreensão e aplicação dos conceitos termodinâmicos por parte dos alunos do terceiro ano do ensino médio ou universitário introdutório, usando contextos comuns que refletem as próprias concepções dos alunos.

Formato: Pré-teste e pós-teste; múltipla escolha.

Duração: 30 min.

Foco: Conhecimento Térmico / Estatístico do conteúdo (temperatura, transferência de calor, mudança de fase, propriedades térmicas dos materiais).

Nível: Colegial; Introdutório na faculdade;

Validação da investigação: validação “prata”

Este é o segundo nível mais elevado de validação da investigação, correspondendo a pelo menos 5 das categorias de validação abaixo.

Resumo da validação da investigação

A pesquisa foi baseada no:

Pensamento do estudante **Sim**

O estudou usou:

Entrevistas de estudantes **Sim**

Análise de especialistas **Sim**

Análise estatística adequada **Sim**

A pesquisa foi conduzida:

Em várias instituições **Sim**

Por grupos múltiplos de pesquisa Não

Publicação revisada por pares **Sim**

Fonte: Physport (n.d., tradução livre).

3.1.1.2.5.3 Validade externa

A **validade externa** se divide em dois campos (Vieira, n.d.), a: (1) *validade da população*; e a (2) *validade ecológica*. A primeira está relacionada com o grau em que as conclusões de uma investigação podem ser generalizadas a outros sujeitos. Para isso é necessário que possuam as mesmas características, além da legitimidade da amostra com a população-alvo. A segunda está relacionada com as condições com que a investigação foi realizada, refletindo sobre as possibilidades de generalizar de uma situação para outra. As condições situacionais equivalentes correspondem a natureza das variáveis, o ambiente físico, o momento do dia e/ou ano, os efeitos causados nos sujeitos pelo próprio investigador ou pelas condições experimentais (Ibid., n.d.).

Foi realizada uma análise (ver no Apêndice C, p. 227) sobre a validade interna e externa quanto ao *design* do teste-piloto e para a aplicação do teste dentro de cada contexto. A aplicação mais ampla em cada contexto seguiu procedimentos equivalentes foram identificados os mesmos riscos.

3.1.1.2.6 Fiabilidade

Quanto à **consistência** das análises e das informações, se tem o aspecto de *fiabilidade* (quantitativo) ou *confiança* (qualitativo). Na abordagem quantitativa, a *fidedignidade* possui duas características fundamentais: a *precisão nos instrumentos de medida* e a *consistência nos resultados*. A consistência nos resultados, nesse caso pode ser alcançada através diferentes métodos:

- *Teste-reteste ou pré-teste e pós-teste* (aplicação do mesmo teste em momentos diferentes) – um teste fidedigno classificará os indivíduos aproximadamente na mesma ordem em duas aplicações;
- *Comparação de um conjunto de medidas entre duas variáveis através do coeficiente de correlação*, considerando que se forem diretamente proporcionais é +1, de modo perfeito (p. ex.: grau de temperatura do ambiente e dilatação da coluna de mercúrio em um termômetro); inversamente proporcionais é -1, de modo perfeito (p. ex.: agulha mais afiada necessita de menos força para enfiar); e quando não houver correlação será 0 (zero);
- *Testes paralelos* (diferentes testes em um mesmo momento) – Para isso são necessários que possua o mesmo(a): *nível de dificuldade; número de itens; área do conteúdo;*
- *Método das duas metades* (um único teste em uma única ocasião) – Para isso é analisada a correlação entre as metades do teste, um para os itens pares ou para os itens ímpares (Fórmula de Spearman-Brown: $r = \frac{2r'}{1+r}$, onde r é a fidedignidade de todo o teste e r' a correlação entre os itens ímpares e pares).

O indicador de *consistência* entre duas medidas de uma mesma coisa pode ainda ser considerada um dos tipos de *fidelidade*, e seu valor sendo expresso pela razão entre duas variâncias (Coutinho, 2011):

$$\text{Fidelidade} = \frac{\text{Variância nas pontuações reais}}{\text{Variância nas pontuações observadas}}$$

Entre algumas fontes de variabilidade do erro, geralmente estão associados a *falta de estabilidade do tempo* (p. ex.: um mesmo instrumento aplicado em momentos diferentes), *inconsistência no conteúdo* e sua interpretação (p. ex.: duas partes

semelhantes de um mesmo instrumento), e a *inconsistência do observador* (p. ex.: mesmo instrumento administrado por duas pessoas diferentes). Em outros casos, podem ainda existir *erros aleatórios*. Dentre as formas de expressar a fidelidade de um instrumento, tem-se:

Quadro 52: Tipos de fidelidade.

Que duas medidas?	Fonte de variabilidade	Tipo de fidelidade
Dois instrumentos separados no tempo	Tempo	Estabilidade
Dois instrumentos simultâneos	Conteúdo, interpretação	Consistência
Duas metades do mesmo instrumento	Conteúdo, interpretação	Consistência, estabilidade
Mesmo instrumento administrado 2 vezes em tempos diferentes	Tempo	Estabilidade
Mesmo instrumento administrado pelo mesmo avaliador em tempos diferentes	Tempo	Consistência
Mesmo instrumento administrado por 2 observadores diferentes	Observador	Consistência

Fonte: Ibid. (2011).

Podemos destacar alguns coeficientes de fidelidade, no caso: o *coeficiente de estabilidade do tempo* (teste-reteste) através do coeficiente de Pearson (superior a 0,80, os alunos tender a pontuar de forma semelhante; *coeficiente de equivalência* (duas versões equivalentes de um teste para um mesmo domínio/construto a seguir); *consistência interna* sendo a média das correlações entre todos os itens ou partes, e o número de itens ou partes (coeficiente de bipartição do teste; *alpha* de Cronbach; coeficiente de Kuder-Richardson; índice de dificuldade; poder discriminativo); *fiabilidade entre observadores* (fiabilidade intra observadores – estabilidade; acordo de observadores ou consenso – replicabilidade; acurácia⁷⁶).

Quanto à consistência de um teste, considera-se que ele “pode ser-se fiável, mas não válido, mas não pode ser válido sem ser fiável” (Amado, 2013, p. 367). Dessa forma, a fidedignidade é o fundamento ou a base da validade, se um teste não é fidedigno, não pode ser válido (p. ex.: um relógio que se adianta ou atrasa as horas não é fidedigno, e, portanto, não é um instrumento válido de medida), ou seja, um

⁷⁶ Indica o grau de rigor da avaliação feita por um instrumento, o que na prática se traduz pela comparação de uma codificação individual com a codificação norma ou padrão reconhecida como “correto” (Coutinho, 2011). Esta associado a noções de *exatidão* e *precisão* numa medição. Na Física também pode representar a proximidade entre o resultado de um instrumento de medida e o verdadeiro valor do que foi medido.

instrumento pode produzir medidas consistentes, porém sem medir aquilo que se propõe a medir (medir consistentemente a coisa errada). Um teste para ser válido ele tem que ser fidedigno.

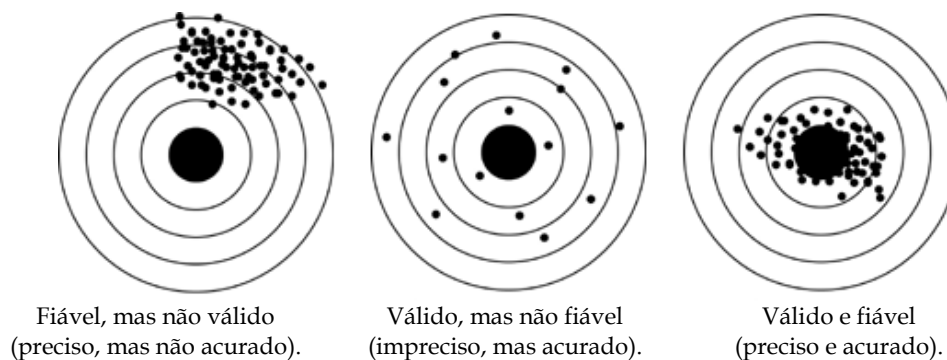


Figura 30: Diferença entre fiabilidade e validade.

Fonte: "QMSS e-Lessons | Validity and Reliability," (n.d.)

Em Amado (2013) esse aspecto de comparação seria a "consistência", entretanto optou-se pelo termo de *fidedignidade* pelo fato de considerar a consistência interna uma qualidade psicométrica inerente a fidedignidade, entretanto levando em consideração os aspectos que devem ser de natureza qualitativa e não quantitativa. Dessa forma, a fidedignidade de um instrumento de medida inclui três qualidades psicométricas: *estabilidade temporal* (ênfase quantitativa); *acordo entre juízes ou observadores* (ênfase quantitativa ou qualitativa - cerca de 80% a 85% de grau de concordância); e *consistência interna* (homogeneidade dos itens em um instrumento de medida ou fiabilidade do instrumento relacionado a consistência de um método - ênfase quantitativa; confiança nas intenções e nos processos metodológicos do investigador - ênfase qualitativa). Um teste além da fidedignidade deve possuir validade. A fiabilidade é uma qualidade que consiste em encontrar os mesmos resultados, quer em pessoas diferentes, quer uma mesma pessoa em momentos diferentes, quer por uma nova pessoa em relação a um protocolo definido, etc. Em outro sentido, também pode-se dizer que está relacionada à qualidade da aplicação da estratégia (De Ketele & Roegiers, 1993, pp. 81-82).

3.1.1.2.7 Neutralidade

A **neutralidade** está relacionada a imparcialidade durante os procedimentos da pesquisa é um aspecto de grande importância para a credibilidade das conclusões, considerando que a interferência do pesquisador em escolhas e numa posição de valor devem de preferência ocorrer no momento inicial e final da pesquisa. Com a valorização da intersubjetividade e confirmabilidade, busca-se alcançar a imparcialidade, ou seja, uma equitativa das diferentes (e, porventura, conflitantes) perspectivas dos implicados ou participantes (Rodrigues, 1991, p. 40, *citado por* Amado, 2013, p. 368). Deve-se evitar o perigo de “tornar-se nativo” (*going native*), sendo incapaz de fazer análises isentas das situações estudadas (Vieira, 1998, *citado por* Amado, 2013, p. 369), com isso torna-se importante executar “o exercício dialético de proximidade e afastamento, para a qual deve ter sido preparado quando se formou como investigador” (Estrela, 2007, p. 27, *citado por* Amado, 2013, p. 369). Diante de teorias prévias e preconceitos do investigador (*research bias*), deve-se fazer um esforço de flexibilidade e subjetividade disciplinada (Maxwell, 1996, p. 90, *citado por* Amado, 2013, p. 368). Considera-se, entretanto, que a reatividade das pessoas investigadas (*reactivity*) não precisa ser eliminada das probabilidades de ocorrências, mas tentar compreendê-las e usá-las, se possível, de modo produtivo para fins da pesquisa (Ibid., 2013, p. 368).

A neutralidade também pode ser contemplada com escolhas conscientes e disciplinadas durante o processo, seguindo critérios a medida em que os dados são produzidos. Apesar de critérios terem sido definidos no início da investigação, esse foi o caminho adotado durante as análises. Contudo, pode-se dizer que este critério foi garantido não apenas pelo investigador, mas como pelo *design* investigativo, na qual os professores, apesar de saberem do âmbito da pesquisa e da finalidade do teste, não obtiveram informações que pudessem influenciar de modo comprometedor os seus modelos de ensino, pois não tiveram contato com os resultados obtidos e nem acesso à chave-de-correção, com isso, entende-se que se torna mais “puro” aferir sobre a qualidade do modelo de ensino e quanto aos níveis de compreensão dos alunos.

3.1.1.3 O Processo de Recolha de Informações

3.1.1.3.1 Aspectos Legais e Éticos na Investigação⁷⁷

A legalidade no contexto português da investigação está amparada na autorização N° 4277/2013 dada pela *Comissão Nacional de Protecção de Dados* (CNPD), e na aprovação do inquérito com registro N°. 0364600001 dada pela *Monitorização de Inquéritos em Meio Escolar* (MIME), de acordo com a *Direcção Geral de Educação* (DGE) ligada ao *Ministério da Educação*. No Brasil a investigação foi registrada na *Comissão de Ética em Pesquisa da Universidade do Estado do Amazonas* (CEP-UEA).

Das questões éticas se tem uma abordagem respeitando a hierarquia da estruturação e organização escolar em cada contexto, partindo de autorizações e encaminhamentos das instituições universitárias responsáveis da investigação para os órgãos de gestão educacional, seguindo para a autorização dos gestores para viabilizar a pesquisa e ser conduzido ao contato com os professores. As formas de comunicação e abordagem seguiram padrão de respeito e consentimento, sendo os professores-participantes e encarregados de educação submetidos a assinarem o *Termo de Consentimento de Livre e Esclarecido TCLE* (em anexo).

Destacam-se algumas recomendações pertinentes da *International Test Commission* (ITC) para o uso de testes educacionais (Cheung et al., 2014a):

- **Permissão para o uso de testes em pesquisa:** a responsabilidade varia entre os de *domínio público*⁷⁸ e de direitos autorais;
- **Permissão para impressão e reimpressão:** somente para aqueles com direitos autorais (não é o caso);
- **Modificação do teste ou de seus componentes:** somente para aqueles com direitos autorais (não é o caso);

⁷⁷ Documentos em anexo.

⁷⁸ Segundo o ITC, os testes de domínio público podem ser utilizados por qualquer pessoa e são considerados propriedades comunitárias e podem ser utilizados sem a permissão do autor, ainda assim, a autora foi contactada via *email* e solicitado seu consentimento para fins acadêmicos (ver no Apêndice G1, p. 280).

- **Uso ético de testes:** os resultados devem ser tratados confidencialmente;
- **Documentação:** apresentar seção de métodos de todos os relatórios de pesquisa permitindo que a comunidade científica avalie a qualidade e os resultados da pesquisa e facilite sua replicação; deve-se fazer menção ao nome do teste, a edição e a data de publicação do instrumento original ou adaptado; Pesquisadores que traduzem ou adaptam um teste deveriam declarar a maneira como estas modificações foram feitas e fornecer evidências de validade e equivalência das formas traduzidas ou adaptadas.
- **Conflitos de interesse:** reconhecer a fonte de recursos (financiadas por agências externas); manter sua dedicação para a objetividade e imparcialidade;
- **Divulgação de edições de pesquisa na prática profissional:** somente prática profissional (não é o caso).

3.1.1.3.2 População e Amostra

Uma representatividade significativa dos alunos foi alcançada pela quantidade de escolas com critério de no mínimo 20% da população, esse aspecto reflete no **tamanho da amostra** alcançado na aplicação do teste TCE, se caracterizando como de **larga escala**.

Os processos da investigação foram desenvolvidos de forma semelhante em dois contextos geográficos, primeiramente no ano de 2014 em Portugal, na *Região Centro*, depois no Brasil em 2015, na *Grande Florianópolis*, capital do Estado de Santa Catarina. Em certa parte, Portugal se estabelece como referência comparativa na avaliação de desempenho reforçado pelos seguintes critérios: (1) ser um país europeu com o indicador do *Programme for International Student Assessment – PISA* mais elevado em relação ao Brasil (apesar de não estar entre os mais destacados países da OCDE); (2) devido ao Brasil ter como meta 7 do *Plano Nacional de Educação (PNE 2014-2024)*, aproximar o *Índice de Desenvolvimento da Educação Básica (IDEB)* da média dos países europeus (Brasil, 2014). A escolha do contexto de estudo possui um *caráter de conveniência* ao investigador devido às limitações de deslocamento em extensos territórios geográficos por força das escolhas no processo de recolha de informações.

3.1.1.3.2.1 Portugal

Um dos parâmetros criteriosos e relevantes foi a escolha de escolas secundárias (equivalente ao ensino médio) que adotam o *Livro Didático ou Manual Escolar de Física A (FA)*⁷⁹, pois correspondem às escolas públicas de regime regular semelhantes às do Brasil e que contemplam a unidade temática que se refere ao objeto de estudo, o conteúdo de termodinâmica. Para determinar o tamanho da amostra foi realizado um levantamento de todas as escolas da Região Centro que obedecem a esse critério.

Em uma sondagem inicial no terreno, foram verificadas **86 escolas**⁸⁰ utilizando três manuais escolares de Física (LDFs), onde o mais adotado representa 49% das escolas, com uma correspondência para os mesmos autores nos manuais de FB incidindo 89,3%.

Quadro 53: Universo e população do estudo.

Universo e População – Escolas Públicas da Região Centro de Portugal	Freq.	%
Universo de Escolas da Região Centro	1620	100,0
Total de Escolas do Ensino Secundário	103*	6,4
Total de Escolas Secundárias com 10FQ	100	100,0
E.S. que possuem o 10º FA	100	100,0
E.S. que possuem o 10º FB	28	28,0
E.S. que possuem o 10º FA e FB	28	28,0
E.S. que possuem o 10º QA	100	100,0
E.S. que possuem o 10º QB	28	28,0
E.S. que possuem o 10º QA e QB	28	28,0

*Três escolas são vocacionais.

Esse tipo de recurso demonstra um aspecto intraescolar comum, de ampla utilização, e que atende a um parâmetro de correspondência entre as características dos sujeitos na amostra, sendo, portanto, uma das variáveis externas que podem influenciar supostamente o desempenho entre aqueles que utilizam esse recurso de forma exclusiva ou mais enfatizada entre os seus hábitos de aprendizagem.

⁷⁹ FB se referem aos manuais escolares dos cursos tecnológicos, e de forma similar serve a interpretação para os *manuais* (Livros Didáticos) de química (QA e QB). Os quantitativos iguais entre FA e QA (entre outros) é devido as componentes disciplinares neste nível de escolaridade serem consideradas partes de uma mesma disciplina denominada Físico-Química (FQ).

⁸⁰ Ver o Apêndice B, p. 217.

Tabela 9: Codificação e descrição dos LDFs recomendados pelo MEC/Portugal e quantitativo de escolas da Região Centro que adotaram na vigência 2008-2014.

Cód.	Física e Química A - CCH (CT - FE)	
	Título, Autores e Editora	%
LDF_1	"Física A 10" - Fernanda Resende, António José Silva e Manuela Ribeiro, Areal Editores, SA.	2,0
LDF_2	"Manual de Física 10º" - Laila Ribeiro, Asa Editores II, AS.	1,0
LDF_3	"Física e Química A 10º ano - A Física do nosso mundo" - Luísa M., Máximo F., Constância Ed.	0,0
LDF_4	"Física, uma Aventura - Física A 10.º Ano" - Vítor Teodoro, Didáctica Editora, Lda.	1,0
LDF_5	"Desafios da Física A 10" - Daniel Marques Silva, Lisboa Editora, S.A./Raiz Editora.	3,0
LDF_6	"Ver + 10.º Ano Física A (Novo)" - Alexandre C., Augusto M. e Francisco C. Plátano Editora, Lda.	1,0
LDF_7	"Eu e a Física - Física e Química A - 10" - Maria J. C., Maria M. Gradim, Noémia M. Porto Editora.	0,0
LDF_8	"Ontem e Hoje - Física A - 10" - Adelaide Bello, Helena Caldeira ⁸¹ . Porto Editora.	25,0
LDF_9	"Física na Nossa Vida - Física e Química A - 10º/11º" - Fernando M., M. Margarida. Porto Editora.	12,0
LDF_10	"Energias 10 - Física A" - Maria Fernanda Barbosa, Maria João Morgado. Santillana-Constância.	6,0
LDF_11	"10F - A" - António J. F., Carlos Fiolhais, Graça Ventura, João Paiva, Manuel Fiolhais. SM - Editorial.	49,0
LDF_12	"Física A - 10.º Ano" - Maria Teresa Marques de Sá. SM - Editorial.	0,0
Física e Química B - CT (CCE/EE/INF - FC)		F
LDF_13	"Ontem e Hoje 10 - Física B" - Adelaide Belo, Helena Caldeira, João Gomes. Porto Editora.	10,7
LDF_14	"10F - B" - António J. F., Carlos Fiolhais, Graça Ventura, João Paiva, Manuel Fiolhais. SM - Editorial.	89,3

*A parte sombreada se refere aos LDFs que incidem entre as escolas da amostra. Fonte: Adaptado de Adoção de Manuais Escolares | Direção-Geral da Educação (n.d.)

A amostra das escolas no contexto português que corresponde a 20% do total das escolas secundárias na Região Centro, sendo a metade adotando o manual escolar (LDF_11) mais adotado na população, praticamente seguindo na mesma proporção.

Quadro 54: Codificação e descrição das escolas públicas na amostra na Região do Centro (Portugal Continental) que foram aplicados o TCE, assim como os respectivos LDFs (código) adotados.

N	Código* das Escolas Secundárias**	LDF
1	E1	LDF_9
2	E7	LDF_9
3	E10	LDF_9
4	E18	LDF_9
5	E36	LDF_8

⁸¹ Também é uma das autoras do programa curricular de Físico-Química A e B, vigente no contexto de aplicação do teste.

6	E37	LDF_8
7	E38	LDF_11
8	E39	LDF_11
9	E40	LDF_11
10	E41	LDF_11
11	E42	LDF_10
12	E43	LDF_8
13	E44	LDF_8
14	E45	LDF_5
15	E46	LDF_11
16	E47	LDF_11
17	E50	LDF_11
18	E54	LDF_11
19	E100	LDF_11
20	E101	LDF_11

Livros Didáticos de Física					Total
LDF_5	LDF_8	LDF_9	LDF_10	LDF_11	
1	4	4	1	10	20

*Codificação de acordo com a listagem de organização das escolas MEC/Portugal da Região Centro.

**Todas as escolas secundárias são de ensino integral.

A amostra no contexto português se distribui entre 1313 alunos do 10º ano, que são predominantes (50,79%) diante dos alunos do 11º ano e 12º ano.

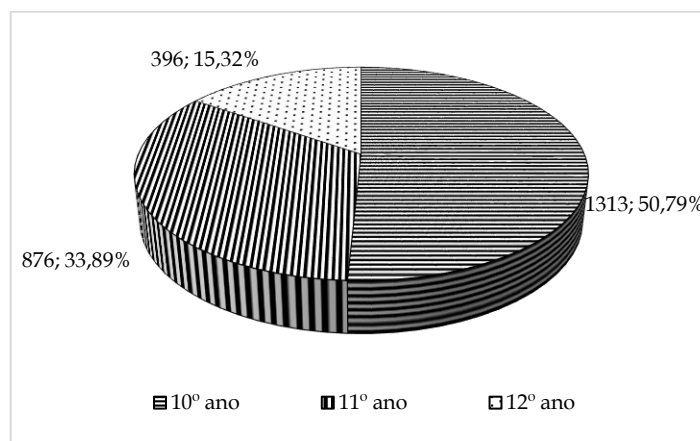


Figura 31: Proporção da amostra portuguesa com referência aos níveis de escolaridade.

A distribuição percentual da amostra com relação as escolas nos permite identificar a representatividade de acordo com essa categoria, que é a principal variável que caracteriza os aglomerados de alunos. Também pode ser observado a informação quantitativa por escola e o LDF adotado.

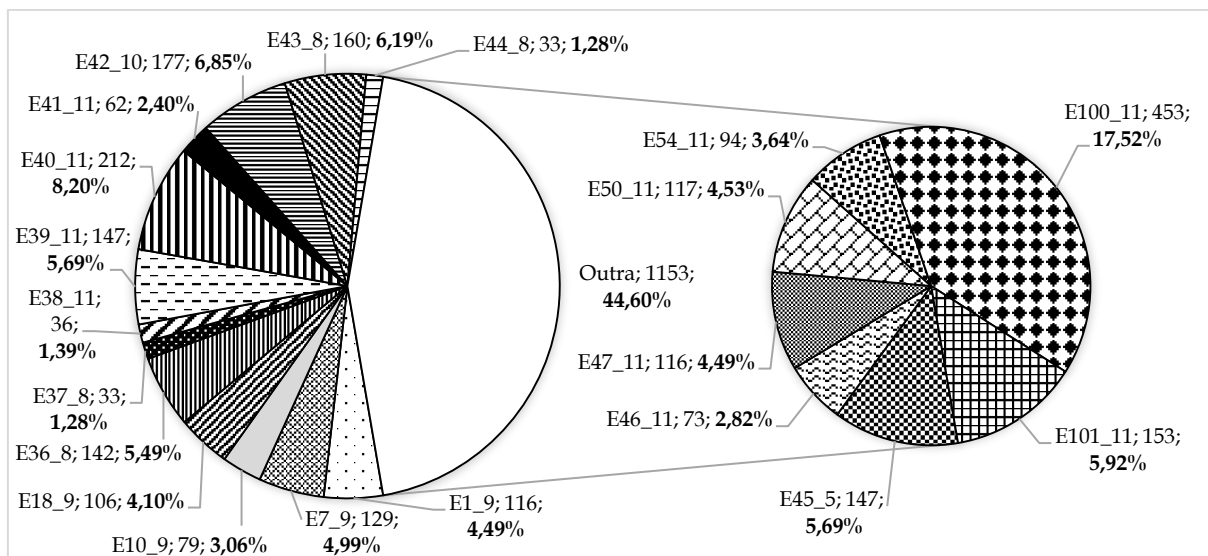


Figura 32: Quantitativo e percentual de alunos por escola no contexto português, juntamente com os livros didáticos de Física adotados.

Para se ter uma noção da distribuição dos grupos puramente relacionado aos LDF adotados nas escolas, sendo esse um parâmetro secundário de referência, tem-se:

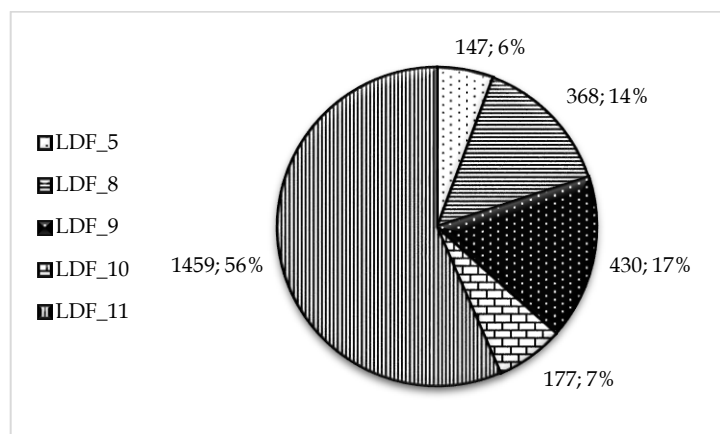


Figura 33: Proporção de alunos nas escolas portuguesas de acordo com os livros didáticos de Física adotados.

Buscando um controle sobre a complexidade de inúmeras variáveis externas (parasitas e moderadoras) da amostra que possam negativamente influenciar as inferências e interpretações, vale ressaltar dois aspectos de coerência como distinção na linha de corte para as comparações das diferentes e comuns concepções térmicas, sendo critérios característicos e inerentes ao *design* da pesquisa em ambos os contextos para uma determinada escola, que são:

- Alunos do 10º/1ºano (GC) não receberam a instrução em Termodinâmica (aplicação do teste no início do ano letivo), e os alunos do 11º/2º ano e 12º/3º ano (GE) que já receberam a instrução e que utilizaram os mesmos LDF (forma de tratamento);
- Grupos experimentais (GE) de desempenho relacionados aqueles *utilizam o LDF com maior incidência* (forma de tratamento) dentro da população, e como grupos de controle (GC) aqueles analisados com *LDFs diferentes* do grupo experimental.

A intenção com isso se volta mais como controle de variáveis do que propriamente avaliar de forma mais precisa um tratamento intraescolar, pois existem diversas variáveis que estão supostamente influenciando o desempenho dos alunos e que não foram observadas internamente em salas de aula. Dessa forma, essa perspectiva não se sustenta suficientemente para uma relação de causa e efeito (não há como se comprometer com o princípio de causalidade fenomenológica devido à falta de precisão, ou seja, não se estabelece um controle maior sobre outras variáveis envolvidas no que diz respeito as causas das dificuldades na aprendizagem dos alunos), mas sim como um parâmetro de controle, organização e sistematização nas análises da pesquisa. Os grupos de desempenho considerados com base nos LDFs possui a finalidade de parâmetro de controle sobre uma informação intraescolar de referência complementar na análise comparativa. De um modo geral, é natural que os desempenhos dos alunos se caracterizem como evidências concretas sobre a qualidade do ensino de Física térmica.

A unidade de amostragem está relacionada aos alunos organizados por escola, ano letivo e nível de escolaridade. Os critérios estabelecidos, de modo geral para os contextos, diferem dos *designs puramente experimentais* por haverem restrições na tentativa de considerar uma **designação aleatória** de indivíduos, tendo em vista que já se encontram previamente organizados em grupos (nas salas de aula, nas escolas), assim, as análises partem de informações específicas da amostra (TCT), e seguem para uma perspectiva analítica posterior (TRI) assumindo como unidade homogênea na amostragem os alunos por níveis de escolaridade, sem distinção entre escolas. Do

mesmo modo, também não foi realizada uma **destinação** (distribuição) **aleatória** dos sujeitos integrantes a grupos de comparação (experimental e de controle).

3.1.1.3.2.2 Brasil

As perspectivas e considerações gerais da população e amostra para Portugal e Brasil diante do *design*, riscos de validade, entre outros, são equivalentes.

A localização da pesquisa no Brasil foi no Estado de Santa Catarina (SC), concentrado na região da Grande Florianópolis. A amostra corresponde a **20,2%** da população das escolas estaduais da Grande Florianópolis, e **36,4%**, considerando somente os dois municípios incidentes na pesquisa.

Tabela 10: Quantidade de escolas estaduais em Santa Catarina e na Grande Florianópolis, em 2015.

Universo de Escolas em Santa Catarina		F	%
Total Estadual		1229	20,7
Total Municipal		3804	64,0
Total Privada		884	14,9
Total Federal		30	0,5
	<i>Total</i>	5947	100,0
Total de Escolas Estaduais na Grande Florianópolis – População		119	9,7
Amostra – <i>com relação as Escolas Estaduais na Grande Florianópolis.</i>		24	20,2
Total de Escolas em Florianópolis e São José (municípios incidentes na amostra)		66	55,5
Amostra – <i>com relação os Município de Florianópolis e São José.</i>		24	36,4

Fonte: Adaptado de planilha fornecida pela *Assessoria de Análise e Estatística* da Secretaria de Estado de Educação de Santa Catarina.

A distribuição das escolas por município demonstra uma concentração na capital Florianópolis (37,8%), seguida de São José (17,6%) e Palhoça (16%).

Tabela 11: Quantidade de escolas estaduais nos municípios da Grande Florianópolis.

N	Municípios da Grande Florianópolis	F	%
1	Palhoça	19	16,0
2	Águas Mornas	2	1,7
3	Angelina	3	2,5
4	Anitápolis	1	0,8
5	Antônio Carlos	1	0,8
6	Biguaçu	17	14,3
7	Florianópolis	45	37,8

8	Governador Celso Ramos	2	1,7
9	Rancho Queimado	2	1,7
10	Santo Amaro da Imperatriz	4	3,4
11	São Bonifácio	1	0,8
12	São José	21	17,6
13	São Pedro de Alcântara	1	0,8
<i>Total</i>		119	100

Fonte: Adaptado de planilha fornecida pela *Assessoria de Análise e Estatística* da Secretaria de Estado de Educação de Santa Catarina em 2015.

As unidades escolares estaduais são organizadas entre *Escolas de Educação Básica* (EEB) na qual podem agregar *Escolas de Ensino Fundamental* (EEF), *Ensino Médio* (EEM) e algumas relacionadas ao *Projeto Escola Pública Integrada* (EPI). No caso, em 2004 o Estado de SC assume as diretrizes do programa das EPIs e são escolas baseadas “no conceito de escola integrada à sua comunidade e aos seus recursos educativos, em currículo em tempo integral e na gestão compartilhada entre Estado e Município” (Ramos, 2011, p. 50).

Nas dez escolas em destaque (com sombreamento – Tabela 12) não houve resposta quanto a recolha de dados, portanto o total da amostra considerado na aplicação do teste TCE foi de 18 escolas. Isso aconteceu por terem sido recolhidas outras informações que envolviam os livros didáticos de Física (professores e alunos) e entrevistas (professores), entretanto, devido aos desdobramentos e delimitações na definidos no percurso da investigação, não foram considerados pertinentes e necessários a esse estudo, porém serão úteis em uma investigação complementar de aprofundamento.

Tabela 12: Codificação e descrição dos LDFs recomendados pelo PNLD-2014 e quantitativos distribuídos no Estado de Santa Catarina (SC/Brasil) na primeira remessa da vigência 2014-2016.

N	Código	Lista dos Livros Didáticos de Física (LDFs) recomendados pelo MEC-2012		
		Título, Autores e Editora	Qtd.	Ord.*
1	1br	“Compreendendo a Física”. Alberto Gaspar. Editora Ática.	202.818	8
2	2br	“Curso de Física”. Antônio Máximo Ribeiro da Luz, Beatriz Alvarenga Alvarez. Editora Scipione.	286.303	6
3	3br	“Conexões com a Física”. Blaidi S.; Glória Martini; Hugo Carneiro Reis; Walter Spinelli. Editora Moderna.	374.928	2
4	4br	“Física - Ciência e Tecnologia”. Carlos M. A. T., Nicolau G. F., Paulo Antonio de Toledo S.. Editora Moderna.	334.886	3
5	5br	“Quanta Física”. Carlos Aparecido Kantor, Lilio Alonso P. J., Luis Carlos de M., Marcelo de Carvalho. B., Osvaldo C., Viviane M.. Editora PD.	145.856	9
6	6br	“Física”. Gualter, Helou e Newton. Editora Saraiva	297.463	5

7	7br	“Física Aula por Aula”. Benigno Barreto Filho, Cláudio Xavier da Silva. Editora FTD.	720.928	1
8	8br	“Física e Realidade”. Aurélio Gonçalves Filho, Carlos Toscano. Editora Scipione.	86.944	10
9	9br	“Física em Contextos Pessoal-Social-Histórico”. Alexander P., Maurício P., Renata A., Talita R.. Editora FTD.	227.735	7
10	10br	“Física para o ensino Médio”. Fuke, Kazuhito. Editora Saraiva.	324.885	4

Fonte: Adaptado do Portal do FNDE - Guias do Livro Didático (n.d.).

*Ordem dos mais disseminados entre as escolas públicas brasileiras.

A amplitude no processo de sondagem das escolas pode ser visualizada (Tabela 13) com detalhes sobre o tipo de modalidade de ensino da instituição escolar. A amostra, naturalmente acabou se definindo pelas instituições na qual obteve-se a colaboração dos professores na pesquisa.

Tabela 13: Codificação e descrição das escolas públicas (de acordo com as modalidades de ensino) na Grande Florianópolis (Brasil) que foram aplicados o TCE, assim como os respectivos LDFs (códigos) adotados – refere ao ano de 2015.

N	Código das Escolas de Ensino Médio	TCE	LDF	M	F	M/F	EMI
1	1BR	X	10br			X	X
2	2BR	X	10br			X	
	3BR		-			X	
3	4BR	X	7br			X	
4	5BR	X	4br			X*	X
	6BR		-			X	
5	7BR	X	7br			X	
	8BR		6br			X	
	9BR		-			X**	X
6	10BR	X	9br			X	
	11BR		-		X		
7	12BR	X	4br			X	
8	13BR	X	9br			X	
9	14BR	X	9br			X	
10	15BR	X	7br	X			
	16BR		7br	X			
	17BR		-	X			
11	18BR	X	2br			X	
12	19BR	X	7br	X			
13	20BR	X	9br	X			
	21BR		6br			X	
14	22BR	X	2br	X*			
	23BR		2br			X	
15	24BR	X	4br	X			
16	25BR	X	6br			X	
17	26BR	X	6br			X	
18	27BR	X	2br	X***			

28BR						7br	X***			
Livros Didáticos de Física						Total	M	F	M/F	EMI
2br	4br	6br	7br	9br	10br					
3	3	2	4	4	2	18	9	1	18	3

Legenda: M = Ensino Médio; F = Ensino Fundamental; EMI = Ensino Médio Inovador.

*Educação de Jovens e Adultos.

**Magistério.

***Possuem modalidade de ensino técnico-profissional.

Quanto aos detalhes da amostra nos níveis de escolaridade e turnos, tem-se:

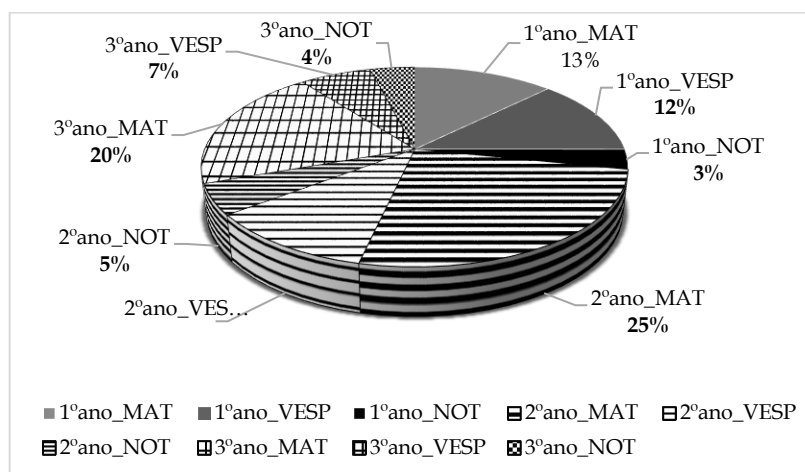


Figura 34: Proporção da amostra brasileira com referência aos níveis de escolaridade e turnos.

Levando em consideração os diferentes turnos, na qual o teste foi aplicado, a amostra no contexto brasileiro se delimita em um total de 835 alunos no 1º ano, 1222 alunos no 2º ano e 900 no 3º ano, que se distribui percentualmente com predominância do 2º ano matutino, correspondendo $\frac{1}{4}$ dos alunos. Na distribuição de alunos por escola, juntamente como a informação dos livros didáticos de Física adotados (parâmetro secundário de referência), lembrando que o livro 7br foi o mais disseminado entre as escolas na vigência 2012-2014 no PNLD.

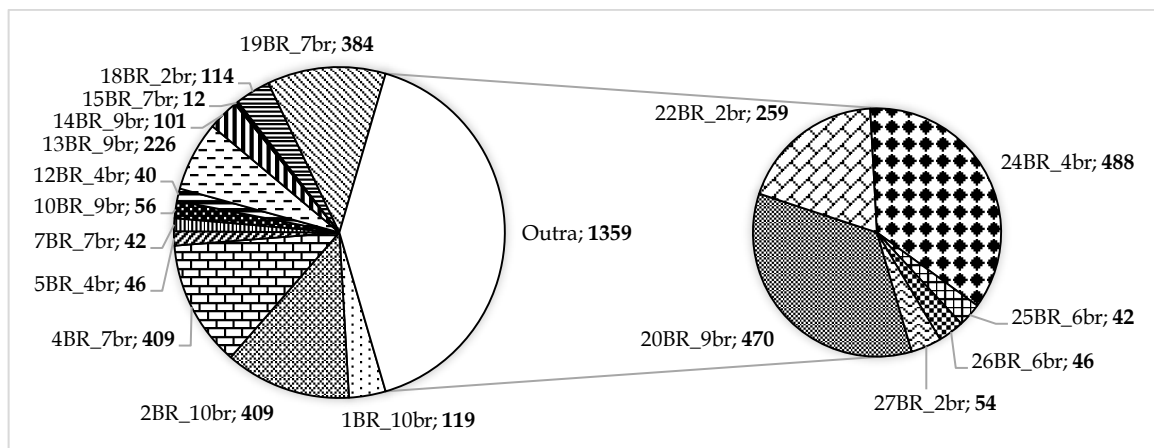


Figura 35: Quantitativo de alunos por escola no contexto brasileiro, juntamente com os livros didáticos de Física adotados.

Também é possível verificar o quantitativo de alunos da amostra por escola no Brasil e em Portugal (ver no Apêndice B3, p. 225). Abaixo (Figura 36) é possível se obter uma melhor compreensão sobre a proporção estimada dos grupos normativos de respondentes tendo como base os Livros Didáticos de Física adotados nas escolas, sendo útil para estimar de forma indireta, possíveis influências relacionadas a essa variável.

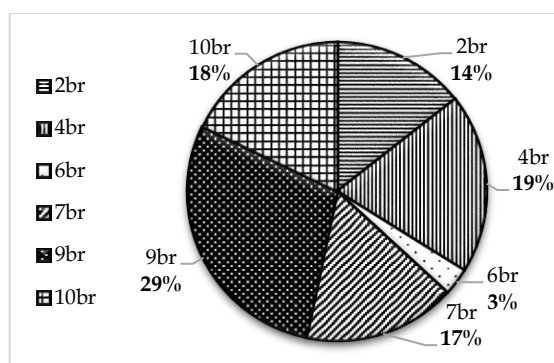


Figura 36: Proporção de alunos nas escolas brasileiras de acordo com os livros didáticos de Física adotados.

3.1.1.4 Distribuição dos Materiais de Coleta de Dados

O estudo buscou o alcance possível de 132 professores (total de professores de Física nas escolas da amostra), em que foram confirmados previamente 68 professores. Para eles foram entregues 132 kits, onde cada kit⁸² equivale a: 1 carta de apresentação; 1

⁸² Incluem documentos formais de solicitação de apoio a direção da escola, e a carta de apresentação com detalhes da investigação à direção da escola e aos professores.

grelha de apreciação dos manuais; 1 termo de consentimento de livre e esclarecido; 30 testes de termodinâmica; 30 termos de consentimento aos encarregados de educação. Existe um alcance de 72 turmas de 10º ano, 55 turmas de 11º ano e 37 turmas de 12º ano, equivalendo um total de 164 turmas. Com uma estimativa de que cada turma possua em média 25 alunos, o alcance máximo foi de 4100 alunos, entretanto, na realidade, não houve uma predisposição de todos os professores e uma perda considerável nessa projeção (Tabela 14).

Tabela 14: Quadro descritivo da amostra quanto ao distrito, concelho, ordenação, alcance de professores, confirmações de apoio, quantidade de kits entregues, quantidade de turmas e Livro Didático de Física (LDF) adotado nas escolas.

Dist.	Concelho	Código das Escolas Secundárias	Prof.	Entreg.	Kits	10º	11º	12º	LDF	
Aveiro	Águeda	E1	5	1	5	2	2	2	9	
	Aveiro	E7	9	-	10	4	4	3	9	
	Estarreja	E10	11	1	9	3	3	3	9	
	Ovar	E18	7	-	9	3	3	3	9	
Coimbra	Cantanhede	E36	9	9	10	4	5	1	8	
	Coimbra	E37	2	2	2	1	1	-	8	
	Coimbra	E38	3	3	3	2	2	-	11	
	Coimbra	E39	6	6	7	3	3	1	11	
	Coimbra	E40	10	2	13	6	6	1	11	
	Coimbra	E41	2	2	3	1	2	0	11	
	Coimbra	E42	10	1	8	6	-	2	10	
	Coimbra	E43	8	8	11	4	5	2	8	
	Fig. da Foz	E44	3	3	3	2	0	1	8	
	Fig. da Foz	E45	10	1	8	3	3	2	5	
	Fig. da Foz	E46	9	2	17	5	6	6	11	
	Lousã	E47	4	4	5	3	2	0	11	
	Montemor	E50	4	4	8	3	3	2	11	
	Soure	E54	6	6	5	2	2	1	11	
Viseu	Viseu	E100	8	8	17	12	0	5	11	
	Viseu	E101	6	6	7	3	3	2	11	
TOTAL	11	-	20	132	68	164	72	55	37	-

3.1.1.4.1 Implicações do método de abordagem: recepção nas escolas

O método de abordagem nas escolas nos dois contextos seguiu o seguinte padrão (*modus operandi*): (1) apresentação formal de *documentos oficiais*⁸³ à direção escolar ou representantes; (2) identificar um professor de Física ou encarregado pedagógico que possa articular com os outros professores, tendo em vista a inviabilidade de contatar pessoalmente todos; (3) recolher informações de maior fiabilidade sobre a quantidade de turmas, número estimado de alunos por turma, quantidade de professores de Física na escola; (4) esclarecer verbalmente os aspectos de colaboração através de uma carta de apresentação e alguns pontos da pesquisa; (5) estabelecer o comprometimento de repassar os resultados da pesquisa aos colaboradores, assim como a disposição por eventual divulgação pública na escola; (6) solicitar os contatos e reforçar todas as solicitações e instruções de colaboração por e-mails (e alguns por telefone) previamente.

Aspectos geográficos⁸⁴ também devem ser levados em consideração por obstáculos óbvios. A determinação do tamanho da amostra em 20 escolas foi um quantitativo justificado pela população de 100 escolas secundárias na Região Centro, sendo considerado aceitável representativamente quanto aos critérios de generalidade dentro da *metodologia científica*⁸⁵, pois a necessidade de obter uma quantidade mínima aceitável para o tratamento estatístico mais significativo, estando previsto às dificuldades de colaboração.

Estava previsto inicialmente coletar informações dos alunos dentro da amostra, porém, no retorno do TCLE dos encarregados de educação foram constatadas algumas resistências para a gravação de áudio, como havia sido previsto, segundo o relato de alguns professores. Essas autorizações serviram de parâmetro para a restrição e escolha de métodos de recolha de dados, na qual acabou por se limitar na aplicação do teste TCE.

⁸³ Condições éticas de pesquisa, de acordo com a autorização 4277/2013 dada pela Comissão Nacional de Proteção de Dados, que representa a Autoridade Nacional de Controlo de Dados Pessoais (Decreto-Lei nº 43/2004 de 18 de Agosto) e a Monitorização de Inquéritos em Meio Escolar, controlado pela Direção-Geral de Estatísticas da Educação e Ciência (Decreto-Lei nº 125/2011 de 29 de Dezembro)

⁸⁴ Com o auxílio do software ArcGis 10.2.

⁸⁵ Em outro aspecto da metodologia da pesquisa, diferentemente deste, relacionado a recolha de dados.

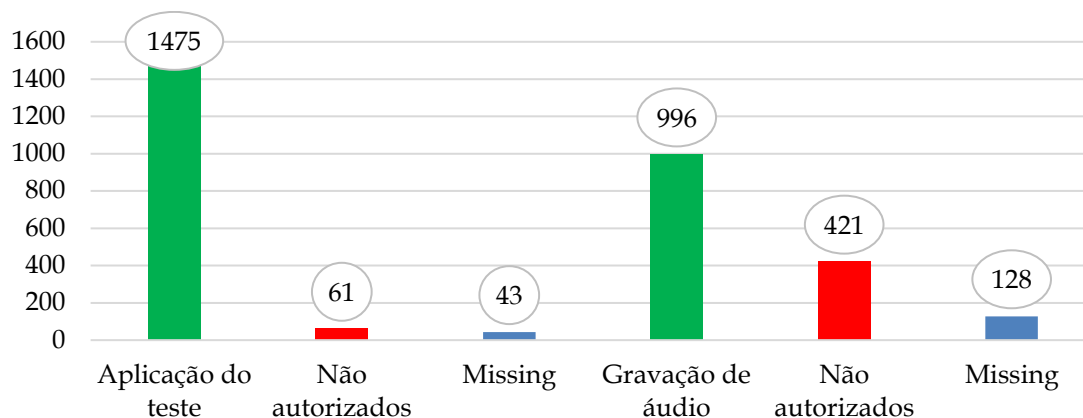


Figura 37: Gráfico quantitativo dos TCLE dos encarregados de educação em Portugal.

Dos TCLE direcionados aos encarregados de educação, obtiveram-se resultados de apenas 1475 autorizações a aplicação dos testes aos alunos, com rejeição de 61, com um total de 43 *missing* supostamente a falta de atenção no preenchimento. Quanto ao item referente a uma eventual gravação de áudio sem ala de aula, foram autorizados 996 com rejeição de 421, e 128 *missing*.

Quanto às turmas de 10^o anos, foram contabilizados 72, mas apenas 56 efetivamente participaram, com isso, dos 1680 testes distribuídos, foram utilizados 1280. Para as turmas de 11^o ano, tem-se um total de 55, com participação colaborativa de 42, onde dos 1260 testes distribuídos, 850 foram respondidos. E por último, de um total de 37 turmas do 12^o ano, participaram 32, onde dos 960 disponibilizados, 395 foram respondidos.

3.1.1.4.2 Algumas Considerações

Os testes foram entregues para cada turma sendo estimado um total de 30 alunos por turma. Dessa forma podemos perceber que o maior desvio está nas turmas de 12^o ano, com o menor número de alunos. O desperdício com os gastos nos materiais poderia ser evitado com um método de sondagem inicial nas escolas com uma maior precisão sobre o total de alunos por turma, porém é uma escolha que compromete ainda maiores gastos, principalmente no deslocamento, tendo em vista que a fiabilidade de recolher dados pessoalmente é maior pessoalmente que por telefone, ou até pelos dados informados na internet, onde geralmente estão desatualizados.

Foi possível observar riscos para a investigação ao condicionar o cumprimento ético das respostas dadas ao TCLE dos encarregados de educação, limitando os métodos de recolhas de informações, tendendo ainda a um comprometimento no tamanho da amostra de alunos participantes, entretanto, muitas dessas dificuldades foram superadas e/ou influenciaram desdobramentos investigativos, como a ênfase dada em uma determinada abordagem (qualitativa/quantitativa) por influência da predominância de um instrumento no processo de recolha de informações, o peso no uso exclusivo do teste aos alunos, por exemplo. Referente a isso, observou-se uma grande rejeição quanto a utilização do recurso de gravação de áudio nas escolas portuguesas, tendo em vista que basta apenas uma única rejeição dentro de uma turma para inviabilizar o processo de *recolha de dados*⁸⁶. Esses parâmetros restringiram esse tipo de abordagem na amostra brasileira, sendo definido a pesquisa com relação a investigação sobre as dificuldades conceituais no teste TCE. No Brasil, foram realizados ajustes, e foi seguido um modo de operação equivalente.

Alguns dos aspectos centrais do processo de recolha de informações resumem-se em:

Quadro 55: Resumo descritivo sobre principais atividades no processo de recolha das informações.

	Portugal	Brasil
Estimativa de visitas e distância média percorrida	Média de 2 visitas para cada uma das 20 escolas da <i>Região do Centro de Portugal</i> (raio médio em cerca de 90 km de Coimbra, percorrendo de carro aproximadamente uns 500 km) que fazem parte de sua amostra (população com cerca de 2,4 milhões de pessoas).	Média de 3 visitas para cada uma das 25 escolas públicas da <i>Região da Grande Florianópolis</i> (raio médio em cerca de 40 km do centro da capital do Estado de Santa Catarina, Brasil, percorrendo de carro aproximadamente uns 200 km) que fazem parte de sua amostra (população em torno de 1 milhão de pessoas).
Retorno do teste TCE	Foi aplicado a 2585 alunos do ensino secundário nos níveis de escolarização do 10º, 11º e 12º ano.	Foi aplicado a 2957 alunos do ensino médio nos níveis de escolarização 1º, 2º e 3º ano.
Considerações no tratamento dos dados	Somando o número de observações no instrumento TCE tem-se um total de 5542, sendo todos em formato de texto em papel, demandando realizar inicialmente uma organização e tratamento dos dados de cada instrumento um a um em uma planilha de <i>Excel</i> para depois serem implementadas as análises. O teste com 26 itens resulta em $26 \cdot 5542$ registros, ou seja, 144.092 registros computados manualmente.	

⁸⁶ Interpretar de modo equivalente as expressões “recolha de dados” e “recolha de informações”.

3.1.2 Procedimentos no Tratamento dos Dados

Quanto às **referências**, os procedimentos *de controle de dados* foram mediados com a utilização do *software* Mendeley para controle das produções e busca de trabalhos correlatos. Quanto às bases de dados, foram organizados, manipulados e articulados com outros *softwares* estatísticos utilizando em especial, o *software* Excel.

Quanto aos **procedimentos de análise de dados** foram realizados cruzamentos de diferentes métodos e técnicas estatísticas dentro da Psicometria para determinar uma maior consistência dos itens e das análises. Dentro da TCT foram determinados os índices de discriminação, índices de dificuldade, coeficientes bisseriais e ponto-bisseriais (aspectos importantes para a consistência dos itens), e a técnica de AGI.

Quanto às **operacionalizações**, na primeira etapa após o recolhimento dos testes, têm-se a computação da base de dados realizada manualmente em planilha *Excel*, sendo registrado o país, a data de realização da prova, a data de nascimento, gênero, a turma, a escola e o Livro Didático de Física adotado na escola, e em especial, as respostas assinaladas de cada teste.

A análise do TCE parte de uma *matriz de respostas*, na qual são considerados diferentes grupos de alunos em cada contexto, estando organizados por diversas categorias e níveis de escolaridade, e com isso avaliar o desempenho através dos escores juntamente com a interpretação qualitativa e pedagógica associada, realizando assim comparações *lado a lado* e colocando-os na mesma métrica.

O instrumento apresenta respostas fechadas e de múltipla escolha, sendo analisado com base na TCT e TRI. Inicialmente a base de dados se configura como um matriz-resposta representada com as letras das alternativas escolhidas em uma planilha de *Excel*, em seguida será convertida nessa mesma plataforma numa matriz-resposta dicotômica admitindo 1 para a resposta certa e 0 para os distratores. Essa matriz servirá de base para os dois campos de exploração das análises.

Na TCT serão realizadas preliminares diretamente da planilha e confirmadas com resultados do *software* BILOG-MG, sendo considerados valores de coeficientes

bisseriais corrigidos. Ampliam-se as análises apresentando os ganhos normatizados e fatores de concentração das respostas, tornando mais eficiente uma interpretação dos resultados obtidos para cada item. A AGI complementa e ilustra as análises preliminares, que serão seguidas com os resultados mais robustos da Psicometria moderna.

De modo geral e sistemático, os seguintes *procedimentos iniciais* no tratamento dos dados, sendo basicamente análises clássicas, são:

Quadro 56: Descrição das etapas iniciais típicas para os procedimentos de análise.

Etapas	Descrição operacional
1	Determinar a frequência e a média das incidências de cada alternativa para cada item.
2	Elaborar a matriz-resposta dicotômica ou binária (1 para a alternativa de acordo com gabarito e 0 para os distratores).
3	Determinar os escores brutos (soma dos acertos entre os 26 itens do TCE).
4	Organizar os escores brutos em ordem crescente (Rol).
5	Calcular a média, o desvio-padrão e a variância dos escores brutos.
6	Calcular a variância para cada item e soma das variâncias de todos os itens.
7	Determinar os grupos de desempenho de nível superior e inferior relativos a 27%. Eventualmente essa porcentagem pode ser sutilmente maior ou menor, pois devem ser considerados todos os respondentes com mesmo escore dentro dessa faixa.
8	Calcular a porcentagem de acerto do grupo superior e inferior correspondente a 27%.
9	Calcular o índice de discriminação do item subtraindo as porcentagens de acerto entre o grupo superior e inferior para cada item.
10	Calcular a consistência do teste (alfa de Cronbach).
11	Calcular o coeficiente de correlação bisserial e ponto-bisserial para cada item (não corrigido e corrigido).
12	Eliminar os itens com correlação ponto-bisserial (KR20, correspondente ao coeficiente de Pearson) próximo de zero ou negativos.
13	Recalcular as etapas anteriores verificando se ocorre um aumento na consistência do teste (alfa de Cronbach).
14	Aplicar diversos procedimentos de análise com a base de dados maximizada e ajustada de acordo com a maior consistência alcançada, para assim construir as interpretações e análises.

Na TRI as análises podem ser estimadas com base nos parâmetros de discriminação, dificuldade, acerto e erro ao acaso a partir do modelo logístico de um, dois, três e/ou quatro parâmetros, sendo escolhido aquele que obteve melhor ajuste. São exploradas diferentes *perspectivas* (devido ter sido utilizado um único instrumento de coleta) para os respondentes de diversos grupos de desempenho (diferentes níveis de escolaridade, escolas, gênero, turnos, LDF adotado), cruzando dados e interpretações obtidas das análises quantitativas (respostas dos itens) e qualitativas (análises prévias do instrumento) – interpretações dos itens; *inventário de conceitos* ou

quadro de tipicidade na relação dos itens com as concepções térmicas envolvidas; elaboração da matriz de referência), cruzamento de conclusões de diversos investigadores ao serem verificados os estudos semelhantes (procedimentos de análises semelhantes), e outros com o uso do mesmo instrumento.

Entre os procedimentos no tratamento dos dados estão incluídos a elaboração de quadros, tabelas e gráficos comparativos dos índices buscando assim constatar a correspondência e consistência dos resultados, para assim estimar os domínios e dificuldades diante do conteúdo térmico envolvido.

Na análise da DIF apresenta uma perspectiva comparativa sobre a funcionalidade dos itens entre os contextos, sendo destacado grupos normativos de gênero.

Através da matriz-resposta binária (acerto 1 e erro 0) ou dicotômica serão calculados alguns parâmetros clássicos para cada item, sendo necessários para estimar os valores dos coeficientes bisseriais (não corrigidos) dos distratores utilizar uma matriz que possibilite calcular seus percentuais modificando os valores de p na função densidade de distribuição normal padronizada $h(p)$ e os respectivos índices de discriminação. Existem alguns programas que foram desenvolvidos apropriadamente para realizar esses procedimentos (p. ex.: Iteman), entretanto, se optou pelo *Excel* ao fato de que contribui para uma apropriação melhor de todos os processos de análise. Os riscos de validade associados a esses procedimentos foram minimizados através da confirmação dos resultados clássicos obtidos pelos *outputs* do *software* BILOG-MG, na qual foi utilizado para algumas análises de verificação com o modelo de três parâmetros logísticos ou 3PL da TRI. Também são conferidos os parâmetros das análises através do pacote de suplemento do Excel denominado *eirt - Item Response Theory Assistant for Excel*.

Analisa-se a base de dados da TCT de acordo com cada contexto separadamente para as interpretações pedagógica dos itens, apesar de ter sido verificada posteriormente em um contexto geral, ou seja, estando Brasil e Portugal na mesma

base de dados. A matriz-resposta dicotômica do contexto geral foi utilizada como base de dados nas análises principais da TRI utilizando o *software* R, sendo complementadas com o *Winsteps*.

3.1.2.1 Correções na matriz-resposta

O compromisso com a clareza representa não apenas bom senso de qualquer pesquisa, como também confiabilidade e rigor científico, e assim possibilita moldar uma demarcação de validade do conhecimento construído, apontando erros e limitações que possam ser evitados ou minimizados durante os desdobramentos do processo investigativo.

Diante 144.092 registros sendo computados manualmente, por mais esforço e compromisso de haja, é natural que se identifique durante a varredura dos dados alguns problemas de registro devido a fadiga pelo demasiado tempo despendido para a realização da tarefa, que no caso corresponde a etapa de tratamento dos dados.. O total de erros de registros computados somam 23, o equivalente a 0,02% do total de respostas, na qual foram corrigidos quando possível ou anulados em casos extremos, tornando o viés praticamente nulo e sem efeitos negativos que possa comprometer as análises. Um dispositivo sugerido, eficiente e adequado para evitar esses erros, que torna otimizado o processo de tratamento e análise dos dados é feito através do *Scantron*. *Scantron* é um dispositivo utilizado para escanear formulários (do tipo cartão-resposta) utilizados em avaliações de larga escala, alguns deles acompanham *softwares* para análise preliminares automática de resultados, porém os custos para a obtenção desse equipamento nem sempre são viáveis economicamente.

A *base de dados* é construída em planilha com auxílio do *Excel*, plataforma útil na análise descritiva dos registros e na aplicação de outras fórmulas, equações e relações matemáticas, correspondentes a aspectos complementares na análise. A matriz-resposta é o principal conjunto de informações do banco de dados, portanto, é fundamental certificar que esteja o mais adequado possível, evitando erros sistemáticos, ou seja, vieses. Para isso, foi verificado se o total de respostas por item correspondia ao total de indivíduos para cada contexto. Em Portugal, por exemplo,

foram contabilizados 2585 respondentes, incluindo todos os alunos de todas as séries de todas as escolas. Este procedimento de varredura possui a intenção de aumentar a fiabilidade da matriz-resposta na versão dicotômica (0 e 1), formato na qual necessário para rodar nos programas de análise (BILOG-MG e R) utilizados. Todas as respostas não válidas foram consideradas erros, e para minimizar discrepâncias na matriz-resposta dicotômica, esses erros foram observados e corrigidos em dois momentos:

- *Na leitura inicial direta do teste ao realizar o registro na construção da base de dados.* Neste processo, todos os tipos de erros dos respondentes foram considerados *missings* (“-”), nisso incluem-se duplas-respostas, alternativas não assinaladas, ou qualquer forma de registro que cause a incapacidade de identificar de forma fiável a alternativa escolhida como resposta;
- *Na leitura da base de dados já construída, certificando eventuais erros de registro.* Entre as causas dos erros, incluem-se aquelas ligadas ao próprio *investigador* devido à falta de atenção por fadiga, equívocos, entre outros. Considera-se este uma *variável moderadora* no estudo, ou seja, que pode influenciar no estudo, mas que pode ser controlado, na intenção de alcançar a *neutralidade* nessas etapas de análise. Neste processo, foram verificados alguns problemas, no caso, 11 erros no item y1, 1 erro nos itens y6, y10, y12, y16, y18, y21 e y23. Incluem-se ainda, equívocos em questões com apenas 4 alternativas (A, B, C e D), na qual verificou-se registro na alternativa “E”. As únicas questões com alternativa “E”, são os itens y9, y11, y16, y18 e y26. Esse tipo de erro ao ser considerado *missing*, evitará discrepâncias nos resultados, de acordo a lógica nos processos de análise. As posições foram colocadas com a justificativa de identificar possíveis padrões entre as escolas envolvidas, porém sem maiores pretensões de avançar em possíveis fatores de correlação.

Apesar de não ser um erro sistemático, testes com elevadas incidências sobre uma mesma alternativa são um risco de validade para o escore, pois sinaliza falta de compromisso do respondente, apesar disso, optou por manter a base de dados de forma íntegra (Tabela 15).

Tabela 15: Correções de equívocos na base de dados.

Itens	y1	y6	y10	y12	y16	y18	y21	y23		
Posições	680, 682,683, 685, 687, 688, 689, 693, 698, 703 e 704	251	1917	2581	308	819	2112	616		
Escolas	1	40	100	45	40	10	7	101		
Erros	-B	BC	CC	CX	S	V	-D	-C		
Correção	B	-	C	C	-	-	D	C		
Itens	y8	y10	y12	y13	y15	y17	y19	y20	y24	y25
Equívocos de registro na letra "E"	2	6	2	1	3	3	2	1	1	2
Correção	-	-	-	-	-	-	-	-	-	-

Em uma perspectiva de análise complementar e futura, pretende-se eliminar da análise aqueles com incidência na mesma alternativa maior ou igual a 21 vezes dentro só 26 itens, ou se houver uma incidência consecutiva de uma mesma alternativa por 10 vezes. Este procedimento possivelmente poderia eliminar ruídos advindos dos respondentes, também considerados variáveis *moderadoras*, e assim tornar a matriz-resposta mais fiável. Abaixo (Tabela 16) tem-se 12 testes identificados e potencialmente recomendados a serem eliminados por esses critérios. Pode-se perceber uma maior ocorrência entre alunos da *escola secundária E18*⁸⁷.

Tabela 16: Elevadas incidências consecutivas de mesma resposta para os itens.

Alternativa	C	B	C	D	A	A	C	C	C	C	C	C
Incidências consecutivas	25	10	12	10	13	15	25	26	22	26	26	26
Posição	69	499	513	596	1101	1502	2188	2192	2193	2196	2197	2199
Escolas	39	54	54	101	36	50	18	18	18	18	18	18

Outro aspecto a ser levado em consideração é o elevado índice de *missing* por item e por teste. No primeiro caso, apesar de não ser uma justificativa suficiente para eliminação, sugere *evidências sobre problemas no item*. Na segunda situação, suposta falta de diligência ou má administração no processo aplicação do teste, ou ainda, indivíduos com pouca disposição a realização do teste. Como a aplicação foi realizada pelos professores colaboradores, esta é uma *variável parasita* pela incapacidade de determinar suas causas, ou de controlar, apesar de que cada um recebeu orientações verbais e uma folha instrucional, na tentativa de minimizar esses efeitos.

Quanto aos itens com grande incidência em *missings*, chama atenção alguns itens em destaque (Tabela 17).

⁸⁷ As escolas foram codificadas, levando em consideração a listagem oficial da Direção Geral de Educação da Região Centro de Portugal.

Tabela 17: *Missing* por item.

Itens	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
<i>Missing</i>	7	16	17	15	16	44	9	34	14	18	32	25	12	28	27	15	24	27	35	29	73	38	43	54	51	41

Quanto aos respondentes com grande incidência de *missings*, destacam-se os das escolas 42 e 43 (Tabela 18).

Tabela 18: *Missing* por teste.

<i>Missing</i>	6	7	21	5	7	9	13	10	11	5	5	5	7	12	10	11	10	12	10
Posições	266	430	477	580	998	1051	1052	1057	1058	1059	1062	1065	1070	1092	1093	1130	1131	1132	1133
Escolas	40	46	54	101	36	42	42	42	42	42	42	42	42	42	42	42	42	42	42
<i>Missing</i>	6	6	8	8	10	9	5	9	5	5	7	10	9	9	9	10	7	15	5
Posições	1134	1138	1139	1140	1143	1145	1161	1254	1580	1623	2000	2073	2274	2279	2335	2347	2371	2401	2171
Escolas	42	42	42	42	42	42	42	38	100	100	100	7	43	43	43	43	43	43	18

Os mesmos procedimentos foram realizados para a amostra brasileira porém, muitos desses possíveis vieses, apesar de identificados, não foram tratados, sendo assim admitido como perspectiva real de caracterização dos respondentes. Sabe-se que alguns psicometristas optam por eliminar testes com elevado risco de enviesamento. A justificativa para isso é que esse procedimento torna a base de dados mais fidedigna, elevando geralmente os índices de consistência do teste, entretanto, também corre-se o risco de exclusão mecânica e inapropriada de respondentes legítimos da população, e assim o viés permanece presente nas análises.

3.1.3 Introdução à Avaliação de Conceitos Térmicos: Resultados Preliminares

Em português, o teste TCE se refere a “Introdução à Avaliação de Conceitos Térmicos”, e é um instrumento classificado como um *Inventário de Conceitos* sendo elaborado originalmente por dois pesquisadores educacionais australianos em 2001 (Yeo & Zadnik, 2001), na qual foi aplicado inicialmente a estudantes secundaristas e iniciantes universitários australianos. O teste possui 26 itens de múltipla escolha, onde as respostas erradas (concepções alternativas) estão identificadas pelos itens dentro de um inventário dividido em 4 grupos de concepções alternativas relacionados a conceitos introdutórios da Física Térmica que são: *calor (A); temperatura (B); transferência de calor e mudança de temperatura (C); e das propriedades térmicas dos materiais (D)*.

Foi decidido pelo teste TCE, pelo fato da linguagem dos itens simular conversas

informais dentro do cotidiano do aluno, e assim tornar os alunos respondentes familiares aos problemas, estando a maioria das alternativas associadas a concepções alternativas, ampliando o poder de análise.

Os resultados preliminares do teste-piloto irão apresentar informações qualitativas e quantitativas minuciosas do instrumento, na qual será expandido e utilizado posteriormente para um contexto mais amplo, se tornando referência de modo geral para todas as análises no estudo. Nessa etapa, foram estabelecidas as seguintes análises:

- Verificação de consistência da chave-de-correção do instrumento por 5 avaliadores e especialistas no EF;
- Análise da versão traduzida do teste TCE para o português, comparando duas outras traduções anteriormente realizadas e versão original. Nesse processo o foco encontra-se na determinação de um inventário de concepções térmicas do teste minimamente fiável para as análises dos resultados em português. Esse inventário ou quadro de tipicidade foi sugerido pelos autores do teste, e ampliada as informações no atual estudo;
- Análise de supostas correspondência de itens no teste, sinalizando eventualmente prováveis informações sobre a dependência dos itens, podendo ser úteis em um estudo complementar e posterior. A intenção está em se apropriar de forma mais precisa do conteúdo do teste e de possíveis causas de vieses interpretativos;
- Análise da coerência interna dos subconstrutos previamente determinados. Esse tipo de verificação possibilita identificar se o teste distribui as informações e se encontra relativamente homogêneo ou se “pesa” mais consideravelmente para um dos subconstrutos inerentes;
- Elaboração da matriz de habilidades como referência sobre as informações de proficiência dos sujeitos e grupos normativos considerados, bem como das competências térmicas básicas na qual o conteúdo dos itens sugerem uma associação dentro de famílias de situações-problema;

- Identificação de conteúdo-padrões dentro da Física Térmica dentro de cada contexto, segundo os documentos oficiais;
- Os resultados do pré-teste e pós-testes relativos a ganhos percentuais de acerto, tamanho do efeito, fator de concentração de análise e desvio na concentração das análises dos itens.

A importância dessas análises está no fato de que o conjunto de procedimentos, bem como dos resultados, acabam por se tornarem parâmetros de referência para as análises posteriores, possibilitando assim um ganho na sistematicidade, projeção e precisão das informações obtidas.

3.1.3.1 Chave-de-Correção do TCE

Tendo sido necessário a determinação da chave do TCE, foi realizado uma análise com 5 avaliadores (4 doutores, e todos formados em Física, com experiência em EF) visando garantir maior *fiabilidade* na interpretação e análise dos escores, assim como numa revisão final da chave devido à observação e discrepâncias dos resultados estatísticos.

Cada avaliador realizou a prova separadamente, sendo livre para optar por uma ou mais alternativas que considerassem pertinente (quando isso ocorreu, o percentual da escolha foi distribuído igualmente entre as alternativas), acompanhado de críticas. Foi considerado um percentual de acordo com o total de opiniões válidas diante da suposta resposta correta.

O resultado mostra a predominância na opinião de uma alternativa para cada item, mas também mostram as forças atrativas de distratores, possivelmente relacionadas a uma má formulação no enunciado do item. Vale ressaltar que um teste de qualidade apresenta bons distratores, se baseando em erros comuns dos alunos (Haladyna & Rodriguez, 2004). Na análise dos resultados esses distratores podem ser classificados em três tipos: *distrator discriminatório* – o grupo de baixo desempenho o escolhe e os de alta performance o evita; *distrator que não é discriminatório*; e o *distrator que raramente é escolhido*, o que indica a implausibilidade (Ibid., 2004).

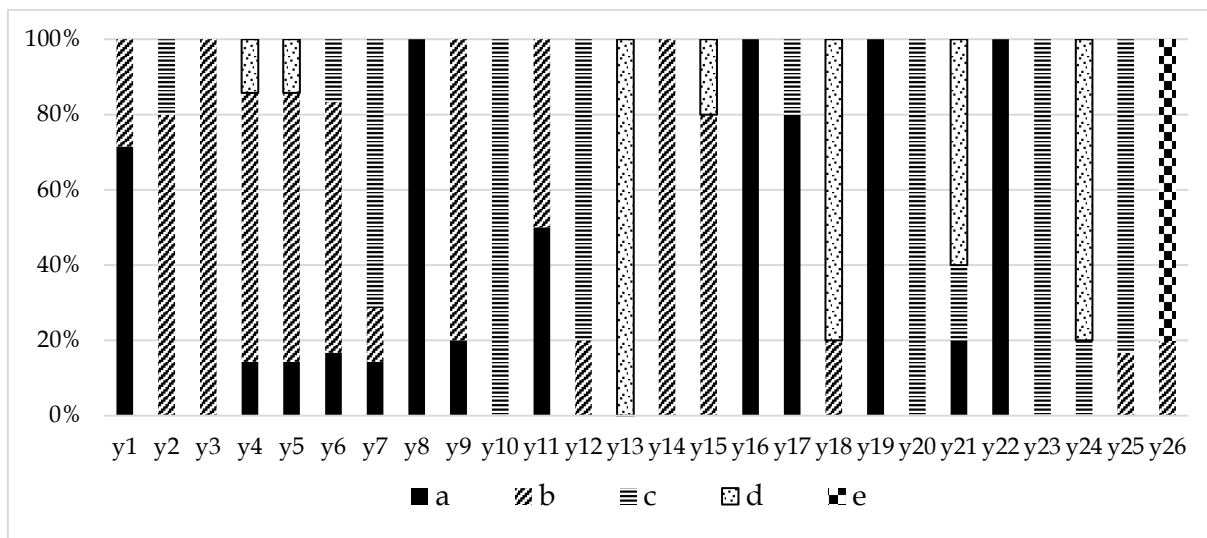


Figura 38: Percentual na convicção da chave-de-correção entre 5 avaliadores.

Considerando que os itens y9, y11, y16, y18 e y26 são os únicos com cinco alternativas, pode-se perceber a força de alguns distratores para cada item. Os outros distratores que não aparecem podem ser considerados alternativas menos plausíveis de serem escolhidas, ao menos, é o que se espera.

Nessa análise prévia pode-se constatar uma consistência de 100% de convicção nos itens y3, y8, y10, y13, y14, y16, y19, y20, y22 e y23. Considerando que uma consistência entre poucos avaliadores, recomenda-se que seja confiável um valor maior ou igual a 80%.

Quadro 57: Itens correspondentes a níveis de consistência no julgamento da chave-de-correção.

Critério de Consistência no Julgamento da Chave-de-Correção	Itens e as alternativas
Alta Consistência: 100% de concordância em uma alternativa entre os avaliadores.	y3b, y8a, y10c, y13d, y14b, y16a, y19a, y20c, y22a, y23c
Média Consistência: maior ou igual a 80% de concordância em uma alternativa, considerando no máximo uma segunda alternativa plausível.	y2b, y9b, y12c, y15b, y17a, y18d, y24d, y25c, y26e
Baixa Consistência: menor que 80% de concordância em uma alternativa, considerando duas ou mais alternativas plausíveis.	y1a, y4b, y5b, y6b, y7c
Péssima Consistência: Menor ou igual a 60% de concordância em uma alternativa, considerando duas ou mais alternativas plausíveis.	y11b*, y21d

Delimitando uma consistência mediana em que as opiniões se dividam entre duas alternativas, em que a diferença seja menor ou igual a 20%, tem-se os itens: y2,

y9, y12, y15, y17, y18, y24, y25 e y 26. Para os de baixa consistência aqueles com duas ou mais alternativas com uma convicção abaixo de 80%: y1, y4, y5, y6, y7. E os de péssima consistência com valor menor ou igual a 60%: y11, y21 (Quadro 56).

A possibilidade de mais de uma resposta, por vezes é causada pela má formulação do item, geralmente relacionada por questões de ambiguidade ou incompletude do enunciado. O item y11, por exemplo, reflete uma baixa qualidade do item por remeter a algumas ideias paralelas.

11. Patrícia pergunta aos amigos: “Se eu colocar 100 gramas de gelo a 0 °C e 100 gramas de água a 0 °C dentro da geladeira, qual dos dois perderá a maior quantidade de calor?”

- a) Cristina diz: “Os 100 gramas de gelo”.
- b) Bernardo diz: “Os 100 gramas de água”.
- c) Marcelo diz: “Nenhum dos dois, porque ambos contêm a mesma quantidade de calor”.
- d) Mateus diz: “Não há resposta, porque o gelo não contém nenhum calor”.
- e) Júlia diz: “Não há resposta, porque não podemos ter água a 0 °C”.

Algumas considerações realizadas pelos avaliadores podem servir de argumentações para aprimoramento dos itens. No caso do item y11, tem-se:

- Avaliador 4, comenta: “O item deve dizer que a temperatura no interior da geladeira é inferior a zero!”.
- Avaliador 1: “Obviamente que somente irão ‘perder’ calor se o ambiente interno estiver abaixo de 0 °C, caso contrário, nenhum dos dois irá ‘perder’ calor, sendo uma informação ausente no item”.

Sabe-se que para uma mesma variação de temperatura entre ambos até o equilíbrio térmico no interior da geladeira, irá perder mais quantidade de calor aquele que estiver em um estado térmico “mais elevado” em uma escala de aquecimento, no caso a água a 0 °C, porém o aluno pode não perceber isso, e associar a uma série de equívocos. Por exemplo, deve-se considerar que para alunos que já possuíram instruções sobre a Física Térmica, poderão associar a maior ou menor perda de energia térmica com o calor específico, considerando que a água necessita em média de uma 1 caloria para variar 1 °C de 1 grama, enquanto o gelo precisa perder em média a metade da quantidade de calor para fazer variar de temperatura a mesma quantidade de massa, portanto, a água irá necessitar perder mais calor. Seguir nesse pensamento faz com que o aluno responda corretamente, porém não evita que ele não tenha cometido

equivocos para chegar nessa conclusão, isso significa dizer que os distratores não devam ser considerados os únicos possuidores na representação de possíveis equivocos que levam os respondentes a assinalarem as respostas, incluindo a resposta correta.

Explicando e justificando a alternativa correta, mesmo com incompletude dita, pode-se dizer que: *Apesar de se considerar importante informar a temperatura interna da geladeira estando abaixo de 0 °C, a intenção seria estimar a temperatura média no interior estivesse abaixo, pois as questões tendem a ser todas intuitivas, o problema é que sendo muito intuitiva pode caracterizar problemas e possibilidades de vieses interpretativos. Apesar disso, quando se diz 'qual dos dois perderá mais calor?', supostamente pode ser subentendido que a temperatura interna (θ_{gelad}) da geladeira é inferior a 0 °C, e ambos os corpos estarão sujeitos a mesma condição térmica, ou seja, estão com a mesma diferença de temperatura ($\Delta\theta = \theta_{gelad} - 0$) quanto ao ponto de equilíbrio térmico. Apesar disso, como estão em estados físicos diferentes e correspondem a uma mesma substância (e mesma quantidade de matéria, 100 g), perderia mais calor aquele que estivesse com uma maior "distância térmica" quanto ao ponto de equilíbrio. Como sabe-se que os dois irão esfriar até atingir o equilíbrio, e que o gelo possui uma menor "distância térmica" quanto a esse ponto de equilíbrio, pode-se afirmar qualitativamente que a água necessitará perder maior quantidade de calor até atingir o equilíbrio sem necessitar de cálculo, apenas por bom senso e um pouco de lógica. Alunos instruídos podem utilizar habilidades cognitivas associadas a recursos quantitativos, no caso, consideraria que a quantidade a mais de calor que se perde equivale aquela necessária para esfriar a água até estar na mesma "distância térmica" dos 100 g de gelo a 0 °C. Para realizar o cálculo que colocaria os dois corpos em um mesmo estado térmico, a água precisaria perder a seguinte quantidade de calor latente, que seria: $Q = m \cdot L_s = 100g \cdot \left(-\frac{80cal}{g}\right) = -8000 cal$, ou seja, as 100 gramas de água a 0 °C esta a -8000 cal mais "distante" que os 100 gramas de gelo a 0 °C, e com isso justifica a letra b.*

De modo geral, considerando em média o consenso entre todas as alternativas possíveis para o teste entre os avaliadores, estando supostamente algumas mais ambíguas que outras, **o nível de confiança ficou em 83,5%** (conforme mostra o tracejado no gráfico). Esse valor está acima dos 80% que são geralmente recomendados (Amado, 2013).

Ao comparar com um estudo recente aplicando o TCE no contexto brasileiro (Louzada & Elia, 2015), constatou-se uma diferença em 2 itens na chave-de-correção (itens y20 e y21). Posteriormente foi obtida a chave pela AAPT que confirmou esse equívoco. Em um outro trabalho (Gonçalves Júnior, 2012) foram apresentadas somente as chave-de-correção dos itens y1, y2, y4 e y6, em que se conferem com o atual estudo.

A análise da chave de correção realizada demonstrou ter sido eficiente a confirmar 100% com a fornecida pela AAPT, além de apontar alguns aspectos para a melhoria do instrumento, e servir de parâmetro moderador nas análises e interpretações sobre os desempenhos dos estudantes. É bom destacar que as recomendações para boas práticas no uso de testes conceituais ou inventários de conceitos é não disponibilizar a chave *online*, pois o acesso pelos alunos pode colocar em risco a validade do instrumento, em um pós-teste, por exemplo. Esse é um problema que pode ser contornável, quando se administra com maior diligência, sem informar previamente aos alunos sobre o teste, além de não permitir consulta via internet. Também se recomenda para que ninguém fique com uma cópia extra do teste.

3.1.3.2 Tradução do Teste TCE

O teste TCE foi *traduzido*⁸⁸ do inglês australiano por *dois juízes nativos*⁸⁹ para cada contexto (português de Portugal, e português do Brasil), desses, 3 são doutores com fluência em inglês e envolvidos dentro de PPGs na área da temática (que envolvem o EF), e 3 deles com graduação e especialistas na temática estando familiarizados esse tipo de instrumento e escala de medida (teste/questionário dicotômico). Diante das análises e discussões, expressões foram modificadas e readequadas a linguagem cotidiana de cada contexto. No caso, para garantir maior

⁸⁸ Testes traduzidos não são testes adaptados, segundo Hambleton (2005) citado por Bosna et al (2015): "os termos adaptação e tradução são distintos, e tem-se preferido o uso do primeiro, uma vez que compreende todos os processos concernentes à adequação cultural do instrumento, para além da mera tradução. A adaptação é um processo árduo e trabalhoso, que envolve vários indivíduos, nível na qual este estudo ainda não alcançou. Nesta fase inicial é preciso tomar cuidado com o clássico erro conhecido como "Traduttore, Traditore" (Tradutor, traidor).

⁸⁹ Doutores portugueses envolvidos na área da pesquisa no EC.

fidedignidade, os problemas linguísticos encontrados na tradução das rubricas das concepções térmicas do artigo original, necessitaram de readaptação para que não violasse o sentido físico.

Em seguida foi realizado um estudo piloto da amostra normativa utilizando a técnica pré-teste/pós-teste, para assim verificar o comportamento do item quanto ao índice de dificuldade, bem como uma análise preliminar de desempenho no que se refere aos ganhos normatizados. O resultado da tradução do teste TCE dentro de cada contexto pode ser comparado duas outras traduções brasileiras (Gonçalves Júnior, 2012; Louzada, 2012, 2015). Também foi encontrado na literatura um versão mexicana em espanhol (Gómez & Hernández, 2010).

3.1.3.2.1 Inventário de concepções térmicas

O inventário se refere a uma correspondência entre as concepções exploradas no teste e os itens. Como as concepções térmicas são considerados equívocos, pode-se dizer que seja um inventário voltado para os distratores, ou seja, ele é um recurso organizador explicativo para os distratores, informando quais são as mais prováveis concepções alternativas relacionadas com as opções atrativas para os grupos de respondentes.

Considera-se que o Inventário de Conceitos seja a relação do item com os conceitos ou conhecimentos que se pretende avaliar. Isso se dá através das respostas corretas e não pelos equívocos cometidos. Entende-se que a tipicidade dos itens está relacionada pelos conceitos corretos, ou seja, aquilo que se pretende medir ou avaliar, na qual se refere ao traço latente do item estando representado em grande parte pela alternativa correta. Admite-se, entretanto, que uma atração aos distratores revela que os traços latentes possuem uma natureza mais relacionada às concepções alternativas, ou seja, o conhecimento prévio disponível é formado com ideias, crenças e entendimentos que não correspondem de forma consistente ao conhecimento científico, havendo relativamente um distanciamento conceitual com o senso comum. Em grande parte, esses atrativos também podem estar relacionados com a natureza do item, e não no sujeito, dessa forma, existem diversas variáveis externas e internas que

podem estar influenciando os sujeitos, com isso, considera-se relevante realizar algumas dessas especulações para ponderar as conclusões diante dos escores.

Primeiramente se tem uma análise do léxico verificando pontos críticos de ambiguidade na tradução, para serem confrontados com outras traduções, discutidos e tomado um juízo de valor. O uso da Hermenêutica na interpretação das rubricas dentro do quadro proposto pelos autores originais possibilitará uma maior fiabilidade na verificação subsequente quanto à tipicidade dos itens. Torna-se possível uma comparação com duas traduções do TCE realizadas anteriormente no contexto brasileiro (Gonçalves Júnior, 2012; Louzada, 2012b, 2015).

Entre alguns pontos mais críticos identificados de imediato, se tem na categoria C, denominada pelos autores como *“Students' conceptions about heat transfer and temperature change”*, que em uma tradução direta seria *“Concepções dos alunos (ou estudantes) sobre a transferência de calor e mudança de temperatura”*. Diante disso, levanta-se a seguinte questão: *Qual a diferença entre concepções sobre calor e sobre transferência de calor?* Parece redundante, um erro tautológico pelo fato do conceito de calor já estar relacionado a uma ideia de processo de transferência, no caso, transferência de energia térmica (e não de calor!), dizer que se perdeu calor remete a ideia que antes se tinha, ou seja, que calor é uma propriedade de um corpo, e isso seria um equívoco. Entende-se que calor esteja relacionado a uma energia em transição, em movimento, havendo em cada extremidade um fluxo de energia térmica diferenciado, considerado que durante o processo termodinâmico sempre há perdas, ou seja, dissipação de energia térmica, pois não existe um sistema real capaz de evitar esse tipo de perda, pois a ideia de *“sistema termicamente isolado”* apenas uma abstração física, que não pode se converter a uma pseudoverdade pelo excesso de repetição e didatização.

Por outro lado, entende-se que a *Termodinâmica* sejam processos, todos os fenômenos de uma forma ou outra são processos, ou seja, nada está parado, nem mesmo um corpo com suposta *“temperatura constante”* (ignorando as flutuações existentes), pois suas partículas estão em agitação térmica, e os processos de interação ocorrem continuamente e configuram seu estado térmico. A única ideia de entender que não seja um processo, seria admitir a possibilidade de que as partículas

constituintes da matéria possam estar estáticas, fato pouco provável de acontecer. Obviamente que se entende a intenção, que seria, no caso, explorar concepções diante dos fenômenos térmicos, na qual um corpo é aquecido ou resfriado. Nisso inclui a ocorrência de uma mudança de fase (calor latente) ou não (calor sensível), em que um corpo fornece energia térmica (relacionada ao movimento/agitação das partículas) e outro a recebe, podendo explorar contextos transitórios em que o fluxo de energia alcance o *equilíbrio térmico* ou não (o fenômeno analisado é um processo de transição de energia térmica).

Diante disso, como sugestão, arrisca-se propor que seria mais adequado o termo “*concepções dos alunos relacionados aos processos termodinâmicos*”, próximo a ideia de “transmitir”, buscando entender que em uma transição pode ocorrer mudança de temperatura ou de fase, sobre tudo de que há uma fonte térmica e um receptor, não focando exatamente em concepções exclusivamente ligadas a uma compreensão diante dos conceitos de calor e temperatura, e nem sobre causas relacionadas as suas propriedades térmicas, mas sim, sobre os conhecimentos que envolvem, sustentam e explicam os fenômenos térmicos de modo geral, dos processos termodinâmicos, propriamente ditos.

Quanto às rubricas, vamos comparar as traduções do trabalho de Louzada (2012b, 2015) sendo (1), e o de Gonçalves Júnior (2012) sendo (2), e o atual estudo sendo (3). Antes disso, cabe ressaltar algumas considerações preliminares diante desses trabalhos:

- Como foi dito, o processo de tradução, não pode ser compreendido como adaptação, porém esse é um equívoco recorrente defendido na “tradução” (1) (e não adaptação!): “[...] *teve que ser devidamente adaptado para a língua portuguesa pelos autores do presente trabalho*” (Louzada, 2015, p. 1508); “*Versão adaptada do instrumento utilizado na pesquisa*” (Ibid., 2015, p. 1508-8) . O autor (2) afirma que “*a simples tradução das questões para uma língua diferente não garante uma equivalência entre esses testes (o original e sua tradução)*” (Gonçalves Júnior, 2012, p. 34), e em outro momento: “*No caso do questionário proposto por Yeo e Zadnik [Yeo e Zadnik 2001], adaptado para aplicação em português*”. Compreende-se que uma mudança no nome de personagens pode ser parte

de uma “adaptação” contextual, mas não transcultural do teste, portanto, sugere-se evitar esse termo;

- A tradução (1) ocorreu por meio de três professores, dois professores de Física e o pesquisador, referente ao inventário e o próprio “questionário”; e a tradução (2) não informa, subentende-se que tenha sido realizado exclusivamente pelo autor;
- O termo “questionário” de conceitos é citado nos dois trabalhos, porém, entende-se que haja uma diferença com o uso do termo “teste”, no caso, os questionários avaliam dimensões afetivas do comportamento ou não cognitivas (McMillan & Schumacher, 1989) ao contrário de um teste conceitual;
- Os autor(es) (1) afirmam que o teste está organizado, considerando que cada item pode estar contido em mais de uma “dimensão”, sendo igual a: “A (14 questões), B (17 questões), C (21 questões) e D (22 questões)” (Louzada et al., 2015). Entretanto a organização correta por dimensão seria: A (14 itens⁹⁰); B (17 itens); C (**23 itens**); D (22 itens), provavelmente tenha cometido esse equívoco ao ter excluído os itens 17 e 18 (porém sem deixar claro o motivo delas serem confusas o suficiente para serem eliminadas!), na qual cada um incide uma única vez na “dimensão” C. A organização por item seria: A (9 itens); B (15 itens); C (16 itens); D (17 itens). O autor (2) não apresenta informações diante esses aspectos;
- Os autor(es) (1) consideram que A, B, C e D sejam dimensões *a priori*, enquanto o autor (2) não faz referência a isso. Entende-se que deva ser estabelecido uma análise criteriosa de pertinência para afirmar e justificar que esses construtos sejam considerados dimensões. Sendo menos rigoroso, entende-se a atribuição do termo se justificativa pela relação dos tipos de concepções pertinentes a uma mesma categoria. Na estatística, quando se fala em dimensões preestabelecidas, é comum serem utilizadas *análises fatoriais confirmatórias*, assim como o uso de cargas fatoriais para determinar dimensões, ou sendo definidas *a posteriori* através de *análises fatoriais exploratórias* com o uso do *Scree plot*, por exemplo. Os elaboradores do TCE em outro artigo, realizaram esses tipos de análise baseadas em estatística inferencial, realizando

⁹⁰ O uso do termo “item”, por vezes se torna confuso, sendo geralmente esta associado a alternativa correta de um teste, porém, quando se retrata de analisar e interpreta-lo na raiz entende-se como um todo (enunciado e alternativas). Foi adotado que para teste será denominado “item” e para questionários o termo “questões”.

ainda correspondência entre as concepções com as alternativas de alguns itens (Chu et al., 2012), reelaborando assim novas subcategorias para o teste, e no caso, mais adequada e pertinente para interpretar os modelos mentais da respectiva amostra;

- O autor (1) afirma que a qualidade do instrumento, quanto a sua *validade de conteúdo* está definida pelo ganho de valores médios e pelo índice de discriminação (referente a correlação ponto-bisserial). Isso caracteriza de certa forma um equívoco, pois a validade de conteúdo se refere à necessidade de o conteúdo abranger suficientemente àquele que corresponde o construto, na qual ele denomina de “dimensão”. A correlação, no caso, é de fato um índice de discriminação, porém se refere a consistência do item diante do teste, ou seja, relacionado à precisão e fiabilidade, e não validade. Para a validade do conteúdo, um dos aspectos seria a necessidade de verificar se as concepções sobre cada domínio cobrem suficientemente cada categoria predeterminada. Vale ressaltar que o teste converge para analisar as respostas corretas (não deveria corresponder a nenhuma concepção alternativa, em tese), e as “dimensões” do inventário se referem aos distratores (remetem a concepções alternativas no inventário), pois as concepções térmicas são equívocos, ou seja, são coisas distintas que precisam ser diferenciadas para evitar confusão. O autor (2), comete esse equívoco, ao denominar as concepções do inventário (“relação conceito/item”) como “conceitos”.

Sabe-se que em uma taxonomia, os conhecimentos podem estar em níveis diferentes de complexidade em suas categorias, ou seja, relacionada a fatos, conceitos, relações, leis e princípios, ou ainda em uma perspectiva sistemática. Se diz que “*a análise é entendida como um modo de definição, e o que se define na análise é uma proposição ou conceito, e não uma expressão ou sinal linguístico*” (Rosa, 2009), e isso pode-se tornar um pouco confuso quando se trata de conhecimentos científicos, levando a pensar que tudo pode ser “definido”, e com isso, “conceituado”, no entanto, nem toda *definição* se trata de um *conceito* propriamente dito. Dentro disso, considera-se ainda que a concepção de um sujeito possa representar suas ideias e percepções diante de um conceito ou um saber, independente de seu nível de complexidade, porém, uma concepção formalmente constituída pode ser caracterizada como um conjunto de conceitos que dá forma ao conhecimento propriamente dito, no caso mais complexo,

abstrato e estruturado. Portanto, quando se falar em fatos, conceito, proposições, leis e princípios, entende-se dentro da taxonomia de objetivos educacionais como objetos do saber em níveis diferenciados de complexidade.

Vejamos alguns problemas encontrados dentro do grupo A, relacionada as *concepções dos alunos sobre o calor*:

- “A3 – *Heat and cold are different, rather than opposite ends of a continuum*”: (1) traduz como “**calor** e frio são diferentes” e (2) como “**calor** e frio são diferentes, em vez de extremos opostos de um contínuo”. Isso parece confuso por essas traduções não representarem uma concepção alternativa, pois de fato **calor e frio são conceitos cientificamente diferentes** em sua natureza e representação física, o calor é energia e frio sensação térmica, portanto as assertivas traduzidas por (1) e (2) não se referem em concepções contrárias às científicas como deveria ser. Entende-se que os autores originais quando se referem a expressão “[...] ser opostos de um mesmo *continuum*”, não podem estar se referindo ao calor em si, mas das sensações térmicas, pois o **oposto de «frio»** dentro de um *continuum* de diferencial semântico relacionado a sensação térmica **seria «quente»** (Osgood, Suci, & Tannenbaum, 1957, p. 34). Pelo fato desses termos reincidirem em outras obras, inclusive a realizada novamente pelos mesmos autores (Chu et al., 2012), é um caso muito provável de diferença na identidade cultural do item por questões de léxico, ou pelo o equívoco de tentar realizar uma tradução direta (um exemplo comum de equivalência seria “hot-dog” e “cachorro-quente”). Muitos livros textos passam a ideia de que um corpo pode receber ou perder calor, nesse sentido, quem recebe calor sente a sensação de quente, e quem perde sente frio, é possível que se tenha intenção de se referir ao frio como “ausência do calor”, porém, entende-se que não sejam opostos de um mesmo *continuum* por fazerem parte de dimensões epistemológicas distintas. Sugestão: “**O quente e frio são diferentes, em vez de extremidades opostas de um *continuum***”.

Com relação ao grupo B, *concepções de alunos relacionados a temperatura*:

- “B3 – *Perceptions of hot and cold are unrelated to energy transfer*”: (1) traduziu como: “Percepções de **calor e frio** não estão relacionadas com transferência de energia”, e (2) como “Percepções de **calor e temperatura** não estão relacionadas com

transferência de energia”. Obviamente que a palavra “percepções” está relacionada a sensação térmica, e entende-se que o sentido da palavra “hot” deva ser “quente” e não “calor”, e “cold” seja frio e não “temperatura”. Considerando que o calor seja uma energia em trânsito, estando no entremeio, se propagando, não estando nem na fonte e nem receptor, apesar de quantitativamente se admita que seja equivalente a energia térmica que se cede, se dissipe (no caso da propagação em um meio), e se recebe. Portanto ela não poderia ser percebida pelo tato, aquilo que se percebe são sensações de energia térmica que se recebe ou que se cede em um processo de “transferência de energia”. Sugestão: **“Percepções de quente e frio não estão relacionados com a transferência de energia”**.

- *“B6 – A cold body contains no heat”*: (1) e (2) traduziram como: “Um corpo frio não contém calor”. Ora, sabendo que o calor não é uma propriedade do corpo, nem corpo frio, nem quente e nem morno deve “conter calor”, ou seja, a assertiva é verdadeira e não configura uma concepção alternativa, nem na sua origem. Porém, entende-se que a concepção envolvida está relacionada com a ideia de quem um corpo com baixa temperatura (que se percebe ele “frio”) esteja ausente da capacidade de ceder energia térmica, mais precisamente, esteja ausente de energia térmica, ou seja, não se compreende que possa existir um outro corpo com temperatura inferior na qual o corpo em questão seja capaz de aquecê-lo, ainda que para quem toca isso aparentemente não seja possível. Como necessidade de ser reelaborada, se propõe: **“Um corpo frio não possui a capacidade de aquecer um outro corpo (somente corpos quentes ‘possuem calor’ e podem aquecer, ou seja, se admite uma concepção ontológica substancialista para o calor)”**.

No grupo C, podem ser destacadas duas concepções quanto a aspectos de conflito em interpretações dos léxicos envolvidos:

- *“C2 – Heat only travels upward”*: (1) traduziu como “calor se movimenta apenas para cima” e (2) como “o calor só se move para cima”. Poderia se considerar que a afirmação está relacionada a todas as formas de propagação do calor, e assim estar correto, porém, não se acredita nisso. Entende-se que o sentido dessa assertiva esteja relacionado a propagação do calor dentro do processo de condução térmica, ou seja, o calor se propaga em um corpo sempre no sentido para cima. A interpretação dessa

concepção pode se confundir com outras formas de propagação do calor, diferenças inclusive de natureza epistemológica, como por exemplo, por radiação térmica, que estaria mais no campo da Mecânica Quântica do que da Termodinâmica propriamente dita (Mattos et al., 2004). Esta é conhecida como “onda de calor” representado por micro-ondas infravermelhas que se propagam independentemente de um meio e em todas as direções, ou seja, não exclusivamente para cima, não parecendo ser essa a intenção do equívoco. Sugestão: **“O calor propaga-se em meios materiais em sentido unicamente para cima”**.

- *“C6 – Objects of different temperature that are in contact with each other, or in contact with air at different temperature, do not necessarily move toward the same temperature. (Thermal equilibrium is not a concept.)”*. (1) e (2) traduziram de forma semelhante, afirmando que os corpos em contato não atingirão a mesma temperatura, e que equilíbrio térmico não é um conceito ou que não possuem esse conceito. Sugere-se que os: **“Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido)”**.

Destacando uma concepção do grupo D, se tem:

- *“D4 – Different materials hold the same amount of heat”*: (1) traduz como “Diferentes materiais retêm a mesma quantidade de calor”. Parece incompleto, pois de fato, diferentes materiais podem “receber” (ou “ceder”) a mesma quantidade de “energia térmica”, dependendo das suas massas. Esse equívoco se refere entender que os todos os corpos possuem a mesma capacidade térmica, e sabe-se que a capacidade térmica é expressa por: $C = m \cdot c$, onde m é a massa e c o seu calor específico. Dessa forma, por mais que os corpos sejam diferentes (possuam diferentes valores para o calor específico), eles podem receber ou ceder a mesma quantidade de calor se houver um ajuste da massa. Obviamente que os alunos não tenham essa compreensão, e que o equívoco seja uma crença generalizada, porém, acredita-se que a concepção precisa ser reelaborada pelo fato dela possibilitar uma situação em que de fato seja possível e coerente cientificamente. A ideia de manter, armazenar ou reter calor, também remete a uma concepção substancialista para o calor, que possa ser uma propriedade térmica.

Diante disso, sugere-se que: “**Diferentes materiais, independentemente de suas massas, sempre irão receber a mesma quantidade de energia térmica (acredita-se que diferentes materiais possuam a mesma capacidade térmica, e o conceito de calor assume uma concepção substancialista)**”.

3.1.3.2.2 Outras possíveis correspondências entre as concepções térmicas

Verificando uma relação entre as concepções no quadro, pode-se constatar aspectos típicos comuns entre algumas delas, sugerindo que muitas dessas concepções não devam ser encaradas como mutuamente exclusivas, ou seja, podem ocorrer ao mesmo tempo, relacionando-se e associando-se entre diversas concepções ontológicas imbricadas.

Quadro 58: Aproximações entre algumas concepções alternativas.

A3, B2 e B3	Concepções [sobre calor e sobre temperatura] que se referem à sensação térmica
B5; D5	100 °C é o ponto de ebulição e a temperatura máxima possível; 100 °C é o único ponto de ebulição
A4; A5; B2	Calor [conceito] = temperatura [conceito]; Calor [é proporcional à] Temperatura; Temperatura (intensidade) = Calor (intensidade)
A1, B6, C4, D2 e D4	Concepções substancialistas sobre o calor
C5	Concepções substancialistas sobre a temperatura
A2	Se calor não é energia, então é um [fluido?] – concepção substancialista?!
C3, C7 e D9	Concepção animista pelo calor ter vontade de subir [aristotélica]; pelos objetos frios tenderem a aquecer por vontade própria (e vice-versa); pela lã ser uma fonte térmica como um mamífero
C8	Concepção sobre o processo termodinâmico [o movimento é lento] está associado a concepção sobre uma propriedade térmica do material [em condutores].
D6, C6	“O gelo encontra-se sempre a 0°C e/ou não é possível alterar a temperatura” D6 , se aproxima de C6, que afirma que os corpos com diferentes temperaturas não tendem ao equilíbrio térmico (os dois rejeitam a lei geral das trocas de calor; e rejeitam a <i>Lei Zero da Termodinâmica</i>). O D6 pode estar associado com as concepções sobre processos termodinâmicos. Rejeitar a <i>Lei Zero da Termodinâmica</i> implica rejeitar o princípio geral das trocas de calor.

Entre algumas prováveis concepções ocultas (não declaradas no documento original), e que podem estar sendo sugeridas no teste, como por exemplo:

- O distrator do item y7 letra “a” poderia estar associado à concepção: “*Um corpo ‘frio’ tem mais influência que um corpo “quente” quando postos em contato, ou seja, sua temperatura tende a se equilibrar com a temperatura do corpo ‘frio’*”;

- O distrator do item y7 letra “b” poderia estar associado à concepção: *“A temperatura final de equilíbrio é igual à média aritmética das temperaturas dos corpos”*;
- O distrator do item y4 letra “b” poderia estar associado à concepção: *“As CNTP são condições de validade geral para o comportamento da água, em especial. E com isso, substituir de uma forma mais geral a concepção alternativa D5”*;
- Os distratores do item y4 letra “c” e “d” poderia estar associado à concepção: *“Quanto maior é o ponto de ebulição, mais rápida a água ferve”*;
- A seguinte concepção poderia substituir às concepções alternativas D7 e D8: *“A temperatura de um corpo que mudou de fase nunca está na temperatura da mudança de fase, se foi aquecido está em uma temperatura mais elevada, e se esfriou está numa temperatura mais baixa”*.

Existem algumas concepções relacionadas ao funcionamento de artefatos e produto modernos e tecnológicos (Thouin, 2004, p. 353), diante disso, o teste apresenta situações que parecem exigir do alunos um conhecimento básico sobre o funcionamento de alguns eletrodomésticos, pois ainda que o foco da situação-problema esteja em outro objeto, o contexto em que se encontra o objeto é um artefato que pode influenciar na sua escolha. O frigorífico da geladeira no item y1, é um exemplo, em que existem dispositivos termoelétricos que controlam as temperaturas e buscam manter o equilíbrio, e ter uma noção sobre isso pode ser um fator a determinar a escolha da alternativa, ainda que só exista uma única alternativa abaixo de 0°C. Uma sugestão para esse «grupo de concepções», seria: *“Concepções dos alunos sobre o funcionamento de artefatos da produção moderna tecnológica”*; e como concepção poderia ser *“Um frigorífico tende a resfriar continuamente um corpo cada vez mais e mais a baixo de 0 °C sem que haja um limite para isso”*, ou *“Um frigorífico tende a resfriar continuamente um corpo a baixo de 0 °C até alcançar sua temperatura mais baixa possível”*, ou *“Um frigorífico tende a resfriar continuamente um corpo, porém nunca irá conseguir colocar o corpo a baixo de 0 °C”*. Essas dúvidas e equívocos se refere a um desconhecimento quanto ao processo de funcionamento das máquinas caloríficas (ou frigoríficas) e dos dispositivos que influenciam sobre o comportamento térmico de um corpo em suas proximidades e/ou em seu interior, por exemplo, dentro de um

frigorífico existem termostatos que regulam a temperatura inferior. Os equívocos também podem estar relacionados ao desconhecimento das propriedades térmicas dos corpos, ou nas duas situações ao mesmo tempo, ou seja, não são mutuamente exclusivos.

A correspondência sugerida pelos autores do TCE entre as concepções térmicas e os itens foram verificadas pelos dois estudos brasileiros, em que buscaram identificar as alternativas que correspondiam as concepções sugeridas, pois no trabalho original não há esse detalhamento. Percebe-se uma sugestão na identificação das alternativas, porém algumas lacunas deixadas, pois alguns itens e alternativas não foram sugeridas, e outras parecem remeter a dúvidas nessas escolhas. Chama atenção que isso sugere a possibilidade de ambiguidades em interpretar aquilo que de fato o item está medindo, seja com as respostas corretas, mas principalmente com relação aos distratores, na qual remetem explicitamente as concepções térmicas equivocadas e atraídas pelos respondentes. Essas são uma das desvantagens com o uso de inventários, sendo um aspecto que já foi apontado, em que os principais problemas estão em itens que os indivíduos respondem apoiados em seus sentimentos gerais e autoconceitos, e por levarem a tendenciosidade e à fraude por usarem respostas como sim, não, concordo, etc., e não indagam sobre respostas que dizem respeito a situações bem-definidas (Erthal, 1987, p. 50).

Para tentar minimizar esses problemas foi utilizada uma técnica de comparação e interpretação item-a-item, para cada uma das alternativas, semelhante a um *rapport*, porém buscando comparar as ligações de sintonia na interpretação e correspondência dos itens/alternativas entre os estudos brasileiros e as concepções sugeridas pelos autores originais, sem garantias que haja uma concordância. Com isso, foi verificado a sincronização e discordância nas correspondências, sendo justificado e explicado as escolhas (ver no Apêndice A, pp. 1-210, dentro do enunciado de cada item), levando em consideração os distratores entre as 109 alternativas disponíveis no teste. Considera-se que essa verificação faça parte do processo de validade interna do construto que está sendo utilizado para medir o desempenho e avaliar o conhecimento dos respondentes diante do instrumento. O quadro adaptado mostra as sugestões

originais e incorpora as sugestões das duas traduções brasileiras, e por último, as sugestões do atual estudo, na qual serão utilizadas para definir qualitativamente as concepções dos grupos normativos. As correspondências entre os distratores foram realizadas de forma completa, e também foram sugeridas novas concepções que por ventura não foram possíveis de serem encontradas (ou concordadas) dentro do quadro sugerido pelos autores originais. Também foi utilizado nesse processo a chave-de-correção do teste, para se certificar e analisar as possíveis ambiguidades.

Quadro 59: Análise do construto – Reelaboração na tradução, comparação e novas sugestões de correspondências entre as concepções e os itens/alternativas.

(Yeo & Zadnik, 2001)	(Louzada, 2012b, p. 43, 2015)	(Gonçalves Júnior, 2012, pp. 76-77)	Sugestões	(Ibid., 2001)	(Ibid., 2015, p. 1508-3)	(Ibid., 2012, p. 168)	Sugestões	
A: Students' conceptions of heat	A: Concepções dos estudantes sobre o calor	A: Concepções sobre calor	A: Concepções de calor dos alunos	Itens (y)	Itens (y)/alt.	Itens (y)/alt.	Itens (y)/alt.	
A1	Heat is a substance.	É uma substância	Calor é uma substância.	O calor é uma substância (<i>concepção substancialista</i>).	10, 22	10a	10d, 21a, 21b, 22c	10a, 10d, 16c, 16e, 21a, 21c, 21b, 22c
A2	Heat is not energy.	Não é energia	Calor não é energia.	O calor não é energia.	22	<i>Ausente</i>	21b, 22c	22c, 21b
A3	Heat and cold are different, rather than opposite ends of a continuum.	Calor e frio são diferentes	O calor e frio são diferentes, em vez de extremidades opostas de um contínuo.	O quente e frio são diferentes, em vez de extremidades opostas de um <i>continuum</i> .	10, 13, 18, 23, 24	23a, 23b	10a, 13b, 15c (?), 18a, 18b, 23a, 24b, 24c	10a, 13b, 15c, 18a, 18b, 23a, 24b, 24c
A4	Heat and temperature are the same thing.	Calor e temperatura são a mesma coisa	Calor e temperatura são a mesma coisa.	Calor e temperatura são a mesma coisa.	15, 18	15a, 15c, 15d	18c, 18e (?)	15a, 15c, 15d, 22b
A5	Heat is proportional to temperature.	Calor é proporcional à temperatura	Calor é proporcional a temperatura.	O calor é proporcional à temperatura.	7, 11, 15	7d, 15c	7b (?), 11c, 15a, 15c, 15d	7b, 11c, 15a, 15c, 15d
A6	Heat is not a measurable, quantifiable concept.	Calor não é mensurável	Calor não é um conceito mensurável ou quantificável.	O calor não é mensurável, conceito quantificável.	7	7a	7a (?), 7d	7d
B: Students' conceptions of temperature	B: Concepções dos estudantes sobre temperatura	B: Concepções sobre temperatura	B: Concepções de temperatura dos alunos	Itens (y)	Itens (y)/alt.	Itens (y)/alt.	Itens (y)/alt.	
B1	Temperature is the "intensity" of heat.	A temperatura é a 'intensidade' do calor	A temperatura é a "intensidade" de calor (ou medida da quantidade de calor).	A temperatura é a "intensidade" de calor.	15	15a, 15c, 15d	<i>Ausente</i>	15a, 15c, 15d
B2	Skin or touch can determine temperature.	O toque pode determinar a temperatura	A temperatura pode ser determinada pelo toque ou pele	Pele ou toque pode determinar a temperatura.	16	<i>Ausente</i>	<i>Ausente</i>	16b
B3	Perceptions of hot and cold are unrelated to energy transfer.	Percepções de calor e frio não estão relacionadas com transferência de energia	Percepções de calor e temperatura não estão relacionadas com a transferência de energia.	Percepções de quente e frio não estão relacionadas com a transferência de energia.	10, 18, 21, 22	21a, 21b, 22b, 22c, 22d	10b, 18a, 18b, 22c	10d, 18a, 18b, 21a, 21b, 22b, 22d
B4	When temperature at boiling remains constant, something is "wrong."	Quando a temperatura de ebulição permanece constante, alguma coisa está 'errada'	Quando a temperatura de ebulição permanece constante, há algo errado.	Quando a temperatura em ebulição permanece constante, algo está "errado".	5	5d	5d	5c, 5d
B5	Boiling point is the maximum temperature a	O ponto de ebulição é a máxima temperatura	A temperatura de ebulição é a maior	Ponto de ebulição é a temperatura máxima que uma	19	<i>Ausente</i>	8d, 19b (?), 19d	19b, 19c, 19d

	substance can reach.	que uma substância pode atingir	temperatura que um corpo pode atingir.	substância pode alcançar.				
B6	A cold body contains no heat.	Um corpo frio não contém calor	Um corpo frio não contém calor.	Um corpo frio não possui a capacidade de aquecer um outro corpo (<i>somente corpos quentes "possuem calor" e podem aquecer, ou seja, se admite uma concepção ontológica substancialista</i>)	7, 10, 11, 22, 26	7a, 7b, 10b, 11d, 22b, 22c, 22d, 26c	7d (?), 10b, 11d, 26c	11d, 22d, 26c
B7	The temperature of an object depends on its size.	A temperatura de um objeto depende do seu tamanho	A temperatura de um objeto depende do seu tamanho.	A temperatura de um objeto depende de seu tamanho.	1, 9, 14	1d, 9e, 14c	1d, 9e, 14c	1d, 9e, 14c
B8	There is no limit on the lowest temperature.	Não há limite para as temperaturas mais baixas.	Não há limite para a temperatura mais baixa	Não há limite para a temperatura mais baixa.	25	25a, 25b	25b	25b
	C: Students' conceptions about heat transfer and temperature change	C: Concepções dos estudantes sobre transferência de calor e mudança na temperatura	C: Concepções sobre transferência de calor e mudança de temperatura	C: Concepções dos alunos relacionados aos processos termodinâmicos	Itens (y)	Itens (y)/alt.	Itens (y)/alt.	Itens (y)/alt.
C1	Heating always results in an increase in temperature.	Aquecimento sempre resulta em uma elevação da temperatura	Aquecimento sempre resulta em um aumento de temperatura.	Aquecimento sempre resulta em um aumento na temperatura. (<i>não compreende o conceito de calor latente</i>)	3, 4, 5	3c, 3d, 4c, 5c, 5d	3c, 3d, 4a, 4c, 5c	3c, 3d, 4c, 5c, 5d
C2	Heat only travels upward.	Calor se movimentava apenas para cima	O calor só se move para cima.	O calor propaga-se em meios materiais em sentido unicamente para cima. (<i>concepção animista, pois o calor possui "desejo" de se propagar em sólidos e fluidos no sentido para cima</i>)	20	20a	20a	20a
C3	Heat rises.	Calor ascende (se eleva) <i>*Ausente em Louzada (2015, p. 1508-3)</i>	Calor sobe.	O calor sobe (concepção aristotélica animista, pois o calor possui "desejo" de estar em cima, é o seu lugar natural).	20	<i>Ausente</i>	20a	20a
C4	Heat and cold flow like liquids.	Calor e frio fluem como líquidos	Calor e frio fluem como líquidos.	Calor e frio fluem como líquidos (<i>concepção substancialista</i>).	10, 13	10d, 13b	10d, 13b	10a, 10d, 13b, 16e, 21b
C5	Temperature can be transferred.	Temperatura pode ser transferida	Temperatura pode ser transferida.	A temperatura pode ser transferida (<i>concepção substancialista, a temperatura é uma substância passível de se deslocar de um corpo para outro</i>).	7, 13	13a	13a, 22b	7d, 13a, 22b
C6	Objects of different temperature that are in contact with each other, or	Objetos a diferentes temperaturas em contato com um outro, ou em	Objetos com diferentes temperaturas em contato entre si, ou em contato	Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar	1, 2, 3, 6, 9, 10, 17, 24	1c, 9a, 24a, 24b, 24c	1c, 2a, 6a, 9a, 9c, 9d, 17b, 17c, 17d, 24a,	1c, 2a, 2c, 2d, 3a, 6c, 6d, 9a, 9c, 9d, 17b,

	in contact with air at different temperature, do not necessarily move toward the same temperature. (Thermal equilibrium is not a concept.)	contato com o ar a diferente temperatura, não necessariamente atingirão a mesma temperatura (equilíbrio térmico não é um conceito)	com o ar em diferente temperatura, não necessariamente atingem a mesma temperatura (não possuem o conceito de equilíbrio térmico)	à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (<i>a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido</i>)			24b	17c, 17d, 24a, 24b,
C7	Hot objects naturally cool down, cold objects naturally warm up.	Objetos quentes naturalmente se esfriam, objetos frios naturalmente se aquecem	Objetos quentes esfriam naturalmente e objetos frios esquentam naturalmente.	Objetos quentes naturalmente se resfriam, objetos frios naturalmente se aquecem. (<i>tange a concepção animista, como se os objetos tivessem vontade própria de se aquecerem ou se resfriarem espontaneamente, independentemente de qualquer fator externo</i>)	3, 13	3a, 13c	3a, 13c	3a, 13c
C8	Heat flows more slowly through conductors making them feel hot.	Calor flui mais lentamente em condutores fazendo com que eles pareçam quentes	Calor flui mais lentamente através de condutores fazendo-os parecer mais quentes.	Calor flui mais lentamente através de condutores fazendo com que os sinta quentes.	25	<i>Ausente</i>	26b	26b
C9	The kinetic theory does not really explain heat transfer. (Explanations are recited but not believed).	A teoria cinética não explica de fato a transferência de calor (explicações são utilizadas, mas não necessariamente "acreditadas")	A teoria cinética não explica de fato a transferência de calor (os conceitos são expressos, mas não se acredita neles).	A teoria cinética na verdade não explica a transferência de energia térmica. (<i>As explicações são relatadas, mas não acreditadas pelo sujeito, ou seja, não se percebe que as diversas formas de energia que se propaga entre as partículas representam a energia térmica em trânsito</i>).	18, 20, 21	20b, 21d	<i>Ausente</i>	18c, 18e, 20b, 21b, 21c
	D: Students' conceptions about "thermal properties" of materiais.	D: Concepções dos estudantes sobre propriedades térmicas dos materiais	<i>Ausente</i>	D: Concepções dos alunos sobre "propriedades térmicas" dos materiais.	Itens (y)	Itens (y)/alt.	Itens (y)/alt.	Itens (y)/alt.
D1	Temperature is a property of a particular material or object.	Temperatura é uma propriedade de um material ou um objeto em particular	<i>Ausente</i>	A temperatura é uma propriedade de um material ou objeto particular. (uma característica específica para cada tipo de material)	9, 14, 16, 24	9d, 14a, 16b, 16c, 24a, 24b, 24c	9d, 14a, 16b, 16c, 24a, 24b, 24c	9d, 14a, 16b, 16c, 24a, 24b, 24c
D2	Metal has the ability to attract, hold, intensify or absorb heat and cold.	Metal tem a habilidade de atrair, reter, intensificar ou absorver calor e frio.	<i>Ausente</i>	Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (<i>concepção substancialista do calor, sendo algo que pode ser atraído, absorvido</i> ,	9, 14, 16, 20	9c, 14d, 16d	9d, 14d, 16d, 16e, 20b, 22d	14d, 16c, 16d, 16e, 20b, 22c, 22d

D3	Objects that readily become warm do not readily become cold.	Objetos que se aquecem rapidamente não se esfriam rapidamente	<i>Ausente</i>	Objetos que facilmente se tornam quentes não são facilmente resfriados.	25	25d	25d	25d
D4	Different materials hold the same amount of heat.	Diferentes materiais retêm a mesma quantidade de calor	<i>Ausente</i>	Diferentes materiais, independentemente de suas massas, sempre irão receber a mesma quantidade de energia térmica (<i>acredita-se que diferentes materiais possuam a mesma capacidade térmica, e o conceito de calor assume uma concepção substancialista</i>).	11	11c	Ausente	11c
D5	The boiling point of water is 100°C (only).	O ponto de ebulição da água é 100 °C (somente)	<i>Ausente</i>	O ponto de ebulição da água é de 100°C (apenas).	4, 8, 19	4d, 8b, 8c, 8d, 19b, 19c, 19d	4d, 8b, 8c, 19d	4d, 8b, 8c, 19b, 19c, 19d
D6	Ice is at 0°C and/or cannot change temperature.	Gelo encontra-se a 0 °C e/ou não pode mudar de temperatura	<i>Ausente</i>	O gelo encontra-se sempre a 0°C e/ou não é possível alterar a temperatura.	1	1b	1b	1b
D7	Water cannot be at 0°C.	Água não se encontra a 0°C (água e sempre líquida)	<i>Ausente</i>	A água no estado líquido não pode estar a 0°C.	2, 11	2c, 2d, 11e	2c, 2d, 11e	2c, 2d, 11e
D8	Steam is more than 100°C.	Vapor se encontra a mais do que 100 °C	<i>Ausente</i>	O vapor d'água está sempre a uma temperatura superior a 100°C.	6, 19	6c, 6d	6c, 6d, 19c,	6c, 6d, 19c
D9	Materials like wool have the ability to warm things up.	Materiais como lã tem a capacidade de aquecer as coisas	<i>Ausente</i>	Materiais como a lã têm a capacidade de aquecer as coisas (concepção animista, pois entende-se que a lã seja semelhante a um mamífero ou humano, sendo uma fonte térmica natural).	17, 23	23b	23b, 26a,	17b, 17c, 17d, 23b, 26d
D10	Some materials are difficult to heat: they are more resistant to heating.	Alguns materiais são difíceis de aquecer: eles são mais resistentes ao aquecimento	<i>Ausente</i>	Alguns materiais são difíceis de serem aquecidos: eles são mais resistentes ao aquecimento.	26	26a, 26b	26d	26d
D11	Bubbles mean boiling.	Bolhas significam ebulição	<i>Ausente</i>	Borbulhar significa ferver.	-	-	-	-
D12	The bubbles in boiling water contain "air," "oxygen," or "nothing."	As bolhas na ebulição da água contêm 'ar', 'oxigênio' ou 'nada'	<i>Ausente</i>	As bolhas na água fervente contêm "ar", "oxigênio" ou "nada".	12	12a, 12b, 12d	12a, 12b, 12d	12a, 12b, 12d

A análise interpretativa encontrou algumas discordâncias com algumas correspondências sugeridas entre os três autores, estando justificado brevemente para cada distrator (ver no Apêndice A, pp. 7-210). As lacunas observadas nos trabalhos foram classificadas como “ausente”, e constatou-se 5 sugestões feitas por Gonçalves Júnior (2012). Apesar de completude alcançada nesse sentido, os problemas de ambiguidades não são totalmente satisfeitos, alguns itens provavelmente necessitam serem aprimorados/reelaborados em seu enunciado (ou quem sabe, descartados), como foi apontado na análise da chave-de-correção, e como estão sugeridos nos parâmetros e índices psicométricos obtidos através das respostas.

Quadro 60: Discordâncias e sugestões para o quadro de tipicidade de concepções térmicas do TCE.

CA's	Discordância com as correspondentes de Yeo & Zadnik (2001)	Discordância com as correspondentes de Louzada (2012)	Discordância com as correspondentes de Gonçalves (2012)	Novas sugestões de Gonçalves (2012)	Novas sugestões
A1				y21a	y16c, y16e, y21a, y21c,
A2		Ausente		y21b	y21b
A3		y 23b		y15c (?)	y15c
A4	y18		y18c, y18e (?)		y22b
A5		y7d			
A6		y7a	y7a (?)		
B1			Ausente		
B2	y16	Ausente	Ausente		
B3		y22c	y10b, y22c		
B5		Ausente	y8d		
B6	y7, y10	y7a, y7b, y10b, y22b, y22c	y7d(?), y10b,		
B8		y25a			
C1			y4a		
C3		Ausente			
C4			Ausente		y16e, y21b
C5				y22b	y22b
C6	y10	y24c	y6a		
C8	y25	Ausente		y26b	y26b
C9		y21d			
D2	y9	y9c	y9d		
D4			Ausente		
D5		y8d			
D9			y26a		
D10		y26a			
Nº de ausências	-	4	2	-	-
Nº de sugestões	-	-	-	5	11
Nº de discordâncias	6	15	12	-	-

Algumas concepções sugeridas pelos autores originais foram reforçadas por outros estudos (y7, y9, y10, y18), entretanto, a análise realizada não constatou correspondência. Por exemplo, a concepção B6, relativa na assertiva errônea de que *um corpo frio não possui a capacidade de aquecer um outro corpo* (ou “*não possui calor*”,

segundo os autores) é sugerida originalmente uma correspondência com o item y7, que diz (ver no Apêndice A7, p. 57): “Susana pega dois copos com água a 40 °C e mistura-os à água de um copo a 10 °C. Qual é a temperatura mais provável da mistura?” Diante disso é defendido uma correspondência com o distrator A e B por Louzada (2015) e D (aparentemente o símbolo “?” indicando uma falta de convicção nessa correspondência) por Gonçalves Júnior (2012), em que as alternativas são 20 °C, 25 °C, 30 °C e 50 °C. Até que ponto faz sentido afirmar a existência dessa concepção para os respondentes das alternativas A, B e C, levando em consideração que eles tenham uma noção de que o equilíbrio térmico deve estar entre 10 °C e 40 °C? Afirmer isso pode ser precipitado, tão pouco seria para a letra D, que sugere exatamente o contrário, no caso, que o corpo frio de 10 °C aqueceu o de 40 °C, resultando em uma temperatura terminal de 50 °C. Portanto, esse é um exemplo que justifica a não concordância com os demais trabalhos.

3.1.3.2.3 Coerência interna: equilíbrio dos subconstrutos no teste

O *quadro de tipicidade* das concepções alternativas térmicas (no caso, 34, de 35 apresentadas, divididas em 4 grupos – A, B, C e D) é amplo e pertinentes a categoria dos indivíduos ao levar em consideração o nível escolar e faixa etária correspondente. Através de uma análise de *coerência interna* é possível verificar o equilíbrio dos encargos pedagógicos relacionados a tipicidade dos construtos, e isso visa compreender se há ou não uma distribuição relativamente homogênea entre as concepções alternativas no instrumento e assim constatar de que forma e intensidade ele avalia. Considera-se, entretanto, que apesar da existência de quatro grupos categóricos de concepções alternativas, a análise do teste com base na TRI (perspectiva fatorial) irá considerar esse conjunto de itens referente à apenas um traço latente predominante (pressuposto da unidimensionalidade), no caso, considera-se *perfil conceptual térmico*.

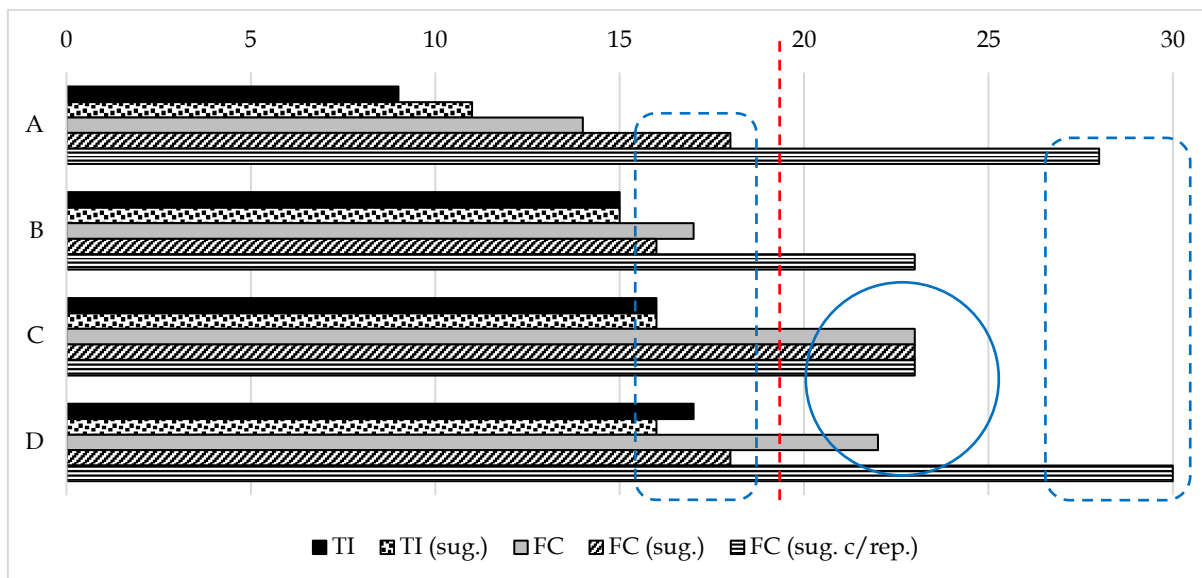


Figura 39: Coerência interna quanto ao equilíbrio de concepções dentro de cada grupo, ou seja, total de concepções distribuídas nos 26 itens e total de itens que possuem concepção dentro de cada grupo (Comparação entre os autores originais e as sugestões nesse estudo).

No gráfico se tem o *Total de Itens* (TI) dentro de cada grupo de concepções, bem como a *Frequência de Concepções* (FC) por grupo, ou seja, quantas vezes as concepções contidas em cada grupo incidem entre os 26 itens. Cada item pode conter mais de uma concepção dentro de cada grupo considerado, e ainda conter concepções repetidas distribuídas entre os distratores. conforme mostra a barra que informa “sugestões com repetição”, por exemplo, o item y7 contém a concepção A1 nos distratores “a” e “d”, com isso, conta-se duas vezes a mesma concepção para o mesmo item.

Analisando as sugestões originais, pode-se dizer que o instrumento foi elaborado de forma a se concentrar nas concepções dos grupos C e D, que se referem os *processos termodinâmicos e propriedades térmicas*. A analisando a quantidade de concepções por item, tem-se uma média de 2,92, ou seja, 3 concepções por item. Apesar disso, uma análise verificativa constatou que de fato o TCE se concentra no grupo C, porém um certo equilíbrio entre o grupo A e D, que pode ser reforçado pelas elevadas incidências repetidas dentro dos itens. Em média, o estudo preliminar considera 4 concepções por item. Comparando, nos dois casos há uma média de 19 concepções por grupo.

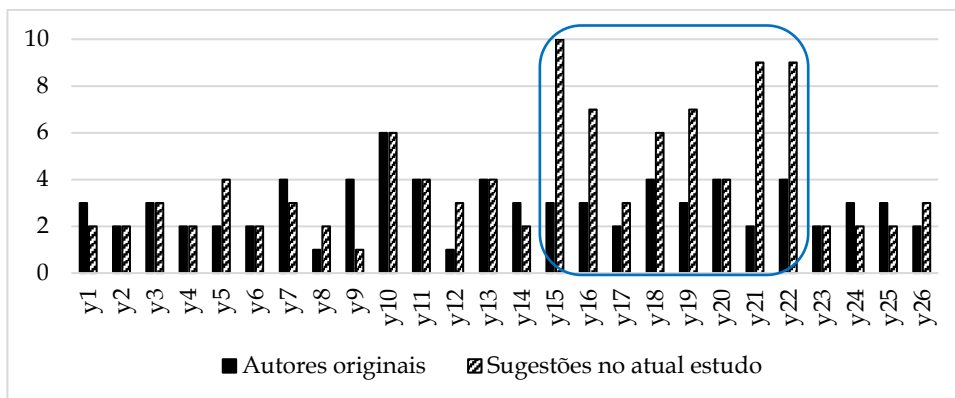


Figura 40: Frequência de Concepções por Item, de acordo com os autores no trabalho original, comparados com a sugestão nesse estudo.

Ao analisar o número de concepções para cada item, é possível associar o gráfico acima (Figura 40) com o tabela abaixo (Tabela 19), evidenciando que algumas concepções acabam por terem incidências dentro do próprio item, entre os distratores (para maiores detalhes, ver Apêndice A, p. 1). Observa-se que o item y10 possui 6 concepções entre os 4 distratores, sendo 2 concepções do grupo A com 3 incidências (A1 nas alternativas “a” e “d”; A3 na alternativa “a”), 1 concepção do grupo B (B3 na alternativa “d”), e 2 concepções do grupo C (C4 na alternativa “a”; C6 na alternativa “d”), sendo o item mais inclusivo nas incidências dos *diferentes grupos de concepções* envolvidas, segundo os autores. Apesar disso, a análise verificou que os itens y15, y16, y19, y21 e y22 possuem mais ocorrência de concepções entre os distratores, se concentrando em dois grupos.

Contrapondo, os itens y8 e y12, possuem apenas 1 concepção ou equívoco entre os distratores, segundo os autores. A análise verificou que o item y9 também possui apenas uma única incidência. Constata-se ainda que o grupo C possui um total de 23 *frequências de concepções* (FC) distribuídas em 16 itens, com o mesmo quantitativo da análise realizada no atual estudo, entretanto, com diferenças qualitativas em destaque para os itens y10 (sem a concepção C6), y16 (considerando a concepção C4 na alternativa “e”), y21 (considerando a concepção C4 na alternativa “b”), y22 (considerando a concepção C5 na alternativa “b”), e ausências nos itens y24 e y25 (ver no Apêndice A24, p. 195, e A25, p. 203, a correspondência especificamente relacionada aos distratores). Essa mesma diferença pode ser contatada entre os outros grupos.

Também foi observado uma quantidade maior de concepções nos itens y21 (5 concepções) e y22 (7 concepções).

Tabela 19: Comparação da distribuição de concepções nos itens do teste entre os autores originais e o atual estudo, no caso se tem a *Frequencia de Concepções por Grupo (FCG)* e *Total de Itens (TI)* que incidem nessas concepções dentro de cada Grupo (G).

G	Itens																										FC G	TI
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26		
A	-	-	-	-	-	-	5; 6	-	-	1; 3	5	-	3	-	4; 5	-	-	3; 4	-	-	-	1; 2	3	3	-	-	14	9
A	-	-	-	-	-	-	5; 6	-	-	1; 3	5	-	3	-	3; 4; 5	1	-	3	-	-	1; 2	1; 2; 4	3	3	-	-	18	11
B	7	-	-	-	4	-	6	-	7	3; 6	6	-	-	7	1	2	-	3	5	-	3	3; 6	-	-	8	6	17	15
B	7	-	-	-	4	-	-	-	7	3	6	-	-	7	1	2	-	3	5	-	3	3; 6	-	-	8	6	16	15
C	6	6	1; 6; 7	1	1	6	5	-	6	4; 6	-	-	4; 5; 7	-	-	-	6	9	-	2; 3; 9	9	-	-	6	8	-	23	16
C	6	6	1; 6; 7	1	1	6	5	-	6	4	-	-	4; 5; 7	-	-	4	-	9	-	2; 3; 9	4; 9	5	-	6	-	8	24	17
D	6	7	-	5	-	8	-	5	1; 2	-	4; 7	12	-	1; 2	-	1; 2	9	-	5; 8	2	-	-	9	1	3	10	22	17
D	6	7	-	5	-	8	-	5	1; -	-	4; 7	12	-	1; 2	-	1; 2	9	-	5; 8	2	-	2	9	1	3	10; 9	23	18
T	3	2	3	2	2	2	4	1	4	6	4	1	4	3	3	3	2	4	3	4	2	4	2	3	3	2	2,92	
T	3	2	3	2	2	2	3	1	3	4	4	1	4	3	4	4	1	3	3	4	5	7	2	3	2	4	4,00	

Nota: Nas linhas dentro de cada grupo estão indicadas as numerações das concepções que se encontram no quadro de tipicidade, na linha inferior o resultado da análise realizada e as sugeridas. O sombreado representa a existência de uma diferença ao comparar com os autores do teste. E em negrito, a forma como foi sugerido originalmente (linha superior dentro de cada grupo G de concepções).

Dentre as diversas concepções, pode-se constatar na versão «original» do TCE que a concepção térmica **C6** é a **mais evidenciada** com 8 incidências entre as 26 questões (y1, y2, y3, y6, y9, y10, y17, y24), e está no grupo dos equívocos relacionados aos *processos termodinâmicos*, e se refere a ideia de que *os corpos com diferentes temperaturas não tendem ao equilíbrio térmico*, ou seja, não se admite a *Lei Zero da Termodinâmica*. As concepções **A3** (y10, y13, y18, y23, y24) e **B6** (y7, y10, y11, y22, y26) possuem 5 incidências e se referem a crença de que *o calor e frio são diferentes, em vez de extremidades opostas de um continuum*, e que *um corpo frio não contém calor*. No grupo dos equívocos relacionados as *propriedades térmicas*, os autores propuseram (intencionalmente ou não) uma maior incidência para as concepções **D1** (y9, y14, y16, y24) e **D2** (y9, y14, y16, y20), e se referem na crença de que *a temperatura é uma propriedade de um material*, e que *o metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio*.

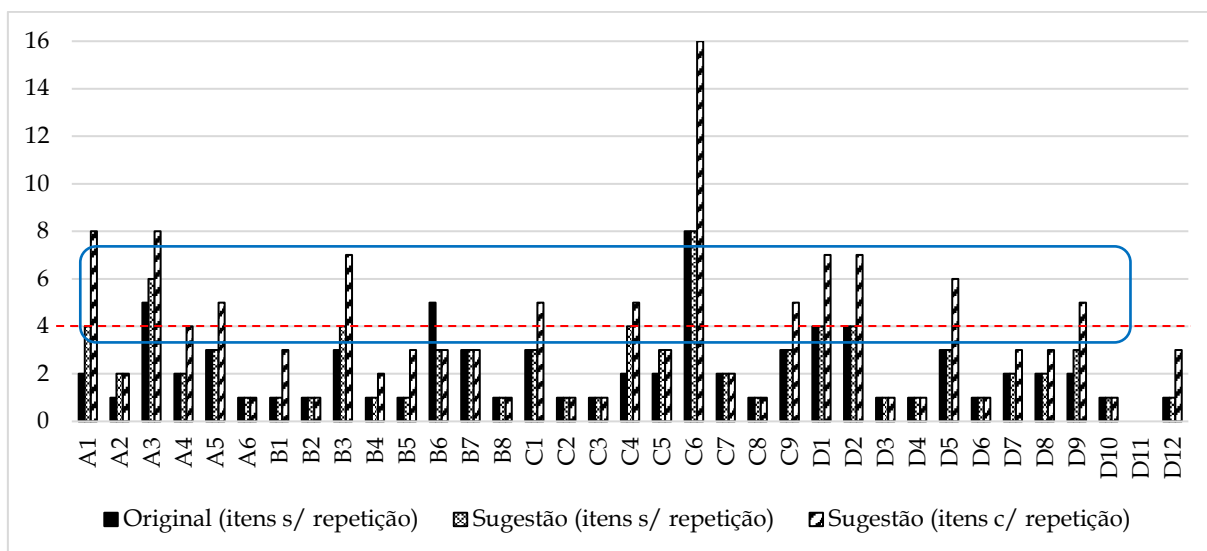


Figura 41: Incidência das concepções no teste.

Após uma análise interpretativa e verificativa na correspondência de itens e as concepções (sem considerar repetições entre os distratores, semelhante como foi apresentado pelos autores do TCE), constatou-se algumas discordâncias com incidências a mais (A1, A2, A3, B3, C4, C5, D9) e a menos (B6). Considerando o quantitativo das incidências de concepções com repetição no mesmo item, sendo distribuída entre os distratores, foi identificada uma **maior ênfase no C6** com 16 ocorrências, seguido de 8 ocorrências **na concepções A1 e A3**, depois 7 **com B3, D1 e D2**. O D11 ficou ausente em todas as perspectivas, e se torna sem propósito para o TCE. De acordo com a análise feita, se tem a média de 4 concepções por item sendo representada pela linha horizontal tracejada.

A análise da coerência interna busca verificar o equilíbrio do construto no TCE, e uma forma de visualizar a distribuição dos itens, de acordo com os autores, pode ser feita por um esquema de conjunto e subconjuntos, na qual nota-se uma **concentração dos itens nos grupos C e D**. Apesar disso, como foi dito, a análise realizada certifica uma concentração no C e depois um certo equilíbrio nos grupos A e D.

É relevante considerar que a elevada incidência de algumas concepções sinaliza que seja uma concepção mais comum, ou seja, presente de modo mais frequente na maioria dos fenômenos e situações térmicas, com isso torna-se um desafio tentar equilibrar de forma mais homogênea as concepções de um teste, tendo em vista que

outras concepções são mais restritas a eventos específicos que outras. A concepção C6, por exemplo, se refere na crença de que “*objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido)*”, ou seja, é um parâmetro bem generalizado, presente em quase todos os fenômenos térmicos, e claramente envolve uma compreensão sobre o **conceito de equilíbrio térmico**, sendo, portanto, considerado um conceito-chave.

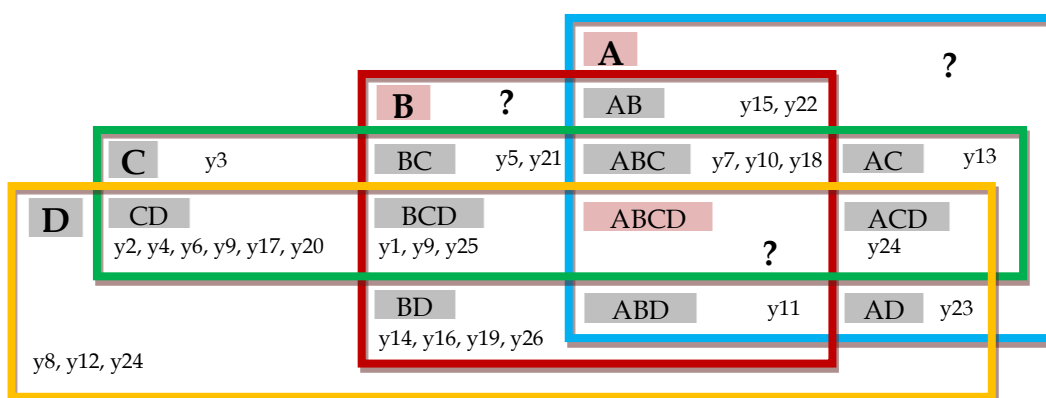


Figura 42: Esquema da distribuição dos itens entre os grupos de concepções, de acordo com os autores do TCE.

3.1.3.3 Elaboração da Matriz de Referência do Teste TCE

A elaboração da matriz de referência está relacionada ao conjunto de competências consideradas dentro da escala de proficiências. Antes disso, foi considerado relevante a elaboração da matriz de habilidades de todos os itens juntamente com a definição da sentença descritora de suas respectivas habilidades. Esse tipo de análise serve para a identificação do principal ou mais destacado traço latente na dimensão cognitiva que o item de forma explícita busca aferir. Apesar disso, considera-se ainda que possam haver outros traços latentes secundários, que em uma análise mais aprofundada pode ser identificada. Após identificar e classificar os objetivos e as sentenças descritoras, torna-se fiável uma validade de construto do teste, afirmando sobre a sua finalidade de acordo com as dimensões do conhecimento envolvido.

A elaboração da sentença descritora de cada item especifica o tipo de capacidade que se exige, nisso se envolve a ação cognitiva que geralmente é expressa por um verbo de ação sendo acompanhada por um determinado conteúdo específico. O conjunto que relaciona a *capacidade* mais o *conteúdo* é considerado o «objetivo específico» (Gérard & Roegiers, 1998). Assume-se uma noção de capacidade podendo estar associada a dimensão cognitiva (relacionada ao saber-fazer cognitivo – habilidades), psicomotora (relacionada ao saber-fazer gestual ou procedimental – destrezas) ou socioafetiva (relacionada ao saber-ser).

Quadro 61: Aspectos delimitadores na elaboração das sentenças descritoras da matriz de competências científicas.

Objetivo Específico « <i>de saberes</i> » é igual a ...	{ <ul style="list-style-type: none"> Capacidade (habilidade ou destreza) + Conteúdo ou Operação Mental + Objeto do conhecimento
Competência	{ <ul style="list-style-type: none"> Objetivo Específico (Capacidade + Conteúdo) + Situações ou Operação Mental + Objeto do conhecimento + Complemento do objeto do conhecimento + Contexto (contexto <i>da área do conhecimento</i> ou contexto <i>da situação real</i> – pessoal ou profissional)
Sentença Descritora: Capacidade (habilidade ou destreza) + Conteúdo + Situação	

A busca de um modelo pedagógico baseada por objetivos educacionais tem seu destaque com o paradigma de *Pedagogia Por Objetivos* (PPO), mas que, entretanto, se concebe suas contribuições na construção das matrizes de habilidades e competências dentro do paradigma de uma *Abordagem Por Competências* (APC). Nesse mesmo sentido, o INEP propõe uma forma de elaborar as sentenças descritoras em que se identifica como «operação mental» a capacidade envolvida e o «objeto do conhecimento» de forma correspondente ao conteúdo específico.

Na análise taxonômica dos itens (Anderson et al., 2001), consideram-se duas dimensões: os *processos cognitivos* e as *categorias do conhecimento envolvido*. Essa análise possibilita especificar a proficiência para uma resposta correta do item, ou seja, o domínio do conhecimento que o item exige e que se pretende verificar na sondagem, sendo considerado com parte principal na qual se interpreta o *subsunçor*, dentro da concepção ausubeliana.

O conhecimento factual no teste se refere conhecer valores de temperatura em situações típicas, como as de mudança de estado físico da água, e outras possíveis temperaturas (y1, y2, y4, y8, y25). Entende-se que o conceito de calor remete a um processo termodinâmico, portanto uma compreensão sobre a transferência térmica envolve ao mesmo tempo conceitos e princípios que devem ser classificados como conhecimento conceptual ou conceitual (y2, y3, y5, y6, y7, y9–y23, y26). No conhecimento processual estão os critérios para a realização de procedimentos adequados, como por exemplo, em quais lugares um líquido pode ferver mais facilmente (y8), ou um corpo pode ser aquecido (y26).

A análise considera o mínimo necessário das habilidades cognitivas para êxito no teste, e em alguns casos sugere que umas capacidades incorporam outras, ou seja, de que existem uma rede de modelos mentais associados, tanto sobre as concepções e equívocos quanto das convicções científicas.

Para uma forma de tentar sistematizar os itens, eles foram organizados com a concepção de competências científicas, no caso, buscando identificar famílias de situações-problemas, ou seja, itens que envolvam conhecimentos associados em situações distintas, e que remetam a uma capacidade de compreender um determinado problema mais amplo e comum. Considera-se que muitas vezes as famílias de situações estão juntas dentro de um mesmo contexto, como por exemplo, na *competência para dirigir um automóvel*, as habilidades podem ser da ordem psicomotora (utilizando os pés, mãos, olhos, ...), cognitiva (sabe a finalidade e funcionamento de dispositivos) ou emocional (saber ter autocontrole em situações de risco, etc.). Para exemplificar algumas habilidades dessa competência, tem-se: *ajustar um retrovisor, ser capaz de ter uma noção lógico-espacial através do retrovisor, saber passar as marchas adequadamente, ajustar a altura do banco, saber frear/acelerar em momentos específicos e necessários, comunicar com os outros motoristas utilizando corretamente os sinalizadores, etc.*, ou seja, nesse caso, todas as habilidades encontram-se dentro de um mesmo contexto, apesar de que alguns devam ser usados em situações específicas e distintas.

A finalidade do TCE não é necessariamente de promover ou verificar competências em si, pois sendo aplicadas em alunos iniciantes em uma circunstância de pré-instrução, se caracteriza como sondagem de concepções (avaliação para a orientação), e para alunos em pós-instrução em caráter verificativo ou acumulativo de conhecimento (avaliação para a certificação). No entretanto, como os *Inventários de Conceitos* são em grande parte elaborados exclusivamente por conceitos introdutórios, sugere-se que se estabeleçam essas restrições, devendo ser no máximo consideradas *competências científicas de base conceitual*. Diante disso, é possível verificar itens que buscam configurar competências científicas comuns pela proximidade e força associativa para responderem determinadas questões relativamente mais amplas, ou seja, se comportam como uma família de situações-problemas que correspondem a uma competência. Concomitantemente à análise taxonômica, identificou diferentes **domínios cognitivos** associados dentro de um mesmo campo epistemológico na qual se reivindica a necessidade de mobilização de recursos para o enfrentamento e compreensibilidade de determinadas situações-problema. A delimitação dessa análise se restringe ao teste, obviamente, porém abre a possibilidade de ser explorada de forma mais exaustiva dentro de um processo de reelaboração e aprimoramento do instrumento, excluindo e/ou propondo a inclusão de novos itens.

Dentro dessa análise, entende-se que a identificação de aspectos comuns para o domínio de um determinado conhecimento científico envolva não somente configurar uma família de situações-problema, mas também a possibilidade de identificar diversos recursos cognitivos necessários para a compreensibilidade de um problema e situação específica, sendo esta uma análise interpretativa preliminar de um processo que possibilita a formulação de competências científicas como categorias inclusivas que abarcam diversas proficiências como subconstrutos de um traço latente predominante, na qual se pretende estimar.

Essa perspectiva de formulação de competência pode ser comparada e explorada com a *perspectiva advinda das análises fatoriais* da TRI, porém, é preciso ressaltar que esses processos são percursos epistemológicos distintos para a construção e validação do conhecimento científico envolvido no teste, mas que não se

anulam, e sim possibilitam uma ampliação da visão diante dos conhecimentos e potencialidades de análise para os resultados do teste. Pode-se dizer ainda que seja um refinamento no processo de validade interna do instrumento.

Partindo do pressuposto que o teste se refira a um traço latente predominante, pode-se denominar como *proficiência térmica em conceitos básicos introdutórios*. É muito provável que essa proficiência não abarque todos os conceitos básicos térmicos necessários dentro para um processo de introdução à Termodinâmica, sendo assim, recomendável aprofundar na validação do construto que satisfaça de forma exaustiva e pertinente essa condição. Entretanto, na tentativa de interpretação e proposição de competências científicas relacionadas ao TCE, realizou-se uma análise interpretativa taxonômica dos enunciados e alternativas que resultaram na descrição de habilidades, conteúdos e contextos envolvidos nos itens, estando-os pertencentes e pertinentes a um recorte epistemológico criterioso do conhecimento térmico para determinados grupos de itens. Assim, essas competências podem ser descritas como domínios cognitivos necessários quanto a:

- *A temperatura de um corpo ou de um ambiente em situações típicas* (**Lembrar** (fatos): y1, y2, y4, y8, y15; **Compreender** (conceitos): y25) – ênfase em *recordar*;
- *A invariância da temperatura em um processo de mudança de fase* (**Lembrar** (fatos): y2, y3; **Lembrar** (conceitos): y5; **Aplicar** (processos): y6; **Aplicar** (conceitos): y7) – ênfase em *recordar*;
- *Aos processos de fervura de acordo com a variação da pressão* (**Aplicar** (processos): y8; **Aplicar** (conceitos): y19) – ênfase em *realizar (aplicar)*;
- *Ao comportamento térmico dos corpos durante a mudança de fase* (**Compreender** (conceitos): y12) – ênfase em *compreender*. Uma competência científica nesse contexto necessitaria de mais habilidades envolvidas, caso contrário se configura como uma capacidade latente incompleta pela falta de itens que remeta a necessidade dentro dessa campo de compreensão. Obviamente que a competência necessita ainda de uma clareza do problema e a apresentação de uma situação concreta, que no caso se restringe deve-se estar textualizada de forma eficiente quanto no que diz respeito a qualidade científico-pedagógica do encargo de legibilidade textual no enunciado do

item, devendo ser um aspecto importante a ser levado em consideração dentro do processo de (re)elaboração ou aprimoramento do item (“engenharia” de construção do item);

- *A propagação do calor (Compreender (conceitos): y11, y13, y16, y20) – ênfase em compreender;*
- *A sensação térmica explicada pela propagação do calor (Compreender (conceitos): y10, y18, y21, y22) – ênfase em compreender;*
- *As implicações no comportamento térmico de acordo com as propriedades térmicas (Compreender (conceitos): y9, y14, y17, y24; Aplicar (conceitos): y23, y26) – ênfase em compreender.*

Como as competências científicas remetem a um conjunto de situações-problemas comuns, podem eventualmente estar associadas na dependência local de itens. Uma das formas de ocorrer é uma sequenciação temporal de situações dentro de um mesmo contexto, como ocorrem nos itens y1, y2, e y3, como também nos itens y4, y5 e y6, sendo dependentes somente os itens subsequentes ao primeiro (no caso, y2 e y3, e y5 e y6). Nesse sentido foram encontrados alguns **indícios** de dependência local entre os itens, quando se supõem que o respondente siga sistematicamente no sentido do primeiro ao último item, sem “pular” questões (Uma breve justificativa para os indícios de dependência encontra-se no Apêndice A, pp. 1-210, dentro de cada item).

Tabela 20: Indícios de dependência local no TCE.

Itens com indícios de dependência	Itens correspondentes a suposta dependência	Sequenciação de situações dentro um mesmo contexto
y2	y1	Sim
y3	y1; y2	Sim
y5	y4	Sim
y6	y4; y5	Sim
y8	y4	Não
y10	y9	Sim
y14	y9; y10	Não
y15	y9; y10; y14	Não
y16	y9; y10; y14	Não
y19	y4; y8	Não
y24	y9; y10; y14; y16	Não
y25	y1	Não
y26	y23	Não

Algumas suposições de dependência estão mais associadas como o conhecimento comum que os itens reivindicam aos respondentes, outras estão relacionadas a uma sequência de situações e respostas dentro de um mesmo contexto. Esses são aspectos qualitativos analisados previamente que podem servir de argumentos subjetivos visando uma análise mais detalhada e comparativa dos resultados quantitativos, caso haja uma necessidade. Também pode ser útil em um diálogo com outros estudos que busque compreender a existência de modelos mentais baseados na associação e respostas para o teste. Ainda assim, seguindo os pressupostos da TRI, os modelos estimam de forma mais otimizada uma forma de melhor explicar a relação entre os itens, porém não devem ser consideradas como argumentos irrefutáveis de que não possam existir dependências, sendo necessário nesse caso aprofundar as análises.

Quadro 62: Matriz das habilidades cognitivas no Teste TCE de acordo com as descrições nas análises taxonômicas.

ANÁLISE DAS DIMENSÕES TAXONÔMICAS (Processo do Conhecimento. / Categoria do Conhecimento.)		Itens
Descritor	COMPETÊNCIA CIENTÍFICA ASSOCIADA A: <i>Conhecer a Temperatura de um Corpo ou de um Ambiente em Situações Típicas</i>	
1	<p>⇒ Lembrar o valor da <u>temperatura de fusão</u> do gelo em situações mais comuns. (Dimensão Processual do Conhecimento).</p> <p>⇒ Factual 1.2 - Conhecimento de detalhes específicos e elementos, no caso, saber que a temperatura de fusão do gelo é em situações mais comuns com valor entorno de 0 °C. (Dimensão Categórica do Conhecimento).</p>	y1, y2 Corpos da mesma substância em contato sendo transformados e misturados
2	<p>⇒ Lembrar que em um processo de resfriamento a <u>temperatura</u> dos corpos pode ficar abaixo de 0 °C até um limite de -273,15 °C. (Dimensão Processual do Conhecimento).</p> <p>⇒ Factual 1.2 - Conhecimento de detalhes específicos e elementos, no caso, das temperaturas prováveis para determinado corpo reconhecendo a temperatura hipotética do <i>Zero Absoluto</i> sendo a menor possível, ou seja, 0 K ou -273,5 °C. (Dimensão Categórica do Conhecimento)</p>	y1, y25 Análise do comportamento de um corpo sendo transformado ou alterando de temperatura
3	<p>⇒ Lembrar o valor da <u>temperatura de ebulição</u> da água em situações mais comuns. (Dimensão Processual do Conhecimento).</p> <p>⇒ Factual 1.2 - Conhecimento de detalhes específicos e elementos, no caso, saber que a temperatura de ebulição da água é em situações mais comuns com valor entorno de 100 °C. (Dimensão Categórica do Conhecimento).</p>	y4, y8 Análise do comportamento de um corpo sendo transformado ou alterando de temperatura
4	<p>⇒ Compreender sobre a concepção de <u>sensação térmica</u> diferenciando-a da noção de temperatura. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 - Conhecimento dos princípios e generalizações, no caso, um valor previsto para a temperatura e condições climáticas ambientais não se refere de forma correspondente uma aferição prevista de sensação térmica, no caso, qualquer medida de temperatura de um sistema deve ser interpretada como uma das grandezas de caracterização de um estado termodinâmico, diferentemente da noção de sensação térmica que se associa com uma percepção tátil de ganho ou perda de energia térmica. (Dimensão Categórica do Conhecimento)</p>	y15 Interpretação comparativa do comportamento térmico diante de diferentes situações postas <i>lado a lado</i>
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: <i>Recordar sobre a Invariância da Temperatura em um Processo de Mudança de Fase</i>		
5	<p>⇒ Lembrar que a temperatura de um corpo permanece constante durante a mudança de fase (seja no aquecimento como no resfriamento), com isso a temperatura é a mesma para a substância nos dois estados físicos de fronteira (p. ex.: gelo + água ou água + vapor d'água). (Dimensão Processual do Conhecimento).</p> <p>⇒ Conceitual 2.2 - Conhecimento dos princípios e generalizações que envolvem a quantidade de calor latente, no caso, que a temperatura permanece constante para substâncias homogêneas. (Dimensão Categórica do Conhecimento)</p>	y2, y3 Corpos da mesma substância em contato sendo transformados e misturados y5, y6, y7 Análise do comportamento de um corpo sendo transformado ou alterando de temperatura
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: <i>Aplicação de Processos de Fervura de acordo com a Variação da Pressão</i>		
6	<p>⇒ Aplicar um processo de aquecimento considerando que o <u>ponto de ebulição</u> da água varia conforme a variação da pressão. (Dimensão Processual do Conhecimento)</p> <p>⇒ Processual 3. 3 - Conhecimento dos critérios para determinar quando utilizar processos adequados, no caso, que a temperatura ou ponto de ebulição diminui com a altitude, e com isso, determinar os lugares mais prováveis em que a água possa ferver mais facilmente. (Dimensão Categórica do Conhecimento)</p>	y8 Análise do comportamento de um corpo sendo transformado ou alterando de temperatura
7	<p>⇒ Aplicar um processo de aquecimento considerando que o <u>ponto de ebulição</u> da água varia conforme a variação da pressão. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 - Conhecimento dos princípios e generalizações, no caso, o ponto de ebulição da água contida na sopa diminui devido ao aumento da pressão no interior da panela, e isso ocorre devido ao aquecimento causar uma maior agitação das partículas e delas estarem confinadas em um recipiente hermético sendo impedidas de expandirem. (Dimensão Categórica do Conhecimento) (y19).</p>	y19 Análise do comportamento de um corpo sendo transformado ou alterando de temperatura

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: <i>Compreensão sobre o Comportamento Térmico dos corpos durante a Mudança de Fase</i>		
8	<p>⇒ Compreender que na <u>água em ebulição</u> as bolhas formadas na parte inferior são moléculas de água que são mais rapidamente vaporizadas. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, quando líquidos (de modo geral) em um recipiente são aquecidos, as moléculas que se encontram na parte inferior e mais próximas da zona de aquecimento são vaporizadas mais rapidamente, e é o que acontece com a água em ebulição. (Dimensão Categórica do Conhecimento).</p>	<p>y12</p> <p>Análise do comportamento de um corpo sendo transformado ou alterando de temperatura</p>
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: <i>Compreensão sobre a Propagação do Calor</i>		
9x	<p>⇒ Compreender o <u>processo de resfriamento</u> entre corpos em estados termodinâmicos distintos, que apesar de estarem na mesma temperatura, seja previsto que perdem uma quantidade de energia térmica distinta até alcançarem um equilíbrio termodinâmico com a vizinhança (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, que o estado termodinâmico para uma substância é o fator que determina de modo geral a quantidade de calor que se perde até que ambos estejam em um mesmo ponto de equilíbrio. Nesse estágio ambos estariam de acordo com a <i>Lei Zero da Termodinâmica</i>. (Dimensão Categórica do Conhecimento).</p>	<p>y11</p> <p>Comparação do comportamento térmico entre diferentes corpos postos <i>lado a lado</i></p>
10	<p>⇒ Compreender sobre o processo das <u>trocas de calor</u> na qual um corpo sólido mais quente transfere energia térmica para o corpo mais frio. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, a 1ª Lei da Termodinâmica garante como princípio geral que a energia térmica se transfere naturalmente de uma fonte mais quente a uma fonte mais fria (Dimensão Categórica do Conhecimento).</p>	<p>y13</p> <p>Análise do comportamento térmico de um corpo por influência do contato com outro</p>
11	<p>⇒ Compreender sobre a concepção de <u>sensação térmica</u> comparando corpos com diferentes propriedades térmicas estando à mesma temperatura (régua de metal e régua de madeira), independente da sensação térmica que se percebe. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, a Lei de Fourier garante que todo corpo que possui maior condutividade térmica tende a conduzir o calor (ganhando ou cedendo energia térmica) mais rapidamente. (Dimensão Categórica do Conhecimento)</p>	<p>y16</p> <p>Comparação do comportamento térmico entre diferentes corpos postos <i>lado a lado</i></p>
12	<p>⇒ Compreender que os <u>processos de convecção</u> de massas de ar aquecidas são em parte explicados pela diferença de densidade, em que as massas de ar quente são menos densas. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, massas de ar quando aquecidas tornam-se menos densas, sobem e evidenciam que a parte superior do forno seja mais favorável para o aquecimento. (Dimensão Categórica do Conhecimento)</p>	<p>y20</p> <p>Compreensão sobre as formas e processos na transferência de energia térmica</p>
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: <i>Compreensão sobre a Sensação Térmica explicada pela Propagação do Calor.</i>		
13	<p>⇒ Compreender sobre o processo das <u>trocas de calor</u> entre diferentes corpos sólidos se baseando pela sensação térmica do tato (quente, frio e morno), obviamente sem que haja mudança de fase. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, o contato físico durante as trocas de calor permite uma percepção de quente quando se recebe energia térmica deste corpo, quando se cede energia se percebe frio. (Dimensão Categórica do Conhecimento).</p>	<p>y10</p> <p>Análise do comportamento térmico de um corpo por influência do contato com outro</p>
14	<p>⇒ Compreender sobre a concepção de <u>sensação térmica</u> entre diferentes corpos sólidos idênticos com diferentes temperaturas, e com isso inferir o porquê de se perceber mais frio ao tocar o corpo de menor temperatura. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, o contato físico permite uma sensação térmica mais intensa quanto maior for a diferença de temperatura entre os corpos, e isso se torna um fator explicativo na percepção de aumento ou diminuição do fluxo de energia térmica quando se comparam corpos com as mesmas propriedades térmicas, sendo a condutividade térmica nesse caso um fator não explicativo. (Dimensão Categórica do Conhecimento).</p>	<p>y18</p> <p>Comparação do comportamento térmico entre diferentes corpos postos <i>lado a lado</i></p>
15	<p>⇒ Compreender que o <u>processo de trocas</u> de calor entre a pele e o suor ocorre no sentido do corpo (quente) para o suor (frio), e sensação térmica é atenuada pela evaporação. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, que as trocas de calor se dão no sentido do corpo para a gota, sendo percebido uma sensação de frio mais intensa devido a evaporação que ocorre. (Dimensão Categórica do Conhecimento)</p>	<p>y21</p> <p>Análise do comportamento térmico de um corpo por influência do contato com outro</p>

16	<p>⇒ Compreender sobre a concepção de <u>sensação térmica</u> com base no entendimento que durante as trocas de calor quem recebe energia térmica possui uma sensação de “quente” (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, que no princípio das trocas de calor o corpo que recebe calor possui a sensação térmica de “quente” e o corpo que cede calor possui a sensação de “frio”, no caso as paredes metálicas da bomba recebe calor do interior por meio das transformações adiabáticas. (Dimensão Categórica do Conhecimento)</p>	<p>y22</p> <p>Compreensão sobre as formas e processos na transferência de energia térmica</p>
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: <i>Compreensão do Comportamento Térmico de acordo com as Propriedades Térmicas.</i>		
17	<p>⇒ Compreender sobre a concepção de <u>equilíbrio térmico</u> comparando corpos com diferentes propriedades térmicas (plástico, metal, madeira, etc.) sem que haja mudança de fase, no caso envolvendo tanto o uso de um termômetro quanto a sensação térmica. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, que a <i>Lei Zero da Termodinâmica</i> garante a existência de equilíbrio térmico entre diferentes corpos sob as mesmas condições climáticas ambientais da vizinhança. (Dimensão Categórica do Conhecimento).</p>	<p>y9, y14, y24</p> <p>Comparação do comportamento térmico entre diferentes corpos postos <i>lado a lado</i></p>
18	<p>⇒ Compreender sobre o processo das <u>trocas de calor</u> comparando corpos com diferentes propriedades térmicas (toalha molhada e seca) em contato físico com mesmos objetos em condições idênticas visando determinar a temperatura externa do ambiente que envolve o sistema. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, corpos com mesma condutividade térmica tendem a variar a mesma quantidade de temperatura em um mesmo intervalo de tempo, com isso, a temperatura ambiente tende a ser igual ou superior a toalha seca de 22 °C, que supostamente estava em equilíbrio térmico no estado inicial. (Dimensão Categórica do Conhecimento)</p>	<p>y17</p> <p>Comparação do comportamento térmico entre diferentes corpos postos <i>lado a lado</i></p>
19	<p>⇒ Aplicar procedimentos ou utilização adequada de objetos com a finalidade de aquecer, manter a temperatura ou que possa resfriar um corpo ou um ambiente, e diante disso saber interpretar sua relação com a sensação térmica. (Dimensão Processual do Conhecimento)</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, isolantes térmicos não aquecem, mas sim dificultam as trocas de calor e com isso exercem uma função de tentar manter a temperatura de um sistema (p.ex.: fornecer informações sobre a finalidade no uso de agasalhos). (Dimensão Categórica do Conhecimento)</p>	<p>y23, y26</p> <p>Análise do comportamento térmico de um corpo por influência do contato com outro</p>
20	<p>⇒ Avaliar a relevância de diferentes perspectivas sobre os procedimentos ou utilização de objetos com a finalidade de aquecer, manter a temperatura ou que possa resfriar um corpo ou um ambiente.</p> <p>⇒ Conceitual 2.2 – Conhecimento dos princípios e generalizações, no caso, isolantes térmicos não aquecem, mas sim dificultam as trocas de calor e com isso exercem uma função de tentar manter a temperatura de um sistema (p.ex.: fornecer informações sobre a finalidade no uso de agasalhos). (Dimensão Categórica do Conhecimento)</p>	<p>y26</p> <p>Análise do comportamento térmico de um corpo por influência do contato com outro</p>
21	<p>⇒ Avaliar a relevância de diferentes perspectivas sobre os procedimentos ou utilização de objetos com a finalidade de aquecer, manter a temperatura ou que possa resfriar um corpo ou um ambiente.</p> <p>⇒ Processual 3. 3 – Conhecimento dos critérios para determinar quando utilizar processos adequados, no caso, saber em quais situações o uso do agasalho poderá ser útil em manter aquecido um corpo. (Dimensão Categórica do Conhecimento)</p>	<p>y26</p> <p>Análise do comportamento térmico de um corpo por influência do contato com outro</p>

A análise sugere que o teste TCE se **concentra** como um instrumento que busca mais enfaticamente **compreender conceitos** introdutórios térmicos (53,8%), como se esperava. Porém traz algumas sinalizações de que o respondente deva lembrar de determinadas temperaturas de mudança de fase, entre outras possíveis. Também afere um conhecimento sobre os procedimentos e condições favoráveis para que ocorram determinados comportamentos térmicos.

Quadro 63: Análise taxonômica cognitiva dos objetivos específicos dos itens no teste TCE, baseado em Anderson et al. (2001, p. 28).

		Dimensão ou Categorias dos Processos Cognitivos						
		Lembrar	Compreender	Aplicar	Analisar	Avaliar	Criar	
Dimensão dos conhecimentos	Factual	1. 1 Conhecimento da terminologia	y1, y2, y4, y8, y25	19,2%				
		1. 2 Conhecimento de detalhes específicos e elementos						
	Conceitual	2. 1 Conhecimento de classificações e categorias						
		2. 2 Conhecimento dos princípios e generalizações	y2, y3, y5, y6, y7	y9-y22, y24	y19, y23, y26	84,5%	y26	3,8%
		2. 3 Conhecimento das teorias, modelos e estruturas	19,2%	53,8%	11,5%			
	Processual	3. 1 Conhecimento de habilidades e algoritmos específicos de um assunto						
		3. 2 Conhecimento de técnicas específicas de assunto e métodos						
		3. 3 Conhecimento dos critérios para determinar quando utilizar processos adequados			y8, y26	7,7%		
	Metacognitivo	4. 1 O conhecimento estratégico						
		4. 2 Conhecimento sobre tarefas cognitivas, incluindo apropriado contexto e condicional conhecimento						
4. 3 O autoconhecimento								

Como foi visto na matriz de habilidades, considera-se que cada item possa avaliar mais de uma habilidade específica ou recurso cognitivo, e entende-se que isso não necessariamente implica interpretar como sendo diferentes traços latentes, pois uma habilidade mais inclusiva pode ser de certa forma ser considerada uma competência (ou parte dela) diante das necessidades em realizar uma tarefa ou

compreender determinada situação-problema dentro de um contexto específico. Isso justifica o fato de surgirem uma classificação repetida de um mesmo item dentro do quadro da análise taxonômica. O item y26 é um exemplo disso, pois apresenta uma alternativa mais inclusiva por ser uma opinião avaliativa de outros julgamentos paralelos que buscam explicar o *porquê de uma boneca de plástico não ser aquecida por um agasalho*.

3.1.4 Teste-piloto: Pré-teste e Pós-teste

O *design* do teste TCE segue a notação clássica de Campbell & Stanley (1963) e pode ser caracterizado como Projeto Pré-Experimental, mais especificamente como *Projeto de Pré-teste e Pós-teste de Um Grupo* (ou *One-Group Pre-Test-Post-Test Design*) (Creswell, 2010, p. 194). O projeto-piloto foi realizado em alunos de 10º ano em Portugal⁹¹, com pré-teste e pós-teste para as mesmas turmas.

Convém destacar que os alunos do 10º ano da «escola 1» se submeteram ao pós-teste para a mesma turma no ano seguinte, pois os 10º e 11º compreendem um ciclo em que os alunos permanecem quase sempre no mesmo grupo (mesma turma). As escolas 7, 10 e 47 aplicaram ainda no mesmo ano, com os alunos já no 11º ano. Ainda que se constate uma pequena diferença quantitativa dos alunos entre o pré-teste e o pós-teste, foi verificado na base de dados pelo gênero e data de nascimento que **correspondem aos mesmos alunos**. O fato é que o conteúdo abordado de termodinâmica se encontra previsto a ser ministrado na segunda metade do total de aulas do 10º ano, e na etapa final, os professores afirmaram ficar demasiado preocupados em cumprir seu planejamento e sobrecarregado de tarefas, e apesar da recomendação de se aplicar nos últimos dias, foi justificada a necessidade de se aplicar no início do ano seguinte, tendo em vista que *a turma continua a mesma* e o professor acompanha a finalização do ciclo 10º e 11º ano.

Esse fator gerou um risco moderado associado ao *design* (Figura 43).

⁹¹ O teste-piloto foi aplicado somente em Portugal pelo fato desse grupo ser sido admitido como grupo de referência comparativa.

Projeto de Pré-teste e Pós-teste de Grupo Único						
Portugal	Grupo 10º ano	O1	X	O2	Grupo 11º ano	
:	:	:	:	:	:	:
Escola 1	Turma A1 (25 alunos)	O1	X	O2	Turma A1 (24 alunos)	
	Turma A2 (29 alunos)	O1	X	O2	Turma A2 (28 alunos)	
Escola 7	Turma A (13 alunos)	O1	X	O2	-	
	Turma B (12 alunos)	O1	X	O2	Turma B (21 alunos)	
	Turma C (29 alunos)	O1	X	O2	Turma C (26 alunos)	
	-	O1	X	O2	Turma D (22 alunos)	
Escola 10	Turma A (23 alunos)	O1	X	O2	Turma A (19 alunos)	
	-	O1	X	O2	Turma B (24 alunos)	
Escola 47	Turma A (13 alunos)	O1	X	O2	Turma A (21 alunos)	
	Turma B (22 alunos)	O1	X	O2	Turma B (22 alunos)	
	Turma C (23 alunos)	O1	X	O2	-	
Total: 189 alunos (eliminar 36)				Total: 207 alunos (eliminar 46)		

Figura 43: Projeto de Pré-teste e Pós-teste de Grupo Único (distribuída em detalhes para cada as quatro escolas portuguesas e suas turmas).

No *Grupo 10º ano* incluem as respectivas escolas secundárias da amostra (codificadas): 1, 7, 10 e 47. A quantidade de alunos que preencheram o pré-teste e pós-teste não foram precisamente a mesma, apesar de serem as mesmas escolas, turmas e supostamente coincidindo os mesmos sujeitos, sobretudo passando pela forma de tratamento considerado (X), no caso, as intervenções de ensino sobre a termodinâmica, independentemente da metodologia, LDF adotado ou quanto a sua forma de utilização, tendo em vista que são diversas, não foram observadas, e, portanto, são variáveis intervenientes consideradas como risco no estudo. São desconsideradas as investigações relacionados à forma de tratamento (X), tendo o *design* uma função mais organizacional e sistemática na diferenciação e caracterização dos grupos normativos, que deve ser levado em consideração. Com isso, a forma de tratamento definida como critério de corte e ajuste na análise é o tempo na qual os alunos são submetidos LDF durante a abordagem de termodinâmica considera que não haja garantias que ele tenha utilizado, e não faz parte o foco da pesquisa.

A ordem temporal estabelecida de visita nas escolas foi de 4 meses, entretanto o tempo no desenho está relacionado a submissão na forma de tratamento⁹², que é cerca de 6 meses entre o pré-teste e o pós-teste. O controle para esta variável é dado

⁹² É irrelevante neste trabalho estender detalhes relacionados à forma de tratamento (X), tendo o desenho uma função mais ilustrativa sobre as necessidades de coleta de dados nas escolas. Entretanto a forma de tratamento definida como critério de corte e ajuste na análise é o tempo na qual os alunos são submetidos a livro didático, levando em consideração que não há garantias que ele tenha utilizado.

pelo indicativo de data de aplicação do teste no próprio instrumento, e recomendações posteriores no retorno de visita às escolas, pois a extensão no tempo entre o pré-teste e o pós-teste remete a perda na verificação dos “efeitos”. O instrumento segue com instruções aos alunos, informando os procedimentos de realização, estimativa de tempo (ver em anexo). Os professores também receberam uma folha com instruções técnicas de aplicação e estimativa do tempo de preenchimento pelos alunos.

Como esta investigação é caracterizada como um estudo transversal e não longitudinal este é um fator de risco iminente a ser considerado para o *design*, com informações mais precisas dentro do *status quo* no contexto histórico da observação.

O *design* dessa etapa quantitativa implica riscos associados a validade interna e externa. Pode-se dizer que a validade interna se define “*como o grau de correspondência entre as observações e medições científicas e a realidade estudada. Ou seja, tem a ver com a possibilidade de excluir conclusões alternativas, através do «controle» de variáveis estranhas*” (Vieira, n.d.).

Quanto à **validade interna**, de 15 riscos analisados, 5 são admitidos para o *design* do Estudo Piloto, que são: *regressão estatística; seleção diferencial; testing (prática do teste ou testagem); interações com a seleção (seleção/instrumentação); e ambiguidade da direção da causa.*

Quanto à **validade externa**, foram analisados 14 riscos pertinentes, sendo 5 admitidos para o *design* do Estudo Piloto, que são: *generalização da população acessível (amostra) para a população-alvo; interação entre variáveis psicológicas (relacionado a personalidade) e o tratamento; interferência de múltiplos tratamentos; interação da história e do tratamento; interação tempo de medição e tratamento.*

A partir de um *checklist*, foi realizada uma análise descritiva de cada um dos riscos associados a validade interna e externa (ver no Apêndice C, p. 227).

3.1.4.1 Quanto a pertinência temática curricular

Em uma perspectiva de *alinhamento curricular vertical*⁹³, de acordo com Norman Lott Webb (1997, 1997, 1999, 2002, 2006), considera-se que a aplicação do teste TCE poderá servir na identificação preliminar ou diagnóstica de um perfil dos alunos sobre suas concepções alternativas diante de alguns conceitos térmicos, e assim, verificar sua adequação/integração diante dos objetivos curriculares, ou caminhos para isso, seja diante dos *recursos didáticos disponíveis* e/ou dos processos pedagógicos existentes. Diante disso, recomenda-se que os conteúdos nos testes possam corresponder a cerca de 50% da unidade que se pretende avaliar, e caso se pretenda um alinhamento, deve ter como base os programas curriculares. Entretanto, essa é uma recomendação quando se tem a intenção de avaliar o conteúdo de um curso ou um ciclo de aprendizagem de forma completa, que não é o caso do Inventário de Conceitos que pretende avaliar geralmente conceitos básicos e introdutórios (conteúdo mais restringido e específico), porém é importante verificar se há uma pertinência com o programa curricular, e dentro disso, buscar garantir o aprendizado nas competências básicas é um fator que favorece o alinhamento do ponto de vista do conteúdo, mas também quanto aos ajustes nos métodos de ensino e processos de aprendizagem.

A temática que se refere o teste TCE é pertinente por estar inserida no programa curricular em ambos os contextos, pois se refere a conceitos específicos indissociáveis ao conhecimento térmico. A abrangência dos aspectos termodinâmicos remete a necessidade no domínio dos alunos tanto na aquisição conceitual que envolve os conhecimentos sobre calor e temperatura, como no desenvolvimento de competências científicas, em que se ampliam em outros domínios do conhecimento dentro das atividades experimentais (quando ocorrem!). Considera-se necessário verificar essas respectivas aquisições devido a sua importância como subsídio as instruções no ensino-aprendizagem, sobretudo, avaliar as concepções mais enfatizadas e as diferenças existentes entre os grupos normativos, possibilitando evidenciar

⁹³ Condição necessária para que dos processos educacionais estejam pertinentes com os objetivos e finalidades preconizados nos documentos formais reguladores dentro do sistema educacional, ou seja, legítimos dentro do *status quo*.

informações diagnósticas relativamente consistentes com as eventuais dificuldades existentes devido a abrangência da amostra.

"Estas, bem como a experiência que se propõe para a realização no laboratório, deverão permitir, além do desenvolvimento de capacidades científicas dos alunos, a consolidação das aprendizagens que a seguir se especificam (ver Objectos de Ensino e Objectivos de Aprendizagem) bem como o aprofundamento dos **conceitos de calor, temperatura** e energia interna que serão aplicados na Unidade 1." (Simões, Simões, Caldeira, Bello, & Pina, 2005, p. 57)

Como foi dito, é bem conhecida no campo de pesquisa do ensino da Física ser um obstáculo comum entre os alunos de nível secundário (e também para alunos ingressantes do ensino superior!) a dificuldade na distinção entre os conceitos de calor e temperatura. Esta necessidade pode ser transversal a outros campos do conhecimento por esses conceitos estarem associados na própria natureza da matéria, de sua constituição e comportamento. Este é um sinal que destaca a necessidade de uma aprendizagem que seja significativamente satisfatória para alcançar uma compreensão mínima e adequada diante dos aspectos fenomenológicos que envolvem esses conceitos, caso contrário se tem o risco de edificar um conhecimento sem base estruturante de compreensibilidade, do tipo a ser capaz de solucionar um problema, obter a resposta correta, mas não compreender o sentido físico das grandezas, de suas propriedades e comportamento da natureza, e assim, ao comprometer o desenvolvimento de competências científicas básicas, se tornar um obstáculo ou até ilusões sobre o seu desempenho em competências mais abstratas dentro da *Termodinâmica*.

No âmbito da aplicação do TCE aos alunos portugueses (ano de 2014) estava em vigor o Programa de Físico-Química A, versão 2005 (ver no Apêndice D1.1, p. 241), na qual são sugeridas **3 unidades temáticas** que abrangem a *Termodinâmica*, com 43 ações distribuídas a serem desenvolvidas dentro de cada unidade. Muitas dessas ações *não são necessariamente mutuamente exclusivas* dentro de cada unidade temática. Isso significa dizer que algumas ações ou habilidades desenvolvidas em uma determinada atividade podem ocorrer ou serem exercidas em outras atividades nas demais unidades temáticas. Por exemplo, "*distinguir calor, temperatura e energia interna*" é uma capacidade básica que se transfere a quase todas as atividades. Também existem

aquelas que coincidem, mas que só mudam do contexto das situações, como “*estabelecer balanços energéticos em sistemas termodinâmicos*” que está presente na unidade “A energia no aquecimento/ arrefecimento de sistemas” ou na própria atividade laboratorial “AL 1.4 – Balanço energético num sistema termodinâmico”. Outra capacidade coincidente é de “*analisar transferências e transformações de energia em sistemas*” presentes na “AL 1.1 – Absorção e emissão de radiação” e “AL 1.3 – Capacidade térmica mássica”, na qual muda-se apenas o contexto ou fenômeno para que se aplique e exerça a mesma ação.

Não considerando a repetição dessas capacidades, elas poderiam ser resumidas em 37 ações. Dentro das 43 apresentadas, existem 18 delas (aproximadamente 50%, sendo aproximadamente suficientes para um *alinhamento curricular*, caso os itens associados no TCE sejam considerados válidos e úteis para isso) que de forma direta ou indireta são exploradas no TCE, que são as números (ver no Apêndice D1.1, p. 241): 1, 2, 9, 14, 15, 20, 21, 22, 23, 24, 27, 36, 37, 39, 40, 41, 42, 43. A única atividade que não apresenta de forma mais consistente uma correspondência é a “AL 1.2 - Energia eléctrica fornecida por um painel”, pois o teste não apresenta item dentro desse contexto.

Após as instruções, é natural considerar que haja uma ampliação dos recursos cognitivos para resolver os itens do TCE, diante disso, pode-se afirmar que algumas das ações sugeridas sejam pertinentes somente para os alunos no pós-teste, como as ações 42 e 43, que se referem a “*Associar o valor, positivo ou negativo, da quantidade de energia envolvida na mudança de estado físico, às situações em que o sistema recebe energia ou transfere energia para as vizinhanças, respectivamente*” e “*Estabelecer um balanço energético, aplicando a Lei da Conservação da Energia*”.

No programa curricular mais recente, versão 2014 (ver no Apêndice D1.2, p. 242), são apresentados uma unidade temática para a Termodinâmica denominada “*Energia, fenómenos térmicos e radiação*”, juntamente com 3 atividades laboratoriais, sendo um total de 37 ações a serem desenvolvidas. Considera-se 12 delas consideradas mais pertinentes ao TCE, que são: 1, 2, 3, 4, 7, 16, 17, 21, 29, 30, 34, 36. Aparentemente,

esse programa curricular de Físico-Química A demonstra ser um modelo mais *genérico* e voltado para o desenvolvimento de capacidades, do que *operacional*, que seria centrado em competências.

3.1.4.2 Resultados do Teste-Piloto

3.1.4.2.1 Ganhos normalizados e tamanho do efeito

Através dos percentuais do pré e pós-teste é possível verificar se houve ganhos percentuais (“ganho normalizado” e “tamanho do efeito”) no acerto e na diminuição de equívocos entre os distratores.

Utilizando as equações apresentadas anteriormente, foi obtido um “tamanho do efeito” de Cohen de **0,67**, considerado entre **médio e alto**, mais precisamente esse tamanho do efeito indica que a média do grupo tratado (pós-teste) está entre o percentil 73 e 76 do grupo não tratado (pré-teste) (Becker, 2000). Também pode se estimar uma diferença entre 38,2% e 43,0% para o grupo sobreposto diante do não sobreposto, que seria o pós-teste sobre o pré-teste (Ver Cohen, 1988, pp. 21–23).

Tabela 21: Média e desvio-padrão amostral do pré-teste e pós-teste com base nos valores dos escores brutos.

Observação	n	X_i	S	$\%X_i$	$\%S$
Pré-Teste	153	8,76	$\pm 3,01$	0,43	$\pm 0,12$
Pós-teste	161	11,09	$\pm 3,89$	0,34	$\pm 0,15$

Considerando $\% \langle G \rangle = 0,09$ e $\% \langle G_{m\acute{a}x} \rangle = 0,16$, o fator de Hake sugere um ganho normalizado “baixo” ($\langle g \rangle < 0,30$) para as 4 escolas, sendo obtido um ganho de **$0,14 \pm 0,11$** . Para os diferentes grupos normativos envolvidos, foram obtidos os seguintes tamanhos de efeito e ganhos normalizados:

Tabela 22: Percentuais no pré-teste, ganhos percentuais dos escores brutos, tamanho do efeito, ganhos normalizados e os livros didáticos de Física correspondentes que foram adotados nas escolas.

Escolas	$\% \langle \text{pré} \rangle$	$\% \langle G \rangle$	“d” de Cohen	$\langle g \rangle$ de Hake	Livro Didático de Física Adotado
E1	0,30	0,09	0,67	0,13	LDF_9
E7	0,40	0,07	0,62	0,12	LDF_9
E10	0,30	0,06	0,42	0,08	LDF_9
E47	0,33	0,12	0,89	0,19	LDF_11
4 escolas	0,34	0,09	0,67	0,14	-

A relação de causalidade relacionado com a instrução, no que se refere às aulas de termodinâmica do 10º ano nas escolas portuguesas, não é considerado a única causalidade desse efeito, mas um indício de que houve ganho, sobretudo de que existem concepções térmicas que resistem a essas instruções, e, portanto, merecem maior atenção. Foi acrescentado no quadro o Livro Didático de Física adotado em cada uma das escolas como parâmetro de comparação, indício de fator de influência no tratamento, mas sem, entretanto, buscar caracterizar como fator de causalidade linear, apesar de que o LDF_11 corresponder 86% das escolas secundárias da *Região Centro de Portugal*, e a 50% das escolas na amostra, ou seja 10 escolas. Também deve ser considerada a possibilidade de uma má formulação nos itens, como foi apontada na avaliação da chave-de-correção.

Seguindo o gráfico clássico e recomendado na literatura para a realização das interpretações de ganhos percentuais e normalizados dos **escores** brutos (considera-se a soma dos acertos de todos os itens, no caso, os 26 itens), de forma correspondente a forma como foi apresentado no trabalho original (Hake, 1998b), entretanto, foi propositalmente enriquecido sendo introduzido com informações dos ganhos dos **percentuais de acerto** (considera-se a soma dos acertos de um item para todos os sujeitos) dos itens na intenção de facilitar a interpretação de modo geral dos ganhos médios entre as 4 escolas, apesar de que necessariamente estejam inclusos os ganhos dos grupos de normatização considerados no estudo. Também foram colocadas duas linhas verticais nas zonas de “*guessing*” do pré-teste para 20%, e 25% somente os itens y9, y11, y16, y18 e y26. Para uma melhor leitura, vale ressaltar ainda que $\langle G \rangle = \langle \text{pós} \rangle - \langle \text{pré} \rangle$ e $\langle G_{\text{máx}} \rangle = 1 - \langle \text{pré} \rangle$, em que o gráfico interno com eixo vertical secundário relacionado ao ganho normalizado se refere a um ângulo ϕ partindo de um ganho máximo, ou seja, supondo um pré-teste com percentual nulo e pós-teste igual a 100%, dessa forma, as retas de atividades representam as médias dos ganhos normalizados. Para as 4 escolas é a «média» das médias dos ganhos normalizados dos itens foi de $\langle \langle g \rangle \rangle = 0,12$, e o ganho normalizado constatado para o teste $\langle g \rangle = \phi = 0,14$.

O item y14 se destaca por apresentar inicialmente um baixo percentual (< 20%, limite de corte do “guessing” do item), porém dar um salto de mais de 30%, isso pode indicar um certo reforço das instruções nos conhecimentos envolvidos nesse item, no caso, sobre a compreensão do comportamento térmico de acordo com as propriedades térmicas (relacionado ao descritor 16). Especificamente remete a capacidade cognitiva do aluno inferir com base na *Lei Zero da Termodinâmica*, considerando que em contato físico com uma cadeira de metal e uma de plástico, compreende-se que as cadeiras estão à mesma temperatura e não devem ser avaliadas com base na sensação térmica. Os itens y10, y13 e y14 também sugerem banhos de médio impacto (conferir os conhecimentos envolvidos desses itens no Apêndice A10, p. 81, A13, p. 105, A14, p. 113).

Para o índice de dificuldade (ID) ou percentual de acerto **do item** (e não do teste⁹⁴), o ganho do percentual bruto vai no sentido favorável do resultado esperado, que seria um domínio sobre um conhecimento específico. Para os distratores, o ganho percentual não é bom, vai no sentido contrário daquilo que se deseja, pois demonstra resistência em equívocos e possíveis padrões de respostas que se caracterizam como modelos mentais no campo do conhecimento físico.

⁹⁴ Hake sugere a média da soma dos escores brutos, e sabe-se que é diferente da soma das médias dos acertos nos itens, apesar de ser um valor aproximado.

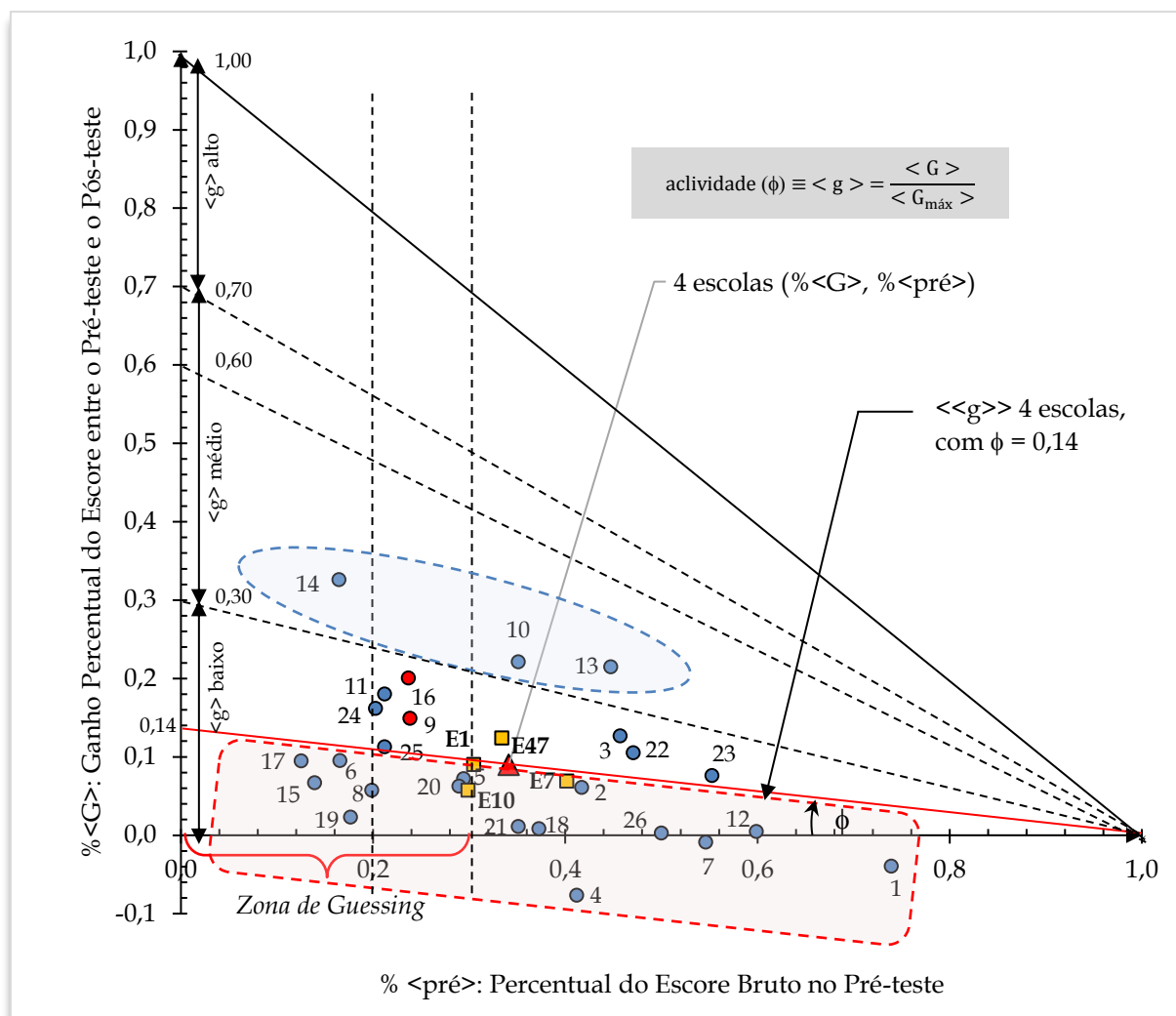


Figura 44: Gráfico %<G> do TCE versus %<pré> no contexto do teste-piloto em relação aos itens, as médias de cada escola, e da média entre as quatro escolas secundárias portuguesas.

Em termos **mínimos de classificação**, tendo o “guessing” dos itens como referência de corte, **12 itens se encontram previamente em uma zona de puro acerto aleatório** (y6, y8, y9, y11, y14, y15, y16, y17, y19, y24), portanto possuem maior necessidade de ganhos após as instruções, enquanto os 14 itens restantes possuem percentual de acerto digamos aceitáveis preliminarmente.

Com o pós-teste, em termos **ideais de classificação** (critério ideal de acerto), **2 itens conseguiram se manter nesse patamar** no pós-teste (y1, y12), apesar de não obterem ganhos após a instrução evidenciando a preexistência de *subsunçores* que favorecem o desenvolvimento de um domínio conceitual diante dos conhecimentos intuitivos da Física Térmica que os envolvem, porém, é importante destacar da necessidade de relativizar os itens que possuem conhecimentos relacionados. O fato

de não haver ganhos pode indicar a ausência na abordagem de conteúdos específicos (incompletude na apresentação e desenvolvimento de alguns conteúdos pelo professor), bem como um certa resistência, como no caso do item y1, que retrocedeu, ao invés de ter obtido ganhos. Dos 12 que precisam sair da zona de *guessing*, somente 4 deles obtiveram ganhos que pudessem sair intensamente (y9, y11, y14 e y16), portanto, os outros itens (y6, y8, y15, y17, y19, y24) devem ter atenção quanto:

- As possíveis dificuldades de aprendizagem em relação ao conteúdo que se exige, que podem ser de natureza epistemológica, conceitual, e/ou relacionadas a concepções que se formam na estrutura cognitiva dos sujeitos diante desses conceitos;
- As tipologias de equívocos que sugerem os distratores, assim como as respectivas concepções envolvidas que são remetidas aos sujeitos;
- A textualização do enunciado ou ainda dos distratores, que no caso, podem haver necessidades de reelaboração, aprimoramento ou descarte;
- As ineficiências nas abordagens metodológicas de ensino diante dos conhecimentos que envolve o teste, considerando a falta de cobertura ou incompletude entre os conteúdos abordados;
- Ao mal uso (ou ausência) de recursos didático-pedagógicos, considerando ainda uma possível má qualidade científico-pedagógica dos recursos adotados e utilizados como apoio;
- As questões relacionadas aos hábitos de aprendizagens dos alunos;
- Aos fatores subjetivos que extrapolam do âmbito educacional para o social, econômico, psicológico, etc.

Com o pós-teste é possível constatar um fluxo de itens diante das mudanças nos percentuais, e naturalmente, as concepções que envolvem os distratores acompanham esse fluxo, diante disso se considera importante acompanhar essas mudanças e buscar investigar sobre uma associação das possíveis causas, devendo serem verificadas por outros métodos de recolha de dados e procedimentos de análises. Sobretudo, se torna mais relevante do ponto de vista pragmático, identificar os itens que permanecem com fortes atrações pelos distratores, mesmo após as aulas de Termodinâmica. Porém, é importante verificar se há uma relação de conjuntura lógica e conceitual quanto as

concepções térmicas que envolvem esses distratores, pois tornam-se pedagogicamente importantes a serem enfrentadas, e eventualmente, superadas.

Em outra perspectiva, o gráfico do ganho percentual bruto ($\langle G \rangle$) do TCE *versus* a porcentagem dos escores brutos no pré-teste ($\langle \text{Pré} \rangle$) mostra que as médias das 4 escolas envolvidas ficaram na zona de baixo impacto (<math>\langle g \rangle < 0,30</math>), sendo a escola secundária E47 a única ganho acima da média entre elas. Destaca-se que na região sombreada da parte inferior existem itens com ausência de ganhos após as instruções das aulas de termodinâmica, e 3 itens sendo regredidos percentualmente (y1, y4, y7), apesar de estarem com percentual no pré-teste acima do “guessing”, e o item y1 estar com um índice de dificuldade considerado ideal para um item com 4 alternativas (ID > 60%).

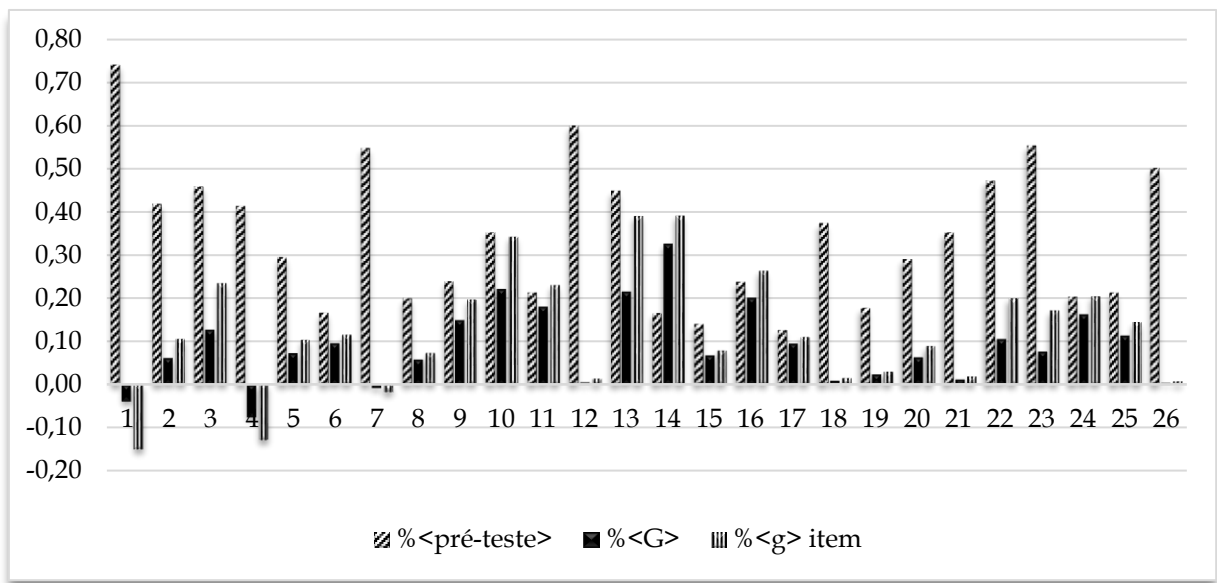


Figura 45: Histograma do índice de dificuldade no pré-teste, ganhos percentuais brutos no pós-teste e ganhos normatizados dos itens.

Isso também sugere uma resistência dos distratores envolvidos para esses itens, assim como o fato de que suas as crenças equivocadas foram reforçadas após as aulas, e isso é inaceitável principalmente para aqueles que permaneceram com baixo percentual de acerto, porém, recomenda-se que informação seja investigada mais profundamente em um estudo complementar, pois as aulas de termodinâmica ministradas nas diferentes escolas não demonstraram terem sido eficientes no sentido produzirem maiores efeitos de *mudança conceitual* nos alunos diante dos conceitos

introdutórios da Física Térmica, podendo ser um fator a comprometer o desenvolvimento de competências científicas de base conceitual.

Considerando a força atrativa de cada item, é possível realizar uma especulação prévia das concepções térmicas (equivocos) associadas a cada distrator. A análise posterior relacionada ao fator de concentração e desvio na concentração de respostas poderá identificar com maior precisão a força atrativa das respostas corretas e dos distratores, sugerindo se há uma relação com um ou mais modelos mentais para cada item, seja correto ou incorreto, bem como se existe uma indeterminação devido a uma distribuição percentual relativamente homogênea.

Item	Distrator	Concepção
y1	b	D6
y2	c	D7
y3	c	C1
y4	a, c	C1; D5
y5	c	B4; C1
y6	c	C6; D8
y7	b	A5
y8	b	D5
y9*	d	C6; D1
y10	a	A1; C4
y11*	e	D7
y12	a	C6
y13	a	C5
y14	d	D2
y15	c	A4; A5; B1
y16*	d	D2
y17	c	C6; D9
y18*	b	A3; B3
y19	d	B5; D5
y20	a	C2; C3
y21	b	B3; C4; C9
y22	d	B3; B6; D2
y23	d	A3; D9
y24	c	A3; D1
y25	d	D3
y26*	c	B6

Figura 46: Especulação prévia dos distratores em destaque e das concepções correspondentes associadas a cada item, sendo, no caso, indícios preliminares de concepções térmicas resistentes após as instruções.

*Itens com 5 alternativas.

O critério de consistência das informações dos itens, na qual caracterizam um perfil de concepções dos sujeitos deve ser reanalisado e comparado na etapa posterior e mais ampla, sendo reforçado através dos índices de discriminação.

Não há um compromisso com a causalidade, podendo estar ou não na instrução, com também na qualidade do item, condições do contexto em que foi respondido ou entre fatores intrapsicológicos dos respondentes, apesar disso esse é um indicador de alerta para que seja verificado se há uma confirmação nas próximas etapas da investigação.

3.1.4.2.2 Fator de Concentração «C» e Desvio na Concentração das Análises «Γ»

Grande parte dos estudos sobre as dificuldades conceituais dos alunos exploram suas falas, associando-as como percepções diante de conceitos e situações-problemas em diversos contextos da Física, quase sempre como formas de representação de seus modelos mentais (ou parte disso). Em abordagens quantitativas, como no caso de análises com base em testes de múltiplas escolhas com resposta única, pode-se admitir que os distratores tendem a se posicionar em determinados modelos mentais que representam concepções equivocadas. Respostas concentradas nesses distratores podem significar uma projeção de determinado perfil conceptual entre os grupos de estudantes (Bao & Redish, 2001). Obviamente, alunos com pouco conhecimento diante das situações propostas podem reagir como se não estivessem nenhum dos modelos, ou como se escolhessem entre uma diversidade de modelos diferentes, com isso, suas respostas estariam próximas de uma distribuição aleatória entre todas as escolhas. No caso do TCE, considera-se que a distribuição e a concentração das respostas dos alunos possam fornecer informações sobre o estado dos alunos diante de suas convicções e modelos alternativos.

Diante das respostas do pré-teste e pós-teste se torna possível apresentar os gráficos *S-C plot* para cada situação e visualizar os prováveis modelos mentais representativos para os alunos, de modo geral entre as 4 escolas, levando em consideração que o fator de Hake sinaliza que as mesmas se encontram classificadas dentro de um mesmo critério no intervalo de ganhos, que no caso é baixo, sendo mais desfavorável para a escola E10. Todas as respostas ficam restringidas entre as linhas de fronteira das curvas de concentração mínima e máxima, conforme Bao & Redish (2001), lembrando que a região aleatória ou zona de “*guessing*” para os itens y9, y11, y16, y18 e y26 é em 20% e o restante 25%. Ao considerar esse critério, o limite da região MM pode ser expandida incluindo o item y10.

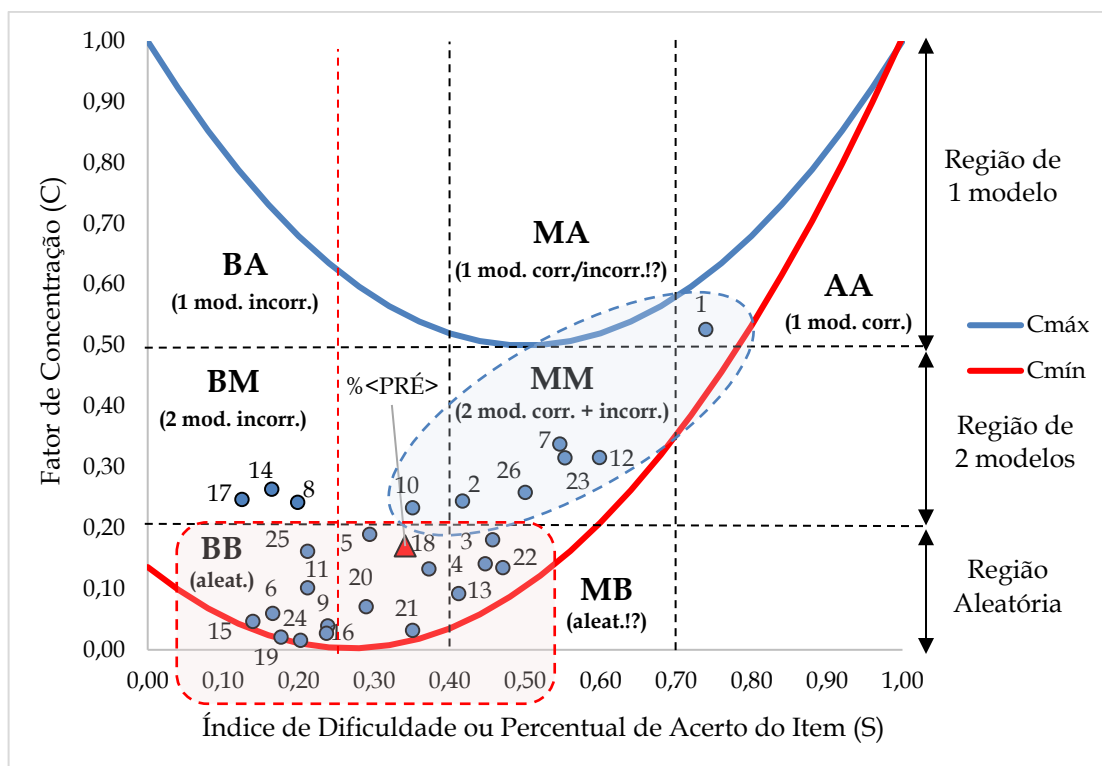


Figura 47: Gráfico S-C plot para o pré-teste.

Observa-se que a maioria dos itens do pré-teste concentram-se em uma região de escolha aleatória, estando alguns itens sugerindo uma adesão a dois modelos mentais, em que um deles é o correto, como mostra a região MM (y2, y7, y, 12, y23, y26), sendo apenas um item considerado com pleno domínio de um modelo correto na região AA (y1). A *concentração média* das 4 escolas no pré-teste encontra-se na zona BB, ou seja, de baixa porcentagem de acertos e baixa concentração na alternativa correta.

Para perceber os indicadores para as maiores necessidades diante das dificuldades dos alunos, observam-se os resultados do pós-teste no gráfico S-C plot. Espera-se uma observação de um deslocamento diagonal, na qual se verifica em grande parte dos itens, apesar de alguns permanecerem na região BB, e se deslocarem mais a direita do que para o aumento de concentração, na qual representa o *vetor de mudança da concentração média* (Bao & Redish, 2001) entre os grupos. Isso indica que sejam itens com fortes resistências a mudanças de modelos, porém, como é uma região mais representativa pela aleatoriedade, convêm valorizar os itens na região MM, de

signalizam de forma mais concreta e significativa os modelos corretos e incorretos dos estudantes.

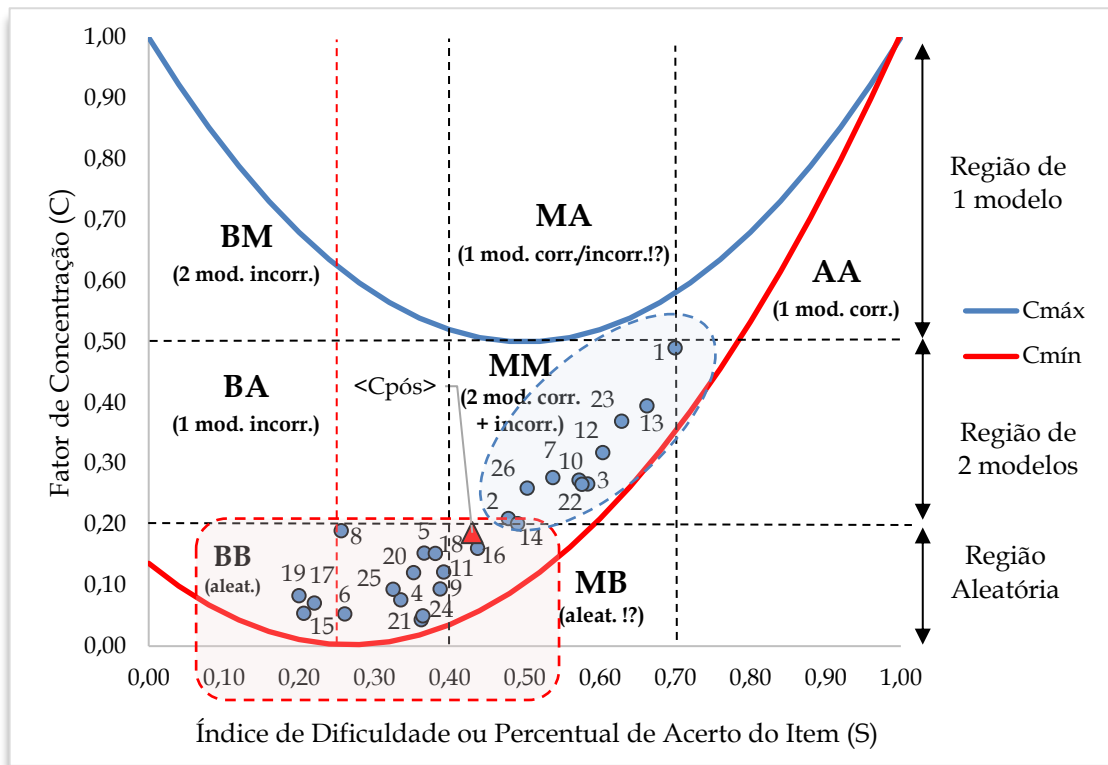


Figura 48: Gráfico S-C plot para o pós-teste.

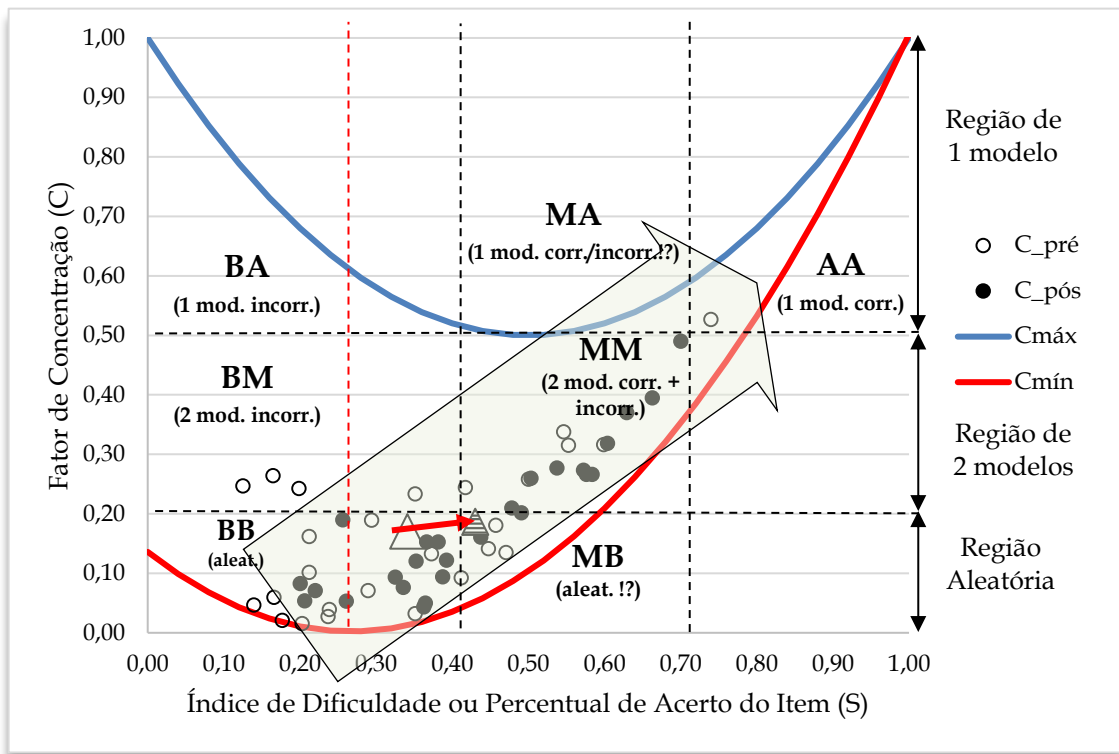


Figura 49: Gráfico S-C plot para o pré-teste e pós-teste.

No pós-teste a *concentração média* está região MB, porém surge a possibilidade de identificação mais modelos corretos devido o deslocamento de mais itens para a região MM, sendo, portanto, a região mais representativa para identificar um padrão sobre as concepções dos estudantes. No gráfico com os resultados do pré e pós-teste, o *vetor de mudança da concentração média* revela um sutil deslocamento na diagonal, representando o aumento na concentração das alternativas corretas e os itens se tornando mais fáceis para os respondentes.

Uma mudança e algumas persistências de modelos dentro das concepções térmicas podem ser visualizadas no quadro. Na análise, considera-se no pré-teste que a região BM (dificuldade baixa e concentração média) para dois modelos incorretos, o modelo incorreto dentro da margem de “*guessing*” deve ser desconsiderado, estando destacado (em negrito) assim apenas um modelo incorreto predominante.

Quadro 64: Identificação dos modelos mentais (corretos e/ou incorretos) em evidência no percentual das respostas, de acordo com a análise do Fator de Concentração (C).

Discrepância	Tipo (S-C)	Implicações dos padrões	Itens/mod. (Pré)	Itens/mod. (Pós)
Nula	I	AA um modelo correto	y1a	-
Alta	I	BA um modelo incorreto	-	-
Média	II	BM dois* possível modelos incorretos	y8 (b , c), y10 (a , d), y14 (a , d), y17 (b , c)	-
Média	II	MM dois** modelos populares (correto e incorreto)	y2 (b , c), y7 (c , a), y12 (c , a), y23 (c , d), y26 (e , c)	y1 (a , b), y2 (b , c), y3 (b , c), y7 (c , a), y10 (c , a), y12 (c , a), y13 (d , a), y14 (b , d), y22 (a , d), y23 (c , d), y26 (e , c)
Sem leitura	-	BB situação quase aleatória	y5, y6, y9, y11, y15, y16, y17, y18, y20, y21, y24, y25	y4, y5, y6, y8, y9, y11, y15, y17, y18, y19, y20, y21, y24, y25
Sem leitura	-	MB situação quase aleatória (!?)	y3, y4, y13, y22	y16

*Considera-se mais relevante o modelo em negrito, devido as taxas mais representativas das porcentagens.

**Considera-se o modelo correto em negrito (primeira letra dentro do parênteses).

Dessa forma, pode ser elaborado um perfil preliminar sobre as convicções do conhecimento térmico para os estudantes nas 4 escolas no teste-piloto. Os itens são apresentados em ordem decrescente quanto ao percentual e fator de concentração nos

acertos ($\%C_{p\acute{o}s}$), assim como o coeficiente de correlação ponto-bisserial (r_{pbs}), que sugere os itens com maior consistência.

Quadro 65: Informações descritivas das convicções científicas dos alunos no TCE em destaque, de acordo os resultados da análise quantitativa do pós-teste (em ordem decrescente, segundo o fator de concentração).

Item	Alt.	$\%C_{p\acute{o}s}$	r_{pbs}	Descritores	Convicção Científica/Descrição da Proficiência
y1	a	49,0	0,16	1; 2	Prever que a «temperatura do gelo» «dentro do congelador de uma geladeira doméstica» «após de um longo tempo está abaixo de 0 °C».
y13	d	39,5	0,41	9	Inferir sobre o sentido de «transferência de energia térmica que ocorre durante o processo de resfriamento de ovos cozidos quando submetidos em água fria»
y23	c	37,0	0,50	18	Inferir que a «sensação térmica de frio da nossa pele pode ser evitada dificultando o processo de trocas de calor com o meio externo, e assim reter a energia térmica que se emite continuamente pelo corpo utilizando um agasalho como isolante térmico»
y12	c	31,8	0,11	7	Inferir por meio da percepção visual que «as bolhas formadas em meio a água fervente representam na maior parte vapor d'água».
y3	b	26,6	0,37	5	Prever que a «temperatura do gelo em equilíbrio térmico» «com a água que se forma com seu derretimento» «permanece a 0 °C, mesmo após cessar o processo de fusão».
y10	c	27,3	0,48	12	Inferir sobre o sentido de «transferência de energia térmica que ocorre durante o processo de trocas de calor entre diferentes corpos sólidos com base na sensação térmica»
y22	a	26,6	0,42	15	Inferir que a «sensação térmica de “quente”» em uma «bomba durante enchimento de um pneu é causada pelo recebimento de energia térmica que advém do interior da bomba».
y7	c	27,7	0,32	5	Prever que a «temperatura de equilíbrio mais provável na mistura de corpos idênticos», «porém com massas diferentes, a temperatura final tende a ser mais próxima da temperatura inicial do corpo com maior massa»
y26	e	26,0	0,49	18; 19	Inferir que «isolantes térmicos não aquecem corpos, apenas dificultam eventuais trocas de calor, sendo eficientes em reter a energia térmica caso o corpo coberto seja um emissor desse tipo de energia em potencial»
y14	b	20,2	0,53	16	Inferir uma «igualdade de temperatura entre corpos com diferentes propriedades com base na <i>Lei Zero da Termodinâmica</i> », «no caso, entrando em contato físico com uma cadeira de metal e uma de plástico, considerando uma independência da sensação térmica, portanto, as cadeiras estão à mesma temperatura e não devem ser avaliadas com base na sensação térmica».
y2	b	21,0	0,29	1; 5	Prever que a «temperatura da água em equilíbrio térmico» «que se forma a partir do derretimento do gelo» «permanece igual a 0 °C, mesmo após cessar o processo de fusão».

Analisando as respostas incorretas, tem-se uma perspectiva quanto as concepções ou equívocos que os alunos aderiram, com isso são sugeridos modelos mentais em destaque na distribuição média das respostas para os grupos, que podem sinalizar dificuldades e necessidades de reforço para as instruções de ensino. Espera-se que o *desvio na concentração de análise* (Γ) possa mostrar uma diminuição na atração de distratores, bem como como uma elevada porcentagem de acerto no gráfico *S- Γ plot*, ou seja, o surgimento de um *vetor desvio na concentração de análise* sugerindo um deslocamento diagonal dos itens no sentido do canto inferior direito. Existem três princípios mais gerais que são recomendados na avaliação de análise de itens (Murphy & Davidshofer, 2005, p. 220, tradução livre): (1) As respostas incorretas devem ser uniformemente distribuídas entre distratores; (2) Os valores de dificuldade do item devem se agrupar em torno de 0,50; (3) As correlações item-total (ponto-biserial) do item devem ser positivas.

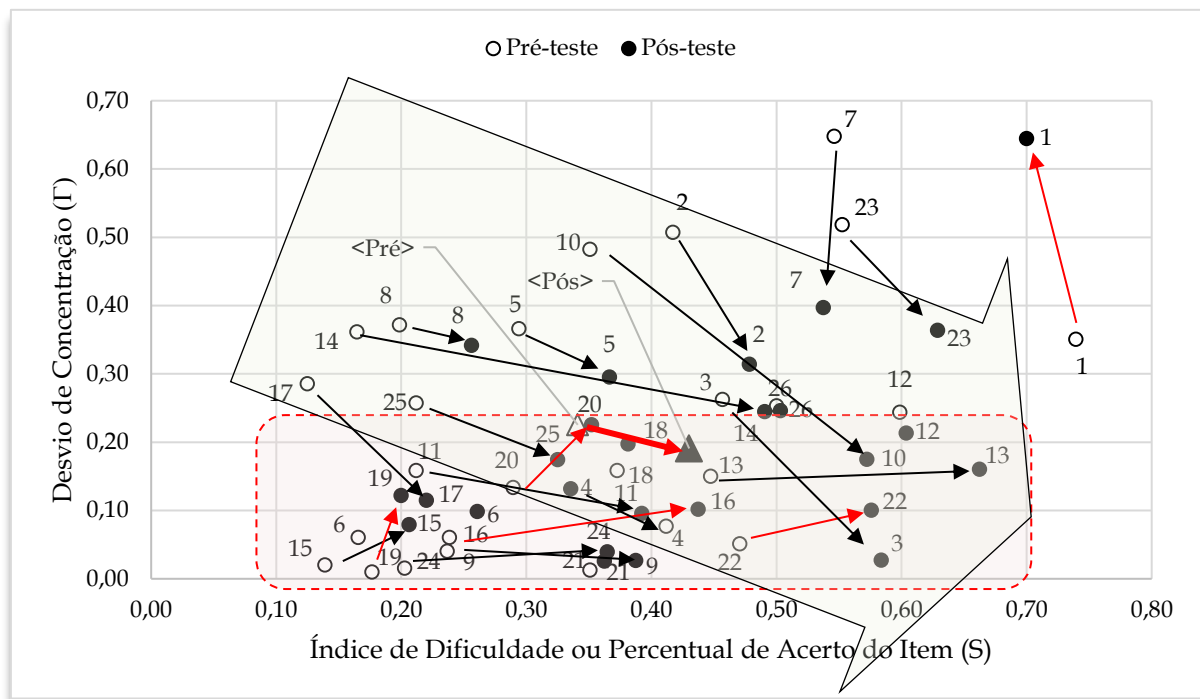


Figura 50: Gráfico *S- Γ plot* para os Grupos do Pré-teste e Pós-teste.

É possível notar que a maioria dos itens se encontra em uma região com um baixo índice no fator de concentração das respostas, isso sugere que muitas delas possam estar sendo escolhidas de forma aleatória, e que a distribuição entre os distratores não destacam veementemente um modelo mental alternativo. O item y1,

apesar de possuir um percentual relativamente elevado comparado aos outros itens, mostra uma resistência no distrator “b”, pois a concentração aumentou, ou seja, os alunos possuem dificuldades em admitir ou compreender que o “gelo pode estar a uma temperatura abaixo de 0 °C”, mesmo após as aulas de termodinâmica. O distrator “d” do item y22, apesar de ser uma concentração elevada, também demonstra resistência pelos alunos, em que se acredita que o “metal provoca aquecimento de uma bomba durante o enchimento de um pneu”.

Alguns itens não demonstraram mudança significativa, nem de ganho de acerto ou diminuição na concentração de respostas incorretas (y8, y12, y18, y21, y26), e muitos que se encontram próximo de uma concentração igual a zero e com um baixo percentual de acerto, demonstram itens confusos, na qual as respostas tendem a se distribuir homoganeamente, sem grandes picos. Se destaca na diminuição da concentração dos distratores no pré-teste os itens y3, y7, y10 e o y23, o restante não demonstraram mudanças significativa em suas perspectivas, apesar receberem as instruções das aulas.

De modo geral, a média do pré-teste se desloca rumo a média do pós-teste no **sentido diagonal a direita**, aumentando o percentual de acerto em 9% e diminuindo o desvio na concentração em 3,5%, isso representa um ganho médio muito baixo, corroborando a análise do Fator de Hake com $\langle g \rangle < 0,30$ (baixo impacto).

Tabela 23: Informações descritivas aos distratores e as concepções térmicas que encontram-se mais evidentes entre os alunos no TCE, de acordo os resultados da análise quantitativa entre o pré-teste e o pós-teste (em ordem crescente dos itens).

Item	Distrator atrativo	% $\Gamma_{\text{pré}}$	% $\Gamma_{\text{pós}}$	Modelos mentais associados as concepções térmicas (Equívocos em destaque)
y1	b	35,1	64,5	D6.
y2	c	50,7	31,4	D7; C6.
y3	c	26,3	2,7	C1.
y7	a	64,8	39,7	<i>*nas trocas de calor o corpo frio é mais difícil de ser aquecido. (a temperatura de equilíbrio tende de estar mais próxima ao corpo mais frio).</i>
y10	a	48,2	17,5	A3; A1; C4.
y12	a	24,4	21,4	D12
y13	a	15,0	16,1	C5
y14	d	36,2	24,5	D2
y22	d	5,1	10,1	B3; B6; D2.
y23	d	51,9	36,4	<i>*confusão sobre o conceito de calor: A3 e D9.</i>

3.1.4.2.3 Considerações na análise da teoria da resposta ao item

De acordo com a TRI, as análises de pré-teste e pós-testes obtiveram diferenças significativas ao se considerar a comparação dos grupos como independentes, $t(306) = -5,214$, $p < 0,001$, porém ressalta-se que o teste estatístico não é apropriado para inferir a significância a nível populacional uma vez que a amostra consiste em grupos dependentes (uma mesma pessoa avaliada em duas situações). Apesar de não ter sido possível garantir a correspondência entre os participantes no pré e pós-teste, na qual possibilitaria calcular os ganhos de acordo com a *mudança normalizada*, considera-se como forma alternativa, fiável e elegante uma visualização do ganho de proficiência através das análises gráficas do padrão de distribuição das proficiências θ (considerando os escores padronizados) em função do teste-reteste.

Tabela 24: Valores dos parâmetros estatísticos descritivos para o pré-teste e pós-teste com base nos valores da TRI.

Parâmetros Descritivos	Pré-teste	Pós-teste
n	153	161
Média ("mean")	0,14	0,59
Desvio-padrão ("SD")	0,69	0,83
Mediana ("median")	0,11	0,61
Desvio-padrão aparado ("trimmed")*exclui valores extremos ou esporádicos	0,15	0,61
Desvio absoluto mediano ("median absolute deviation" ou "mad")	0,56	0,94
Mínimo	-1,24	-1,22
Máximo	1,96	2,64
Intervalo ou "range"	3,2	3,86
Assimetria ou "skew"	-0,06	-0,1
Curtose ou "kurtosis"	-0,18	-0,64
Erro padrão ou "SE"	0,06	0,07

Pode-se observar um ganho na distribuição mais homogênea dos estudantes, sobretudo nos intervalos de $1,0 < \theta < 2,0$ no pós-teste, estando, de modo geral, os alunos com proficiência um pouco acima da média. O gráfico sugere que para a zona com $\theta < -0,5$ haja uma baixa densidade, ou seja, pouca representatividade para avaliar o perfil de proficiências técnicas dos sujeitos tanto no pré-teste como no pós-teste.

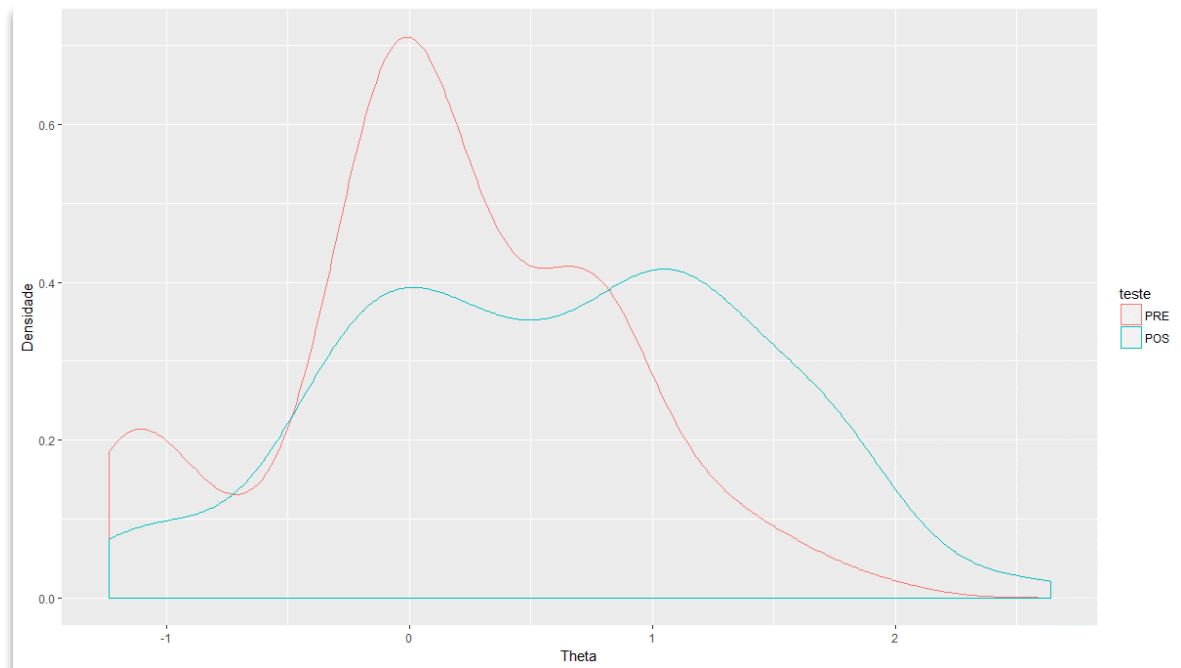


Figura 51: Gráfico da distribuição da densidade das respostas em função das habilidades dos estudantes no pré-teste e pós-teste.

No diagrama de caixa (“*boxplot*”) é possível notar que o 1º quartil permaneceu quase o mesmo após as instruções (zona representada predominantemente pelos alunos do grupo inferior – ABAIX) estando todos com $\theta < 0$, e que o 3º quartil do pré-teste passou para o 2º quartil do pós-teste, e isso é um ganho.

A *amplitude interquartilica* (correspondente a diferença entre o 3º e o 1º quartil) determina o comprimento das caudas do “*boxplot*”, com isso, mostra uma diminuição na assimetria das respostas com relação ao pré-teste. Nisso, observa-se que no 2º quartil (mediana) foi elevada a porcentagem de alunos com proficiência acima de $\theta = 0$ (aproximadamente), mostrando que após a instrução, os alunos da *zona intermediária* (correspondente aos 50% entorno da média) foi ampliada apresentando uma distribuição nas suas proficiências no intervalo de $-0,05 < \theta < 1,45$ (variação de 1,5 valores aproximadamente) enquanto antes era de $-0,20 < \theta < 0,70$ (variação de 0,90 valores aproximadamente), com um ganho entre as medianas em cerca de 0,5 valores na pontuação normalizada. Os alunos do grupo superior (ACIM) também foi ampliado na sua distribuição de proficiência do intervalo de $0,70 < \theta < 1,75$ (variação de 1,1 valores aproximadamente) para $1,2 < \theta < 2,3$ (variação de 1,1 valores aproximadamente), bem como na elevação do limite superior de 1,75 para 2,70

(aproximadamente), considerando que no pré-teste surgiu um aluno *outlier* com proficiência próximo de $\theta = 2,00$.

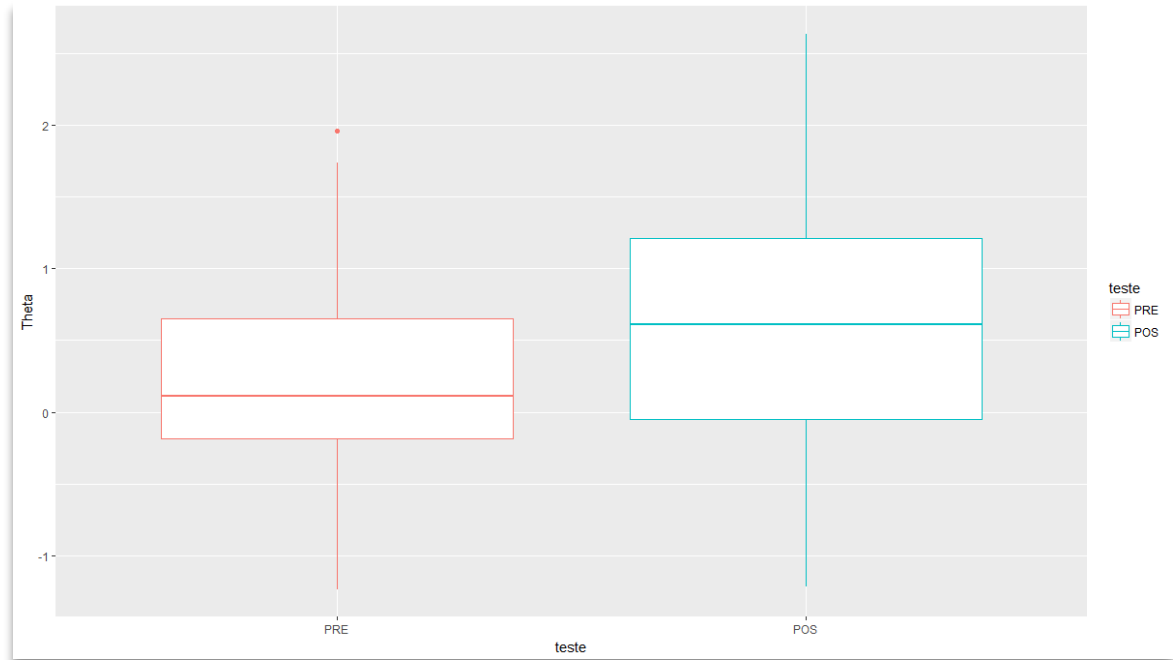


Figura 52: Gráfico dos diagramas de caixa para o pré-teste e pós-teste, de acordo com os níveis de habilidade.

No pós-teste obteve-se um alfa de Cronbach de 0,636, considerado baixo, porém aceitável para análises e *algumas medidas projetivas* (Murphy & Davidshofer, 2005, p. 150), porém mais adequado para avaliar e comparar *médias de grupos*. Cruzando os valores obtidos na análise da TRI do pacote *irt* (suplemento do *Excel*) para os modelos de dois e três parâmetros logísticos é possível afirmar que seja um **teste difícil para os alunos**, sendo o item y1 (com $b < -1,0$) o único considerado fácil, e os mais difíceis sendo y6, y8, y15, y17, y19 e o y25 (com $b > 2,0$), reforçando assim as análises clássicas anteriores.

CAPÍTULO 4 - RESULTADOS

4.1 FASE DESCRITIVA - Análises da Teoria Clássica dos Testes

4.1.1 Considerações Gerais

A fase descritiva se refere a ênfase dada nas estatísticas descritivas, na qual envolvem cálculos relativamente mais simples, como médias, desvio-padrão, variância, sendo essencial para uma sondagem exploratória e uma leitura inicial panorâmica da avaliação. Esse tipo de avaliação é mais geral e pertinente aos modelos da TCT que se concentram no teste de modo geral, enquanto para a TRI se exploram análises mais robustas e inferenciais multivariadas, extraindo informações mais precisas sobre os itens e das relações entre eles. Isso justifica de certa forma estruturar a pesquisa distinguindo a fase descritiva para a TCT e a fase analítica inferencial para TRI. Entre outras perspectivas descritivas estão inclusas medidas de amplitudes, percentis (ou centil), quartis e decis, além de medidas de distribuição (achatamento e assimetria de distribuição) e análises gráficas (histogramas, boxplots, etc.) que também podem ser realizadas como instrumentos na avaliação mais simples sobre os itens. As análises clássicas exploradas nessa etapa buscaram dessa forma extrair informações gerais sobre o teste e os respectivos itens, utilizando basicamente os procedimentos adotados no teste-piloto.

Visando uma ampliação, refinamento e fiabilidade nas conclusões e interpretações, estão inclusos aspectos avaliativos mais peculiares que abrem terreno para uma discussão mais detalhada das evidências, e assim, fornecer subsídios sobre as características e influências que determinam os perfis conceituais (e/ou conceptuais) tendo como base as análises psicométricas. De modo geral, seguem:

- Apresentar de forma clara o tipo de *design* experimental da investigação, apontando as peculiaridades entre os diferentes *grupos de respondentes* e um possível parâmetro de tratamento relativamente comum entre os contextos, expondo na discussão as principais diferenças e riscos de validade eminentes.

- Aplicar e analisar os procedimentos da TCT buscando identificar seus principais índices discriminativos, de fiabilidade e de dificuldade, e, a partir disso, identificar *grupos de itens* com características relativamente comuns dentro de determinadas categorias e níveis, que possam sugerir informações relevantes quanto à representatividade de *grupos de respondentes* e maior consistência das evidências que caracterizam os perfis de proficiências;
- Delimitar zonas específicas dentro dos escores nas amostras que busquem representar de forma mais consistente as informações dos perfis de proficiência, e assim, sinalizar aspectos de domínio do conhecimento relacionado ao teste, bem como das concepções térmicas e dificuldades conceituais diante dos conhecimentos envolvidos.

As análises clássicas partem do registro, organização a aplicação de fórmulas com o auxílio do *software Excel*. Para isso, a base de dados dicotômica foi tratada, sendo utilizadas diversas funções do programa, e algumas fórmulas sendo introduzidas nas células da planilha para se efetuarem os cálculos. Posteriormente são utilizados outros softwares para a realização de outras análises específicas, como também na confirmação dos resultados nas análises prévias.

A tabela 25 ilustra os dados organizados com a chave-de-correção (Key) com as respectivas respostas, ou indivíduos ordenados (Rol), os 26 itens codificados de y1 a y26, os erros ("0") e acertos ("1"). O X_i se refere ao escore bruto ou total de acertos do sujeito i , sendo os valores máximos obtidos igual a 23 (Brasil) e 24 (Portugal). E ao fim estão alguns dos principais parâmetros da TCT, o *índice de dificuldade* (ID), os grupos de indivíduos que acertaram ("A") o item, mas que se encontram em uma margem inferior correspondente 27% abaixo (ABAIX) e 27% acima (ACIM) do contínuo de desempenho. A diferença entre eles corresponde ao *índice de discriminação* (DISCRIM). O coeficiente de correlação bisserial (BISS) é um valor que varia de -1 a 1, além de indicar uma consistência do item com relação ao teste, também pode ser considerado um índice de discriminação por sugerir diferentes perfis de composições dos grupos de respondentes entre os de baixo e alto desempenho. Muitos outros parâmetros clássicos não estão na ilustração, como a: *variância amostral* (σ) nas respostas dada a

cada item (1 ou 0); soma da variância de cada item com relação à amostra ou *soma da variância amostral* dos itens ($\sum \sigma_i$); soma da variância do escore ou pontual total de cada sujeito da amostra ou *soma da variância amostral* dos escores dos sujeitos ($\sum \sigma_i$); o alfa de Cronbach (α) e o correspondente alfa de Kuder-Richardson (KR20); erro padrão de medida (*Standard Error Measurement* – SEM), além de médias, desvios em relação à média, desvios-padrões, diferenças de quadrados, escores normatizados (Z), função de densidade, entre outros.

Tabela 25: Matriz-Resposta Dicotômica (ilustração parcial das respostas) dos Alunos Brasileiros do 1º, 2º e 3º ano do Ensino Médio na Grande Florianópolis.

Chave	A	B	B	B	B	B	C	A	B	C	..	A	D	A	C	D	A	C	D	C	E	X_i
ROL	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	..	y17	y18	y19	y20	y21	y22	y23	y24	y25	y26	
1	0	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	1	0	0	0	..	0	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	1	0	0	0	0	0	..	0	0	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0	0	..	0	0	0	0	0	0	0	0	0	0	1
5	0	0	0	0	0	0	0	0	1	0	..	0	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	..	0	0	0	0	1	0	0	0	0	0	1
7	0	0	0	0	0	0	1	0	0	0	..	0	0	0	0	0	0	0	0	0	0	1
8	0	0	1	0	0	0	0	0	0	0	..	0	0	0	0	0	0	0	0	0	0	1
9	0	0	0	0	1	0	0	0	0	0	..	0	0	0	0	0	0	0	0	0	0	1
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
2957	0	1	1	1	1	1	1	1	1	1	..	1	0	1	1	0	1	1	1	1	1	23
ID	0,46	0,34	0,37	0,29	0,21	0,23	0,54	0,29	0,27	0,30	..	0,22	0,16	0,26	0,19	0,30	0,30	0,31	0,22	0,25	0,25	
ABAI	0,28	0,17	0,20	0,17	0,12	0,16	0,37	0,17	0,18	0,13	..	0,13	0,08	0,15	0,10	0,19	0,15	0,12	0,12	0,14	0,13	
ACIM	0,62	0,52	0,56	0,39	0,33	0,31	0,68	0,42	0,39	0,50	..	0,31	0,26	0,36	0,28	0,41	0,48	0,54	0,32	0,40	0,40	
DISCR	0,35	0,34	0,36	0,22	0,21	0,15	0,31	0,26	0,21	0,36	..	0,17	0,18	0,22	0,18	0,22	0,34	0,42	0,20	0,26	0,27	
BISS	0,34	0,39	0,39	0,27	0,32	0,22	0,31	0,34	0,29	0,43	..	0,23	0,34	0,29	0,28	0,23	0,39	0,49	0,29	0,36	0,38	

Os procedimentos de tratamento e análise para a base de dados de Portugal são semelhantes, considerando apenas da existência dos diferentes turnos de atividades escolares, sendo a amostra no Brasil com turnos matutino, vespertino e noturno, e em Portugal de forma padrão a educação em tempo integral, que corresponde aos turnos matutino e vespertino. Apenas para complementar, essa questão da educação integral vem sendo defendida no Brasil (e em outros contextos) com o argumento de que um aluno com “mais tempo” na escola possui maiores chances de sucesso na aprendizagem, ao contrário do que a Finlândia vem defendendo, e ainda assim

mantendo seus alunos com elevados índices de desempenho. Quanto a esse aspecto, PNE vigente no Brasil prevê até 2024 atingir a meta de 50% das escolas ofertando educação de tempo integral e atendendo cerca 25% dos alunos (Brasil, 2014, p. 33).

Na base de dados de alunos do Brasil, no que se refere aos intervalos inferiores e superiores igual a 27%, o limite inferior ficaria na posição ou indivíduo no rol igual a 789,39, por razões lógicas arredonda-se para 799. O desempenho do indivíduo correspondente a essa faixa possui **5 acertos**, convencionou-se estender o limite agregando todos os indivíduos com mesmo desempenho, ou seja, até o último respondente com a mesma nota, no caso a posição 840. O limite superior ficaria em 2158,61, arredondando 2159 com desempenho igual a **9 acertos**, se estende o intervalo partindo do sujeito 2157 no rol. O limite inferior estimado inicialmente de 27% acaba por corresponder a $A_{27\%}^- = 840$ respondentes e o limite superior a $A_{27\%}^+ = 801$, e um grupo intermediário de $A_{46\%}^\pm = 1316$.

A amostra em Portugal foi igual a 2585 alunos do ensino secundário. Nesse caso o limite inferior no rol ficou em 697,95, sendo arredondado para 698. O escore nesse ponto são de **8 acertos** e se estende até a posição 921, o limite superior fica em 1887,5, ou seja, 1888, com **escore 12**, e se estende até 1712 (vindo debaixo para cima do rol). Os limites ficam com $A_{27\%}^- = 921$ (quantidade de alunos do grupo de desempenho inferior) e $A_{27\%}^+ = 874$ (quantidade de alunos do grupo de desempenho superior), e um grupo intermediário com $A_{46\%}^\pm = 790$ respondentes.

A tabela elaborada relativa ao quantitativo de alunos por escola dentro de cada nível de pontuação nos dois contextos ilustra alguns aspectos interessantes, como a representatividade da amostra por escola, dos alunos que mais se destacam por nível de escolaridade (e por turno, no contexto brasileiro), e pode-se observar um baixo quantitativo de aluno do turno noturno no grupo superior (ACIM), porém, grande parte daqueles que conseguem maiores pontuações são de níveis mais avançados, como era de se esperar (ver Apêndice E4, p. 268). Também é possível observar que os níveis de desempenho 7 e 8 possuem respondentes em todas as escolas, de acordo com a média, assim como a diminuição de indivíduos acima da média geral. Pode-se notar

que os turnos matutino e vespertino também são os que mais se destacam, sendo os do matutino aparentemente com uma certa vantagem de modo geral, enquanto o turno noturno se concentra mais entorno da média. Chama atenção a existência de alunos isolados que se distanciam bastante do desempenho dos outros, incluindo o turno noturno. As escolas 20BR e 22BR demonstram se destacar perante as outras por concentrarem uma quantidade maior de indivíduos com desempenho acima da média e em uma região de destaque com os mais elevados escores entre os alunos.

Dentre os valores obtidos para os parâmetros clássicos (ver no Apêndice E3. p. 266), constata-se que os escores brutos para os **respondentes brasileiros se encontram abaixo da média esperada** para todos os itens, já indicando que o *teste de modo geral foi difícil*. Para os **portugueses** também ocorre de forma semelhante, entretanto, as médias se apresentam *sutilmente superior que a dos brasileiros*.

Um aspecto interessante está na média dos escores brutos para os alunos dentro de cada item, em que se constatou que para os itens y14 e y24, os respondentes portugueses desse grupo de acerto alcançaram a média do teste ($\geq 50\%$, ou seja, no mínimo 13 acertos), mesmo sendo um item com porcentagens baixas de acerto (ID iguais a 0,35 e 0,29), entretanto, com coeficientes bisseriais mais elevados (0,69 e 0,62). Isso significa um perfil de composição em que predominam alunos de alto desempenho, ou seja, esses respondentes demonstram compreender o conhecimento envolvido.

Para os critérios considerados ideais na porcentagem de acerto, considera-se que itens com 5 alternativas deva-se alcançar 60% de acerto, e com 4 alternativas 63% (Thompson & Levitov, 1985), porém não se constata nenhum item no contexto brasileiro nessa faixa, e em Portugal se identificam três itens que estão dentro dessas expectativas (y1, y7 e y12). Alguns dos «**itens mais fáceis**» para o **Brasil** foram, em ordem decrescente o y7 (54%), y12 (53%), e o y1 (46%), enquanto para **Portugal** foram o y1 (74%), y7 (60%), y12 (60%), y23 (59%), y13 (58%), y3 (53%), y2 (51%) e y22 (50%) (ver no Apêndice A, pp. 1-210).

De modo geral, pode-se afirmar que os valores dos parâmetros clássico dos itens possuem a finalidade de contribuir na construção e caracterização de um **perfil de**

proficiências térmicas (conhecimento térmico que o aluno supostamente domina) e um **perfil de concepções térmicas** (equívocos relacionados aos fenômenos térmicos mais representativos) para cada contexto. Obviamente, quando se fala em *perfil de proficiência térmica*, se refere a uma ideia em sentido inverso sobre o *perfil de concepções térmicas*, sendo o primeiro voltado para analisar aquilo que o aluno possui maior domínio, com isso a análise se concentra na *resposta correta* e nos *descritores* referentes as proficiências dos alunos. No segundo, busca-se construir uma imagem diante das dificuldades baseadas nos *equívocos* que os alunos evidenciam através do percentual médio suas respostas, com isso a análise se concentra nos conhecimentos térmicos errôneos dos *distratores*, bem como nas *concepções equivocadas típicas* diante desses conhecimentos. Isso significa dizer que quanto mais consistente for a identificação das concepções térmicas, menor será o domínio que o aluno tem sobre o conhecimento necessário envolvido para que ele responda corretamente, e vice-versa. Considera-se que essas análises sejam complementares diante o conhecimento avaliado e medido no teste dentro das análises clássicas, na qual fornecem subsídios interpretativos às análises da TRI.

Já no campo procedimental da TRI, entende-se que as análises alcançam um maior aprofundamento ao serem estimados níveis dos processos mentais dos respondentes através de probabilidades de acertos e erros, subsidiando assim de forma ainda mais consistente as interpretações e definições para os perfis de proficiência e concepções térmicas. Nesse sentido favorecem identificando as necessidades no campo da aquisição conceitual dos alunos (seja para grupos de respondentes ou até indivíduos), e com isso, enfatizar aspectos necessários para a instrução do ensino dentro de um processo avaliativo mais amplo, que visem em especial, alcançar uma aprendizagem significativa de acordo com as condições apresentadas por David Ausubel.

Para a obtenção de uma melhor compreensão e para uma maior fiabilidade desses perfis foram delimitadas *zonas de maior representatividade e consistência*, que no caso, permitem uma análise interpretativa mais precisa para conhecimento sobre esses perfis.

4.1.2 Design do Teste TCE: Portugal e Brasil

O *design* do teste TCE constitui uma base de dados de três níveis de escolaridade para os alunos de Portugal em outra semelhante para o Brasil, com destaque em vermelho para as observações (O1 e O5) na qual os alunos responderam antes de receberem a instrução das aulas de termodinâmica, sendo os Grupos de Controle (GC 10º ano e GC 1º ano), e as aulas consideradas a forma tratamento (X). A segunda observação foi desconsiderada nessa perspectiva (O2), sendo no caso o pós-teste, considerado na análise prévia do teste-piloto para as escolas secundárias do 10º ano em Portugal.

Apesar do *design* considerar uma determinada forma de tratamento, sabe-se que esses tratamentos são distintos e caracterizam risco de validade. Dessa forma foi estabelecido apenas como uma forma racional de compreender a configuração do estudo, pois não há interesse em avaliar o tratamento, mas sim no perfil das dificuldades conceituais de cada grupo e comparativamente entre eles. Ainda assim, se espera que os grupos experimentais sejam sujeitos com maior maturidade devido ao fato de estarem em níveis de escolarização mais avançado, com isso se pressupõe que possuam maior nível de desempenho e competência científica.

Projeto de Comparação de Grupo Estático Não Equivalentes - Portugal	
Portugal	GC 10º ano O1 _____ X
	GE 11º ano X _____ O3
	GE 12º ano X _____ O4
Projeto de Comparação de Grupo Estático Não Equivalentes - Brasil	
Brasil	GC 1º ano O5 _____ X
	GE 2º ano X _____ O6
	GE 3º ano X _____ O7

Figura 53: Design do teste TCE - projeto de comparação de grupo estático não equivalentes em Portugal e no Brasil.

Os riscos de validade interna e externa para o *design* de Portugal e Brasil tendem a ser equivalentes, seguindo os critérios de homogeneidade nos tratamentos, entretanto, a descrição dos riscos é relativizada devido à base de dados admitir diferentes grupos em níveis de escolaridade diferentes, principalmente pelas peculiaridades no contexto brasileiros entre os perfis dos alunos nos diferentes turnos.

Com isso, para a validade interna, de 15 riscos analisados, 5 são admitidos para o *design*, que são: *interações com a seleção (seleção/história)*; *regressão estatística*; *seleção diferencial*; *interações com a seleção (seleção/instrumentação)*; e *ambiguidade da direção da causa*.

Para a validade externa, de 14 riscos analisados, 7 são admitidos para o *design*, que são: *história*; *maturação*; *interações com a seleção (seleção/maturação)*; *interação entre variáveis psicológicas e o tratamento*; *interferência de múltiplos tratamentos*; *interação da história e do tratamento*; e *interação tempo de medição e tratamento*.

A partir de um *checklist*, foi realizada uma análise descritiva para cada um dos riscos associados tanto para a validade interna como para a validade externa (Apêndice C, p. 227).

4.1.3 Zonas de Representatividade e de Consistência

Do total de 2957 de respondentes brasileiros nos três níveis de escolaridade (1º, 2º e 3º ano do ensino médio), são exatamente os níveis de escore 6, 7 e 8 que possuem a maior concentração de alunos (44,5%), sendo a mediana e a moda iguais a 7 na distribuição, e a média 7,12 (aqueles que acertaram até 8 itens), ou seja, a distribuição apresenta uma leve assimetria positiva. A interpretação de uma assimetria positiva significa que a maioria dos alunos ficaram abaixo da média da amostra, bem distante da média do teste, que seriam 13 acertos. Como foi visto no teste-piloto, o “*guessing*” do teste são de 7 acertos, ou seja, a média dos alunos encontra-se na zona de acerto aleatório, podendo sinalizar uma maior preocupação diante às dificuldades e necessidades encontradas.

Torna-se ainda mais preocupante ao ser levado em consideração o *escore bruto corrigido* ($X_{\text{corrigido}_j}$) de cada sujeito (j), ou seja, eliminando as estimativas para os *números de acerto aleatório* (A_{aj}), e com isso elaborar a relação matemática da *média dos escores corrigidos do teste* ($\bar{X}_{\text{corrigido_teste}}$) para o TCE. Para isso, deve-se considerar a subtração de pontos estimados para os erros entre os itens com 4 e 5 alternativas (itens com $k = 5$, são: y9, y11, y16, y18, y26), que seria:

Para 1 sujeito, tem-se: $X_{corrigido_j} = A_j - A_{a_j} = A_j - A_{a,k=4_j} - A_{a,k=5_j}$

Para toda a amostra ($N = 2957$), tem-se:

$$\bar{X}_{corrigido_{teste}} = \frac{\sum_{j=1}^N X_{corrigido_j}}{N} = \frac{\sum_{j=1}^N (A_j - A_{a_j})}{N}$$

$$\bar{X}_{corrigido_{teste}} = \frac{\sum_{j=1}^N A_j - \sum_{j=1}^N A_{a,k=4_j} - \sum_{j=1}^N A_{a,k=5_j}}{N}$$

$$\bar{X}_{corrigido_{teste}} = \frac{\sum_{j=1}^N A_j - \sum_{j=1}^N E_j \cdot \frac{1}{4-1} - \sum_{j=1}^N E_j \cdot \frac{1}{5-1}}{2957}$$

Os resultados mostram que $\bar{X}_{corrigido_{teste}} = 1,16$, ou seja, os alunos possuem um domínio de até 2 acertos entre os 26 itens, além do que, a média do grupo de desempenho inferior (até 5 acertos) foi de -2,82 pontos, e até aqueles que alcançaram a média de 7,12 (até 8 acertos) a média corrigida dos escores foi de -0,49, isso significa uma evidência preliminar de que esses alunos em média não possuem nenhum domínio sobre o conteúdo avaliado no teste.

Ao eliminar esses alunos que não apresentam domínio, e considerar somente aqueles a partir de 9 acertos, ou seja, aqueles que fazem parte do grupo superior na amostra brasileira, foi obtida um *média de escore corrigido do teste* de 5,60, ou seja, ainda mais distante da média esperada para o teste que seria de 13 acertos. Isso reforça de forma mais precisa a dificuldade média dos alunos diante do teste.

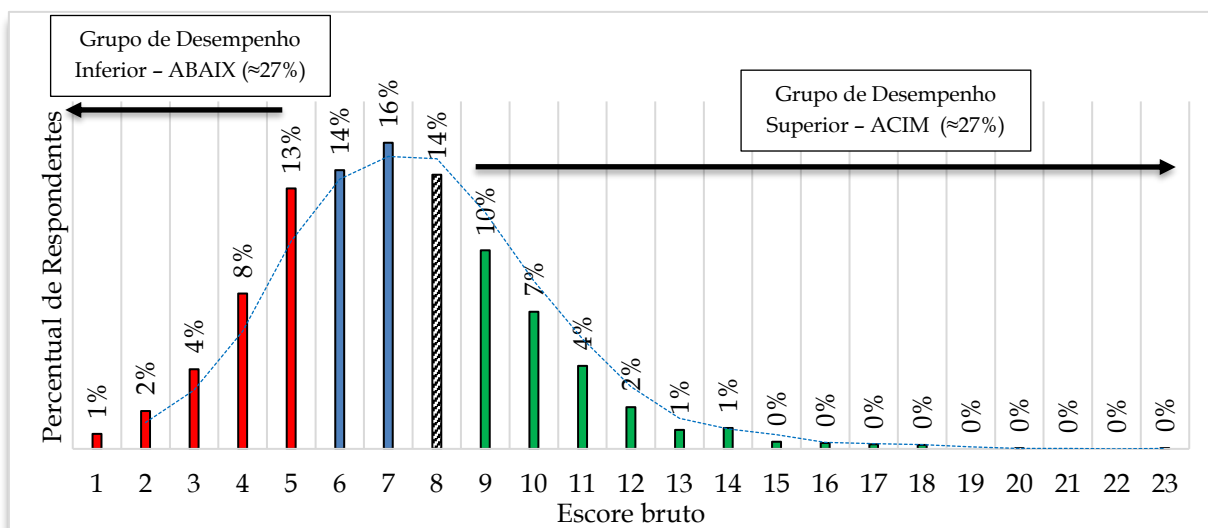


Figura 54: Distribuição percentual de respondentes brasileiros de acordo com seu respectivo escore.

Observando o gráfico semelhante essas distribuições de acordo com os turnos, é possível perceber uma média relativamente próxima entre os grupos (Figura 55).

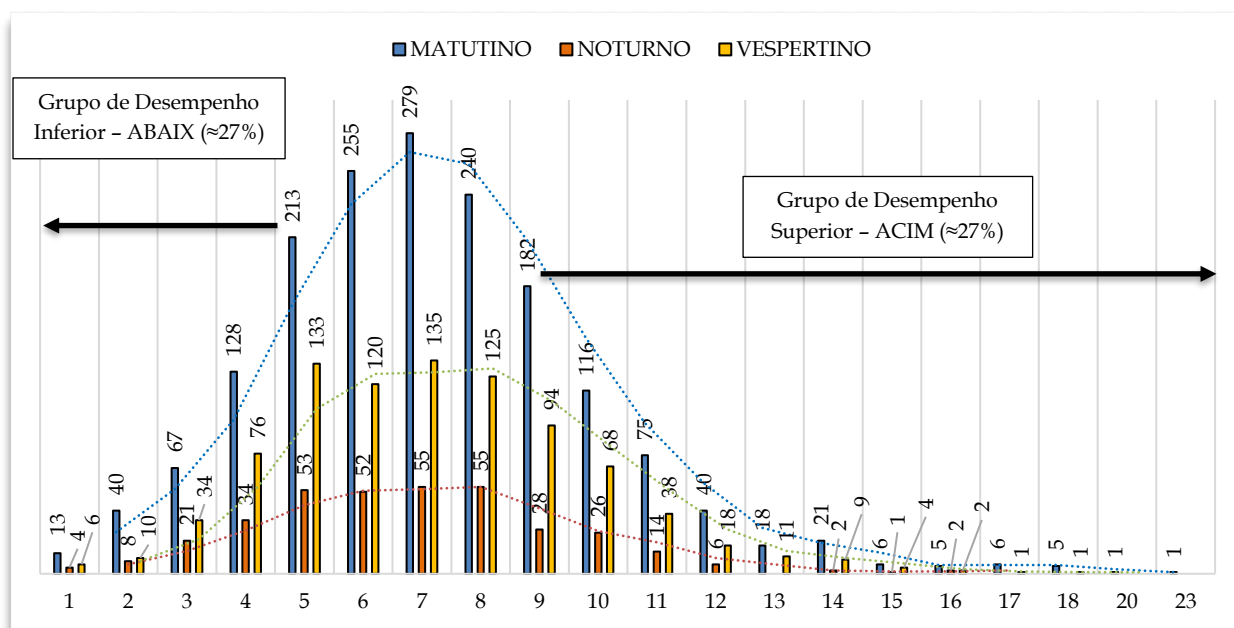


Figura 55: Distribuição quantitativa de respondentes brasileiros por escores de acordo com os turnos.

Convém lembrar que esta análise engloba além das diferenças entre os turnos, também alunos com níveis de escolaridade diferentes (1º, 2º e 3º ano do ensino médio no Brasil), e obviamente se espera que haja diferenças com relação a essas variáveis, entretanto, a análise inicial possui a finalidade de proporcionar uma visão ampla do desempenho dos alunos, porém identificando algumas de suas características, e com isso, os itens serão analisados inicialmente sob essas perspectivas utilizando a TCT e TRI de forma complementar. As análises comparativas dessas variáveis serão

realizadas mais adiante também por meio de modelos da TRI, na qual incluem procedimentos de análise baseadas na *Função Diferencial dos Itens* (DIF), que se refere a diferença no parâmetro de dificuldade entre um grupo de referência e um grupo focal.

De modo geral, os alunos de baixo rendimento correspondentes aproximadamente aos 27% inferior (ABAIX) da amostra possuem um escore de 5 acertos, ou seja, abaixo do nível de probabilidade do “*guessing rate*” que seriam de 7 acertos ($\approx 27\%$ de acertos no teste). Isso evidencia que parte dos sujeitos do grupo intermediário. Na realidade também são evidentemente respondentes de baixa proficiência, com escores pouco fiáveis representativamente. Isso reforça uma necessidade de verificar uma zona de respondentes que seja mais representativa para as interpretações dos escores. O gráfico (Figura 56) mostra que o ponto de “*guessing rate*” representa cerca de 59% de todos os respondentes, sendo uma porcentagem predominante de alunos com baixa proficiência, e 73% encontram-se abaixo da média que seria de 8 acertos (diferentemente da média esperada do teste que seriam de 13 acertos).

Observa-se ainda que, somando os alunos que atingem até 12 acertos, alcança-se cerca de 97% do total de respondentes, com isso, apenas 3% vão subindo na escala de desempenho até pouquíssimos alcançarem um limite de 23 acertos. Esta região representa bem respondentes dentro da amostra que se caracterizam com um *perfil de proficiência térmica* dentro daquilo que se espera que haja **domínio sobre o conhecimento envolvido**, ou seja, são respondentes com alto índice de acertos, e assim, deve-se acreditar que suas respostas sejam mais fidedignas quanto às dificuldades que não foram tratadas adequadamente durante as instruções, assim como aquelas que foram. Entende-se que o grupo da amostra de respondentes que sugerem de forma mais consistente internalizarem equívocos, são aqueles enquadrados numa faixa de desempenho que configura um *perfil de concepções térmicas* dos sujeitos (dentro de cada contexto), e estão obviamente dentro de um mesmo *continuum*, com isso, pode-se dizer que possuem predominantemente níveis de proficiências de média para baixa. A distinção desses grupos é em parte didática, refinando dentro dos grupos de inferiores, intermediários e superiores, mas

fundamentalmente na intenção de concentrar as análises em um campo de maior fiabilidade interpretativa.

Dessa forma, pode-se dizer que as informações mais representativas do grupo diante de suas concepções devam se concentrar entre 59% e 97% (intervalo de 37%), correspondendo ao intervalo de $7 < X_i \leq 12$ para os valores dos escores. Apesar disso, os atrativos de distratores para os alunos de alta performance, ainda que sejam poucos, torna-se relevante de serem verificados, pois alunos que supostamente dominam o conteúdo não deveriam ser atraídos fortemente por distratores. Essas informações podem ser extraídas ao se observar as *Análises Gráficas dos Itens* - AGIs (verificar nas análises clássicas dentro no painel estatístico dos itens que se encontra no Apêndice A, pp. 7-213).

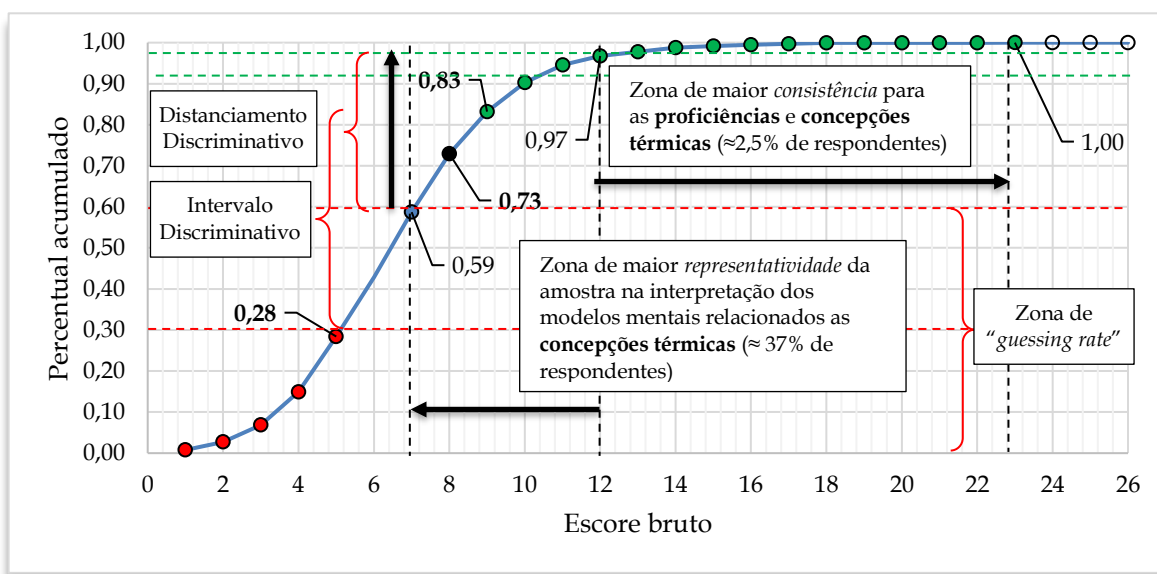


Figura 56: Percentual acumulado de respondentes *versus* Escore bruto - Brasil.

Para uma análise mais detalhada, pode-se considerar que o *perfil de concepções térmicas* para cada contexto deva se concentrar dentro de uma maior abrangência de respondentes, ou seja, em uma **zona de maior representatividade dos respondentes da amostra**, enquanto o *perfil de proficiências térmicas* com informações mais fiáveis na **zona de consistência**. Remete-se ao termo “proficiente” sugerindo relacionar sujeitos supostamente em um níveis mais elevados de competências científicas para o conhecimento térmico, adotando a sugestão do **modelo Dreyfuss** (Dreyfus, & Dreyfus, 1986), que é adotado pela OCDE (Keevy & Chakroun, 2015), na qual se inclui o teste

de larga escala para avaliação de desempenho de estudantes em perspectiva comparativa internacional PISA. Sabe-se essa distinção em última análise não exista de fato, pois o perfil de proficiências é um *continuum* de habilidades que agrega a todos os sujeitos, tanto os mais proficientes quanto os menos proficientes, sendo os menos proficientes com maiores dificuldades conceituais e com concepções equivocadas associadas de forma mais intensa. Apesar disso, acredita-se que essa sutil distinção seja conveniente para uma lente de análise interpretativa e discriminatória, como foi dito, pois é óbvio que a zona de consistência dos escores (acima de 97%) também pertença ao perfil das concepções térmicas (entre 59% e 97%), e vice-versa.

Dessa forma, se estabelece como critério de consistência para o **perfil de proficiências térmicas**, fundamentalmente o parâmetro de discriminação (e não simplesmente em intervalos de escore bruto), na qual é responsável em diferenciar grupos de indivíduos que dominam o conteúdo com aqueles que não dominam. A perspectiva do perfil de proficiências torna-se mais rica agregando na análise de forma conjunta o parâmetro de dificuldade dos itens. Nesse caso, a forma como os itens se dispõem diante desses parâmetros são considerados suficientes e determinantes para uma conclusão preliminar sobre o perfil de proficiências térmicas no teste para cada contexto, de acordo com a TCT. Pode-se admitir que a representatividade desse perfil se associe com uma maior concentração dos itens, e sendo mais consistentes naqueles identificados como mais discriminativos. Nas análises da TRI, o perfil de proficiências térmicas é caracterizado por outros parâmetros e por análises estatísticas que possibilitam informações ainda mais consistentes.

Para o **perfil de concepções térmicas**, torna-se conveniente estabelecer como linha de corte a faixa de acertos baseadas no “*guessing rate*”, e definir um *distanciamento discriminativo* sendo o intervalo percentual dentre aqueles que provavelmente «**não dominam o conteúdo**» (contidos na zona de “*guessing*” do teste) e entre aqueles de maior desempenho que provavelmente «**dominam o conteúdo**», considerado como limite de corte os 2,5% (aproximadamente) mais elevado dentro do *grupo superior de indivíduos que estão em destaque*, sendo este um critério correspondente ao das curvas de distribuição normal para um intervalo de confiança aproximadamente igual a 95%,

apesar de não ser a mesma curva. Esse percentual de alunos no grupo superior (ACIM) que são atraídos por distratores merece uma atenção especial, pois a sua relação com o grupo intermediário possibilita atingir uma maior consistência na determinação do perfil de proficiência e de concepções relacionadas ao conhecimento térmico que o teste explora.

Em Portugal o «acerto ao acaso», de modo geral, permanece por se tratar do mesmo teste. Os níveis de maior concentração de respondentes são o 8, 9 e 10 (33%), um terço da amostra, sendo a mediana igual a 10, moda 9 e média com o valor de **10,3 pontos** (aqueles acertaram até 11 itens). Comparando com o Brasil, a assimetria de Portugal é mais sutil, além da concentração ser mais branda, indicando que os sujeitos se encontram mais bem distribuídos.

Considerando os escores corrigidos dos sujeitos de forma semelhante ao que foi feito no contexto brasileiros, foi obtido uma média geral para o teste de **5,32 pontos**. Para o grupo inferior (até 8 acertos) a média foi de 0,28, ou seja, estima-se que esses alunos não dominam nem 1 único item do teste, sendo representado por **35,6% de toda a amostra portuguesa**. Reunindo todos os alunos de baixo desempenho até os que alcançaram a média (até 11 acertos) a média foi de 2,37 pontos (66,2% da amostra), e para os alunos restantes do grupo superior (acima de 11 acertos) a média foi de 11,10 pontos, sinalizando que esse seja um grupo mais representativo e consistente para as análises, apesar de ainda estarem abaixo da média esperada para o teste, que seriam de 13 acertos no mínimo, ou seja, é um teste considerado difícil também para os alunos portugueses.

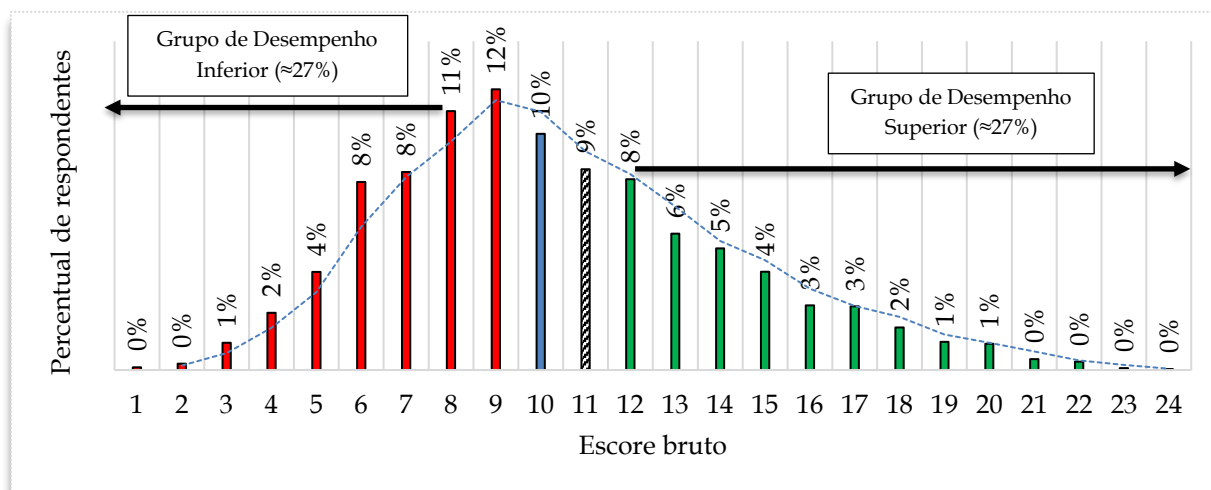


Figura 57: Distribuição percentual de respondentes portugueses de acordo com seu respectivo escore.

Pode se observar que o grupo de baixo desempenho corresponde a cerca de 48% do total de alunos portugueses, dentro disso, cerca de 25% encontram-se na margem do “guessing rate”. Na parte superior, 97% dos alunos alcançam 18 pontos, correspondendo a um intervalo de escores brutos de $7 < X_i \leq 18$, isso contempla uma fração de 11% dos alunos de baixo desempenho. Dessa forma, a representação das concepções dos respondentes encontra-se principalmente entre 25% e 97%, tendo uma diferença de 72% correspondendo ao dobro do Brasil.

Torna-se possível explorar do gráfico algumas perspectivas discriminatórias, na qual se pode denominar de *intervalo discriminativo* do percentual acumulado do teste em geral para todos os respondentes portugueses com valor de 38,7%, indo de indivíduos (grupo inferior) com pontuação igual a 8 até aqueles com pontuação igual a 12 (grupo superior), bem menor que a dos brasileiros que é de 54,8% (entre 5 e 9 pontos). Isso significa que para os alunos portugueses há um certo equilíbrio entre os que são capazes de responder corretamente o teste, enquanto no Brasil, somente acerta quem for de fato mais proficiente, ou seja, aqueles que estão mais distantes dos alunos de baixo desempenho ou que tenham acertado uma menor quantidade de itens.

É importante afirmar que o *intervalo de discriminação* corresponde ao índice de discriminação dos itens, mas não pode ser confundido com o *distanciamento discriminativo* que possibilita uma correspondência com a *zona de maior representatividade* da amostra para o perfil de concepções térmicas, na qual se exclui a

zona de “guessing rate” (de “chute” ou acerto ao acaso do teste, e não do item) e da pequena parcela de indivíduos que alcançam um maior desempenho, pertencentes ao grupo superior em destaque. A exclusão da parcela de indivíduos na parte superior extrema não deve ser considerada uma condição rígida, mas sim um critério de corte e comparação devido às supostas diferenças para o perfil desses indivíduos. Sobretudo, essa análise deve servir de complementariedade visando estabelecer uma melhor consistência na determinação tanto das proficiências quanto das concepções que os alunos possuem diante do conteúdo do teste.

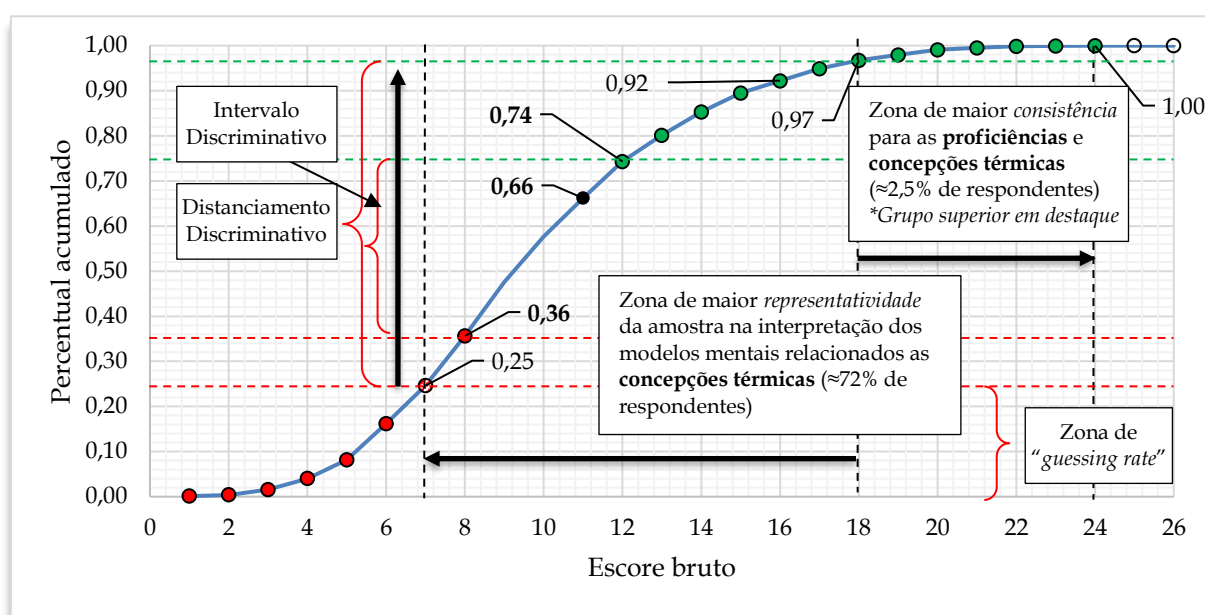


Figura 58: Percentual acumulado de respondentes versus Escore bruto - Portugal.

Comparando os índices de dificuldade entre Brasil e Portugal (ver no Apêndice E3, p. 266), pode-se constatar que o percentual de acertos em grande maioria está entre os portugueses. É possível observar um tipo de proporcionalidade entre os acertos e erros ilustrado, o que pode demonstrar a existência de um certo padrão de resposta para alguns itens. Essa razão de proporcionalidade pode ser notada, por exemplo nos itens y1, y2 e y3. Também chama a atenção o atrativo comum de alguns distratores entre os contextos, pois esses são os que representam a resistência de concepções alternativas dentro de cada grupo da amostra. Entre alguns itens desconcertantes quanto a esse percentual está o item y14, sendo o distrator “d” mais atrativo que a resposta correta para ambos os grupos (> 50,0%).

As análises verificaram uma correspondência elevada entre os percentuais de equívocos dos *grupos na zona de maior representatividade* (REPRES) dentro de cada contexto e os *grupo superior* (ACIM) em destaque. Por exemplo, o item y1 possui na *zona de representatividade* de $7 < X_i \leq 12$ um total de 1124 respondentes (38,2%), de um geral de 2944 alunos brasileiros (com 13 *missings*). As análises mostram que as proporções de respondentes dentro de cada grupo é similar a média percentual das respostas atrativas desses respondentes dentro de cada escore.

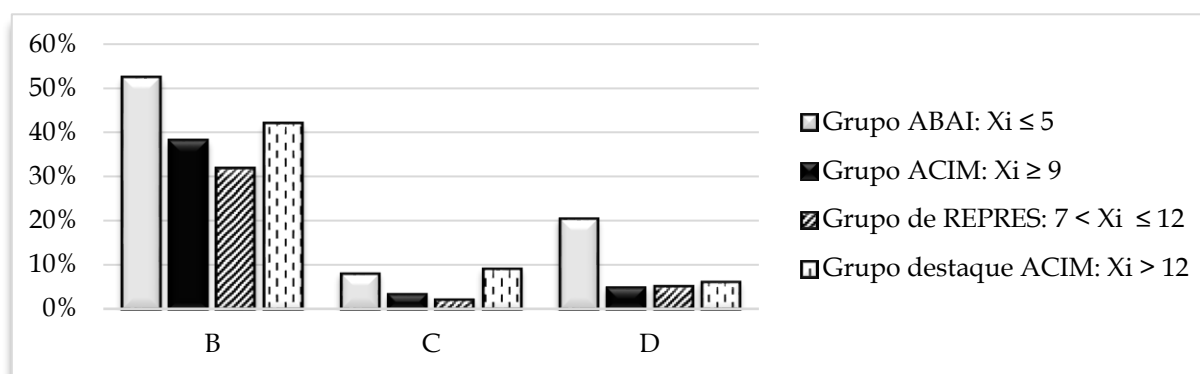


Figura 59: Média dos percentuais dos distratores para os níveis de escore (X_i) dentro dos intervalos ou zona dos diferentes grupos considerados para o item y1.

Fica evidente identificar o distrator mais atrativo, e esse padrão de análise também serve para reforçar uma compreensão sobre a disposição dos perfis de composição de grupos para cada alternativa entre os distratores dos itens. A média percentual dos distratores B, C e D indica uma atração em média maior para B e D. Apesar disso, a letra B demonstra ser mais atrativa para os grupos ACIM, incluindo aqueles que estão na zona de maior representatividade da amostra e os que estão em destaque atingindo os escores mais elevados. Os modelos mentais diante das concepções térmicas devem ser associados com o quadro de Inventário das Concepções do TCE.

Nesse caso, o distrator “b” demonstrou influência entre os brasileiros, atraindo respondentes com 17 e 23 pontos, conforme mostra a AGI (ver no Apêndice A1, p. 8). A concepção (D6) envolvida se refere a ideia de que **o gelo se encontra sempre a 0°C**, e que sua temperatura não se altera. Essa concepção não demonstrou muita influência (19%) entre alunos portugueses, ficando na proporção da margem do “*guessing*”.

Para uma perspectiva comparativa preliminar diante a representatividade dos escores, ao observar a frequência de respondentes em cada contexto em relação à média, pode-se observar que **a quantidade de respondentes abaixo da média é predominantemente preenchida pelos alunos brasileiros**, e entre aqueles acima da média predominam-se os portugueses, apesar da diferença entre as amostras ser de 372 alunos. Essa é uma informação que acarreta implicações diretas para interpretações advindas de uma base de dados para o contexto geral, como é o caso de uma escala de proficiência que contemple alunos brasileiros e portugueses conjuntamente.

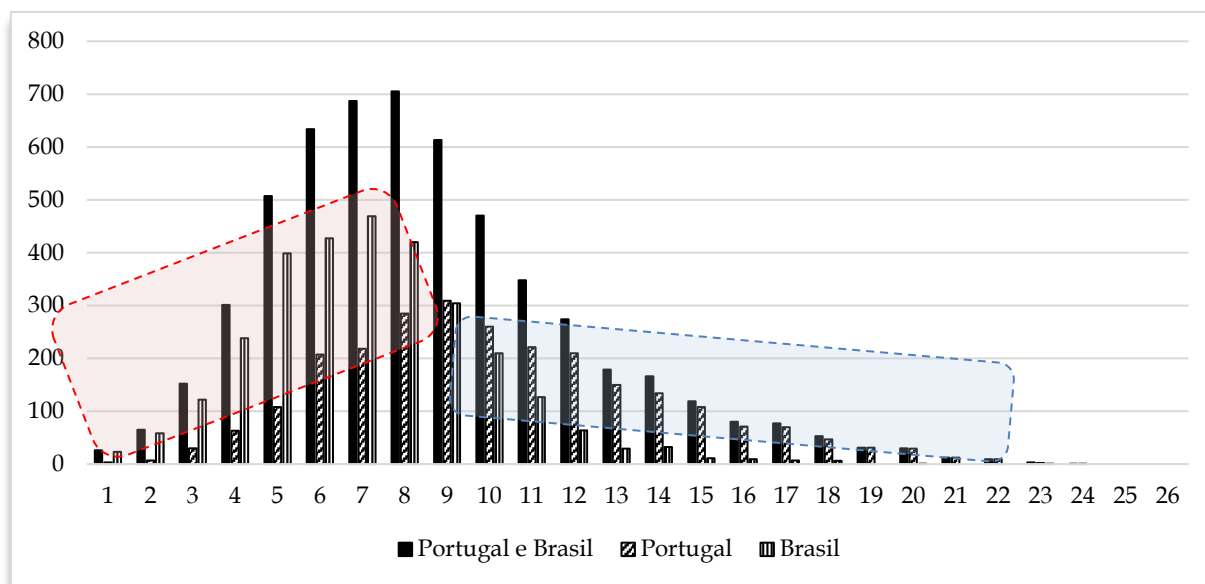


Figura 60: Frequência de respondentes brasileiros e portugueses dentro de cada nível de escore.

4.1.4 Perfil das Proficiências Térmicas

Essas análises se baseiam de forma conjunta diante das características psicométricas dos itens, das médias dos escores dentro de cada contexto de modo geral e dos grupos normativos específicos considerados, e dos respectivos ganhos percentuais. Adiantando, é possível dizer que no Brasil, devido aos baixos ganhos percentuais, pode-se afirmar que o perfil de proficiências, no que diz respeito ao domínio sobre o conteúdo do teste, encontra-se relativamente homogêneo entre os diferentes níveis de escolaridade, apesar de haverem algumas diferenças estatisticamente significativas quanto às médias entre alguns turnos. Enquanto isso,

em Portugal constata-se uma melhor discriminação, facilitando a identificação de itens que representam dificuldades e domínios de conhecimento dentro de um *continuum* de proficiências.

Como a construção de um perfil de proficiências envolve diversas faces, seja do ponto de vista dos índices psicométricos quanto na distribuição dos respondentes, níveis de escolaridade, entre outros, foram estabelecidos alguns critérios para facilitar a compreensão dentro cada uma dessas faces, assim como para estabelecer uma demarcação de validade e restrição do conhecimento. Com isso, para delimitar os resultados obtidos, os critérios mínimos visam proporcionar maior consistência e caracterizar de forma mais fiável um perfil de proficiências dentro das análises clássicas.

Os critérios preestabelecidos que visam **configurar um domínio diante do conteúdo no teste** de forma representativa os respondentes dentro de cada contexto, ou seja, informações que proporcionem determinar regiões ou conjunto de itens na qual sugerem uma maior apreensão por parte dos alunos, são:

- Comparar os itens quanto aos **índices de dificuldade e discriminação**, considerando todos os níveis de classificação.
- Destacar os itens que atendam a **critérios mínimos** (acima do *guessing* dos itens – 20% e 25%) e **ideais** de acerto (60% e 62,5%) no teste.
- Destacar os itens quanto a sua discriminação, considerando os níveis de classificação **bom e muito bom**, pois são considerados dentro de uma zona de maior consistência.
- Destacar os itens que se **mantiveram acima do “guessing” dentro da análise dos ganhos normatizados** no contexto Brasil.
- Destacar os itens na **região com ($\langle g \rangle \geq 0,30$) dentro da análise dos ganhos normatizados** no contexto Portugal.
- Identificar **características comuns** entre os itens para Brasil e Portugal, considerando os critérios preestabelecidos.

- Comparar de itens no contexto Portugal que **coincidem com os resultados do teste-piloto**.
- Analisar o **fator de concentração** dos itens nos dois contextos, de modo geral, e para cada grupo normativo de acordo com os diferentes níveis de escolaridade, e assim, identificando por consistência os itens que se comportam de forma comum e diferenciada. Dessa forma é possível identificar evidências que configuram o perfil de proficiências em cada contexto, sugerindo os modelos mentais corretos e incorretos, ou seja, conhecimentos supostamente adquiridos, bem como as dificuldades existentes, que serão informações de sustentação nas análises dos distratores e das concepções errôneas.

Dentro do perfil de proficiências, os critérios que visam **configurar as dificuldades conceituais e concepções errôneas diante do conteúdo no teste**, são:

- Identificar os **distratores mais atrativos** dentro do percentual absoluto de erro, na qual **competem fortemente** com a alternativa correta;
- Destacar os distratores que apresentam **percentual acima da alternativa correta**.
- Verificar a **consistência da força atrativa** através dos índices discriminatórios.
- Destacar os distratores e as respectivas concepções relacionadas que possuem forças atrativas dentro da **zona de maior representatividade e de consistência entre os respondentes**, de acordo com a AGIs.
- Destacar os **distratores de itens que regrediram percentualmente** na análise dos ganhos normatizados.
- Identificar os **subconstrutos com percentual mais intensos e consistentes**, bem como as **concepções relacionadas dentro desse contexto** que mais se destacam.
- Comparar *lado a lado* as **concepções que mais se destacam dentro de cada contexto**, destacando aquelas que percentualmente **sobressaem de um contexto em relação a outro**, sem considerar uma mesma métrica.

Sugere-se que as análises interpretativas no texto que configuram tanto os domínios sobre o conteúdo no testes quanto das dificuldades/equívocos sejam realizadas verificando as informações contidas nos itens comentados que se encontram

no Apêndice A (pp. 7-210), pois eles possuem informações qualitativas e psicométricas (painel estatístico dos itens) detalhadas dos itens e dos distratores especificamente, incluindo índices clássicos corrigidos e gráfico de desempenho através dos diferentes níveis de acerto.

4.1.4.1 Perfil diante dos Índices de Dificuldade e de Discriminação dos Itens

A noção sobre um perfil das proficiências se refere a tentar identificar os modelos mentais relacionados aos níveis de domínio sobre o conhecimento, de acordo com as respostas corretas. Entende-se que esses níveis dentro da TCT possa ser sugerido pelos parâmetros de **dificuldade** e **discriminação**, na qual quanto mais fácil o item maior será o domínio diante do conhecimento que o item propõe a saber, e quanto maior a discriminação, melhor é a qualidade do item para afirmar sobre essas proficiências. Determinam-se assim dois critérios relevantes para afirmar sobre o **perfil de proficiências** dos respondentes em cada contexto:

- A identificação de uma zona de aglomerado de itens indicará as características de dificuldade e discriminação do perfil que são predominantes ou **mais representativas** dentro de cada grupo diante do conhecimento envolvido;
- Os itens nas zonas com parâmetros de discriminação considerado bom ou muito bom indicarão a **maior consistência** sobre o domínio que os alunos possuem diante do conhecimento envolvido.

Uma forma de comparar *lado a lado* as características psicométricas supostamente comuns entre os alunos nos diferentes contextos está em identificar os valores desses parâmetros dentro de critérios correspondentes. Sabe-se que há uma limitação em comparar dois grupos distintos em um mesmo contínuo de dificuldade na TCT, sendo um dos problemas bem conhecidos, que no entanto pode ser superado com os procedimentos da TRI, em que os procedimentos estatísticos mais robustos garantem de forma mais fiável uma comparação desses grupos dentro de uma mesma métrica. Apesar dessa ressalva, foram identificados na TCT esses valores comuns, que podem servir de comparação especulativa, bem como os valores de cada contexto sendo mais adequado na comparação dentro de seu próprio grupo e não especificamente entre eles.

Destacando as principais linhas de corte de admissão dos itens, é possível uma comparação preliminar das dificuldades de cada contexto. As linhas tracejadas inferiores correspondem ao “guessing” do item, sendo 20% para itens com 4 alternativas, e 25% para itens com 5 alternativas. A linha de 50% se refere a predominância de acertos em termos absolutos, sendo muitas vezes um critério utilizado para indicar a tendência ou força atrativa, sem considerar, entretanto, o desempenho dos sujeitos no teste. As linhas tracejadas superiores são critérios percentuais consideradas os termos ideais sendo 60% para itens com 4 alternativas e 62,5% para itens com 5 alternativas (y9, y11, 16, y18, y26). Diante disso é possível perceber no contexto português 1 (y1) item predominantemente com percentual ideal, e 3 no limiar do corte (y7, y12, y23). Somente 2 itens (y7, y12) no contexto brasileiro possui uma predominância absoluta (> 50%). Com relação aos itens, os que estão mais abaixo da linha de “guessing” são 4 itens no contexto brasileiro (y11, y14, y15, y18) e em Portugal apenas 1 item (y19), enquanto existem 5 itens no limiar para os brasileiros (y5, y13, y16, y20, y26), e 1 item no limiar para os portugueses (y9). Observa-se que praticamente todos os itens para os portugueses foram mais fáceis que para os brasileiros, sendo 3 itens com dificuldades quase iguais (y8, y9, y17), com exceção de um dos itens (y19), que demonstrou ser mais fácil para os brasileiros. A média das dificuldades de todos os itens para os brasileiros foi de 0,27, enquanto para os portugueses foi de 0,40.

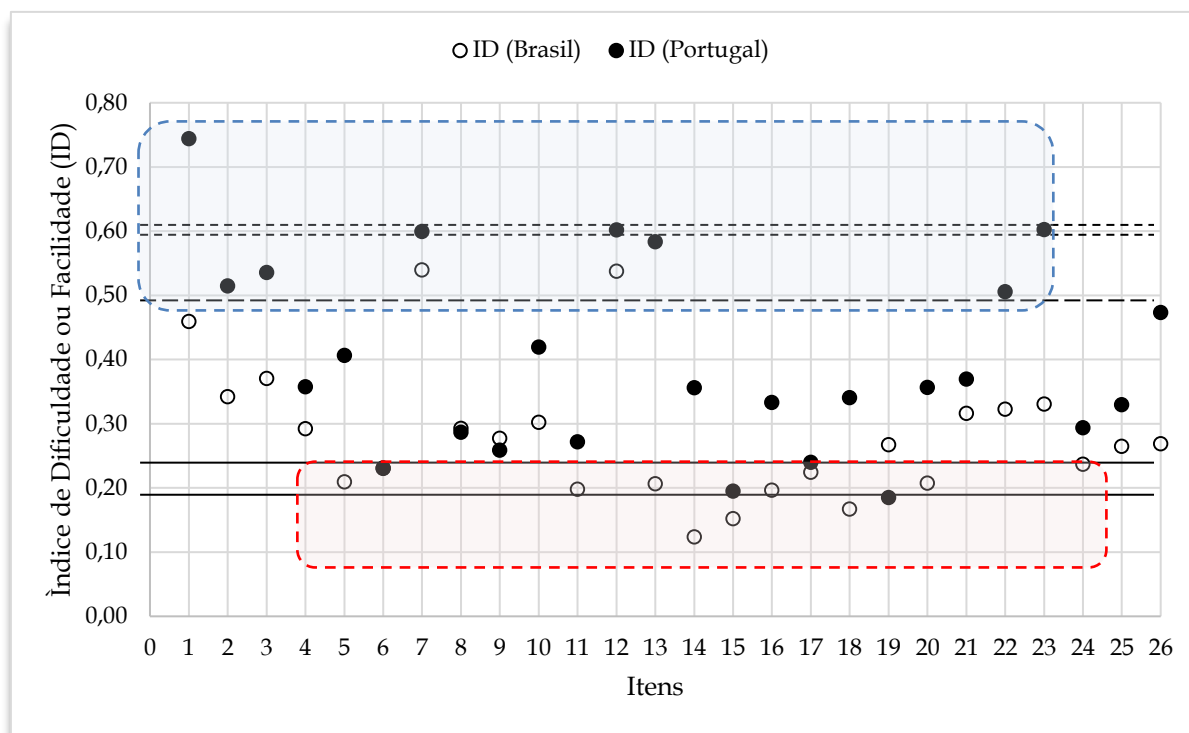


Figura 61: Comparação da dificuldades (percentual de acertos) dos itens entre Brasil e Portugal, tendo como referência as linhas de corte de admissão, segundo os critérios de "guessing", percentual de predominância absoluta (50%), e do percentual ideal mínimo a ser alcançado.

Sabe-se que a combinação dos valores dos parâmetros de dificuldade e discriminação dos itens possibilita identificar propriedades psicométricas mais úteis quanto as suas qualidades, e de acordo com os critérios preestabelecidos para os valores desses parâmetros é possível comparar diferentes grupos de tipicidade de itens dentro de cada contexto. Em uma perspectiva comparativa *lado a lado*, tem-se a seguinte configuração de grupos de itens, assim como aqueles com características coincidentes identificados:

Quadro 66: Identificação de grupos de itens dentro de cada contexto de acordo com os critérios de dificuldade (índice de dificuldade - ID), assim como itens coincidentes.

Tipologia do item	Proporção esperada	Limite aproximado	Brasil	Portugal	Itens coincidentes
Muito fáceis	10%	de 0,75 a 0,95	-	-	-
Fáceis	20%	de 0,55 a 0,74	-	y1, y7, y12, y13, y23	-
Normais/ Medianos	40%	de 0,45 a 0,54	y1, y7, y12	y2, y3, y22, y26	-
Difíceis	20%	de 0,25 a 0,44	y2, y3, y4, y8, y9, y10, y19, y21, y22, y23, y25, y26	y4, y5, y8, y9, y10, y11, y14, y16, y18, y20, y21, y24, y25	y4, y8, y9, y10, y21, y25
Muito difíceis	10%	de 0,05 a 0,24	y5, y6, y11, y13, y14, y15, y16, y17, y18, y20, y24	y6, y15, y17, y19	y6, y15, y17

Fonte: Elaborado com critérios baseados em Cerdá (1984) e Erthal (1987).

Quadro 67: Identificação de grupos de itens dentro de cada contexto de acordo com os critérios de discriminação (índice de discriminação - DISCR), assim como itens coincidentes.

Tipologia do item	Valores ⁹⁵	Brasil	Portugal	Itens coincidentes
Muito bom	$DISCR \geq 0,4$	y23	y3, y10, y13, y14, y16, y22, y23, y24	y23
Bom (sujeito a aprimoramento)	$0,3 \leq DISCR < 0,4$	y1, y2, y3, y7, y10, y22	y2, y5, y18, y20, y25, y26	y2
Marginal (sujeito a reelaboração)	$0,2 \leq DISCR < 0,3$	y4, y5, y8, y9, y13, y16, y19, y21, y24, y25, y26	y4, y8, y9, y11, y17	y4, y8, y9
Inadequado (sujeito a rejeição)	$DISCR < 0,2$	y6, y11, y12, y14, y15, y17, y18, y20	y1, y6, y7, y12, y15, y19, y21	y6

Fonte: Elaborado com critérios baseados em Rabelo (2013).

Associando os dois índices, se obtêm:

Quadro 68: Identificação comparativa de grupos de itens dentro de cada contexto de acordo com os critérios de dificuldade (ID) e de discriminação (DISCR).

Brasil						
DISCR \ ID	Muito fáceis	Fáceis	Normais/ Medianos	Difíceis	Muito difíceis	
Inadequado (sujeito a rejeição)	-	-	y12	-	y6, y11, y14, y15, y17, y18 e y20	
Marginal (sujeito a reelaboração)	-	-	-	y4, y8, y9, y19, y21, y25 e y26	y5, y13, y16, y24	
Bom (sujeito a aprimoramento)	-	-	y1; y7	y2, y3, y10 e y22	-	
Muito bom	-	-	-	y23	-	
Portugal						
DISCR \ ID	Muito fáceis	Fáceis	Normais/ Medianos	Difíceis	Muito difíceis	
Inadequado (sujeito a rejeição)	-	y1; y7; y12	-	y21	y6; y15; y19	
Marginal (sujeito a reelaboração)	-	-	-	y4; y8; y9; y11	y17	
Bom (sujeito a aprimoramento)	-	-	y2; y26	y5; y18; y20; y25	-	
Muito bom	-	y13; y23	y3; y22	y10; y14; y16; y24	-	

Na comparação das características dos dois parâmetros, de acordo com os critérios de corte, é possível constatar 3 itens que dentro da mesma região para as duas amostras, os itens y4, y8 e y9 sendo **marginais/normais** e **difíceis** para ambos os contextos.

De forma similar ao quadro, o gráfico ID *versus* DISCR possibilita uma visualização mais fácil para os diferentes grupos de itens (Vendramini, Silva, & Canale,

⁹⁵ Foi adotado cores correspondentes para melhor visualizar cada grupo dentro dos quadros e nos seus respectivos gráficos comparativos.

2004), assim como de aspectos característicos comparativos entre os contextos, seja pela diferença quanto por questões de similaridade.

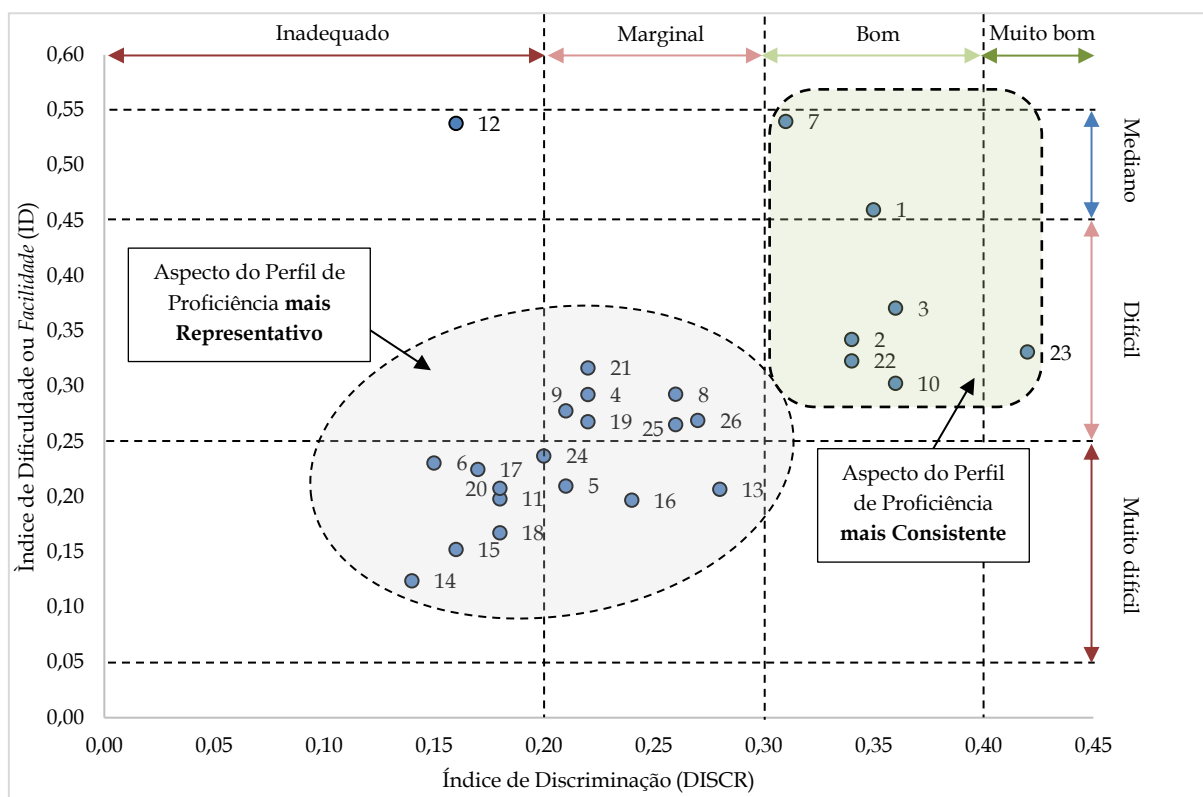


Figura 62: Representação gráfica de dispersão dos itens, segundo os índices de discriminação e de índice de facilidade (ou dificuldade) na amostra da Grande Florianópolis (Brasil).

De modo mais amplo é possível determinar que no Brasil, existe uma certa concentração de itens nas regiões caracterizadas como *inadequados e muito difícil*, *marginais e difíceis*, e, *marginais e muito difíceis*, isso significa dizer que grande parte das informações do teste para os brasileiros estão numa zona de baixa consistência, sendo difícil afirmar em grande parte sobre o domínio dos alunos diante do conhecimento técnico que o teste apresenta. Ainda assim, na zona de *discriminação bom e muito bom* encontram-se os itens (y1, y2, y3, y7, y10, y22, y23) que de forma mais fiável representam os conhecimentos em que os alunos brasileiros mais dominam no teste (26,9%), na qual pode-se afirmar os aspectos de determinam o perfil de proficiências desse grupo.

Em Portugal, a concentração dos itens é menor por estarem mais bem distribuídos quanto a discriminação, porém mais aglutinados nas regiões de difícil e muito difícil, seguindo o padrão do contexto brasileiro, apesar de conterem uma

quantidade maior de itens com boa discriminação, e serem admitidos 5 itens fáceis (y1, y7, y12, y12 e y23), enquanto **no Brasil não há item fácil**. Isso sugere que o teste tenha sido mais fácil para os alunos portugueses de modo geral.

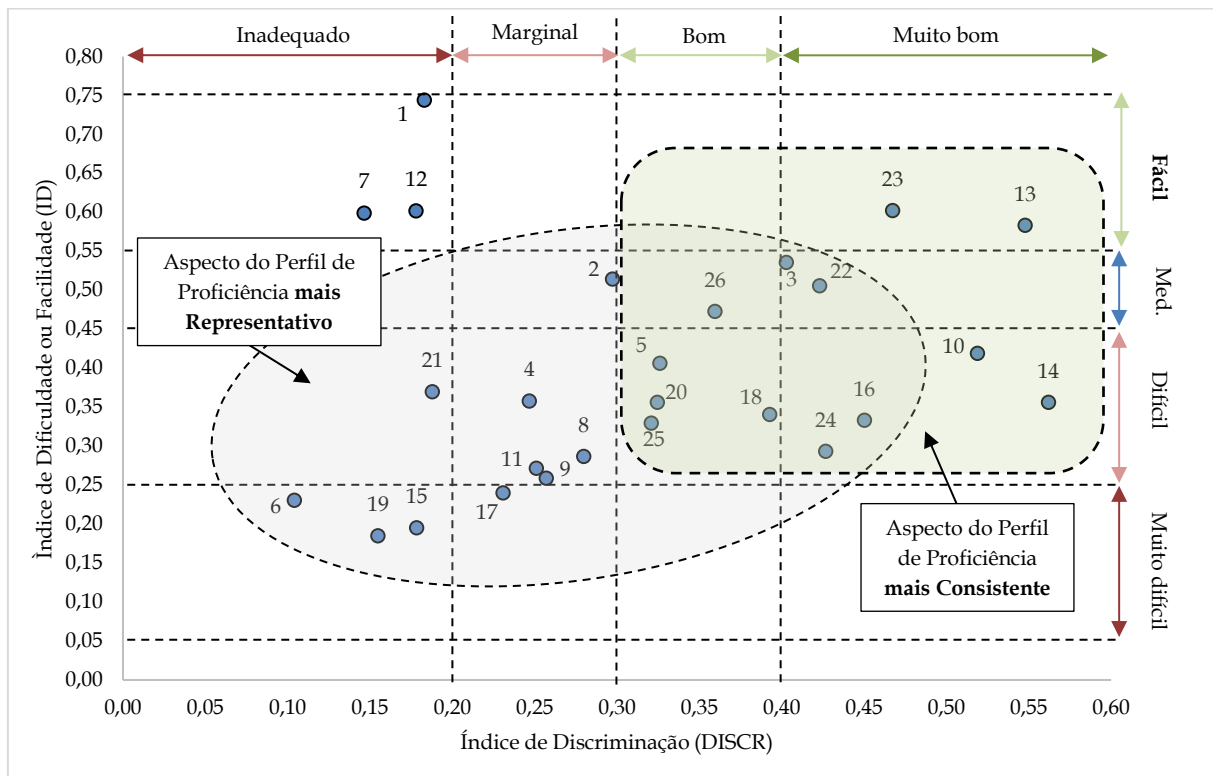


Figura 63: Representação gráfica de dispersão dos itens, segundo os índices de discriminação e de índice de facilidade (ou dificuldade) na amostra da Região do Centro (Portugal).

O cruzamento da matriz ID *versus* DISCR ainda possibilita a visualização e comparação de grupos de itens específicos entre os contextos, como por exemplo, existem 6 itens com *discriminação boa* (y1, y2, y3, y7, y10, y22) e apenas 1 item *muito bom* (y23), porém sendo difícil. Portugal apresenta nessas mesmas regiões 5 itens com *discriminação boa* (y5, y18, y20, y25, y26) e 8 itens *muito bons* (y3, y10, y31, y14, y16, y22, y24), sendo dois deles considerados fáceis.

Nos dois contextos não foram obtidos itens com alto valor de facilidade ou de proporção de acerto ($0,75 \leq ID \leq 0,95$: *muito fáceis*), e pode ser notado que Portugal apresenta uma quantidade maior de itens com bons níveis de discriminação comparado ao Brasil. Chama atenção o comportamento do item **y12**, que se comporta de modo discrepante (*outlier*) diante dos outros itens para os alunos brasileiros, porém esse mesmo item se encontra numa região do gráfico equivalente para o contexto

português. Verifica-se que o grupo com discriminação mediana ou marginais (sujeitos a reelaboração) no Brasil é bem maior que em Portugal. O item **y7** merece destaque por apresentar um elevado percentual de acertos nos dois contextos, porém, em Portugal possui um valor de discriminação mais baixo, ou seja, não apresenta um elevado poder para distinguir alunos de alto e baixo desempenho, apesar de que no Brasil esse índice também não seja considerado de grande magnitude. O item **y1** possui uma dificuldade intermediária no Brasil, porém acima da maioria dos outros itens, e sendo considerado o item mais fácil para os alunos portugueses. O item **y23** por sua vez apresenta o maior poder de discriminação para a amostra brasileira, isso justifica por ser o item com maior valor para o coeficiente de correlação bisserial desse grupo de respondentes ($r_{biss} = 0,49$), apesar de apresentar características psicométricas semelhantes para os dois contextos.

4.1.4.1.1 Especificidades no Perfil diante das Médias dos Escores

Com relação ao contexto brasileiro, ao levar em consideração no *design* que todas as escolas do 1º ano sejam o grupo do pré-teste (alunos que não receberam instruções ou aulas de termodinâmica), e que os grupos do 2º ano e 3º ano sejam grupos pós-teste (receberam as instruções, supostamente), é possível obter uma análise mais geral dos ganhos tendo como referência a variável nível de escolaridade, porém sem levar em conta as diferenças existentes entre os diferentes turnos.

O fato é que nunca se tem acesso aos dados de toda a população, mas sempre na busca de obter uma amostra significativamente representativa, por isso é que não temos certeza sobre as diferenças em determinadas características dos indivíduos, e por esse motivo devemos utilizar os testes de hipóteses para verificar essas diferenças. Diante disso, faz-se o tratamento estatístico do teste-t *student* para grupos independentes com o pressuposto de haja diferenças significativas entre as variâncias, sendo um procedimento minimamente necessário a ser verificado preliminarmente para constatar a existência ou não de eventuais diferenças entre as médias dos escores nos diversos grupos normativos considerados.

No caso do contexto português, o grupo correspondente do pré-teste é o 10º ano, sendo, portanto, o nível de escolaridade sugerido como referência nas comparações, conforme o *design* estabelecido.

Entre os diferentes níveis de escolarização, é possível notar uma sutil discriminação que decresce do 1º ano no turno matutino ($\bar{X}_i = 0,277$), passando pelo vespertino ($\bar{X}_i = 0,267$) e depois noturno ($\bar{X}_i = 0,228$). A análise do teste-t *student* para amostras diferentes com previsão de variâncias diferentes mostra que **não há diferenças significativas entre as médias do 1º ano matutino e 1º ano vespertino** ao nível de significância de 5% ($t = 1,36$; unicaudal à direita; $t_{\text{crítico}} = 1,65$; $t < t_{\text{crítico}}$; Não se rejeita $H_0: \mu_{1^\circ \text{ mat}} = \mu_{1^\circ \text{ vesp}}$), porém **há diferenças significativas com o 1º ano noturno**. Para as médias dos escores no 2º ano, também não houve diferenças significativas do turno matutino com o vespertino, **somente do 2º ano matutino com o 2º ano noturno**. Para o **3º ano não houve diferenças** significativas entre os turnos (ver no Apêndice E5, p. 270).

Considerando o aglomerando as escolas de 1º ano, 2º ano e 3º ano, juntando todos turnos, é possível verificar, de modo geral, **diferenças significativas entre as médias do 1º ano e 2º ano**, mas **não entre o 1º ano e o 3º ano** (ver no Apêndice E7, p. 271). Uma interpretação pedagógica diante disso sugere afirmar que, em média, a maioria dos alunos não desenvolveu satisfatoriamente as competências científicas de base conceitual dentro da Termodinâmica devido a uma falta ou ineficiência dentro do processo de aquisição conceitual no nível introdutório da Física Térmica, sendo agravado ao avançar entre os níveis de escolarização ao invés de ser reforçado.

Um aspecto secundário estabelecido para relativizar as análises das médias e do ganhos normatizados para os grupos de 2º e 3º ano (e os correspondentes 11º e 12 ano), pode-se observar as médias para os diferentes grupos de normatização considerados, que no caso são as escolas que adotaram determinados Livros Didáticos de Física (LDFs). Interessante notar que o livro didático de Física mais difundido entre as escolas de ensino médio no Brasil, dentro do contexto histórico da pesquisa, foi o que obteve a menor média, correspondendo a 6,5 acertos (≈ 7 acertos), abaixo do “*guessing rate*”,

ou seja, os alunos não dominam o conteúdo e obtiveram escore através de acerto aleatório. De modo geral, todas as médias são baixíssimas.

Diante disso, pode-se dizer que para os grupos de alunos que adotaram os mesmo livros didáticos de Física, sabe-se que 2br obteve maior média ($\bar{X}_{2br} = 0,317$), enquanto o 7br sendo o mais disseminado obteve a média mais baixa entre os 6 que contemplam a amostra ($\bar{X}_{7br} = 0,252$). O teste-t *student* **constatou diferenças significativas do grupo de escolas (14% da amostra) que adotaram o 2br com todos os grupos de escolas** que adotaram os outros LDF, sendo a maior diferença com a média do 7br e a menor diferença com o 6br (ver no Apêndice E6, p. 270). Essas evidências indicam a necessidade da realização de uma investigação complementar mais profunda diante da qualidade científico-pedagógica do LDF 2br como possível forma de tratamento a acarretar efeitos de causalidade no desempenho dos alunos. Essa qualidade científico-pedagógica se refere aos encargos pedagógicos contidos nos LDF que exercem de forma mais precisa a *função de desenvolvimento de capacidades e competências* (Gérard & Roegiers, 1998), nomeadamente aquelas que estão à serviço da aprendizagem, mas que de certa forma, poderá estar influenciando na organização, estruturação, abordagem, desenvolvimento, sequenciação, e avaliação do ensino, dependendo do perfil docente, mas que entretanto, também depende dos hábitos de aprendizagem. Essas são algumas das variáveis que devem ser verificadas e controladas nesse contexto investigativo complementar.

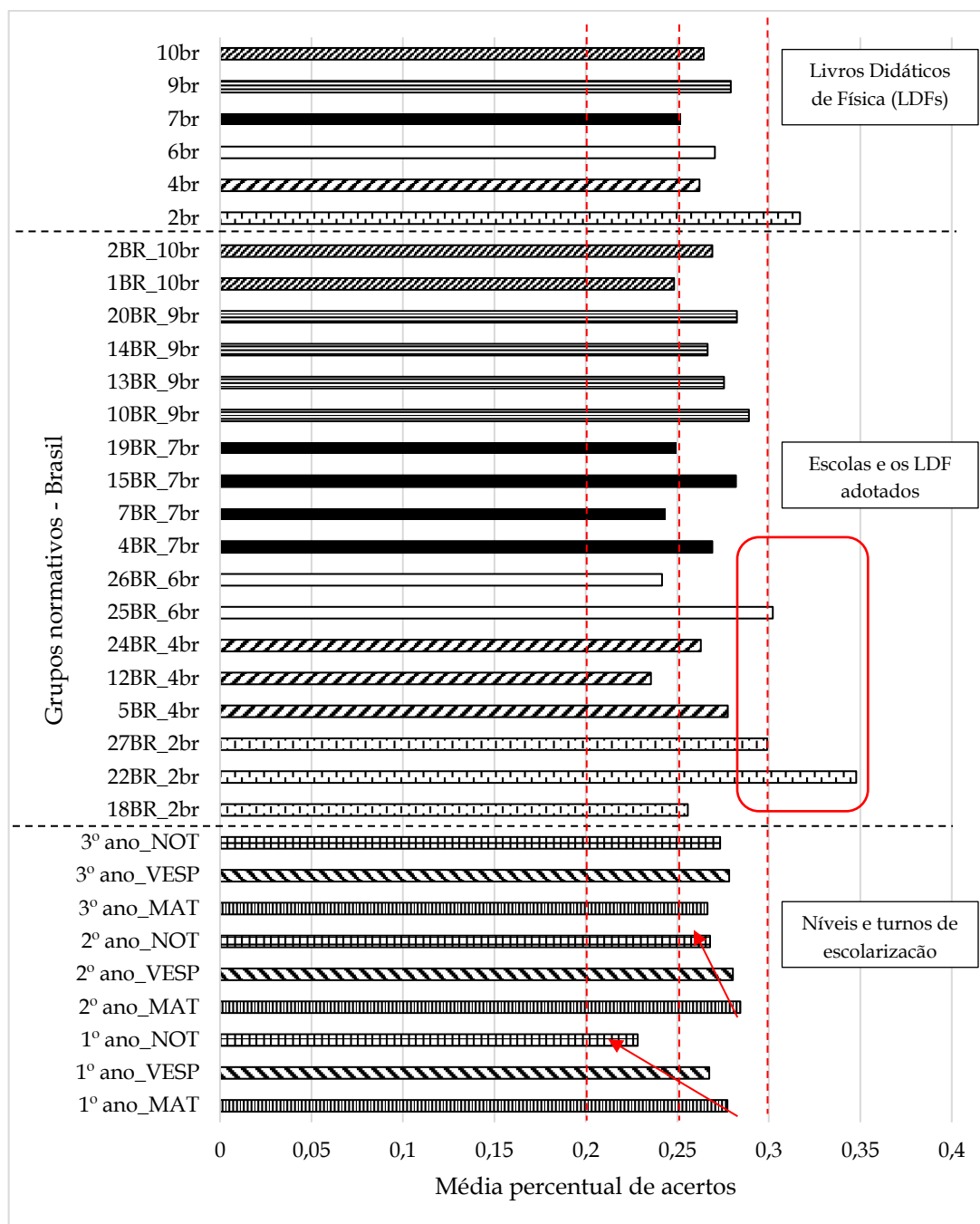


Figura 64: Média percentual dos escores brutos dos principais grupos normativos do estudo no contexto Brasil.

Pode-se destacar as escolas 22BR e 27BR com o livro didático 2br e a escola 25BR com o livro de Física 6br por estarem com uma média atingindo os 30%, que infelizmente, em termos absolutos ainda se encontram bem distantes de um patamar considerado do ideal. Essas diferenças são facilmente observadas através do desempenho das escolas brasileiras ao subtrair suas porcentagens com relação à média

geral entre elas ($\bar{\bar{X}}_i = 0,274$), mostrando de forma mais precisa percentualmente quais são as escolas que estão acima e abaixo da média.

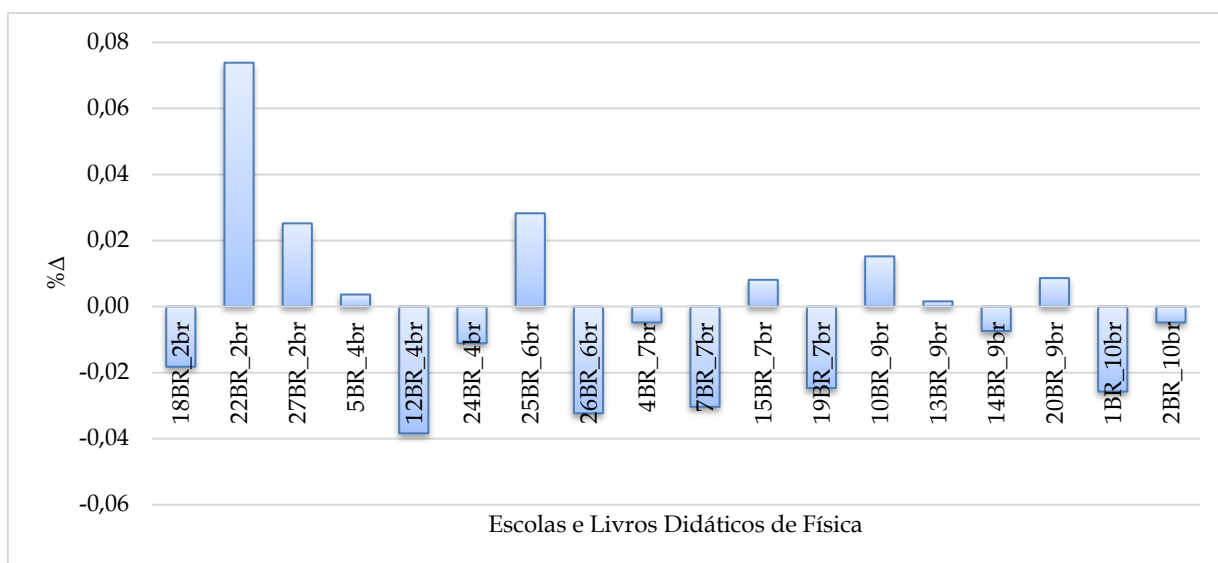


Figura 65: Variação percentual dos acertos das escolas em relação à média da amostra no contexto Brasil.

Em Portugal o cenário muda um pouco comparativamente, de forma mais consistente e com qualidade, mas não demonstra ser muito promissor de modo geral quanto aos critérios minimamente ideais de qualidade e expectativas favoráveis do ensino em Termodinâmica.

Considerando os diferentes grupos normativos, de forma equivalente ao Brasil, é possível preliminarmente observar uma maior discriminação percentual na média entre o 10º e o 11º, e entre 11º e o 12º ano, como era de se esperar, pois alunos em níveis de escolarização mais elevado devem possuir naturalmente maior nível de proficiência ou de habilidades diante de um mesmo conteúdo que supostamente tiveram contato e foram instruídos. De acordo com as análises verificativas, o teste-t *student* constatou diferenças significativas entre as médias dos escores do 10º e 11º ano, e bem mais entre o 10º e 12º ano.

A média dos escores para os alunos nas escolas que adotaram o LDF 8 possui uma sutil vantagem em relação ao 11, sendo o LDF 11, o mais adotado em Portugal continental, bem como para a população considerada no estudo.

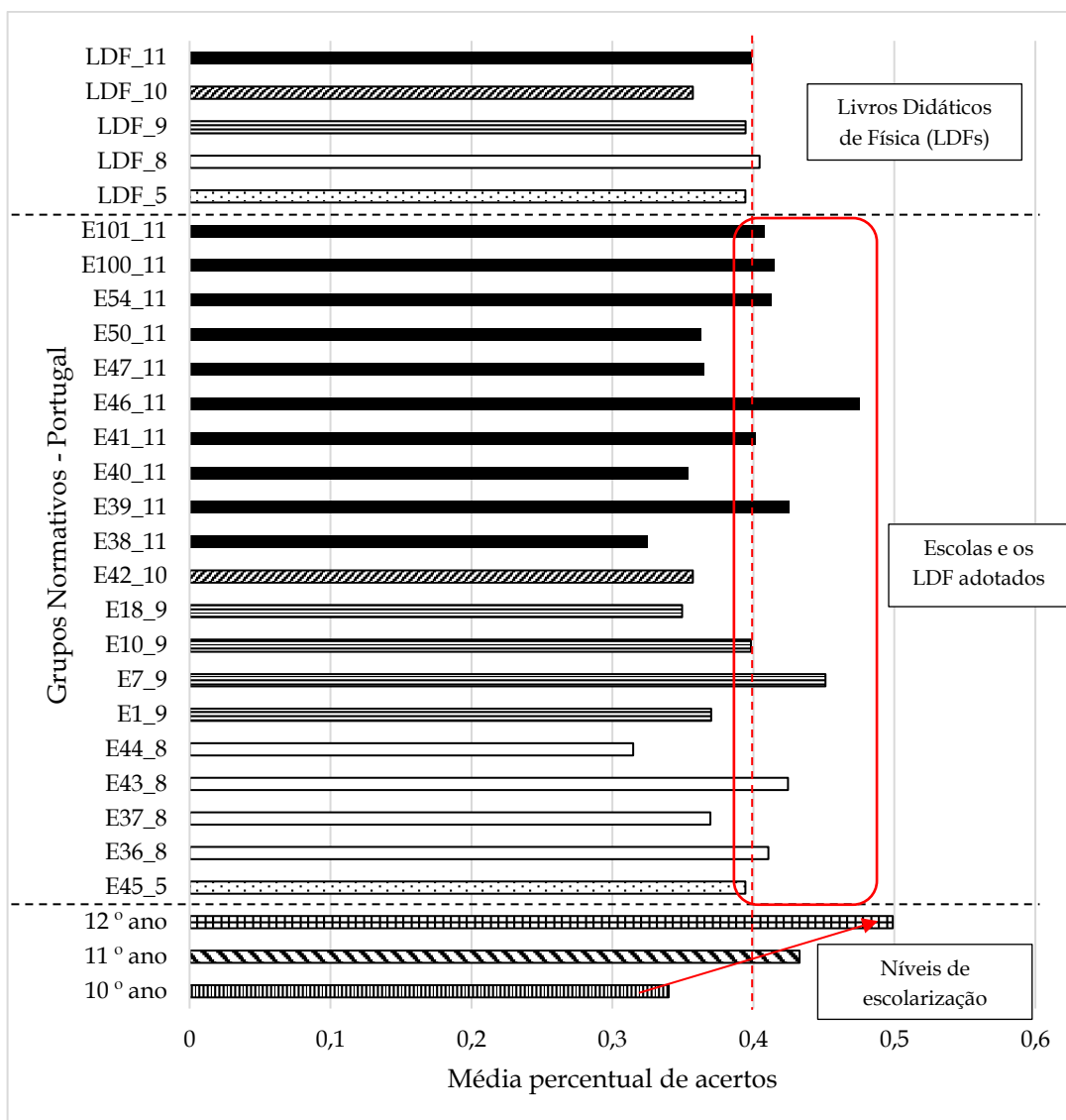


Figura 66: Média percentual dos escores brutos dos principais grupos normativos do estudo no contexto Portugal.

Realizando o teste-t *student* para amostras independentes supondo variâncias diferentes entre os grupos de escolas do LDF 8 e 11, observa-se que $t(0,663) > 0$ (uni caudal à direita), com isso tem-se: $H_0: \mu_8 = \mu_{11}$ ou $H: \mu_8 > \mu_{11}$. Assim, sendo $t < t_{\text{crítico}}$ ($0,663 < 1,648$), não rejeita-se H_0 ao nível de 5% de significância, portanto, as médias dos dois grupos não são significativamente diferentes, em outras palavras, supostamente os livros didáticos 8 e 11 adotados não sugerem diferenças entre as médias dos escores de desempenho no teste TCE. Da mesma forma não há diferenças entre o LDF 8 e 9 ($t = 0,909$; $t_{\text{crítico}} = 1,647$), nem entre o 8 e 5 ($t = 0,720$; $t_{\text{crítico}} = 1,650$), havendo diferenças significativas somente entre os grupos normativos relacionado ao LDF 8 e 10 ($t = 3,124$; $t_{\text{crítico}} = 1,650$), ou seja, ou o LDF 10 possui de fato baixa

qualidade, ou esse é apenas um reflexo de um contexto pontual pelo fato de ser representado somente pela escola E42, que corresponde a 6,85% da amostra.

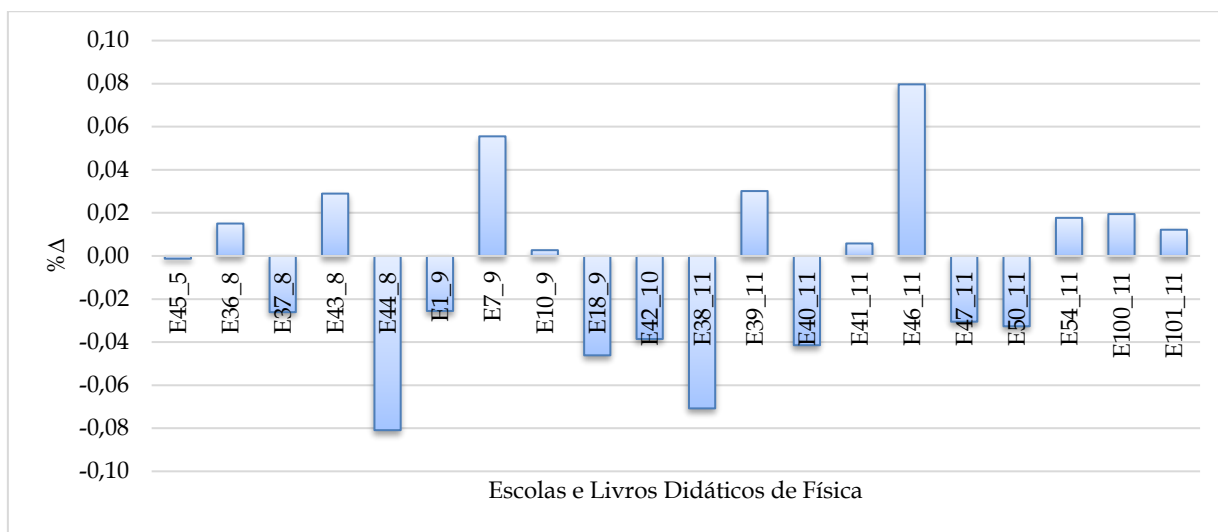


Figura 67: Variação percentual dos acertos das escolas em relação à média da amostra no contexto Portugal.

Na figura 67 é possível verificar as 10 escolas que ultrapassaram o patamar de 40% de acertos, conforme mostra a média dos valores desse grupo normativo na figura 66. Diante da variação percentual (%Δ) entorno da média entre elas ($\bar{\bar{X}}_i = 0,396$), grande parte desses grupos encontra-se abaixo da média, sendo as escolas E7 e E46 as que mais apresentam maior variação em relação à média, ao isolar a categoria sobre os LDFs, isso mostra o potencial dessas escolas na determinação da média, podendo estar relacionada a fatores pedagógicos internos, provavelmente sem relação com os LDF envolvidos, pois o teste-t *student* não demonstra diferenças significativas entre as médias desses grupos ($t = -0,487$; $t_{\text{crítico}} = -1,647$).

4.1.4.1.2 Ganhos nos Percentuais de Acerto

Ao considerar, de certa forma grosseira, que o 1º ano seja um grupo pré-teste, e os outros pós-teste, como foi dito, é uma admissão que deve remeter a novas interpretações da curva original proposta por Hake, pois consideram-se diversos grupos distintos para cada nível de escolaridade, apesar de manter como referência essa categoria. No caso, tanto para o 2º como para o 3º ano, admite-se o 1º ano como sendo o pré-teste (alunos que ainda não haviam se submetido ao tratamento). Isso remete a considerar que os ganhos percentuais brutos (%<G>) e ganhos normalizados

($\langle g \rangle$) devam evoluir em uma linha vertical fixa, fazendo correspondência a uma abscissa comum dentro do gráfico, em que o valor é a média geral do 1º ano ($\bar{X}_i = 0,267$). Com essas hipóteses, foi possível constatar que os ganhos foram quase que nulos, sendo o ganho do 2º ano superior a do 3º ano, contrariando as expectativas.

Tabela 26: Ganhos normalizados com relação as médias gerais do 2º e 3º ano, tendo como referência a média geral do 1º ano.

$\langle g \rangle$ 2º ano	$\langle g \rangle$ 3º
$0,019 \pm 0,04$	$0,003 \pm 0,01$

A análise dos ganhos percentuais no gráfico $\% \langle G \rangle$ versus $\% \langle \text{pré} \rangle$ do 2º ano e 3º ano permite visualizar que diante da média dos alunos do 1º ano **não houve mudança após as aulas de termodinâmica.**

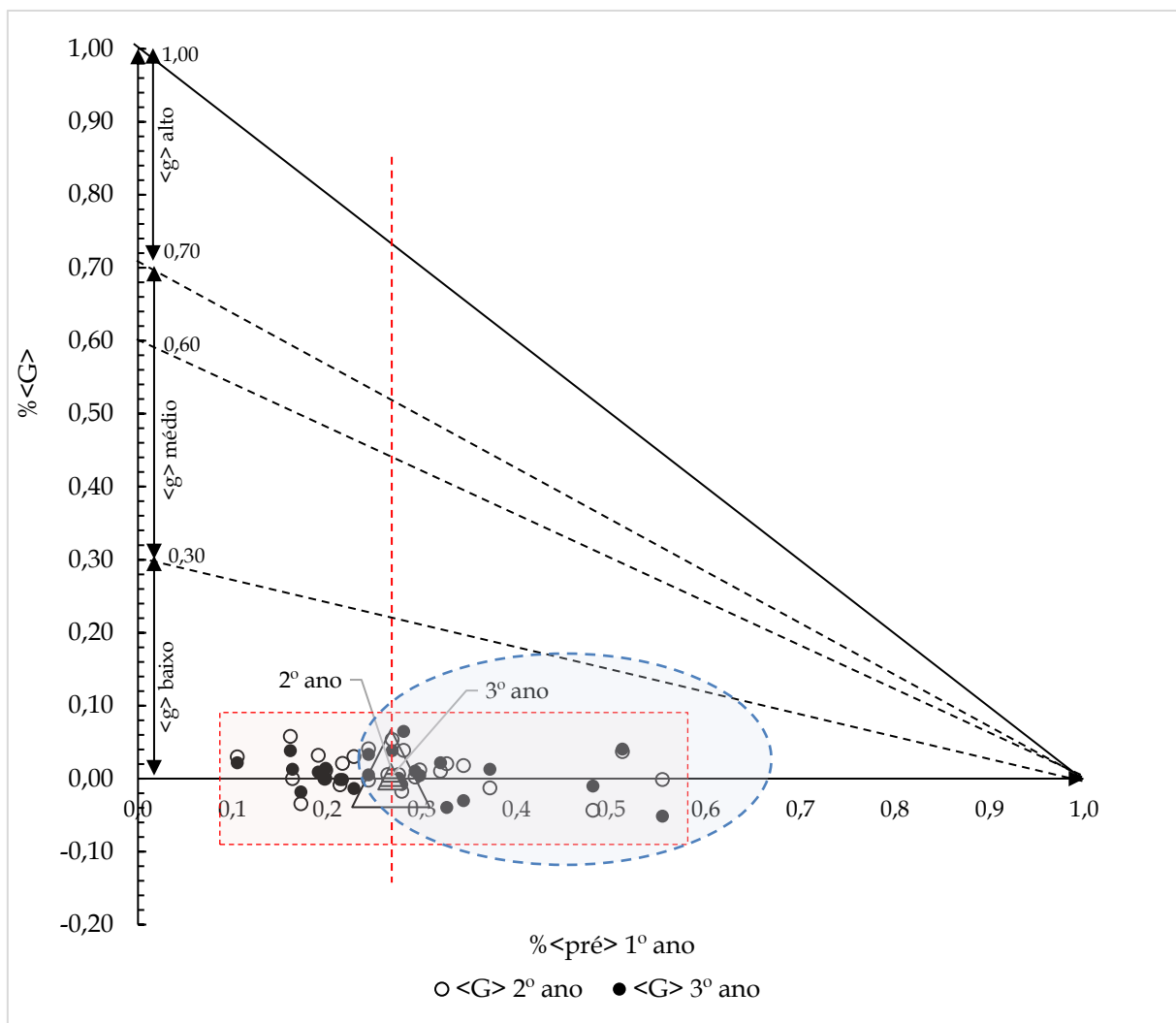


Figura 68: Gráfico $\% \langle G \rangle$ do TCE versus $\% \langle \text{pré} \rangle$ no contexto Brasil, comparando o 1º ano com os ganhos do 2º e 3º ano.

No sentido apresentar maiores detalhes do gráfico, observa-se que além da média dos ganhos normalizados dos grupos, conforme Hake apresenta em seu trabalho, estão inclusos os ganhos normalizados dos itens. Entende-se que esse tipo de inclusão somente faz sentido porque existem apenas dois grupos a serem comparados, pois se fossem mais grupos, a inclusão das informações desses itens tornaria o gráfico demasiadamente confuso, porém, não é o caso.

De modo geral, eles praticamente permanecem com as mesmas dificuldades que antes das aulas, considerando ainda perda percentual no acerto de alguns dos itens para o 3º ano, isso é um sinal de grandes necessidades para o aprendizado e ineficiência para o sistema de ensino, sendo uma evidência que sugere grandes grandes preocupações pedagógicas para a amostra. Apesar disso, deve-se ressaltar que essas perdas em parte representam um efeito previsto na concepção de Ausubel, no que se refere ao processo de *obliteração subsunativa*. Diante disso, especialistas recomendam quando não haja um distanciamento acima de 6 meses entre o pré e o pós teste. Como foi dito, o pré-teste não ocorreu nesse âmbito, considera-se uma análise de grupos distintos, sendo é um risco interpretativo assumido, assim como do tempo demasiado entre as aulas de termodinâmica para os alunos do 2º e 3º ano, apesar disso, é natural de haja em média níveis mais elevados para as habilidades cognitivas.

Um aspecto positivo minimamente a ser considerado seria dos **itens que se mantêm em níveis acima do “guessing”**. Para o 2º ano tem-se os itens: y1, y2, y3, y4, y7, y8, y9, y10, y12, y19, y21, y22, y23, y24, y25, y26. E para o 3º ano: y1, y2, y3, y4, **y5, y6, y7, y8, y9, y10, y12, y17, y19, y20, y21, y22, y23, y24, y25, y26**, sendo 5 deles a mais que no 2º ano (y5, y6, y10, y17, y20).

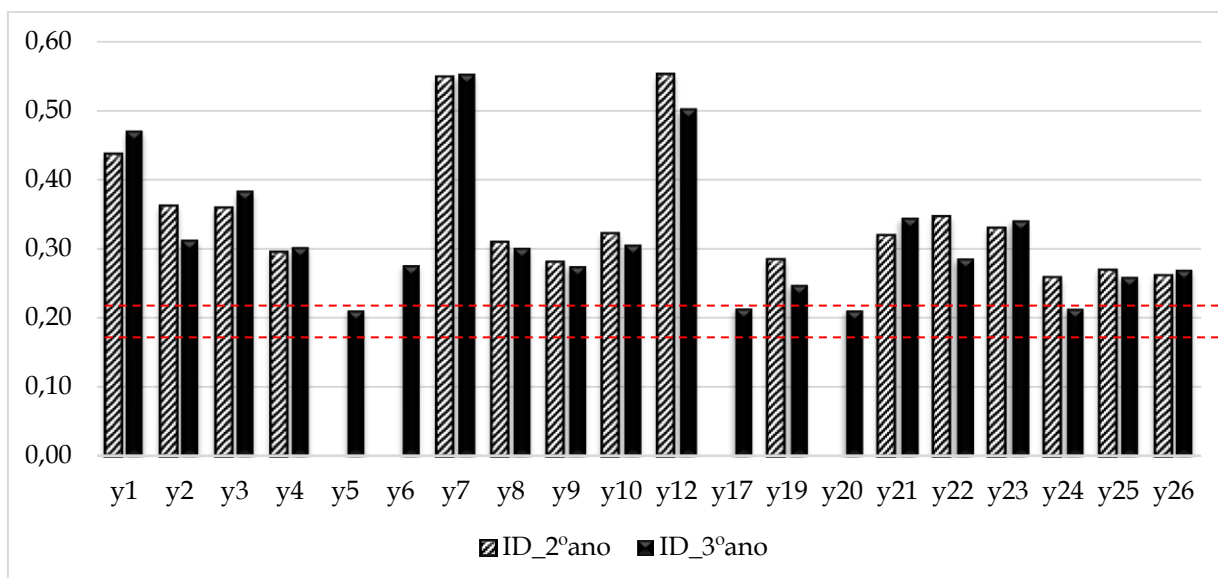


Figura 69: Índices de Dificuldades do 2º e 3º ano - Brasil.

Quanto aos itens do 2º ano que **regrediram percentualmente** ao serem comparados com as médias dos alunos do 1º ano, constata-se uma perda relativamente baixa, tendo em média de -1,5. Isso pode demonstrar tanto resistência dos distratores, como lacunas na instrução de ensino, ou problemas no itens. Supondo que as causas se concentrem nos distratores, pode-se destacar aqueles que estavam previamente na zona de “guessing” ($\leq 0,25$), e que ainda obtiveram fortes perdas (y5, y15). Também chama a atenção na perda de 4,4% do item y1, apesar de permanecer como um item de dificuldade normal.

Para o 3º ano, a perda foi em média -1,8% entre os itens que regrediram, sendo 5,1% para o item y12 (dificuldade normal) e 4% para o item y22 (difícil). Observa-se uma redução percentual mais intensa no 3º ano que no 2º ano, ou seja, os distratores parecem ser mais atrativos, e isso contraria o ideal esperado, que seriam os alunos do 3º serem mais proficientes. Comparando com o 2º ano, é possível notar que o índice de dificuldade são similares para alguns itens que possuem fortes distratores (y1, y5, y12, y26).

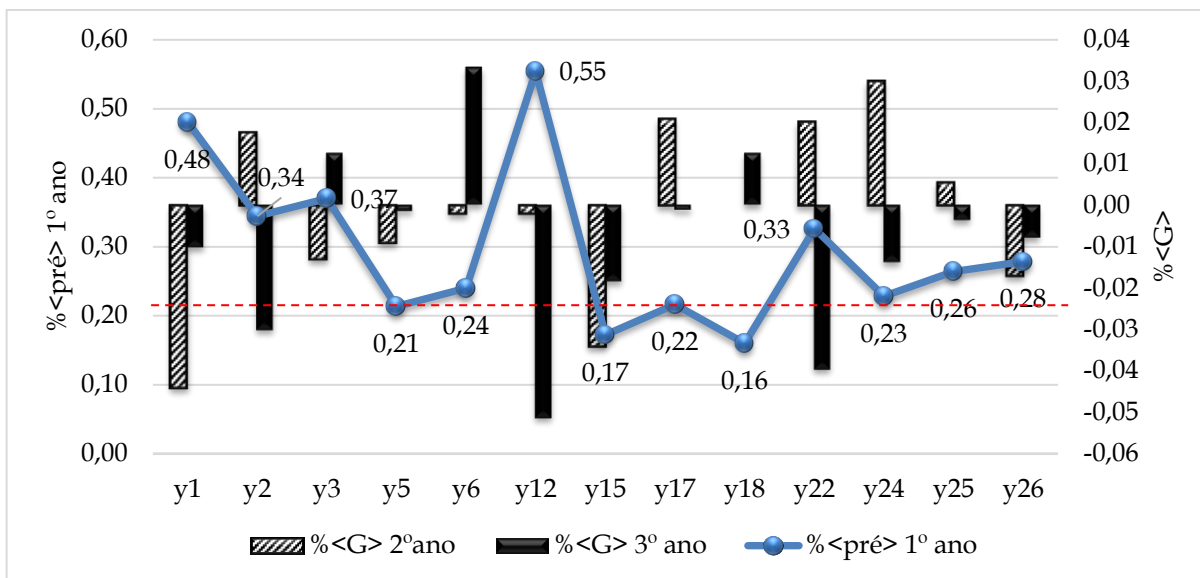


Figura 70: Detalhamento dos ganhos percentuais do 2º e 3º ano (pós-teste) em relação ao percentual do 1º ano (pré-teste).

Analisando os ganhos percentuais no contexto português, tem-se o 10º ano como referência de grupo como pré-teste, e os outros como pós-teste, e a análise segue de forma similar à anterior. Os ganhos normatizados mostram na mesma lógica e os valores evidenciam um ganho progressivo com o avanço dos níveis de escolaridade, porém os impactos são considerados baixos ($\langle g \rangle < 0,30$).

Tabela 27: Ganhos normatizados com relação as médias gerais do 2º e 3º ano, tendo como referência a média geral do 1º ano.

$\langle g \rangle$ 11º ano	$\langle g \rangle$ 12º
$0,14 \pm 0,36$	$0,24 \pm 0,40$

A análise dos ganhos percentuais no gráfico %<G> versus %<pré> do 10º ano e 11º ano permite visualizar que diante o desempenho dos alunos do 1º ano **houve poucas mudanças após as aulas de termodinâmica** de um ano para o outro, estando os ganhos médios normatizados em uma região de baixo impacto. Segundo Hake, isso sinaliza preocupações pedagógicas nesse contexto, devendo com isso ser investigado em uma perspectiva complementar.

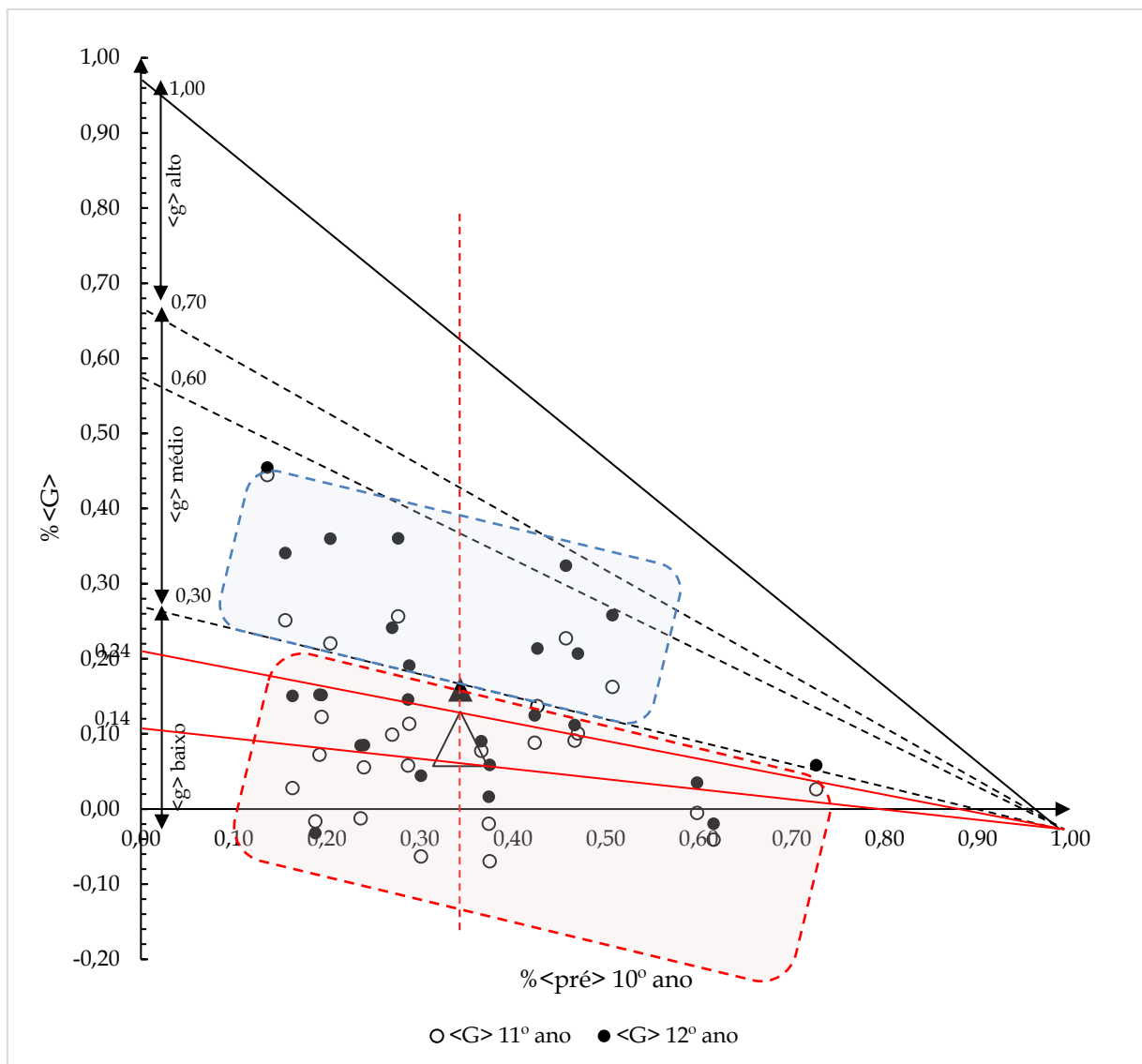


Figura 71: Gráfico %<G> do TCE versus %<pré> no contexto Português, comparando o 10º ano com os ganhos do 11º e 12º ano.

Percebe-se que os ganhos dos itens tendem a se acumular na zona de médio impacto, porém a concentração continua na região de baixo impacto. Um critério positivo a ser considerado na delimitação de um perfil de proficiências seriam os itens na região de acima de $\langle g \rangle = 0,30$, ou seja, informações pedagógicas que configuram um certo domínio dos alunos no conhecimento envolvido.

Convém ressaltar que os ganhos normalizados são tradicionalmente expressos no gráfico para comparar ganhos de populações ou amostras diferentes, e não dos itens, apesar de que os ganhos normalizados dos itens estejam sempre implícitos. Como o propósito segue na identificação de ganhos dos grupos considerados, mas

também dos itens, pois eles sinalizam a apropriação de conteúdos e de dificuldades existentes, portanto não há motivos que impeçam de fazer e explorar essa perspectiva.

Seguindo o critério estabelecido, é possível observar que os ganhos normalizados dos itens para o 12º ano foram maiores que para o 11º ano, com dois itens a mais (y16, y18), estando com médio no impacto, com exceção do item y13 para o 12º ano, na qual alcançou a inclinação que corresponde a um alto impacto. O item y13 remete a uma compreensão sobre a propagação do calor (Descritor 9), especificamente quanto ao sentido da «transferência de energia térmica que ocorre durante o processo de resfriamento de ovos cozidos quando submetidos em água fria», ou seja, saber qual o corpo que é aquecido e esfriado quando postos em contato. Repare que alguns desses itens para os quais houve ganho, já eram considerado relativamente normais/medianos para os alunos no 1º ano (y3, y13, y22, y23), e pode-se destacar aqueles que correspondiam a um baixo percentual, mas que tiveram elevados ganhos normalizados (y14), sendo alguns obtidos somente no grupo de alunos do 12º ano (y16, y18, y24).

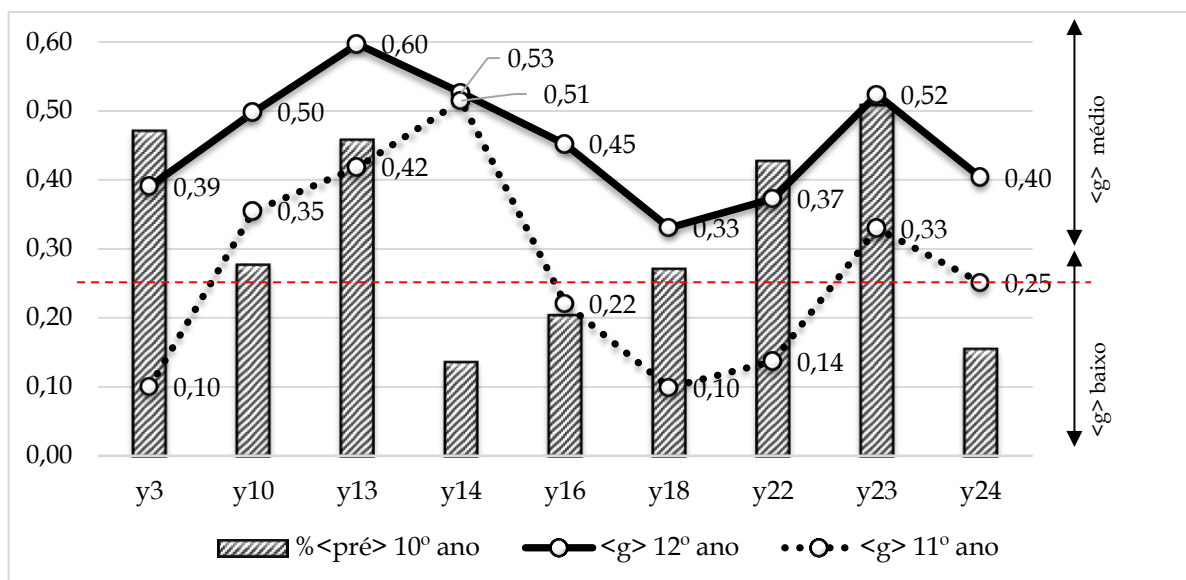


Figura 72: Ganhos percentuais normalizados do 11º e 12º ano - Portugal.

Para finalizar, quando se analisa as médias percentuais gerais para o Brasil e Portugal, ao considerar a média dos escores de todos os alunos em cada contexto e verificar se há diferenças significativas, **constata-se uma diferença a favor de Portugal**

($t = -34,86$; $t_{\text{crítico}} = -1,65$), ou seja, rejeita-se a hipótese nula, e afirma-se que há diferenças a um nível de significância de 5% (ver no Apêndice E7, p. 271).

4.1.4.1.3 Fator de Concentração nas Análises «C»

A verificação dos valores do *Fator de Concentração* dos itens (ver no Apêndice E8, p. 271) é um tipo de análise que fornece informações relacionadas de regiões gráficas específicas que sinalizam a existência de evidências representativas de 1 ou 2 modelos mentais predominantes, dependendo do número de incidências das respostas (“picos”) em determinada alternativa, podendo ser corretos ou incorretos. Dessa forma, entende-se que os modelos corretos são informações que se associam com a aquisição de um conhecimento mais consistente por parte dos alunos, e os modelos incorretos associados aos distratores e concepções equivocadas mais atrativas e representativas a cada contexto, porém, torna-se necessário verificar de que forma esses modelos são representados para os diferentes grupos normativos que formam a amostra, pois, essas distinções torna-se uma forma de controle e maior rigor nas análises, evitando com isso conclusões gerais enviesadas.

A concentração de itens no contexto brasileiro é bastante reduzido, com 3 itens na região MM, porém sendo comuns entre os alunos nos diferentes níveis de escolaridade, na qual são os y1, y7 e y12, isso pode sugerir a existência dos modelos corretos e incorretos mais representativos na caracterização de perfil de proficiências e de concepções térmicas para esse contexto. Na região BM tem-se a existência de um item comum aos alunos do 2º e 3º ano, sendo o y14, que por sua vez deve sugerir a presença de dois modelos incorretos predominantes. De modo geral, pode-se se notar que a maioria dos itens estão na zona de acerto ao acaso, e isso demonstra muitas dificuldades dos alunos brasileiros ao responder o teste TCE entre todos os níveis de escolaridade. É possível notar que a média das concentrações (<C_1ºano>; <C_2º ano>; e <C_3º ano>) dos diferentes níveis estão praticamente sobrepostos, com uma sutil superioridade do 2º ano, porém, não sendo significativo. Isso reforça na afirmação de que, em média, os conhecimentos entre os diferentes alunos são praticamente os mesmos, e que não houve progressão na aquisição dos conceitos básicos e

introdutórios da Física Térmica após as aulas de Termodinâmica, e que houve pouco reforço desses conteúdos na sequenciação do ensino-aprendizagem.

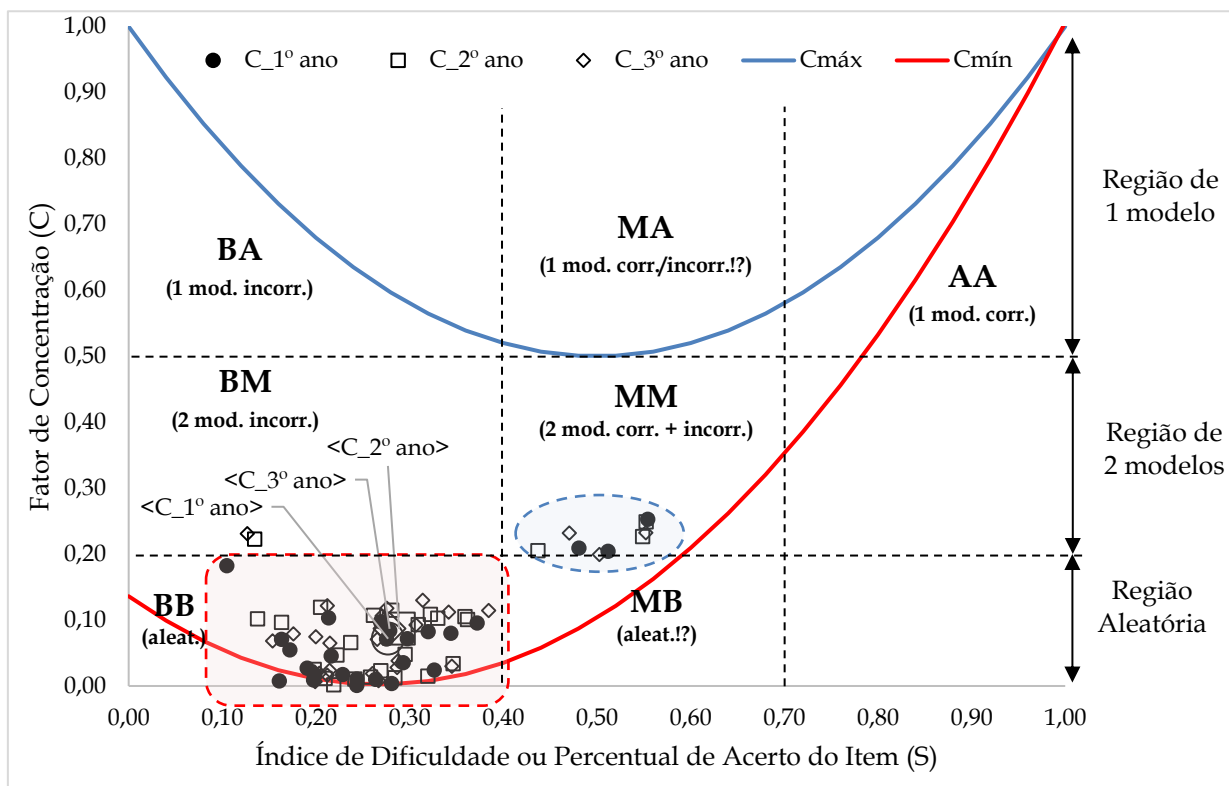


Figura 73: Gráfico S-C plot para o 1º, 2º e 3º ano (Brasil).

As evidências levam a sugerir, de acordo com a psicologia de Ausubel, que, se houve algum tipo de aprendizado, esse conhecimento sofreu um processo de obliteração subsunsiiva intensa, sendo muito provável que não tenha ocorrido uma aprendizagem significativa nesse campo do conhecimento, acarretando comprometimento no processo de desenvolvimento de competências científicas de natureza cognitiva para o conteúdo da Física Térmica, pois o teste sonda a aquisição de conceitos físicos básicos, introdutórios e edificadores. Com isso, acaba se tornando um objeto de preocupação e interesse para a superação das dificuldades envolvidas.

No contexto português, constata-se uma elevação no aglomerado de itens nas regiões MM e AA, em especial entre os alunos do 11º e 12º ano, isso mostra um aumento na consistência de modelos mentais corretos, bem como na identificações de modelos incorretos comuns entre os diferentes níveis de escolaridade, nomeadamente naqueles itens que permanecem na região MM.

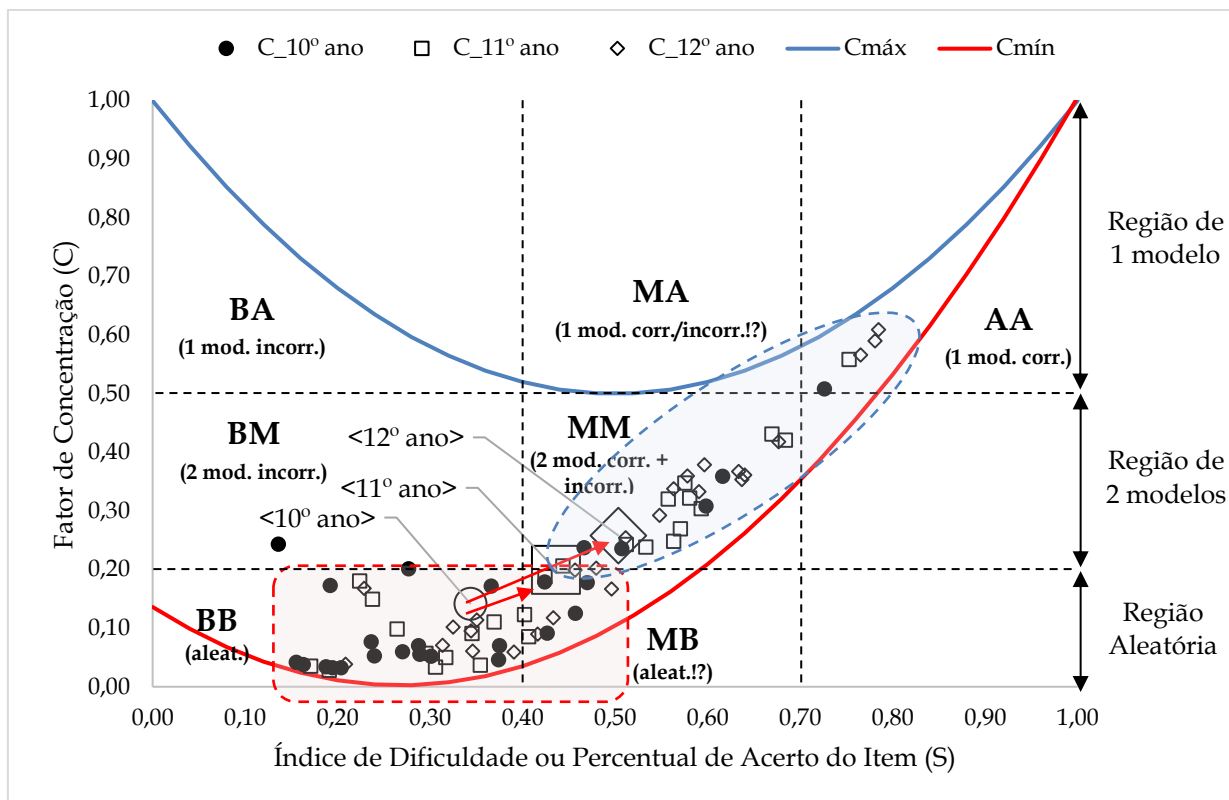


Figura 74: Gráfico S-C plot para o 10º, 11º e 12º ano (Portugal).

Fazendo uma análise geral comparativa, partindo dos resultados do teste-piloto com os resultados em Brasil e Portugal nos diferentes níveis de escolaridade, é possível chegar em uma perspectiva conclusiva dentro das análises clássicas quanto aos detalhes que configuram um perfil de proficiências e de concepções térmicas dentro de cada contexto, observando a influência dos modelos mentais em cada nível, bem como se houve progresso ou não após as instruções de ensino nesse contexto da Física Térmica, levando em consideração os três últimos anos de escolaridade da educação básica.

Sabe-se que a região AA remete a um único modelo mental predominante, fato que ocorre quando existe uma concentração na alternativa correta. A região MM caracteriza dois modelos populares para os grupos, sendo uma correta e outra incorreta, e a região BB uma falta de concentração ou de predominância no acerto em uma das alternativas, se configurando como zona de “guessing”, sendo praticamente classificada da mesma forma os itens que se posicionam na região MB, porém sendo um item mais fácil. A região BM representa o predomínio de dois modelos mentais incorretos.

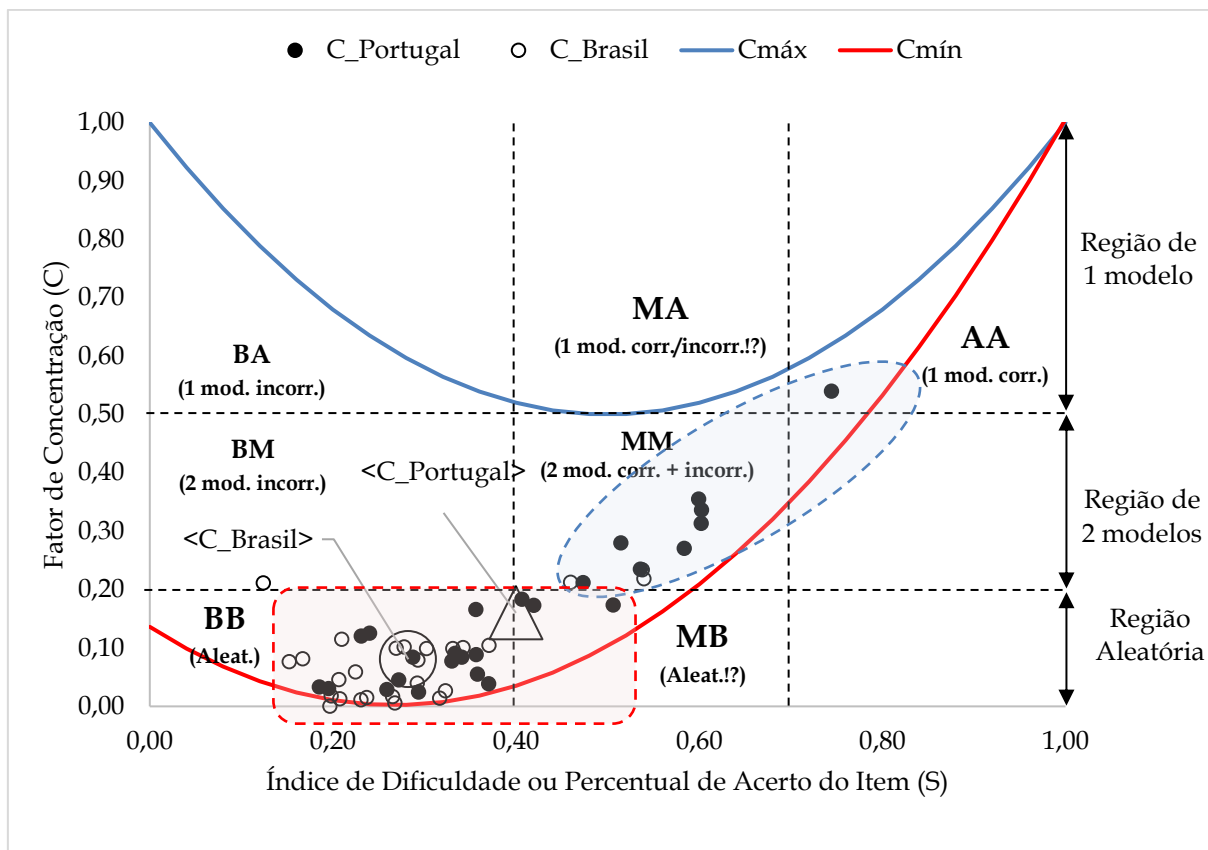


Figura 75: Gráfico S-C plot comparativo para Portugal e Brasil, considerando todos os alunos.

Diante disso, pode-se notar que em média, no Brasil, nenhum dos níveis apresenta um item em região de consistência relativo a um modelo mental correto (AA), estando possível um modelo correto em 3 itens, de acordo com a região MM (y1, y7, y12), e para a consistência na força atrativa de distratores, o item y14 encontra-se na região BM, que remete a identificação de dois modelos mentais incorretos. Em Portugal, a análise mostra um modelo correto AA (y1), e 7 itens com modelos corretos e incorretos na região MM (y2, y3, y7, y12, y13, y23, y26).

A caracterização desses modelos mentais são obviamente formados e influenciados pelos perfis de proficiências dos grupos que formam a amostra em cada contexto, ou seja, para afirmar de forma mais categórica e precisa aquilo que os alunos dominam diante dos conhecimentos envolvidos, considera-se que depende de qual aluno estamos nos referindo, e que um perfil baseado na média de seus acertos e escores deve conter suas restrições peculiaridades. Comparando os modelos nos diferentes grupos nos dois contextos, e observando a permanência de itens em uma determinada região torna-se possível identificar aspectos comuns, assim como itens

que se comportam de forma diferenciada para os diferentes alunos em cada contexto. Entre os itens relativamente comuns que se encontram na região aleatória, tem-se: y4, y5, y6, y8, y9, y11, y15, y16, y17, y18, y19, y20, y21, y24 e y25, sendo os itens y16 e y18, com modelos corretos para os alunos do 12º ano. Entre itens que sugerem a presença de modelos mentais corretos e incorretos para ambos os contextos, ou seja, entre aquilo que remete a ser de domínio comum, sem garantir entretanto que estão na mesma métrica de proficiência, são os itens: y1, y7 e y12. Itens que demonstram uma diferença na compreensão e domínio, são: y2, y3, y10, y13, y14, y22, y23 e y26.

É notório que o fator de concentração da análise sugere uma quantidade de modelos mentais associando-o com a porcentagem de acertos. Entretanto, as evidências mostram que é preciso fazer ponderações diante os critérios que determinam as regiões de predominância dos modelos mentais, de acordo com Lei Bao e Redish, pois existem itens com percentual de acerto abaixo de 40% que sugerem a existência de 2 modelos mentais, mesmo tendo concentração baixa. Uma sugestão é estabelecer o critério da zona de *guessing* dos itens. Isso iria expandir a região MM e diminuir a região de baixo acerto. Com isso deve-se verificar se de fato a que a região BB apresenta uma região aleatória, pois eventualmente pode apresentar dois modelos, um correto e um incorreto. Uma baixa concentração nem sempre também parece sugerir uma inexistência de modelo mental predominante, tendo em vista que nessas condições é possível considerar e observar um percentual de acerto acima de 50% na região MB, apesar de não ser muito comum.

Os resultados mostram que um item demonstra ser mais aleatório à medida que o fator de concentração se aproxima de zero, porém, é preciso verificar o percentual de acerto para afirmar de forma mais convicta sobre os modelos mentais predominantes, bem como a influência desses modelos pelos diferentes grupos normativos, que possuem características distintas, fazendo que esses sejam modelos mentais médios para o grupo como um todo, assumindo riscos de legibilidade na interpretação, caso não se observe os resultados parciais dos diferentes grupos inerentes.

A predominância ou elevado percentual das alternativas que são aparentemente comuns entre as diferentes características culturais dos alunos se

associam ao conteúdo dos itens, e assim, sugerem uma maior consistência diante das dificuldades dos alunos em uma perspectiva geral, podendo se tornar uma informação pedagógica que justifica a interpretação de características psicométricas de independência do item nos modelos da TRI. Porém, é necessário cautela e verificar se essa evidência não está ligada ao enunciado elaborado para o item. Isso ocorre no item y20.

Mesmo em itens com distribuição mais homogêneas com concentração baixa, é possível identificar uma sutileza comum na porcentagem de uma das alternativas entre os diferentes contextos, como pode ocorrer na alternativa que representa o acerto (y21). Isso também pode ser uma evidência influenciada por características de alunos com maior desempenho dentro do grupo de respondentes, porém isso não é considerado ideal para um bom item, e ao mesmo tempo pode estar relacionado a uma compreensão típica desses determinados alunos representativos proporcionalmente, que são provavelmente os mais proficientes. No Apêndice E10 (p. 273) encontra-se um quadro geral dos resultados do fator de concentração.

4.1.4.2 Perfil diante das Concepções Térmicas em Destaque

4.1.4.2.1 Forças Atrativas dos Distratores na Análise Gráfica dos Itens

A noção sobre um perfil das concepções se refere a tentar identificar os modelos mentais relacionados aos níveis de domínio sobre o conhecimento, de acordo com os distratores. Esses distratores estão representados no TCE por concepções ou equívocos típicos diante do conhecimento envolvido para cada item, porém observa-se que determinadas concepções podem estar relacionadas em diferentes situações e contextos para o problema. Diante disso, considera-se relevante, do ponto de vista analítico-interpretativo da organização e disponibilização dos itens, compreender duas formas na quais as concepções estão associadas aos distratores são mais incidentes: (1) a primeira seria considerando a *independência dos itens*⁹⁶, ou seja, adotando uma perspectiva com base no âmbito dos resultados e dificuldades globais

⁹⁶ Caracteriza-se como um pressuposto que deve ser considerado quando se adotam grande parte dos modelos logístico da TRI. A dependência local, entretanto pode ser objeto das análises clássicas, e servirem para perspectivas complementares das análises.

do teste, ou de predominância de um fator explicativo geral, que incorpore de forma lógica e satisfatória todas as informações e elementos representativos de equívocos contidos nas análises, sem, entretanto, considerar a influência relativa entre eles; (2) o segundo seria considerando a *dependência dos itens*, as relações lógicas, condicionais e sequenciais de raciocínios influenciados pela ordem, graus de dificuldades e da própria natureza epistemológica da concepção térmica que está subjacente e presente nas diferentes situações-problemas. Entre outras formas de analisar essas dificuldades seria através do perfil de desempenho dos diferentes grupos normativos. Para fatores externos, tem-se as influências relacionadas ao ensino (perfil docente), aprendizagem (hábitos na aprendizagem) e dos meios *físicos* (infraestrutura escolar, ...) e *socioculturais* (exercício de práticas sociais comuns, compartilhamento de normas/valores, ...).

Parte-se do entendimento de que os níveis em que esses modelos mentais encontram-se distribuídos dentro da TCT possam estar sugeridos por zonas específicas nos intervalos de escore bruto (de representatividade da amostra e de maior consistência). Nisso, deve-se levar em consideração os percentuais absolutos de resposta, sendo complementados com informações psicométricas discriminatórias dos distratores. Entende-se que um maior número de evidências de concepções entre as respostas erradas, assim como aqueles equívocos mais atraentes entre os respondentes de alto desempenho, sejam os principais critérios para estabelecer um perfil de concepções dos grupos dentro de cada contexto. Dessa forma, quanto mais amplo for o conjunto das diferentes concepções e maior o peso atribuído a essas concepções segundo os níveis de escore, mais substancial será o perfil de concepções. Esse critério segue o raciocínio em pressupor que aquele que acerta mais, seja mais proficiente. Entretanto, vale ressaltar que esse não é um argumento válido para os modelos da TRI, pois a proficiência é construída levando em consideração o acerto aleatório, isso significa dizer que nem sempre aquele que acerta mais é o mais proficiente, apesar disso, esse tipo de análise torna-se um parâmetro enriquecedor dentro das análises no âmbito da TCT.

É natural que um *perfil de concepções* deva indicar evidências de algumas causas relacionadas às dificuldades dos alunos diante de situações-problemas específicos. Inevitavelmente isso remete a uma baixa proficiência em média para os grupos

envolvidos, e assim vice-versa. Determinam-se assim alguns critérios relevantes para afirmar sobre o **perfil de concepções** dos respondentes em cada contexto:

- Compreendendo a *natureza epistemológica* das concepções das alternativas distratoras dos itens mais atrativas dentro da zona de representatividade e consistência das amostras;
- Selecionando os distratores de acordo com as *forças atrativas* na AGI (com o grupo ACIM) e associando-os adequadamente com as médias percentuais dos grupos de respondentes dentro zona de representatividade e consistência das amostras;
- Relativando a força atrativa da quantidade de incidências das concepções com um *peso de relevância diferenciado* de acordo com as diferentes zonas de proficiência dos respondentes, sendo peso nulo na zona de “*guessing*”;
- Distinguindo a representatividade das concepções mais atrativas levando em consideração a relação entre a soma dos pesos das incidências com o total de concepções distribuídas no teste *dentro de cada subconstruto*, ou seja, considerando as concepções mais consistentes aquelas com maior valor para o *peso de relevância diferenciado* de acordo com o quantitativo ou intensidade das incidências em cada subconstruto.

Para isso, na determinação do perfil conceptual, foram exploradas as informações contidas nos itens comentados (ver no Apêndice A, p. 1), em especial, na *Análise Gráfica dos Itens* (AGIs), dando atenção ao grupo superior (ACIM), bem como os parâmetros gráficos contidos na escala de desempenho com base no score. Considera-se que a atração de respondentes de alto desempenho por distratores (k é o número de alternativas, portanto, $k-1$ é a quantidade de distratores) com percentual na zona de “*guessing rate*” seja ocorrido por descuido (erro ao acaso), e com isso configura-se uma categoria de poucos respondentes, porém, supostamente sendo alunos que dominam o conteúdo e não deveriam ser atraídos por distratores.

Na análise foi considerado como forte atrator o distrator que exerce um “pico percentual” mais elevado proporcionalmente que os outros na zona de maior desempenho dentro de cada contexto (ver no Apêndice A, p. 8, dentro das AGIs, os percentuais que representam quantitativamente a força atrativa), no que se refere a

violação da curva do tipo 2, que ao invés de decrescer, o percentual que representa a curva se eleva. Percebe-se, entretanto, que seja relevante não apenas o número de violações como é sugerido (Batenburg & Laros, 2002), mas sim a sua representatividade dentro na zona de nível de escore mais elevado com relação a alternativa fora da zona de “*guessing rate*”, juntamente com a média dos percentuais que sugerem a força atrativa do distrator. A força atrativa nesse contexto também deve levar em consideração as proporções dos picos com relação aos padrões anteriores, se segue relativamente a uma tendência dentro de um desvio padrão aceitável e mais significativo, com isso, deve-se levar em conta a quantidade diferenciada de respondentes dentro de cada nível de escore, conforme a AGI apresenta. Para atender ao propósito desse estudo, é possível contornar essa necessidade através da verificação de reincidências das concepções que envolvem esses distratores entre os dois grupos, e assim ser considerado um dos critérios minimamente suficiente para contribuir na construção de um perfil conceptual relativamente consistente dos respondentes. É importante ainda dizer que a correspondência com os valores estimados na TRI dentro da CCI, possibilita verificar as probabilidades de acerto e erro desse indivíduos nessas zonas de desempenho.

Os modelos mentais dos grupos que se apresentam de forma mais consistente reincidem entre os distratores acima da zona de “*guessing rate*” (distância discriminativa), e estão destacados na análise entre colchetes e negrito na tabela de identificação das violações da AGI (ver no Apêndice A, p. 1). Na mesma tabela, os asteriscos e itálicos simbolizam uma referência aos grupos que estão na zona de “*guessing rate*”. Os critérios para a análise baseiam-se nas incidências das concepções térmicas que envolvem os itens que estão sendo atraídos, onde a AGI ilustra como “picos” percentuais, e assim, se consideram as repetições dessas concepções dentro do mesmo item para cada contexto separadamente, e sem repetições na correspondência e incidência nos dois contextos. Na incidência de concepções relacionados aos distratores com percentual acima do “*guessing*” do item ($> 0,25$ para os itens com $k = 4$, e $> 0,20$ para itens com $k = 5$), foi estabelecido **peso igual a 1**, na repetição de concepções entre distratores acima do “*guessing*” com aqueles abaixo **peso igual a 0,5**,

e na correspondência entre duas concepções de distratores com percentual dentro do *guessing*, **peso igual a 0,25**. Para exemplificar, no item y9, tem-se (ver no Apêndice A9, p. 71):

Tabela 28: Item y9 como exemplo na análise dos distratores mais atrativos através dos picos mais significativos dentro dos escores, segundo a AGI.

Itens (k)	Análise Gráfica do Item: Distratores mais Atrativos nos Contextos (BR e PT) para os Grupos Superiores (ACIM) dentro de Níveis Específicos de Escore Bruto						Concepções Alternativas Associadas (BR/PT)
	BR	%; >0,25 ₍₄₎ e >0,20 ₍₅₎	(picos sig. em escore ≥9)	PT	%; >0,25 ₍₄₎ e >0,20 ₍₅₎	(picos sig. em escore ≥12)	
y9 ₍₅₎	d; e*	0,36; 0,10*	17; 16*	d; e	0,25; 0,21	22; 21	[C6], [D1]; B7* / [C6], [D1]; B7

Nota: Asterisco e itálico representa os grupos na zona de “guessing” do item; “;” separa os alternativas, as porcentagens, os escores de pico para cada contexto e as concepções para cada alternativa; em colchetes correspondências de concepções térmicas entre os dois contextos; “/” separa os dados dos dois contextos; k é o número de alternativas que o item possui.

É um item com $k = 5$ (4 distratores, a zona do “guessing” é $\leq 0,20$), com percentual do distrator D igual a 36%, tendo um pico no escore 17, e com as concepções alternativas C6 e D1 que coincidem com os alunos portugueses (em colchetes e negrito), dessa forma se atribuí um peso igual a 1. A concepção B7 é referente a uma atração ocorrida no grupo superior dos dois contextos, porém representa um percentual baixo de respondentes (na zona do “guessing” do item) no contexto brasileiro, diferentemente ao de Portugal, com isso se atribuí peso de 0,50. Quando os dois estão no *guessing* entende-se como uma atração correspondente de percentual baixo, por isso se atribui peso igual a 0,25. Portanto, para o Brasil a pontuação do item y9 será de 2,5 (C6 = 1,0; D6 = 1,0; B7= 0,5), e para Portugal 3,0 (C6 = 1,0; D6 = 1,0; B7= 1,0). Verificando uma consistência das concepções entre os dois contextos (em negrito e colchetes), não conta-se a repetição da concepção dentro de cada item, porém, soma-se com a reincidência da concepção em outro item. Dessa forma é possível montar um perfil das concepções térmicas nas três perspectivas.

Como foi dito, além de admitir que a zona dos grupos superiores (ACIM) possua uma maior consistência dentro das amostras de modo geral, é prudente ainda considerar que a quantidade de incidências das concepções também agregue valor de consistência, ou seja, as evidências (percentuais de respostas) relacionadas a concepções com **maior número de incidências nas respostas** dentro do grupo de desempenho superior são ainda mais consistentes.

Quadro 69: Critérios relativos de maior consistência quanto ao número de incidências das concepções térmicas identificadas no grupo superior (ACIM).

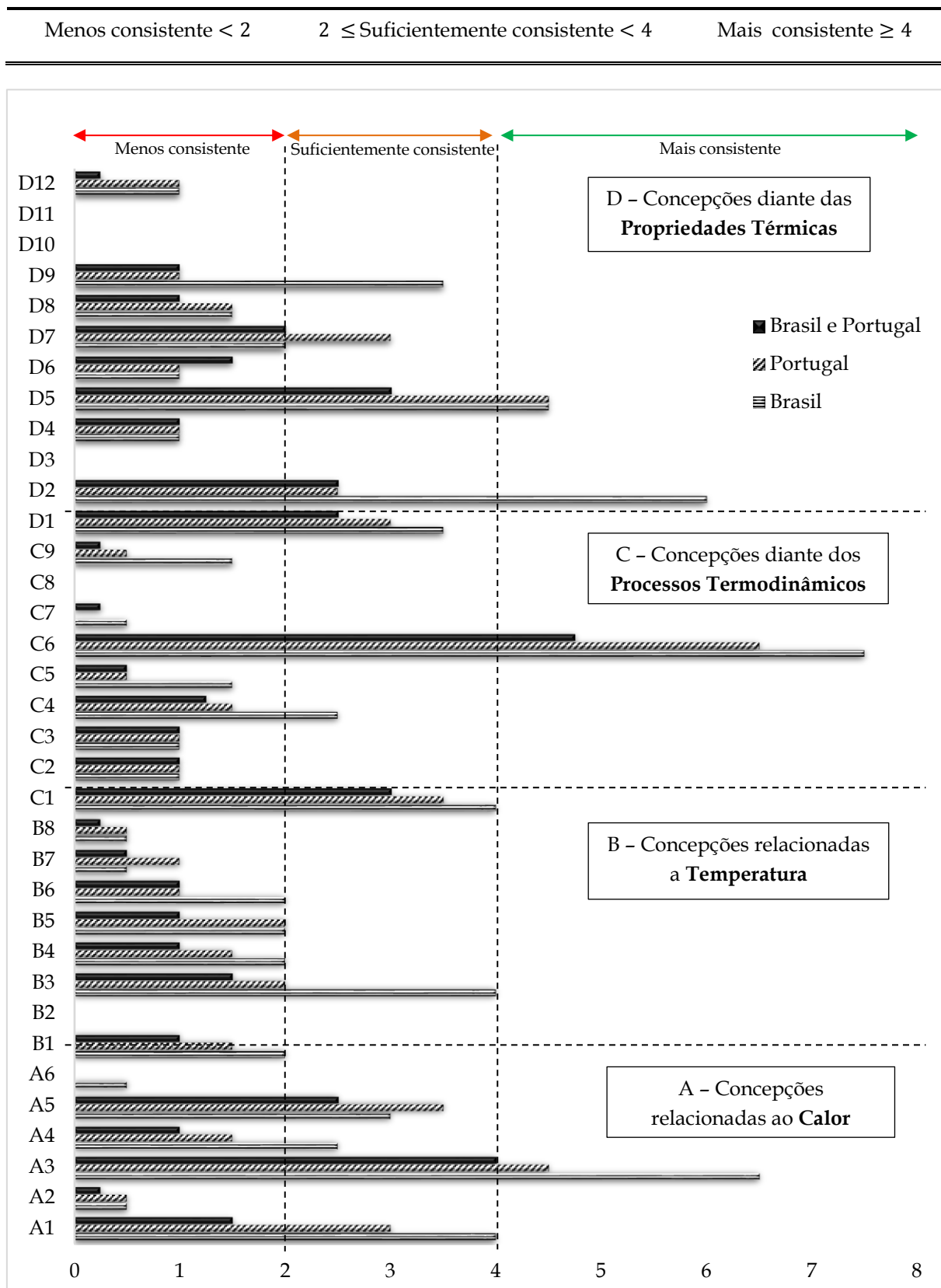


Figura 76: Perfil de concepções térmicas de Brasil e Portugal de acordo com o «peso das incidências» das concepções para os grupos ACIM.

Dessa forma, foram delimitados critérios mais refinados de consistência relacionados aos *peso das incidências*, na qual considera-se relativamente *menos consistente* as evidências combinadas com incidências relativas ao peso **menores que dois**, de *consistência suficiente* aquelas **maiores ou iguais a dois** até próximo ao dobro desse valor, e valores **maiores ou iguais ao dobro** da consistência minimamente consistente podem ser consideradas *mais consistentes* por serem constatadas mais vezes.

Verificando, pode-se dizer que, de modo geral, para os alunos **brasileiros**, o perfil de concepções térmicas se distribui entre 30 das 34 ($30/34 = 88,2\%$) concepções, e 28 para os portugueses ($28/34 = 82,4\%$), porém deve-se verificar as consistências e incidências de cada concepção dentro do grupo ACIM.

É possível observar a **ausência de algumas concepções para ambos os contextos** (B2, C8, D3 e D10), e algumas que estão ausentes **exclusivamente para os alunos portugueses** (A6, C7). Também é possível perceber de imediato que os alunos brasileiros apresentam concepções mentais equivocadas mais frequentemente que os alunos portugueses.

Diante disso, foi possível constatar na análise que os portugueses possuem uma menor quantidade de equívocos em evidências de modo geral, sendo de menor quantidade aqueles de maior consistência, sugerindo que eles possuem maior domínio no conhecimento envolvido no teste. A quantidade elevada de concepções de baixa consistência indica a necessidade de procedimentos de análises mais robustas ou do uso de novos métodos e bases de dados para fazer o cruzamento, ou seja, complementar e comparar com estudos posteriores. Isso também sugere uma deficiência no instrumento, na qual deveria supostamente repetir no mínimo duas vezes cada concepção para que haja uma consistência mínima na análise, sobretudo que essas concepções estejam em um zona de maior fiabilidade, preferencialmente.

Também foi evidenciado um **equilíbrio na incidência de 5 concepções** entre os dois contextos (B5, C2, C3, D4 e D9). Isso sugere que esses equívocos se aproximam relativamente a dificuldades predominantes entre os grupos, porém não deve ser

afirmado categoricamente isso, pois não as análises não garantem esse tipo de correspondência.

Como foi visto anteriormente, todas essas concepções estão distribuídas com quantidades diferenciadas dentro de cada um dos itens. Dessa forma, convém reconhecer que essas incidências, ao serem comparadas com a análise de distribuição dos subconstrutos do teste, **constata-se que segue um padrão similar**, ou seja, a densidade das concepções nos subconstrutos parecem ser determinantes na proporção das incidências das concepções das respostas, revelando assim um risco de viés interpretativo acarretado pela falta de equilíbrio entre os subconstrutos. Com isso, convém avaliar cada concepção separadamente, tendo como referência o número de vezes que ela se encontra distribuída no instrumento.

Na tentativa de minimizar esse possíveis efeitos, e para se obter uma percepção mais realista, foi **elaborado um gráfico com a relação de proporção das respostas** conforme os pesos das concepções térmicas atribuídas a cada contexto para os grupos ACIM, com o total de incidências dessas concepções distribuídas no teste para cada subconstruto.

Proporcionalmente, pode-se dizer que o **perfil brasileiro das concepções térmicas** mais representativas (> 50%) relativo a quantidade de suas incidências no grupo superior (ACIM) se distribuem como:

- Subconstruto A - **50%** com 3 concepções diante do **conceito de calor**, que são: A3 (81%), A4 (63%) e A5 (60%);
- Subconstruto B - **62,5%** com 5 concepções diante do **conceito de temperatura**, que são: B1 (67%), B3 (57%), B4 (100%), B5 (67%) e B6 (67%).
- Subconstruto C - **33,3%** com 3 concepções diante da compreensão sobre os **processos térmicos**, que são: C1 (80%), C2 (100%), C3 (100%),.
- Subconstruto D - **50%** com 6 concepções diante da compreensão sobre as **propriedades térmicas**, que são: D2 (86%), D4 (100%), D5 (75%), D6 (100%), D7 (67%), D9 (70%).

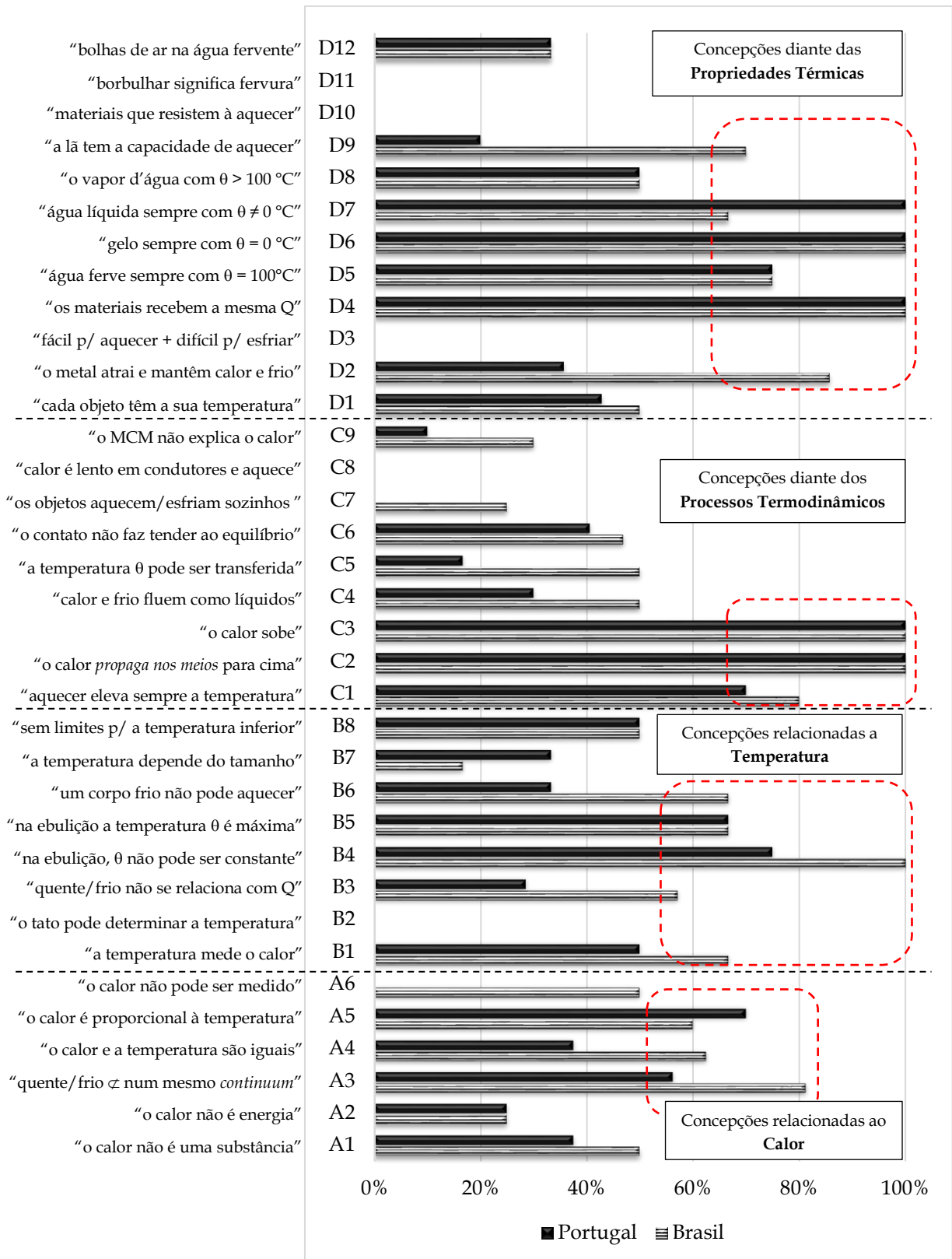


Figura 77: Perfil de concepções térmicas com base na relação entre o total de concepções dentro de cada subconstruto e o peso das incidências das concepções evidenciadas nas respostas do grupo superior (ACIM) em cada contexto.

Comparando os gráficos, pode-se dizer que para os **alunos brasileiros** existem 30 concepções em destaque, dessas, há um total de **17 concepções** ($17/34 = 50,0\%$) que reincidem mais a que dos portugueses (A1, A3, A4, A6, B1, B3, B4, B6, C1, C4, C5, C6, C7, C9, D1, D2 e D6 – em negrito abaixo). Considerando que a consistência nesse caso se refira a quantidade das incidências das concepções e sua representatividade percentual nos respectivos subconstrutos, tem-se a seguinte classificação:

- **Mais consistentes (≥ 4 pesos das incidências)** – 7 concepções (20,6%), estando em negrito aquelas com peso das incidências superior a dos alunos portugueses: **A1, A3, B3, C1, C6, D2 e D5**. Concepções em *ordem decrescente* de consistência e sua representatividade percentual dentro dos subconstrutos: **C6 (47%), A3 (81%), D2 (86%), D5 (75%)**; seguido de **A1 (50%), B3 (57%) e C1 (80%)** estando no mesmo nível.
- **Suficientemente consistentes (≥ 2 e < 4 pesos das incidências)** – 10 concepções (29,4%), estando em negrito as 5 mais incidentes que a dos alunos portugueses: **A4, A5, B1, B4, B5, B6, C4, D1, D7 e D9**. Concepções em *ordem decrescente* de consistência e sua representatividade percentual dentro dos subconstrutos: **D1 (50%) e D9 (70%)** no mesmo nível; depois **A5 (60%)**; **A4 (63%) e C4 (50%)** no mesmo nível; e por fim, **B1 (67%), B4 (100%), B5 (67%), B6 (67%) e D7 (67%)** no mesmo nível.
- **Pouco consistentes (< 2 pesos das incidências)** – 13 concepções (38,2%), estando em negrito as 5 mais incidentes que a dos alunos portugueses: **A2, A6, B7, B8, C2, C3, C5, C7, C9, D4, D6, D8 e D12**. Concepções em *ordem decrescente* de consistência e sua representatividade percentual dentro dos subconstrutos: **C5 (50%), C9 (30%) e D8** no mesmo nível (30%); **C2 (100%), C3 (100%), D4 (100%), D6 (100%), D12 (36%)**; e, **A2 (25%), A6 (50%), B7 (17%), B8 (50%) e C7 (25%)**.

Já o **perfil português das concepções térmicas** mais representativas ($> 50\%$) relativo a quantidade de suas incidências no grupo superior (ACIM) se distribuem como:

- Subconstruto A – **33,3%** com 2 concepções diante do **conceito de calor**, que são: **A3 (56%), A5 (70%)**;

- Subconstruto B – **25,0%** com 2 concepções diante do **conceito de temperatura**, que são: B4 (75%), B5 (67%).
- Subconstruto C – **33,3%** com 3 concepções diante da compreensão sobre os **processos térmicos**, que são: C1 (70%), C2 (100%), C3 (100%).
- Subconstruto D – **33,3%** com 4 concepções diante da compreensão sobre as **propriedades térmicas**, que são: D4 (100%), D5 (75%), D6 (100%), D7 (100%).

Para os alunos **portugueses** o perfil de concepções térmicas se distribui entre 28 de 34 (82,4%), na qual existem **3 equívocos** ($3/34 = 8,8\%$) que se destacam diante do quantitativo incidente entre os brasileiros (A5, B7 e D7), isso sugere que comparativamente o perfil de concepções térmicas dos alunos brasileiros constitui de uma quantidade maior de equívocos. Pode-se dizer que os grupos ficam entre elas, como:

- **Mais consistentes (≥ 4 pesos das incidências)** – 3 concepções (8,8%), que foram: A3, C6 e D5. Em *ordem decrescente* de consistência: C6 (41%), A3 (56%) e D5 (75%) no mesmo nível.
- **Suficientemente consistentes** (8 concepções): A1, **A5**, B3, B5, C1, D1, D2 e **D7**. Em **ordem decrescente** de consistência, estando em negrito as duas concepções com incidência superior a dos alunos brasileiros: C1(70%) e **A5 (70%)** no mesmo nível; A1 (38%), D1 (43%) e **D7 (100%)** no mesmo nível; D2 (36%), B3 (29%) e B5 (67%) no mesmo nível.
- **Menos consistentes** (17 concepções): A2, A4, B1, B4, B6, **B7**, B8, C2, C3, C4, C5, C9, D4, D6, D8, D9 e D12. Em **ordem decrescente** de consistência, estando em negrito a concepção com incidência superior a dos alunos brasileiros: A4 (38%), B1 (50%), B4 (75%), C4 (30%) e D8 (50%) no mesmo nível; B6 (33%), **B7 (33%)**, C2 (100%), C3 (100%), D4 (100%), D6 (100%), D9 (20%) e D12 (33%) no mesmo nível; A2 (25%), B8 (50%), C5 (17%) e C9 (10%) no mesmo nível.

O contraste entre os perfis de concepção térmica entre os dois contextos sugere que os alunos brasileiros e portugueses apresentam diversas dificuldades semelhantes, porém, com uma intensidade diferente, sinalizando que os brasileiros apresentam

desvantagens ou maiores dificuldades de compreensão, tendo como parâmetro comparativo o quantitativo das incidências nas faixas preestabelecidas.

De modo geral, foi constatado que nas respostas dentro da zona de maior representatividade e consistência dos grupos há uma maior cobertura dos subconstrutos pela amostra brasileira, ou seja, o perfil de concepções térmicas é constituída por mais as concepções diante dos conhecimentos que envolve o teste, em média 49%, enquanto os portugueses cobrem em média 31,2% do rol das concepções. Pode-se se destacar uma maior cobertura das concepções do subconstruto B, relacionado as concepções diante do conceito de temperatura.

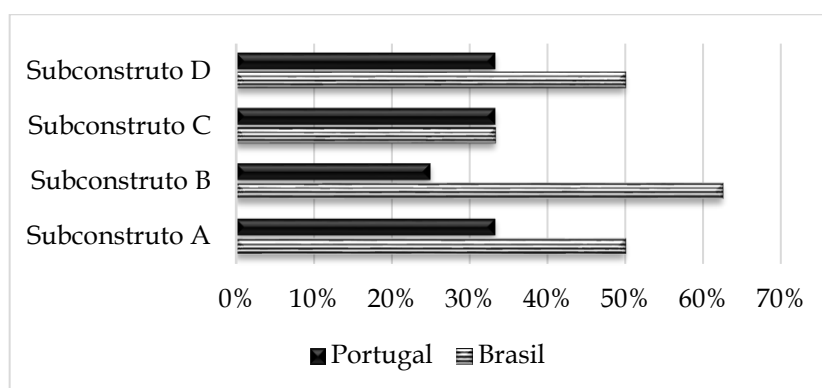


Figura 78: Perfil de concepções térmica do contexto do Brasil e de Portugal conforme a sua representatividade nos subconstrutos.

Diante da representatividade de cada concepção em cada subconstruto, é possível ainda verificar a sua porcentagem de cobertura individualmente, mas também um contraste de sua intensidade de incidências entre os contextos.

Analisando separadamente cada contexto e aplicando os mesmos critérios, pode-se constatar peculiaridades em uma comparação *lado a lado*, entre elas, supostas dificuldades de compreensão comuns e em níveis equivalentes, porém com intensidades diferenciadas, por exemplo, as concepções C6, A3 e D5 são mais consistentes para ambos os grupos, ou seja, tanto alunos brasileiros como portugueses apresentam evidenciam esses equívocos, mas para os brasileiros as incidências foram quantitativamente mais intensas. Chama a atenção o fato de que todas as dificuldades que os portugueses possuem com relação as concepções térmicas, os brasileiros também as possui, com o acréscimo de 2 concepções a mais (A6, C7), que no caso se

refere na crença de que “o calor não é mensurável, conceito quantificável”(item y7d), e de que os “objetos quentes naturalmente se resfriam, objetos frios naturalmente se aquecem” (item y3a e y13c). As diferenças nesses concepções, entre outras, do ponto de vista qualitativo podem sugerir uma explicação da eventualidade existência de DIF dos itens correspondentes para os grupos, ou seja, as diferentes dificuldades dos itens para os grupos podem estar relacionadas a concepções distintas.

Tabela 29: Contraste entre os Perfis de Concepções Térmicas para Brasil e Portugal, conforme o percentual de cobertura da concepção no seu subconstruto, das intensidades nos pesos de incidências e aspectos comuns.

Classificação	Brasil	Mais intenso que em Portugal	Portugal	Mais intenso que no Brasil	Brasil e Portugal
Mais consistentes	C6 (47%), A3 (81%), D2 (86%), D5 (75%) ; seguido de A1 (50%), B3 (57%) e C1 (80%) estando no mesmo nível	C6, A3, D2, D5, A1, B3, C1	C6 (41%), A3 (56%) e D5 (75%) no mesmo nível.	-	C6, A3, D5
Suficientemente consistentes	D1 (50%) e D9 (70%) no mesmo nível; depois A5 (60%) ; A4 (63%) e C4 (50%) no mesmo nível; e por fim, B1 (67%), B4 (100%), B5 (67%), B6 (67%) e D7 (67%) no mesmo nível	D1, A4, B1, B4, B6	C1(70%) e A5 (70%) no mesmo nível; A1(38%), D1(43%) e D7(100%) no mesmo nível; D2 (36%), B3 (29%) e B5 (67%) no mesmo nível.	A5, A7	D7
Pouco consistentes	C5 (50%), C9 (30%) e D8 no mesmo nível (30%); C2 (100%), C3 (100%), D4 (100%), D6 (100%), D12 (36%) ; e, A2 (25%), A6 (50%), B7 (17%), B8 (50%) e C7 (25%) .	C5, C9, D6, A6, C7	A4 (38%), B1 (50%), B4 (75%), C4 (30%) e D8 (50%) no mesmo nível; B6 (33%), B7 (33%), C2 (100%), C3 (100%), D4 (100%), D6 (100%), D9 (20%) e D12 (33%) no mesmo nível; A2 (25%), B8 (50%), C5 (17%) e C9 (10%) no mesmo nível.	B7	C5, C9, D8, C2, C3, D4, D6, D12, A2, B8

Essas concepções podem ser associadas na interpretação das análises das CCI da TRI para ambos os contextos ao considerar especificamente os respectivos itens correspondentes (ver no Apêndice E11, p. 274), no que diz respeito às habilidades θ de alunos com maior desempenho em cada contexto.

4.2 FASE ANALÍTICA E INFERENCIAL - Análises da Teoria da Resposta ao Item

4.2.1 Considerações Gerais

A intenção de construir uma escala de proficiências de uma amostra geral (Brasil e Portugal) visa estabelecer uma métrica comum de proficiências entre os sujeitos, e assim, a partir das habilidades θ dos sujeitos e dos grupos normativos estimadas pelo modelo logístico dentro dessa escala, torna possível e de forma mais precisa realizar uma análise comparativa direta entre os perfis de proficiências dentro de cada contexto. Isso ocorre devido à robustez dos modelos matemáticos, capacidade de múltiplas análises em softwares apropriados, conseguindo atender a critérios estatísticos previamente estabelecidos, de acordo com os pressupostos da *Teoria da Resposta ao Item* (TRI).

Uma das justificativas e vantagens de serem agrupados os dados nos dois contextos é o fato de que a construção de uma escala de proficiências se torna mais rica de informações ao serem considerados itens em diferentes níveis de dificuldades, sendo essa uma necessidade que ao ser atendida, caracteriza uma qualidade da escala ao ampliar o campo de informações precisas dentro do *continuum* dos traços latentes.

Na TRI, os dados das respostas dos estudantes são analisados considerando os acertos e erros a cada item. Assim, para modelagem dos dados de resposta na formação da escala de proficiência foram empregados os modelos da TRI para itens dicotômicos, conforme a natureza do teste TCE que é de *múltiplas alternativas de respostas simples* (“*multiple-choice single-response*”) ou de *múltipla escolha*.

O fato é que a TRI, como método, avalia diferenças individuais dos sujeitos ao longo de um *continuum* latente, identificando no nível da escala de proficiência onde um item é mais informativo independentemente da habilidade do sujeito (de Ayala, 2009). A TRI também permite avaliar discrepâncias entre observações empíricas e previsões do modelo (estas são denominadas resíduos), para acessar se as inferências feitas a partir do modelo são precisas (Glas, 2016). Os índices S-X² de Orlando e Thissen (2000) foram utilizados para verificação da adequação dos itens ao modelo.

De modo geral, para utilização dos modelos de TRI é preciso que atendam os seguintes pressupostos: (1) de *monotonicidade* (a probabilidade de acertar determinado item cresce proporcionalmente ao aumento da habilidade do participante); (2) *independência local* (os itens do teste não influenciam um no outro para a obtenção da resposta do sujeito); (3) e de *unidimensionalidade* (o conjunto de itens reflete a variação de único traço latente) devem ser atendidos. Estes foram investigados através da identificação de itens com discriminação negativa, análise de componentes principais da *matriz de correlação dos resíduos* do modelo escolhido e da inspeção do *Scree plot* feito a partir da matriz de correlações tetracóricas (DeMars, 2010).

Dessa forma, as etapas seguem na seguinte ordem:

- A base de dados considera uma amostra geral, ou seja, as respostas de todos os alunos do contexto de Brasil e Portugal, estando incluso os alunos de todos os níveis de escolaridade;
- Realizar a análise clássica de consistência dos dados pela alfa de Cronbach e a análise dimensional tendo como base a matriz de correlação tetracórica e gráfico *Scree Plot* de autovalores;
- Análise de quatro modelos da TRI, visando determinar e adotar nas análises psicométricas o modelo mais ajustado;
- Ajustar o modelo adotado eliminando itens de baixa consistência;
- Construir um painel estatísticos com as informações psicométricas relevantes da TCT e TRI de cada item, incluindo as respectivas *Curva Característica do Item* (CCI) e *Curva de Informação do Item* (CII) dentro de cada contexto separadamente e para o contexto geral (ver no Apêndice A, p. 1), assim como a *Curva Característica do Teste* (CCT) e *Curva de Informação do Teste* (CIT). Para apresentar as CCI e CII de cada contexto separadamente foram utilizados três modelos estatísticos de *Função Diferenciada do Item* (DIF⁹⁷) para que fiquem na mesma métrica, e assim se tornem comparáveis dentro de uma mesma métrica;
- Identificar e comparar as dificuldades significativamente diferentes de itens entre os contextos através da análise de DIF;

⁹⁷ Técnicas de Mantel-Haenszel, regressão logística e de DIF uniforme (modelo Rasch).

- Construir a escala de proficiências em concepções térmicas através do modelo geral e obedecendo o rigor dos critérios estatísticos preestabelecidos para as probabilidades de acerto entre os níveis de proficiência, considerando uma única base de dados para os dois contextos, ou seja, estando na mesma métrica, e assim, podendo ser comparáveis estatisticamente;
- Interpretar e comentar pedagogicamente os índices e parâmetros de cada item (incluindo o que é fácil/difícil? E por que é fácil/difícil?), os resultados na escala de proficiências, e apresentar sugestões/recomendações e implicações dentro da psicologia educacional cognitivista de Ausubel, na qual foi adotada como eixo de referência nas análises qualitativas das evidências psicométricas.

Diante disso, preliminarmente, quatro modelos de TRI foram testados, sendo o modelo escolhido aquele que apresentou melhor ajuste aos dados. Entre os quatro modelos, tem-se o: modelo Rasch (somente varia a dificuldade dos itens e a discriminação é restringida a 1, levando em consideração o $D = 1$, e não 1,7); modelo de dois parâmetros (2PL; dificuldade e discriminação variam); modelo de três parâmetros (3PL; dificuldade e discriminação variam além da assíntota inferior da curva característica do item variam); e o de quatro parâmetros (4PL, dificuldade, discriminação, assíntota inferior e superior da curva característica do item variam).

Esses quatro modelos foram avaliados por intermédio do teste de hipótese estatística na comparação de valores de índices de ajuste geral do modelo por meio da verossimilhança (*logLik*), do critério de informação de Akaike (AIC; *Akaike Information Criterion*), critério de informação de Akaike corrigido (AICc; *corrected Akaike's Information Criterion*), do critério de informação bayesiano (BIC; *Bayesian Information Criterion*), e critério de informação bayesiano ajustado para o tamanho da amostra (SABIC; *Sample Size Adjusted Bayesian Information Criterion*). Eles são usados quando comparam dois ou mais modelos (não necessariamente alinhados), com valores menores indicando melhor ajuste (Beaujean, 2014). Também foram consideradas a estatística M2 para ajuste de itens dicotômicos de Maydeu-Olivares e Joe (2006) em que um modelo é considerado satisfatório quando possui *comparative fit index* (CFI) e

o *Tucker-Lewis Index* (TLI) maiores que 0,90 e o *Root Mean Square Error of Approximation* (RMSEA) menor que 0,08.

Por fim, foram realizadas análises para detecção de *Funcionamento Diferencial dos Itens* (“*Differential Item Functioning*” ou DIF), que basicamente se refere a verificar se existem diferenças entre os parâmetros de dois grupos, geralmente compara-se o parâmetro de dificuldade (b) (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, p. 116).

$$H_0 = \Delta b = b_F - b_R = 0 \text{ (ausência de DIF)}$$

$$H_1 = \Delta b = b_F - b_R \neq 0 \text{ (DIF)}$$

Este procedimento visa identificar os itens cuja probabilidade de acertos difere entre distintos subgrupos de uma amostra, cujas pessoas **possuem o mesmo nível de habilidade na variável mensurada** (de Ayala, 2009).

Um dos métodos utilizados foi o de Mantel-Haenszel (MH), um procedimento que compara as respostas (certas e erradas) do «grupo de referência» e do «grupo focal» em relação a um item. No estudo, o DIF visa comparar e identificar em quais itens existem diferenças significativas quanto as dificuldades existentes entre a amostra Brasil (grupo focal) com Portugal (grupo de referência). Nesse procedimento calcula-se um índice que expressa o quociente ou proporção entre a probabilidade de acertar o item no «grupo focal» *versus* a probabilidade de acerto quando comparado com o «grupo de referência» (Finch & French, 2015). O outro método utilizado denomina-se *TSW Likelihood Ratio Test* desenvolvido por Thissen, Steinberg, and Wainer (1988, citado por de Ayala, 2009) baseado na comparação do ajuste de dois modelos TRI (no inglês “IRT”) usando a estatística de teste da razão de verossimilhança. No método TSW é investigado se há uma diferença significativa no ajuste geral do modelo quando um item é restringido para ter a mesma dificuldade entre os grupos (**modelo restringido**) versus quando o item é livre para ter diferentes dificuldades e discriminação nos subgrupos (**modelo não-restringido**).

Para realização das análises foi utilizado o *Software R* (R Core Team, 2017) em que as funções implementadas pelos pacotes *psych* (Revelle, 2015), *mirt* (Chalmers, 2012), e *difR* (Magis, Béland, Tuerlinckx, & Boeck, 2010) foram utilizadas.

4.2.2 Consistência e Dimensionalidade

A matriz de correlações tetracóricas dos 26 itens apresentou uma consistência interna com um **alfa de Cronbach igual a 0,75** para amostra total, amostra brasileira e portuguesa. Esse valor, dentro do quadro de classificação para os coeficientes de correlação sugere ser suficiente para medidas de grupo, mas com *certa restrição a indivíduos*.

A análise do *Scree plot* (Figura 79) e dos autovalores da matriz de correlação tetracórica (ver no Apêndice E12, p. 275) apontaram uma dimensão predominante explicando a variância conjunta dos dados. A partir dos resultados foram utilizados modelos unidimensionais via *Teoria da Resposta ao Item* (TRI) para modelagem das respostas dicotômicas ao teste de proficiência em termodinâmica.

William Revelle tornou disponível no *software* livre R o pacote *psych*⁹⁸, subpacote *biserial*⁹⁹. Baseado nesses pacotes foram realizadas análises de correlação tetracórica da matriz de respostas, e com isso foi possível verificar a existência (ou não) de um traço latente predominante.

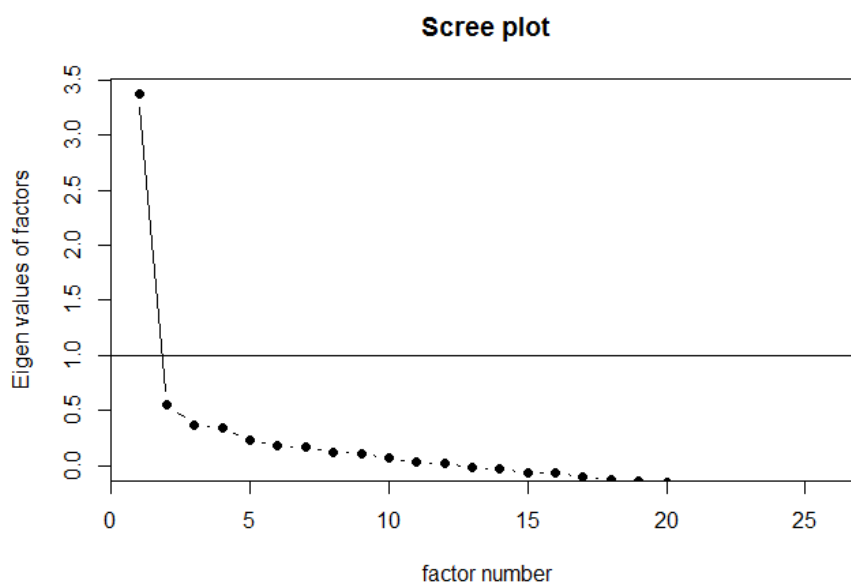


Figura 79: Gráfico para a análise do *Scree plot* e autovalores (“*eigen values*”) para retenção de fatores da matriz de correlação tetracórica.

⁹⁸ *Procedures for Psychological, Psychometric, and Personality Research.*

⁹⁹ *Tetrachoric, polychoric, biserial and polyserial correlations from various types of input* (ver: <http://127.0.0.1:14331/library/psych/html/tetrachor.html>).

4.2.3 Definição do Modelo Logístico

Segundo a Tabela abaixo, a comparação dos modelos aninhados mostrou **melhor ajuste do modelo de quatro parâmetros (4PL)**. Na comparação da verossimilhança dos modelos (Rasch x 2PL, 2PL x 3PL, 3PL x 4PL) via teste χ^2 todos os resultados obtidos foram significativos ao se adicionar o novo parâmetro ao modelo.

Tabela 30: Índices de ajuste, parâmetros e variância explicada dos modelos Rasch, 1PL, 2PL, 3PL, 4PL.

Modelo	logLik	AIC	AICc	SABIC	BIC	Parâmetros	Variância Explicada
Rasch	-84718,7	169491,3	169491,6	169584,3	169670,1	27	8,00%
2PL	-83704,5	167512,9	167513,9	167691,9	167857,1	52	12,00%
3PL	-83409,2	166975,0	166976,7	167243,0	167490,8	78	37,30%
4PL	-83302,8	166813,7	166817,7	167171,7	167502,2	104	61,10%

Assim procederam-se às análises utilizando o modelo de quatro parâmetros logísticos (4PL). Esse é um modelo que surgiu na literatura em 1981 (Barton & Lord, 1981) e tem se dado pouca atenção a ele. Ele considera a assíntota superior na estimação dos parâmetros (“*upper*”), referindo-se a sujeitos de alta competência errarem por descuido itens fáceis (Hambleton & Swaminathan, 1985). Não há muitos estudos relativos à utilização desse modelo (Pereira, Andrade, & Tavares, 2004) seja pela afirmação de que não oferece maiores vantagens diante do modelo 3PL (José Muñiz, 1997), dificuldade para a estimação e por oferecer poucos benefícios práticos. Atualmente já se encontram pacotes disponíveis em grande parte dos *softwares* sobre TRI, incluindo alguns livres, como o R. Considera-se, entretanto, que se assumem alguns riscos inferenciais ao utilizar um modelo de análise que não corresponda ao modelo sugerido (Loken & Rulison, 2010), e existem supostamente dois motivos para isso:

Primeiro, as sugestões para a sua aplicação foram bastante isoladas, de modo que não existe um consenso claro sobre a necessidade ou a utilidade desse modelo (Barton & Lord, 1981; Hambleton & Swaminathan, 1985); e em segundo lugar, o ajuste das 4PL foi considerado difícil, já que até mesmo estimar a assíntota inferior nas 3PL pode ser difícil (Baker & Kim, 2004; Embretson & Reise, 2000). No que se segue, argumentamos que existem várias aplicações potenciais na educação e psicologia que pode exigir uma 4PL e argumentamos que é bastante direto encaixar as 4PL usando uma abordagem bayesiana. Em seguida, use um estudo de simulação

para demonstrar que as inferências podem ser seriamente comprometidas quando o modelo usado para análise não corresponde aos dados do modelo gerador (ou seja, quando o 2PL ou 3PL é usado para analisar dados gerados a partir de uma 4PL). (Loken & Rulison, 2010, *tradução livre*)

4.2.4 Ajustando o Modelo Geral a partir dos itens: Brasil e Portugal

Os parâmetros dos itens são apresentados na Tabela 31. De forma iterativa, itens que apresentaram baixa discriminação ou discriminação negativa foram excluídos e novamente os parâmetros foram estimados. Na primeira rodada de análise os itens y6, y11, y12, e y21 foram excluídos devido aos coeficientes ponto-bisseriais terem sido *próximos de zero* (ver no Apêndice F1, p. 278). Na segunda rodada, os itens y7 (discriminação = 0,18) e o y4 (discriminação = 0,51) foram excluídos pelos mesmos critérios resultando na versão final de 20 itens.

Tabela 31: Os 4 parâmetros dos 26 itens associados a proficiência em conhecimentos introdutórios da Termodinâmica, segundo o TCE.

Itens	a	b	g	u
y1	7,80	-0,83	0,00	0,75
y2	2,70	-0,68	0,00	0,59
y3	0,83	0,28	0,00	1,00
y4	0,50	3,41	0,21	0,91
y5	3,31	0,62	0,19	0,56
y6	-0,01	-290,40	0,09	0,92
y7	0,27	-3,65	0,06	0,76
y8	1,80	2,37	0,26	1,00
y9	9,44	1,66	0,24	0,73
y10	6,52	0,77	0,23	0,78
y11	0,73	2,96	0,13	0,98
y12	0,12	-0,71	0,12	0,98
y13	9,53	0,56	0,18	0,87
y14	5,48	1,21	0,12	1,00
y15	1,17	2,90	0,12	0,99
y16	9,96	1,06	0,18	0,74
y17	5,37	1,47	0,21	0,49
y18	7,65	0,92	0,17	0,60
y19	7,66	2,52	0,22	1,00
y20	3,85	0,97	0,21	0,57
y21	0,12	8,80	0,13	0,95
y22	4,17	0,76	0,30	0,75
y23	3,96	0,63	0,28	0,90
y24	14,23	1,46	0,21	1,00
y25	1,79	1,97	0,23	1,00
y26	1,64	1,43	0,24	1,00

Nota: Itens em negrito foram excluídos devido ao parâmetro de discriminação negativo ou baixa discriminação; a = parâmetro de discriminação; b = parâmetro de dificuldade; g = parâmetro de acerto ao acaso ("guessing"); u = parâmetro de erro ao acaso ("upper").

Convém ressaltar que a exclusão desses itens resulta em um *guessing rate* de 5 itens ($19 \cdot 0,25 + 1 \cdot 0,20 = 4,95 \approx 5$ **itens**, equivalente a 19,23% do teste), ou seja, os alunos de modo geral, mesmo não dominando absolutamente nada no teste possuem uma probabilidade de acertarem 5 itens. Lembrando que anteriormente, ao se considerar os 26 itens, o *guessing rate* era de 7 itens (26,92% do teste).

Nota-se que a exclusão dos seis itens melhorou o modelo de medida, **o novo modelo 4PL com 20 itens foi responsável pela explicação de 78,2% da variância dos padrões de resposta** (verossimilhança = - 63131,7; AIC = 126423,4; AICc = 126425,8; SABIC = 126698,8; BIC = 126953), quando em comparação 61,1% dos 26 itens. As dificuldades dos itens variaram de -0,84 a 3,00 ($M = 1,01$; $DP = 0,96$). Na Tabela 3 é possível visualizar o ordenamento dos itens de acordo com a dificuldade.

Tabela 32: Os 4 parâmetros dos 20 itens associados a proficiência em conhecimentos introdutórios da Termodinâmica, segundo o TCE (em ordem crescente de dificuldade).

Itens	a	b	g	u
y1	6,61	-0,84	0,00	0,75
y2	3,10	-0,68	0,00	0,58
y3	0,87	0,19	0,00	0,96
y13	8,69	0,56	0,18	0,87
y5	3,45	0,61	0,20	0,55
y23	4,30	0,62	0,29	0,90
y10	6,42	0,75	0,23	0,77
y22	4,74	0,76	0,30	0,75
y18	7,04	0,91	0,17	0,59
y20	3,81	0,96	0,21	0,57
y16	8,94	1,05	0,18	0,74
y14	5,48	1,20	0,12	1,00
y26	1,67	1,43	0,24	1,00
y24	14,01	1,45	0,20	1,00
y17	4,70	1,55	0,21	0,51
y9	9,52	1,65	0,24	0,73
y25	1,72	1,98	0,23	1,00
y8	1,93	2,37	0,26	1,00
y19	7,48	2,52	0,22	1,00
y15	1,04	3,00	0,12	0,98

Nota: g = parâmetro de acerto ao acaso; u = parâmetro de erro ao acaso.

Seguindo os critério de classificação para os parâmetros de discriminação (a) e dificuldade (b) para a TRI, é possível identificar preliminarmente algumas características do itens para cada contexto, bem como de itens com mesmas tipologias, lembrando que o modelo de DIF pressupõe uma mesma métrica para os grupos, podendo ser feita uma interpretação comparativa direta, diferentemente de como foi feita com os mesmos índices na TCT.

Comparativamente, o item y3 e y26 estão classificados de forma semelhante, e é possível identificar dos itens y1 e y2 para ambos os contextos, porém, Portugal se destaca com mais itens com *alta e muita alta discriminação*, isso significa um melhor consistência.

Para o grupo geral, ou seja, considerando uma única base de dados para ambos os grupos, na qual foi construída uma escala de proficiência térmica válida para ambos os contextos, se observa que os itens y1, y2 e y3 são os mais fáceis, sendo predominante os itens classificados como *difícil e muito difícil*.

Quadro 70: Identificação comparativa de grupos de itens dentro de cada contexto de acordo com os critérios de dificuldade (b) e de discriminação (a).

Brasil (modelo 4PL com 20 itens)						
a \ b	Muito fáceis	Fáceis	Medianos	Difíceis	Muito difíceis	
Nenhuma						
Muito baixa						y17
Baixa						y15
Moderada			y3			y19
Alta						y22; y25
Muito Alta		y1	y2	y10; y13; y23	y5; y8; y9; y14; y16; y18; y20; y24; y26	
Portugal (modelo 4PL com 20 itens)						
a \ b	Muito fáceis	Fáceis	Medianos	Difíceis	Muito difíceis	
Nenhuma						
Muito baixa						
Baixa						
Moderada			y3			
Alta			y5			y8
Muito Alta	y1	y2	y10; y13; y22; y23	y9; y14; y16; y17; y18; y20; y24	y15; y19; y25; y26	

Brasil e Portugal (modelo geral 4PL com 20 itens)					
a \ b	Muito fáceis	Fáceis	Medianos	Difíceis	Muito difíceis
Nenhuma					
Muito baixa					
Baixa					
Moderada			y3		y15
Alta					y26
Muito Alta		y1; y2		y13; y5; y23; y10; y22; y18 y20; y16; y14	y24; y17; y9; y25; y8; y19

4.2.5 Análise de Funcionamento Diferencial dos Itens

Como foi dito anteriormente, esse procedimento de análise representa uma tendência de funcionamento diferenciado de um item em dois em grupos distintos, de modo a favorecer determinado grupo. O interessante é que essa análise possibilita uma verificação da ocorrência de viés nos itens em relação a variáveis como o sexo, escolaridade, região do país, entre outras. Sabe-se que a base de dados de Brasil e Portugal contempla diferentes níveis de escolaridade, sexo e regiões específicas, entretanto, a análise realizada considerou em termos gerais a ocorrência do DIF, na intenção de identificar quais os itens que ocorre DIF com relação ao parâmetro de dificuldade entre os dois contextos, ou seja, se determinado item beneficia um dos contextos. Para entender as análises, considera-se um item com DIF para sexo, por exemplo, assim, ele irá apresentar expectativas de respostas diversas entre homens e mulheres, ou seja, **probabilidades de acertos diferentes para as pessoas com o mesmo nível de habilidade/mesmo valor de θ** (Hutz et al., 2015, p. 118).

Para análises de DIF, três diferentes métodos foram utilizados para identificar uma maior consistência nas análises, o procedimento Mantel-Haenszel de acordo com o valor alfa e delta, o modelo Rasch se concentrando no parâmetro da dificuldade (DIF uniforme), e também sendo comparados com a utilização de um modelo restringindo o parâmetro de dificuldade e de discriminação para serem iguais entre a amostra brasileira e portuguesa. Inicialmente se tem os resultados DIF utilizando o teste Mantel-Haenszel (Mantel & Haenszel, 1959, p. 744), na qual eles destacam que *“reconhecendo a facilidade com que as associações podem ser influenciadas pela natureza do problema, pode ser importante que a medida do risco relativo seja significativamente diferente”*

(trad. livre), sendo o “risco” nesse caso, relacionado ao âmbito da epidemiologia no campo das ciências da saúde, mas que entendimento pode ser transferido de forma análoga para a educação, entre outros. As análises mostram que esses “riscos” estão representados por um DIF significativamente relevante entre os itens y8, y9, y13 e y19 para ambos os contextos.

Tabela 33: Resultados da análise de funcionamento diferencial dos itens utilizando o método Mantel-Haenszel.

Item	Estatística	<i>p</i>	$\hat{\alpha}_{MH}$	MH (logits)	Delta MH	Tamanho de efeito	Grupo favorecido
y1	57,97	0,00		1,6964	-1,242	Moderado	Portugal
y2	0,01	0,92		0,9915	0,0201	Pequeno	Brasil
y3	1,99	0,16		0,9073	0,2287	Pequeno	Brasil
y5	9,78	0,00		1,2517	-0,5276	Pequeno	Portugal
y8	107,16	0,00		0,4848	1,7014	Grande	Brasil
y9	135,02	0,00		0,43	1,9831	Grande	Brasil
y10	22,57	0,00		0,7156	0,7863	Pequeno	Brasil
y13	181,49	0,00		2,6015	-2,2468	Grande	Portugal
y14	29,25	0,00		1,5721	-1,0631	Moderado	Portugal
y15	25,64	0,00		0,6604	0,9749	Pequeno	Brasil
y16	0,70	0,40		0,9369	0,1531	Pequeno	Brasil
y17	64,28	0,00		0,5494	1,4075	Moderado	Brasil
y18	3,96	0,05		1,1659	-0,3608	Pequeno	Portugal
y19	249,36	0,00		0,2963	2,8584	Grande	Brasil
y20	0,76	0,38		1,0662	-0,1506	Pequeno	Portugal
y22	2,11	0,15		1,1024	-0,2291	Pequeno	Portugal
y23	25,97	0,00		1,4162	-0,8178	Pequeno	Portugal
y24	55,24	0,00		0,5746	1,3023	Moderado	Brasil
y25	40,36	0,00		0,6364	1,062	Moderado	Brasil
y26	8,0256	0,01		1,2095	-0,447	Pequeno	Portugal

Após algumas adaptações no teste χ^2 (Holland & Thayer, 1988) o alfa de Mantel-Haenszel pode ser estimado pela expressão (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005, p. 127):

$$\hat{\alpha}_{MH} = \frac{\sum_{k=1}^m \frac{A_k \cdot D_k}{N_k}}{\sum_{k=1}^m \frac{B_k \cdot C_k}{N_k}}$$

A lógica do procedimento Mantel-Haenszel é a seguinte: **se um item não apresenta DIF, a razão entre o número de respondentes que acertam o item e os que**

falham deve ser a mesma nos dois grupos comparados ao longo de todos os níveis de pontuação, com isso, tem-se:

$$H_0: \frac{A_k}{B_k} = \alpha \cdot \frac{C_k}{D_k}, \text{ sendo } \alpha = 1 \text{ para todo } k \text{ (ausência de DIF)}$$

$$H: \frac{A_k}{B_k} = \alpha \cdot \frac{C_k}{D_k}, \text{ sendo } \alpha \neq 1 \text{ para todo } k \text{ (DIF)}$$

, sendo A_k e B_k os números de respondentes de acerto e erro para o grupo de referência, e C_k e D_k se refere ao grupo focal, e N_k o número total de respondentes de acerto e erro no nível k .

Considerando que o estimador varie no intervalo: $0 < \hat{\alpha}_{MH} < \infty$, um valor $\hat{\alpha}_{MH} = 1$ representa a hipótese nula, com isso, valores próximo a 1 considera-se um DIF pequeno, e se $\hat{\alpha}_{MH} > 1$ o DIF favorece o grupo de referência (Portugal), e $\hat{\alpha}_{MH} < 1$ favorece o grupo focal (Brasil)

O DIF na escala delta das diferenças pode ser obtido através da conversão entre logits e unidades delta, é (Holland & Thayer, 1988):

$$\text{Delta MH} = -2,35 \ln(\hat{\alpha}_{MH})$$

Nessa escala, valores encontram-se entorno de zero. **Valores negativos indicam que o item favorece o grupo de referência** (Portugal) e os **valores positivos o grupo focal** (Brasil). Foi considerado um **DIF pequeno** um módulo inferior a uma unidade delta, acima e próximo de 1 unidade um **DIF moderado**, e próximo e acima de 2 unidades um **DIF grande**.

Diante das análises, pode-se afirmar que existem 9 itens favoráveis a Portugal e 11 itens para o Brasil, entretanto existem alguns **diferenças pequenas**, na qual pode-se dizer que sejam praticamente equivalentes. As pequenas diferenças a favor do Brasil foram 5 itens (y2, y3, y10, y15, y16) e 6 itens para Portugal (y5, y18, y20, y22, y23 e y26). As **diferenças moderadas** resultam em 2 favoráveis a Portugal (y1, y14) e 3 para o Brasil (y17, y24, y25). Entre os valores de **maior consistência e magnitude** com relação ao DIF se tem 1 item favorável a Portugal (y13) e 3 itens ao Brasil (y8, y9, y19).

Uma análise complementar semelhante pode ser realizada através do *software* Winsteps, utilizando o modelo Rasch (1PL, restringindo a discriminação “a”, considerando o parâmetro “b”, e ignorando o parâmetro “c”), e assim, verificando com maior precisão a existência de **DIF uniforme** ao restringir o parâmetro de dificuldades dos itens para os alunos em cada contexto. A forma comum do modelo Rasch DIF pode ser articulada utilizando a função logit (Jusoh & Nor, 2014):

$$\text{Logit} = \{P(X_{n,i} = 1|\theta_n, g)\} = \theta_n - \delta_i + \gamma_i \cdot G$$

, onde $X_{n,i}$ é o escore para o indivíduo i ($1 = \text{correto}; 0 = \text{incorreto}; i = 1,2,3, \dots, I; n = 1,2,3 \dots N$), θ_n é o escalar do parâmetro de proficiência (habilidade) para o indivíduo n , δ_i é o parâmetro de dificuldade do item i ; γ_i é o índice do parâmetro de DIF para o item i , e g é o indicativo do grupo de referência ou grupo local, sendo $G = 1$ ou $g = R$ (grupo de referência), e $G = 0$ ou $g = F$ (grupo focal). Na equação, o γ_i é a diferença de dificuldade do item entre o grupo focal e o grupo de referência (i. e., $g = \delta_i F - \delta_i R$).

Para interpretar a tabela elaborada a partir dos dados extraídos das análises do programa *Winsteps*, segue-se as seguintes orientações:

- Na primeira coluna estão identificados os 20 itens analisados do modelo geral ajustado, considerando uma comparação entre cada contexto, Brasil e Portugal;
- A coluna “*person class*” (“classes de pessoas”) indica os grupos a serem comparados, de acordo com o parâmetro de dificuldade analisado.
- Na terceira coluna, da esquerda, tem-se o “*diff. measure*”, que se refere aos parâmetros de dificuldades para o grupo 1 (Brasil) e a sexta coluna para o grupo 2 (Portugal), considerando o restante dos parâmetros constante. É preciso considerar que o:
 1. DIF seja suficientemente grande em tamanho para ser notável (geralmente tamanho $> 0,5$ logits);
 2. DIF seja suficientemente significativo para não ter acontecido por acaso (geralmente $t > 2,0$).

Alguns critérios próximos a esses também são recomendados (Zwick, Thayer, & Lewis, 1999), na qual se tem $1 \text{ logit} = 2,35 \text{ unidades Delta } (\delta)$, conforme se adota no *Educational Testing Service - ETS* (“escala Delta”) desde os trabalhos clássicos

desenvolvidos em Princeton para conversão do estimador de Mantel-Haenszel (Holland & Thayer, 1988; Holland & Thayer, 1985). No caso do teste *t-student*, tem-se a adaptação e desenvolvimento partir de outros trabalhos (Satterthwaite, 1946; Welch, 1947).

- A diferença entre as probabilidades de acerto dos dois grupos está representada na coluna “*DIF contrast*” (“constrate das diferenças no parâmetro *b*”), cujo sentido, positivo ou negativo, oscila somente em função da ordem de inserção dos grupos. Assim, o foco é o valor dessa diferença, e não se positiva ou negativa.

A interpretação irá seguir o sistema de classificação ETS: C-, B-, A, B+, e C+.

Tabela 34: Classificação do tamanho do efeito de DIF, segundo a ETC.

ETS DIF Category	Com “DIF Contrast” e “DIF Statistical Significance”	
C = moderada a grande	$ DIF \geq 0,64 \text{ logits}$	Prob ($ DIF \leq 0,43 \text{ logits}$) $\leq 0,05$ (2-sided) <i>aproximadamente: $DIF > 0,43 \text{ logits} + 2 * DIF \text{ S.E.}$</i>
B = sutil a moderado	$ DIF \geq 0,43 \text{ logits}$	Prob ($ DIF = 0 \text{ logits}$) $\leq 0,05$ (2-sided) <i>aproximadamente: $DIF > 2 * DIF \text{ S.E.}$</i>
A = insignificante	-	-

C-, B- = DIF favorece o grupo de referência; C+, B+ = DIF favorece o grupo focal

Fonte: “Mantel-Haenszel DIF and PROX are Equivalent!” (n.d., *tradução livre*).

- A coluna Mantel-Haenszel apresenta a significância da diferença encontrada entre os grupos, cuja análise indicará ou não a presença de DIF se $p < 0,05$.

O mais importante na resolução de DIF é que podemos comunicar esse processo de forma simples, clara e precisa, ou seja, compreensível para o público. A importância se dá por minimizar quaisquer sentimentos de injustiça ou preconceito por diferentes grupos. Para isso, seguem algumas das recomendações no caso de evidências significativas de DIF, seriam:

- Ignorar como inerente ao sistema de medição. No processo de construção de testes, os especialistas em conteúdo devem **tentar equilibrar o DIF nos principais grupos**, mas isso nunca pode ser feito perfeitamente.

- **Remova o item** (talvez reescrito). Isto é, quando o DIF é considerado um fator de confusão que supera a intenção do item.
- Trate os dados de um grupo DIF como **dados em falta**.
- **Divida o item**. Faça o item dois itens, cada um com dados ativos para um grupo e dados faltantes para o outro grupo. Isso mantém as mesmas pontuações brutas, mas produz diferentes medidas para cada grupo.

Tabela 35: Análise de DIF dos 20 itens por País pelo Modelo Rasch.

Item	Person Class (focal)	Diff. Measure	Diff. S.E.	Person Class (referência)	Diff. Measure	Diff. S.E.	Diff. Contrast	Class. ETS	Joint S.E.	Rasch-Welch			Mantel-Haenszel	
										t	d.f.	Prob.	Prob.	Size
y1	Brasil	-1,02	0,04	Portugal	-1,72	0,05	0,70	C+	0,06	11,43	Inf	0,00	0,00	0,78
y2	Brasil	-0,47	0,04	Portugal	-0,57	0,04	0,09	A	0,06	1,58	Inf	0,11	0,01	0,20
y3	Brasil	-0,61	0,04	Portugal	-0,66	0,04	0,05	A	0,06	0,89	Inf	0,37	0,95	-0,02
y5	Brasil	0,26	0,05	Portugal	-0,06	0,04	0,32	A	0,06	5,00	Inf	0,00	0,00	0,42
y8	Brasil	-0,22	0,04	Portugal	0,55	0,05	-0,78	C-	0,06	-12,3	Inf	0,00	0,00	-0,66
y9	Brasil	-0,14	0,04	Portugal	0,72	0,05	-0,86	C-	0,06	-13,4	Inf	0,00	0,00	-0,68
y10	Brasil	-0,27	0,04	Portugal	-0,12	0,04	-0,15	A	0,06	-2,49	Inf	0,01	0,00	-0,36
y13	Brasil	0,28	0,05	Portugal	-0,89	0,04	1,17	C+	0,06	18,33	Inf	0,00	0,00	0,98
y14	Brasil	0,93	0,06	Portugal	0,19	0,04	0,74	C+	0,07	10,25	Inf	0,00	0,00	0,48
y15	Brasil	0,68	0,05	Portugal	1,13	0,05	-0,45	B+	0,08	-5,97	Inf	0,00	0,00	-0,29
y16	Brasil	0,35	0,05	Portugal	0,33	0,05	0,02	A	0,07	0,32	Inf	0,75	0,02	-0,18
y17	Brasil	0,17	0,05	Portugal	0,84	0,05	-0,67	C-	0,07	-9,88	Inf	0,00	0,00	-0,42
y18	Brasil	0,56	0,05	Portugal	0,27	0,04	0,29	A	0,07	4,27	Inf	0,00	0,00	0,26
y19	Brasil	-0,09	0,04	Portugal	1,21	0,05	-1,30	C-	0,07	-18,9	Inf	0,00	0,00	-0,93
y20	Brasil	0,28	0,05	Portugal	0,19	0,04	0,09	A	0,07	1,36	Inf	0,17	0,01	0,20
y22	Brasil	-0,38	0,04	Portugal	-0,53	0,04	0,14	A	0,06	2,40	Inf	0,02	0,04	0,13
y23	Brasil	-0,42	0,04	Portugal	-0,98	0,04	0,55	B+	0,06	9,14	Inf	0,00	0,00	0,40
y24	Brasil	0,09	0,05	Portugal	0,52	0,05	-0,43	B-	0,07	-6,57	Inf	0,00	0,00	-0,57
y25	Brasil	-0,08	0,05	Portugal	0,33	0,05	-0,41	A	0,06	-6,36	Inf	0,00	0,00	-0,33
y26	Brasil	-0,1	0,04	Portugal	-0,37	0,04	0,27	A	0,06	4,36	Inf	0,00	0,00	0,32

Para um interpretação do DIF uniforme deve-se admitir um probabilidade de 50% de acerto para indivíduos com habilidade diferentes estimadas, assim, aqueles com habilidades inferiores ou com menor medida de dificuldade (“*Diff. Measure*”) possui vantagens por ser mais fácil de acertar. Por exemplo, no item y1, os resultados apresentam habilidades θ diferentes para alunos portugueses e brasileiros para que tenham 50% êxito ou chance de acerto, no caso, os portugueses apresentam $\theta = -1,72$ enquanto os brasileiros $\theta = -1,02$, sendo a diferença significativa ($p < 0,00$), ou seja, o item y1 é mais fácil de ser acertado para alunos portugueses do que brasileiros, existe um funcionamento diferencial do item favorecendo o contexto português. No item y8, os brasileiros possuem vantagem estimada, com $\theta = -0,22$ é possível garantir 50% de probabilidade de acerto, enquanto os portugueses precisam de um $\theta = 0,55$. A explicação para os DIFs identificados é um campo que extrapola o escopo dessa investigação, porém serve de evidência a estudos correlatos ou complementares.

Cruzando os valores de DIFs mais expressivos nas duas análises, pode-se afirmar que o Brasil possui um facilidade moderadora no item y24 diante de Portugal, e uma vantagem maior para os itens y8, y9 e y19. O item y13 sugere ser uma maior facilidade para os alunos portugueses diante aos brasileiros.

Tabela 36: Cruzamento no tamanho do efeito de itens com DIF nas duas análises realizadas para o grupo focal (Brasil) e grupo de referência (Portugal).

Grupo favorecido	Tamanho do Efeito		
	Pequeno	Moderado	Grande
Itens favoráveis ao Brasil	y2, y3, y10 , y15, y16	y17, y24 , y25	y8, y9, y19
Itens favoráveis a Portugal	y5, y18, y20, y22, y23, y26	y1, y14	y13
Grupo favorecido	Tamanho do efeito		
	A	B- (Brasil)/B+ Portug.)	C- (Brasil)/C+ (Portug.)
Itens favoráveis ao Brasil	y10, y25	y24	y8, y9, y17, y19
Itens favoráveis a Portugal	y2, y3, y5, y16, y18, y20, y22, y26	y15, y23	y1, y13, y14
Grupo favorecido	Itens coincidentes nas análises de DIF		
Itens favoráveis ao Brasil	y10	y24	y8, y9, y19
Itens favoráveis a Portugal	y5, y18, y20, y22, y26		y13

A maioria dos itens que se encontram com um baixo valor de DIF indica que não há diferenças significativas entre os respondentes nos dois contextos, isso indica que esses itens se comportam de forma muito semelhante para ambas as amostras, de

modo geral. Sabe-se que podem haver DIFs mais específicos quanto aos níveis de escolaridade e sexo, porém não foram delimitados esses aspectos na análise.

Conforme pôde ser visualizado o teste de proficiência apresentou grande incidência de itens com funcionamento diferencial de acordo com o país de origem (Portugal e Brasil), o que não assegura a invariância da medida. Para possibilitar a comparação dos desempenhos entre os dois países sem violação da invariância da medida foram realizadas **análises multigrupo restringindo** todos os itens para terem «dificuldades e discriminações iguais» nas duas subamostras. Esse *modelo restringido* foi então comparado com o *modelo não restringido*, e escolhido para as interpretações o modelo mais ajustado aos dados.

O **modelo não restringido** apresentou melhor ajuste aos dados do que o modelo restringido, sendo essa diferença significativa a nível $p < 0,001$, porém nota-se que os valores de ajuste são muito próximos e o tamanho das amostras tenha identificado diferenças significativas, ainda que tenha sido sutil as diferenças. Ressalta-se que apesar do modelo restringido apresentar um pior ajuste ainda assim ele atende aos critérios de ajustes observados pelos índices de ajuste geral CFI, TLI e RMSEA.

Tabela 37: Comparação no processo de ajuste entre o modelo multigrupo restringido e o não restringido.

Modelo	AIC	AICc	SABIC	BIC	logLik	CFI	TLI	RMSEA
Restringido	127580,1	127585,4	127993,2	128374,5	-63670	0,95	0,93	0,01
Não Restringido	127439,8	127449,4	127990,6	128499	-63559,9	0,95	0,94	0,01

Diante disso, a estimativa dos parâmetros realizada pela análise multigrupo resultou nos seguintes resultados para o modelo não-restringido.

Tabela 38: Estimativa dos parâmetros logísticos para o modelo não-restringido diante da análise multigrupo entre a amostra Brasil e Portugal.

Modelo não restringido (20 itens)									
Portugal					Brasil				
Itens	a	b	g	u	Itens	a	b	g	u
y1	0,19	-5,74	0,00	1,00	y1	9,62	-0,80	0,01	0,59
y2	1,22	1,74	0,43	0,98	y2	8,53	-0,49	0,00	0,50
y3	1,40	0,94	0,36	1,00	y3	1,18	-0,03	0,00	0,74
y5	0,53	1,10	0,07	0,99	y5	7,18	1,29	0,19	0,42

y8	1,10	2,26	0,20	1,00	y8	1,74	2,45	0,26	0,99
y9	12,96	1,13	0,20	0,63	y9	4,14	2,25	0,26	1,00
y10	3,08	0,23	0,13	0,82	y10	10,59	1,08	0,23	0,74
y13	1,96	-0,35	0,02	0,96	y13	12,03	1,12	0,15	0,58
y14	4,81	0,74	0,15	1,00	y14	5,83	2,17	0,11	1,00
y15	3,49	2,02	0,17	0,99	y15	2,07	0,89	0,11	0,28
y16	9,98	0,59	0,18	0,74	y16	10,08	1,56	0,17	0,65
y17	3,03	0,88	0,17	0,48	y17	0,00	2,90	0,16	0,85
y18	11,07	0,43	0,21	0,59	y18	5,33	1,81	0,15	0,60
y19	6,38	2,22	0,17	1,00	y19	0,67	4,38	0,22	0,96
y20	2,06	0,66	0,23	0,64	y20	2,84	2,94	0,21	0,99
y22	0,89	0,29	0,11	1,00	y22	1,76	1,35	0,26	0,62
y23	1,33	-0,12	0,15	1,00	y23	8,90	1,07	0,25	0,77
y24	13,95	0,98	0,17	0,94	y24	10,41	2,25	0,22	1,00
y25	1,52	1,69	0,23	1,00	y25	1,56	2,66	0,23	1,00
y26	1,75	1,30	0,36	1,00	y26	1,97	2,09	0,23	1,00

Reagrupando os itens de acordo com os critérios que classificam os itens conforme os valores para o parâmetro de discriminação obtidos no modelos não-restringido e no modelo geral, considerando os 20 itens, pode-se constatar:

Tabela 39: Identificação de grupos de itens dentro de cada contexto de acordo com os critérios de discriminação da TRI, assim como itens coincidentes dentro de cada grupo e para os dois em geral.

Valores ¹⁰⁰	Classificação Discriminativa	Tipologia do item	Brasil	Portugal	Itens Coincidentes	Portugal e Brasil
$a = 0,0$	Nenhuma	Péssimo (sujeito a rejeição)	y17	-	-	-
$0,0 < a \leq 0,35$	Muito baixa	Muito Ruim (sujeito a reelaboração)	-	y1	-	-
$0,35 < a \leq 0,65$	Baixa	Ruim (sujeito a reelaboração)	-	-	-	-
$0,65 < a \leq 1,35$	Moderada	Bom (sujeito a aprimoramento)	y19; y25	y5; y8; y22; y23	-	y3; y15
$1,35 < a \leq 1,70$	Alta	Muito bom (sujeito a aprimoramento)	-	y2; y3; y25; y26	-	y26
$a > 1,70$	Muito alta	Ótimo (disponível a reaplicação)	y1; y2; y3; y5; y8; y9; y10; y13; y14; y15; y16; y18; y20; y22; y23; y24; y26	y9; y10; y13; y14; y15; y16; y17; y18; y19; y20; y24	y9; y10; y13; y14; y15; y16; y18; y20; y24	y1; y2; y13; y5; y23; y10; y22; y18; y20; y16; y14; y24; y17; y9; y25; y8; y19

Fonte: Elaborado com critérios adaptados de Rabelo (2013, p. 136).

Considerando o parâmetro de dificuldade dos itens, tem-se:

¹⁰⁰ Foi adotado cores correspondentes para melhor visualizar cada grupo dentro dos quadros e nos seus respectivos gráficos comparativos.

Tabela 40: Identificação de grupos de itens dentro de cada contexto de acordo com os critérios de dificuldade da TRI, assim como itens coincidentes dentro de cada grupo e para os dois em geral.

Valores	Proporção esperada	Tipologia do item	Brasil	Portugal	Itens coincidentes	Brasil e Portugal
$b \leq -1,28$	10%	Muito fáceis	-	y1	-	-
$-1,28 < b \leq -0,52$	20%	Fáceis	y1; y2	-	-	y1; y2
$-0,52 < b \leq 0,51$	40%	Medianos	y3	y10; y13; y18; y22; y23	-	y3
$0,51 < b < 1,28$	20%	Difíceis	y10; y13; y15; y22; y23	y3; y9; y14; y16; y17; y20; y24	-	y13; y5; y23; y10; y22; y18 y20; y16; y14
$b \geq 1,28$	10%	Muito difíceis	y5; y8; y9; y14; y16; y17; y18 y19; y20; y24; y25; y26	y2; y5; y8; y15; y19; y25; y26	y5; y8; y19; y25; y26	y26; y24; y17; y9; y25; y8; y19; y15

Ao levar em consideração a combinação dos dois parâmetros logísticos, observa-se que a amostra brasileira agrupa um maior número de itens difíceis.

Tabela 41: Identificação comparativa de grupos de itens dentro de cada contexto e para os dois juntos, de acordo com os critérios e valores estimados para os parâmetros de dificuldade (a) e de discriminação (b) para TRI.

Brasil (modelo não-restringido com 20 itens)						
a \ b	Muito fáceis	Fáceis	Medianos	Difíceis	Muito difíceis	
Nenhuma						y17
Muito baixa						
Baixa						
Moderada						y19; y25
Alta						
Muito Alta		y1; y2	y3	y10; y13; y15; y22; y23	y5; y8; y9; y14; y16; y18; y20; y24; y26	
Portugal (modelo não-restringido com 20 itens)						
a \ b	Muito fáceis	Fáceis	Medianos	Difíceis	Muito difíceis	
Nenhuma						
Muito baixa	y1					
Baixa						
Moderada			y22; y23			y5; y8
Alta				y3		y2; y25; y26
Muito Alta			y10; y13; y18;	y9; y14; y16; y17; y20; y24		y15; y19
Brasil e Portugal (modelo geral com 20 itens)						
a \ b	Muito fáceis	Fáceis	Medianos	Difíceis	Muito difíceis	
Nenhuma						
Muito baixa						
Baixa						
Moderada			y3			y15
Alta						y26
Muito Alta		y1; y2		y13; y5; y23; y10; y22; y18 y20; y16; y14	y24; y17; y9; y25; y8; y19	

Apesar do ajuste ter possibilitado uma melhor discriminação para os itens, observa-se que a maioria dos itens são considerados de fato *difíceis* e *muito difíceis*, com exceção dos itens y1, y2 e y3. Comparando com os valores dos parâmetros estimados dentro do modelo restringido (ver no Apêndice A1, p. 9; A2, p. 19; e, A3, p. 29), no caso, estando restringido o parâmetro de discriminação e de dificuldade, é possível também verificar a intensidade de alguns eventuais DIFs entre os contextos comparando as CCIs, assim como das diferenças com os valores do modelo não-restringido.

4.2.6 Curva Característica do Teste - CCT

A análise desse tipo de curva é considerada descritiva, ou seja, mais pertinente a TCT, porém, foi propositalmente colocada nessa seção na intenção de facilitar as interpretações comparativas que tendem a seguirem de forma lógica e sequencial, fazendo-se perceber as aproximações com a perspectiva da TRI.

Considerando uma única base de dados com informações dos dois contextos (n=5542 alunos), a linha de corte do grupo inferior fica com 6 acertos e o superior com 10 acertos, sendo a moda e a média igual 8 acertos, e a média sendo 8,60. Nota-se que 96% de todos os alunos alcançam no máximo 15 acertos, ou seja, um pouco a mais da média esperada para o teste, sendo portanto um teste difícil para o conjunto de respondentes nos dois contextos, sendo uma pequena parcela a atingir no máximo 24 acertos.

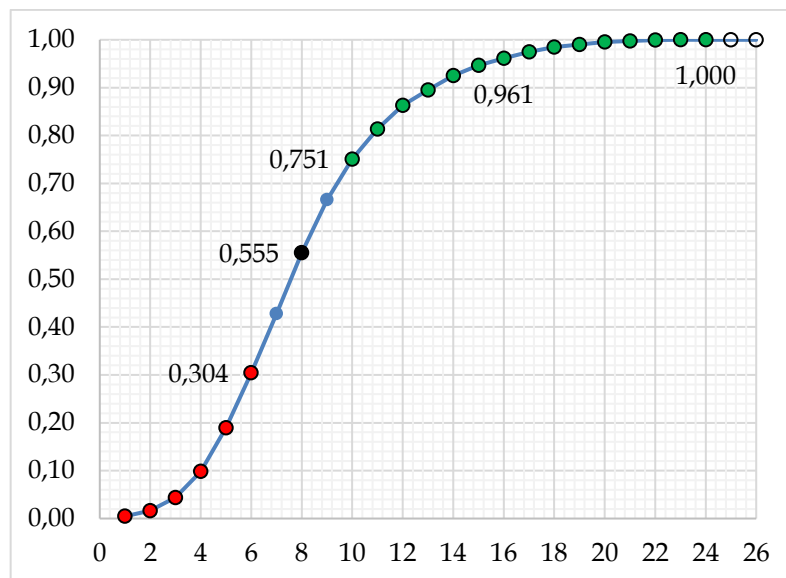


Figura 80: Curva do percentual acumulado versus escore bruto – Brasil e Portugal

Na TCT uma pontuação verdadeira representa a soma dos acertos entre os itens, além do modelo prever também a existência do erro. Como se sabe, essa pontuação é também denominada de *escore bruto*, na qual é possível uma transformação da escala de pontuações verdadeiras de forma similar a CCI na TRI, podendo ser aplicada não apenas por sujeitos com determinada habilidade θ como para o teste em si, levando em consideração todos os níveis θ . A pontuação verdadeira (V) pode ser representada por (Arias, 2005, p. 304-307):

$$V = \sum_{i=1}^n P_i(\theta)$$

Considerando as pontuações verdadeiras para o teste, torna-se possível a projeção da *Curva Característica do Teste* (CCT), na qual representa a relação funcional entre a *pontuação verdadeira* (parâmetro estimado pela TCT) e a *escala de habilidades* (parâmetro estimado pela TRI), levando em consideração o número de itens no teste. Não há uma forma explícita para essa curva, com isso não se podem utilizar parâmetros para descrevê-la, porém ela pode refletir informações dadas pelos valores médios estimados dos parâmetros de cada item da mesma forma para os modelos 1PL, 2PL, 3PL e 4PL. O interesse em situações práticas para essa curva está em interpretar as pontuações no teste dentro de cada habilidade θ ou nível de proficiência (conjunto de habilidades), sendo um procedimento de equiparação das pontuações dos testes.

De modo geral a CCT é uma função monótona crescente, em que, assintoticamente tende ao número de itens, podendo ser interpretada intuitivamente da mesma forma que a CCI. No contexto dos *testes referidos ao critério* pode ser obtido a *proporção de acertos* (π) ou a *pontuação do domínio* dividindo a pontuação verdadeira (V) pelo número de itens (n):

$$\pi = \frac{1}{n} \cdot \sum_{i=1}^n P_i(\theta)$$

Pegando os dados obtidos pelo modelo 4PL, podem ser obtidos as pontuações verdadeiras e valores médio para os parâmetros a, b, g e u, considerando n = 20:

Tabela 42: Relações entre θ , V e π .

θ	-3,000	-2,000	-1,000	0,000	1,000	2,000	3,000	4,000	5,000
V	3,659	3,742	4,236	5,548	9,268	13,525	15,388	15,928	16,129
π	0,183	0,187	0,212	0,277	0,463	0,676	0,769	0,796	0,806

Tabela 43: Valores médios para os parâmetros a, b, g e u.

Parâmetros	a	b	g	u
π	5,276	1,102	0,180	0,813

As curvas obtidas mostram que a soma das probabilidades de acerto cresce em cada domínio ou nível de proficiência conforme se elevam as habilidades dos respondentes, como era de se esperar. Considerando os *20 itens no modelo 4PL*¹⁰¹, a proficiência em média dos sujeitos que respondem corretamente a metade das questões ficou com $\theta = 1,135$, ou seja, com habilidades dentre os níveis 1 e 2. Isso mostra que o teste exige muitas habilidades, de acordo o conhecimento envolvido, e é, portanto, considerado difícil para os respondentes. A curva de proporção de domínio mostra que os alunos abaixo da média tiveram um percentual de acertos muito próximo ao *guessing* esperado (de modo geral) que seria média 0,243, e que com relação à média das proporções de acerto de todos os itens há um erro ao acaso de 19%.

¹⁰¹ No tópico relacionado a análise da TRI, foi realizado um ajuste e eliminado alguns itens restando 20. A CCT na TCT foi elaborada com base nessa análise para que possa facilitar a comparação e interpretação.

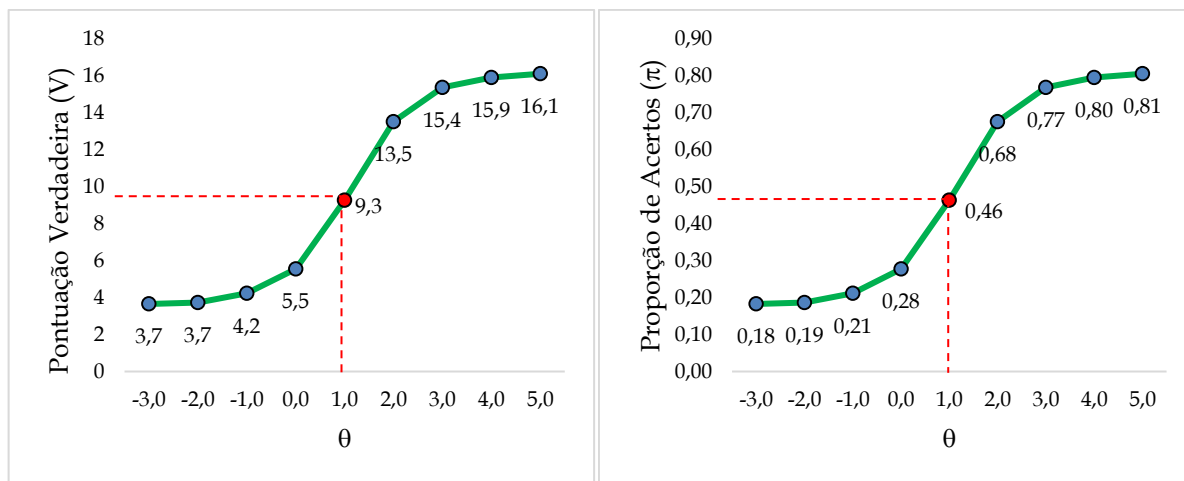


Figura 81: Curva Característica do Teste na TCT. Figura 82: Curva de proporção de acertos ou de pontuação de domínio na TCT.

Como pode ser observado, essas curvas apresentam perspectivas gerais do teste, seja da relação entre a pontuação verdadeira e as habilidades, como das proporções de acerto diante as habilidades médias obtidas, com isso podem servir de auxílio na leitura das CCI de um modelo geral dentro da TRI, que incluem os dois, Brasil e Portugal. A CCT para cada contexto colocada *lado a lado* se torna possível identificar para qual país o teste foi mais difícil, nesse caso, a assíntota superior estará mais elevada para a população que considerar o teste mais fácil, obviamente, pois irá tender a alcançar pontuações mais elevadas diante de uma mesma habilidade θ estimada. Entretanto, uma comparação semelhante também pode ser feita utilizando a *Curva de Informação do Teste* dentro da TRI.

4.2.7 Curva de Informação do Teste - CIT

Esta seção busca verificar e apresentar graficamente evidências que se configuram como aspectos de adequação do teste para cada contexto, de acordo com a curva de erro padrão associada a curva de informação. No sentido de ilustrar o nível de dificuldade de modo global entre os respondentes portugueses e brasileiros, foi realizada uma análise através do *software* BILOG-MG utilizando o modelo 3PL buscando a disponibilização da curva de informação do teste para os 2957 respondentes. Nisso, para fazer seguir um padrão de análise semelhante as análises de DIF, os itens foram calibrados através dos 2585 respondentes portugueses como grupo de referência, e utilizado para estimar a *Curva de Informação do Teste* (CIT) no Brasil e

comparar em uma mesma métrica com CIT no contexto de Portugal. Observando as duas curvas, pode-se perceber um teste supostamente com maior dificuldade para o grupo focal (Brasil), pois apresenta uma inclinação da curva para a direita acima da média ($\theta = 0$), revelando que existe uma concentração de informação em itens difíceis. Para Portugal, a média se encontra mais próxima de zero, ao contrário do Brasil. A *curva de erro padrão de medida* surge de modo a contrapor a *curva de informação* do teste, pois ambas estão intimamente ligadas e são inversamente proporcionais, quanto menos informação, maior será o erro padrão estimado, e vice-versa.

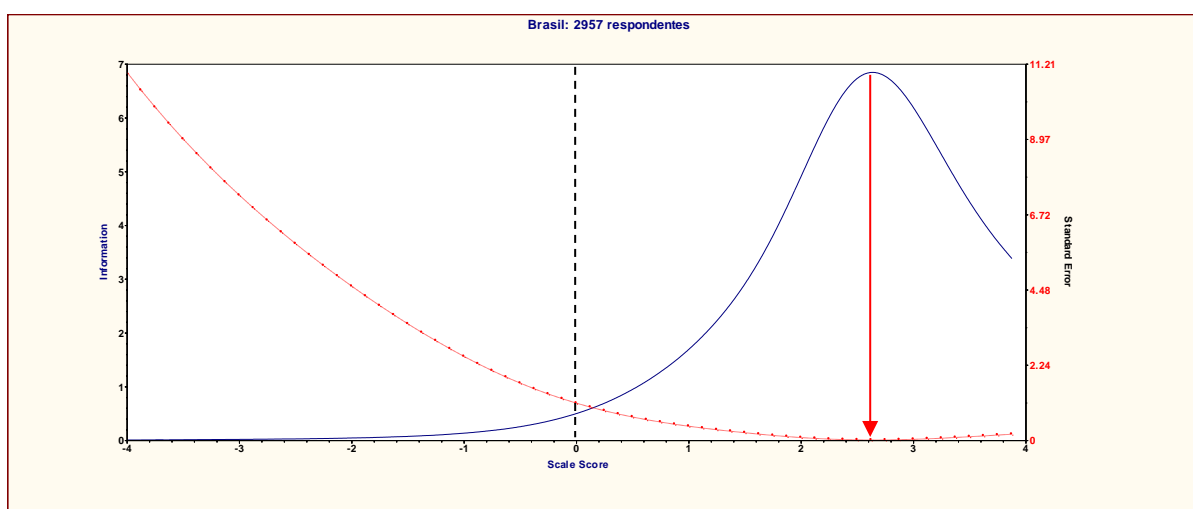


Figura 83: Curva de Informação do Teste (Brasil) através do modelo de 3PL, estimado pelo BILOG-MG.

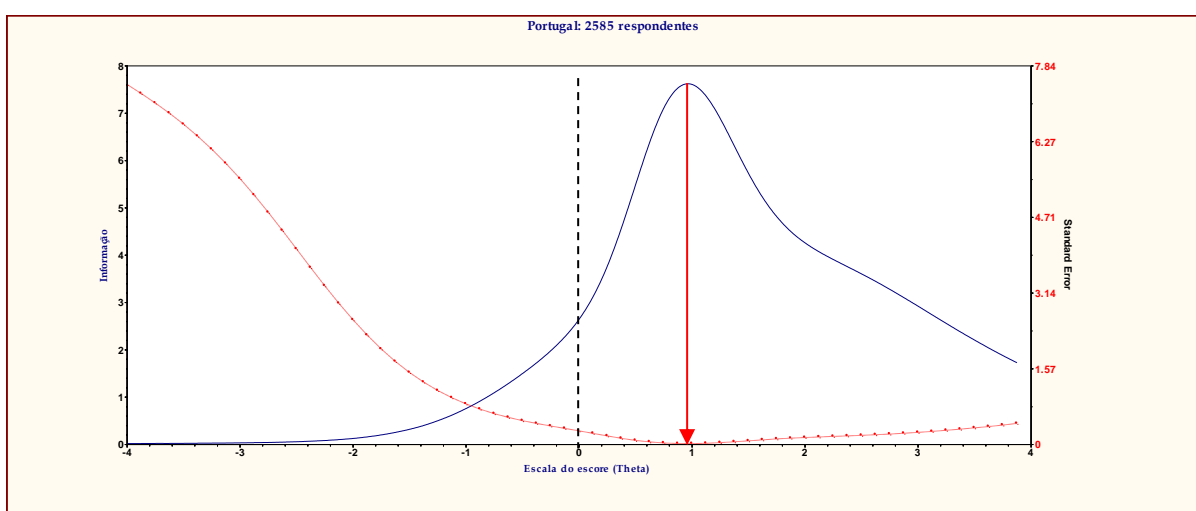


Figura 84: Curva de Informação do Teste (Portugal) através do modelo de 3PL, estimado pelo BILOG-MG.

Na sequência das análises é possível de forma precisa, utilizando o modelo 4PL, identificar os itens que fazem parte da escala de proficiências que podem justificar e esclarecer a ausência de itens mais fáceis para ambos os contextos, na qual se torna evidente os itens que possibilitam e validam a interpretação na escala, bem como das limitações existentes nesse tipo de análise.

4.2.8 Construção das Escalas de Proficiência sobre as Concepções Térmicas

A construção da escala de proficiências com base no teste TCE foi elaborada através da estimação dos parâmetros do modelo logístico 4PL, considerado como o melhor ajuste para a interpretação dos dados de modo geral. Esse aspecto se justifica pelo fato de buscar uma validade e interpretação para a escala em ambos os contextos, estando a escala disponibilizada para outras investigações caso haja interesse na avaliação diagnóstica sobre concepções térmicas de estudantes e construção de séries históricas de desempenho dentro desse domínio. Esses estudos têm como referência informações de psicométricas de alunos brasileiros e portugueses, estando em níveis de escolaridade relacionadas ao último ciclo da educação básica, no caso, 1º, 2º e 3º do ensino médio para o Brasil e o 10º, 11º e 12º ano em Portugal. Espera-se que os níveis de proficiência para alunos em séries mais avançadas alcancem níveis mais elevados tendendo assim na identificação e calibração de itens fáceis, medianos e difíceis, buscando com isso construir uma escala mais equilibrada para as interpretações pedagógicas, e definir de forma mais concreta os níveis de proficiência.

O processo de validação da escala é um aspecto complexo que depende de diversas etapas, e a busca para atender esse processo resulta em indícios de validação na qual partem desde a tradução do TCE para o contexto português e brasileiros, passando pela validade de conteúdo, constructo e critérios, assim como para o *design* em sua aplicação e os procedimentos adotados.

Para compreender o padrão de dificuldade dos itens, assim como o padrão de distribuição das habilidades estimadas das pessoas foi realizada a análise do chamado *Wright Map* ou Mapa item-pessoa, disponível recentemente no *software* R (M. David & Irribarra, 2016). Esse tipo de gráfico possibilita verificar em um mesmo *continuum* do

traço latente a distribuição das dificuldades dos itens, assim como a distribuição das habilidades das pessoas. Diferentemente das figuras 83 e 84, na qual foram isoladas as bases de dados de cada contexto, e aplicado o modelo 3PL, nesse gráfico têm-se a base geral contendo informações dos dois contextos e utilizando os resultados das análises do modelo 4PL, na qual irá representar informações da escala de proficiências técnicas com evidências de validade para Brasil e Portugal.

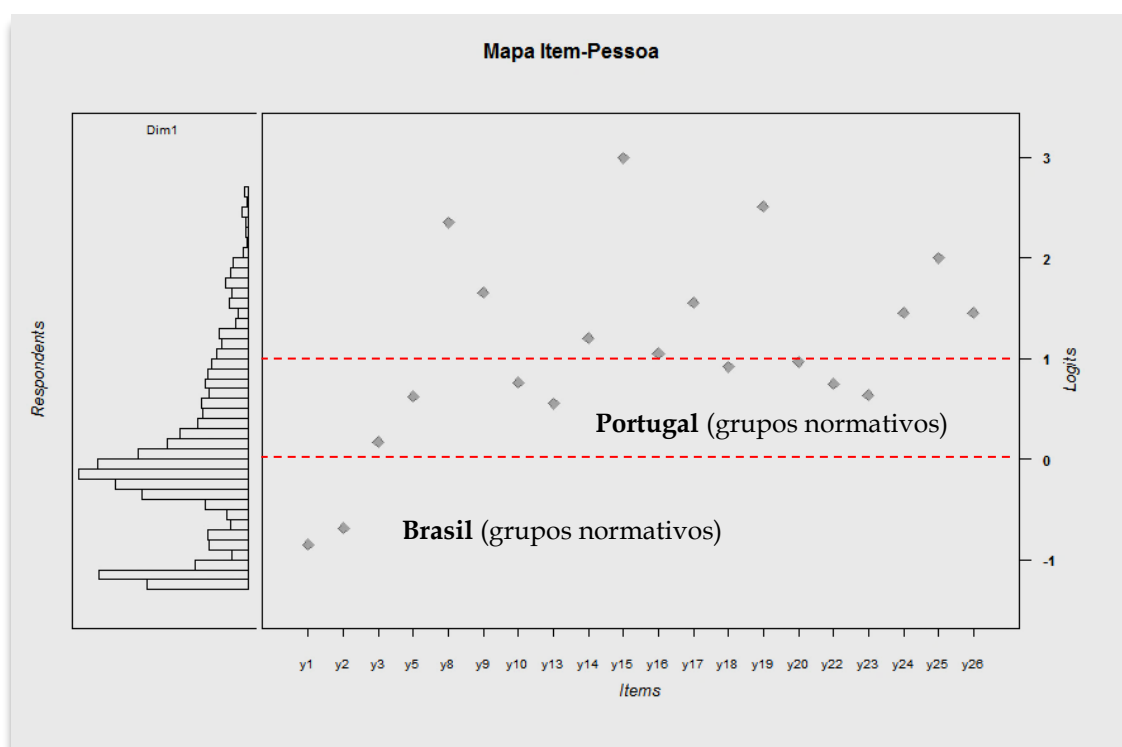


Figura 85: Mapa item-pessoa. Distribuição das habilidades das pessoas e das dificuldades dos itens pelo modelo 4PL.

A distribuição acima não corresponde ainda a escala de proficiências, para isso, é preciso seguir os procedimentos de acordo com os critérios para a identificação de *itens-âncoras*, na qual fornecem as informações mais consistentes e relevantes na interpretação da escala.

Na análise, alguns itens não obedeceram todos os critérios de consistência e com isso serão denominados *itens quase-âncoras* (ver no Apêndice E13, p. 276). Os itens-âncoras responsáveis pela construção da escala foram 10 itens (em ordem crescente de dificuldade): y1, y13, y23, y10, y16, y14, y26, y24, y9, y19. Os itens quase-âncora foram 3 itens (em ordem crescente de dificuldade): y3, y22, y8. Os outros 7 itens que não obedeceram no mínimo a dois dos três critérios de consistência não foram

considerados: y2, y5, y18, y20, y17, y25, y15. Além de poucos itens fáceis e âncoras, outro inconveniente foi que grande parte dos itens-âncoras encontram-se acima do nível 1, e nenhum grupo normativo entre os dois contextos conseguiram se posicionar nessa região da escala, estando o Brasil abaixo da média.

Através da figura 86 é possível observar que a distribuição das baixas habilidades não acompanha as dificuldades dos itens, indicando uma necessidade de itens mais fáceis para melhor avaliar estes participantes. Nota-se também o “gap” entre a dificuldade do item y2 e y3 provocando um aumento do erro de medida no intervalo aproximado entre -0,8 e 1,2 da escala de habilidade (Figura).

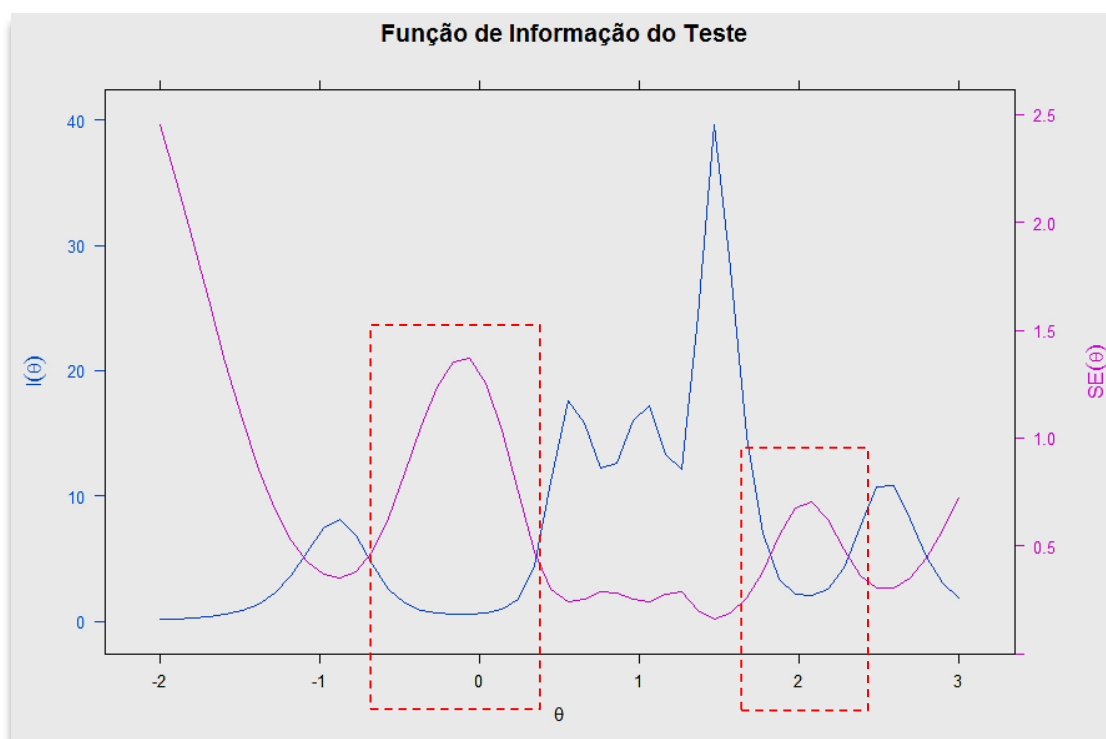


Figura 86: Função de Informação do Teste e erro padrão de medida ao longo da escala de proficiência.

Para uma interpretação direta das pontuações dos diferentes grupos normativos dentro da escala, se apresenta *lado a lado* a descrição das proficiências dos itens-âncoras no que se refere as capacidades cognitivas dos alunos, essas capacidades dizem respeito as habilidades que configuram a competências dentro do conhecimento técnico para todos os grupos dentro do intervalo ou nível de proficiências.

Os resultados mostram que de forma absoluta, os todos os grupos normativos do contexto brasileiro possuem em média habilidades entre os níveis de -1 a 0 (Média_BR: $\theta = -0,377$), na qual a escala informa que os respondentes possuem apenas um domínio diante do conhecimento relativo ao item y1. A curva de erro padrão de medida no entanto revela uma falta de consistência ou de informação para grande parte do intervalo da escala entre esses níveis, ou seja, **a escala de proficiências térmicas não possui precisão para interpretar e estimar as habilidades dos grupos normativos e sujeitos posicionados nesse intervalo.**

Na intenção de formular uma interpretação para esse nível no que diz respeito a uma competência científica do conhecimento térmico dentro dos níveis de desvio-padrão na qual a escala foi construída, inclusive na tentativa de comparar com as competências previamente formuladas, com base na interpretação qualitativa dos agrupamentos familiares de itens na matriz de habilidades, esse tipo de análise se torna ainda mais comprometido pela ausência de itens-âncoras.

Verificando os valores θ para o Brasil, observa-se que o turno noturno possuem uma posição mais baixo dentro da escala, em ordem crescente de grupos mais proficientes seria 3º, 2º e depois 1º ano. Essa mesma sequência serve para o turno vespertino, mas se inverte para o turno noturno, apesar disso, para os níveis de escolaridade de modo geral, segue em ordem crescente estando o 1º ano com nível de proficiente, seguido do 3º ano, estando o 2º ano com pontuação mais elevada. Apesar de todos os grupos relacionados as escolas e formados por livros didáticos de Física adotados estarem no mesmo nível de proficiência, também é possível verificar sutis diferenças nas pontuações dentro da escala, considerando que se tornaria mais relevante se esses grupos estivessem em níveis de proficiências diferentes. Ainda assim, torna-se uma informação complementar para as análises preliminares das concepções térmicas, visando identificar vantagens dos diferentes grupos normativos com base no percentual de respostas, que encontra-se na análise dos distratores.

No contexto português, a média geral ficou entre os níveis de 0 a 1 (Média_PT: $\theta = 0,431$). Considera-se que todos os grupos dentro de cada nível de proficiência possuam de forma relativamente consistente habilidades relacionadas as informações de todos os itens-âncoras que se encontram nesse nível, mas também dos níveis

anteriores. Em média, todos os grupos normativos portugueses se encontram nesse nível de proficiência, na qual corresponde aos itens: y3, y13, y23, y10, y22.

Observa-se que as proficiências dos alunos se apresentam em ordem crescente do 10º para o 12º ano, como era de se esperar, pois a medida que os alunos avançam em níveis de escolaridade, espera-se que se tornem mais proficientes de modo geral. Apesar disso, sabe-se que o teste conceitual é muito específico, e não possui caráter acumulativo. Esse é um diferencial que deve ser levado em consideração. Entretanto, como se refere a conceitos básicos e introdutórios, acaba por se caracterizar como recursos cognitivos permanentemente presentes à medida em que o conhecimento se torna cada vez mais inclusivo, diversificado e complexo. A falta de atenção ou de discutir com maior propriedade esses conceitos essenciais pode ser um problema, principalmente se o conhecimento avança sem estar claro para o estudante esses conceitos térmicos. Por isso, uma das funções principais na realização desse tipo de teste é identificar essas dificuldades e realizar um diagnóstico sobre as proficiências e modelos mentais incorretos, em especial, as concepções alternativas relacionadas, para assim serem revisadas e aprofundadas, caso necessário, e levadas em consideração na proposição e progressão das situações de aprendizagem. Acredita-se que a natureza epistemológica das dificuldades associadas a essas concepções possa sinalizar de forma mais coerente o campo epistemológico para os processos de ensino, seja do ponto de vista metodológico, incluindo a utilizando a escolha e formas de recursos físicos e/ou virtuais, quanto ao campo teórico-conceitual a ser abordado e desenvolvido. Entre os diferentes grupos, observa-se que as escolas com os livros didáticos de Física 11 e 9, obtiveram praticamente a mesma média, enquanto aquelas relacionadas ao livro 10 obteve a média mais baixa.

É importante ressaltar que para compreender a tabela 44 e realizar uma leitura adequada e clara, deve-se ter atenção que ela está constituída de informações pedagógicas referentes exclusivamente de itens-âncoras no sentido crescente das dificuldades, por esse motivo, os demais itens são posicionados no *continuum* da escala, mas sem suas respectivas informações pedagógicas, pois a escala busca *precisar o que de fato o aluno domina*, e isso somente é garantido pelos itens com maior consistência, na qual os parâmetros psicométricos estimam probabilidades que

atendem satisfatoriamente aos critérios estatísticos impostos para a classificação de item-âncora (ver p. 265, relacionado ao tópico “Fundamentos para a construção de uma escala de proficiências”). Assim, entende-se que o fato de um respondente (ou grupo normativo) se posicionar acima ou abaixo de um item não âncora, não garante seu domínio ou dificuldade propriamente dito devido ao erro padrão associado, conforme foi comentado anteriormente, revelando regiões de inconsistências na escala que precisam ser melhoradas.

Tabela 44: Escala de Proficiências Térmicas.

Níveis de Proficiência	Ítems-âncora (e quase-âncoras)	Competências Associadas (com referência aos seus descritores na análise taxonômica)	Descrição das Proficiências	Ítems e Grupos Normativos (Portugal)	Ítems e Grupos Normativos (Brasil)	b e Z (ou θ)	Indícios de dependência local do item
De -1 a 0	y1	Conhecer a Temperatura de um Corpo ou de um Ambiente em Situações Típicas - [Descritores 1 e 2]	⇒ Prever que a «temperatura do gelo» dentro do congelador de uma geladeira doméstica «após de um longo tempo está abaixo de 0 °C».	-	-	-0,84	-
	-	-	-	y2	y2	-0,68	y1
	-	-	-	-	15BR	-0,587	-
	-	-	-	-	26BR	-0,578	-
	-	-	-	-	1N	-0,560	-
	-	-	-	-	1BR	-0,553	-
	-	-	-	-	4BR	-0,544	-
	-	-	-	-	12BR	-0,532	-
	-	-	-	-	LDF_7br	-0,505	-
	-	-	-	-	LDF_10br	-0,503	-
	-	-	-	-	2N	-0,490	-
	-	-	-	-	14BR	-0,486	-
	-	-	-	-	24BR	-0,471	-
	-	-	-	-	3N	-0,467	-
	-	-	-	-	LDF_4br	-0,464	-
	-	-	-	-	19BR	-0,463	-
	-	-	-	-	2BR	-0,453	-
	-	-	-	-	18BR	-0,426	-
	-	-	-	-	7BR	-0,425	-
-	-	-	-	1V	-0,411	-	
-	-	-	-	1	-0,392	-	

-	-	-	-	5BR	-0,391	-	
-	-	-	-	3	-0,382	-	
-	-	-	-	Média_BR	-0,377	-	
-	-	-	-	2V	-0,376	-	
-	-	-	-	3M	-0,374	-	
-	-	-	-	2	-0,364	-	
-	-	-	-	3V	-0,353	-	
-	-	-	-	LDF_9br	-0,350	-	
-	-	-	-	2M	-0,335	-	
-	-	-	-	LDF_6br	-0,331	-	
-	-	-	-	1M	-0,330	-	
-	-	-	-	10BR	-0,310	-	
-	-	-	-	20BR	-0,307	-	
-	-	-	-	13BR	-0,299	-	
-	-	-	-	LDF_2br	-0,185	-	
-	-	-	-	27BR	-0,116	-	
-	-	-	-	25BR	-0,083	-	
-	-	-	-	22BR	-0,014	-	
-	-	-	-	E37	-	0,080	
-	-	-	-	10	-	0,101	
-	-	-	-	E38	-	0,118	
-	-	-	-	E18	-	0,142	
De 0 a 1	y3	Recordar sobre a Invariância da Temperatura em um Processo de Mudança de Fase - [Descritor 5]	⇒ Prever que a «temperatura do gelo em equilíbrio térmico» «com a água que se forma com seu derretimento» «permanece a 0 °C, mesmo após cessar o processo de fusão».	-	-	0,19	y1; y2
-	-	-	-	E47	-	0,246	-
-	-	-	-	E40	-	0,265	-

-	-	-	E42	-	0,270	-
-	-	-	LDF_10	-	0,270	-
-	-	-	E1	-	0,272	-
-	-	-	LDF_8	-	0,294	-
-	-	-	E50	-	0,308	-
-	-	-	E41	-	0,380	-
-	-	-	E45	-	0,390	-
-	-	-	LDF_5	-	0,390	-
-	-	-	Média_PT	-	0,431	-
-	-	-	LDF_11	-	0,438	-
-	-	-	LDF_9	-	0,441	-
-	-	-	E100	-	0,459	-
-	-	-	E101	-	0,513	-
-	-	-	E43	-	0,530	-
-	-	-	E54	-	0,548	-
-	-	-	E10	-	0,552	-
y13	Compreensão sobre a Propagação do Calor - [Descritor 9]	⇒ Inferir sobre o sentido de «transferência de energia térmica que ocorre durante o processo de resfriamento de ovos cozidos quando submetidos em água fria», nessa situação que ocorre uma «transferência de energia térmica dos ovos para a água, ou seja, do corpo mais quente (maior temperatura) para o mais frio (menor temperatura)».	-	-	0,56	-
-	-	-	E36	-	0,571	-
-	-	-	y5	y5	0,61	-
y23	Compreensão do Comportamento Térmico de acordo com as Propriedades Térmicas - [Descritor 19]	⇒ Inferir que a «sensação térmica de frio da nossa pele pode ser evitada ao dificultar o processo de trocas de calor com o meio externo, e assim reter a energia térmica que se emite continuamente pelo corpo utilizando um agasalho como isolante térmico», dessa forma «o agasalho serve para reduzir a perda de calor com o ambiente».	-	-	0,62	-
-	-	-	E39	-	0,643	-
-	-	-	11	-	0,674	-

	y10	Compreensão sobre a Sensação Térmica explicada pela Propagação do Calor - [Descritor 13]	⇒ Inferir sobre o sentido de «transferência de energia térmica que ocorre durante o processo de trocas de calor entre diferentes corpos sólidos com base na sensação térmica», no caso, «a lata de alumínio sobre a mesa de apoio retira um pouco de energia térmica por meio de calor na região da mesa em que houve contato físico, diferenciando a sensação térmica de frio entre essa região da mesa e o restante. De modo geral, se refere a uma compreensão de que o calor transita no sentido do corpo mais quente para o mais frio, ou seja, parte da energia térmica foi transferida da mesa para a lata de refrigerante».	-	-	0,75	y9
	y22	Compreensão sobre a Sensação Térmica explicada pela Propagação do Calor - [Descritor 16]	⇒ Inferir que a «sensação térmica de “quente”» em uma «bomba durante enchimento de um pneu é causada pelo recebimento de energia térmica que advém do interior da bomba».	-	-	0,76	-
	-	-	-	E7	-	0,800	-
	-	-	-	E46	-	0,900	-
	-	-	-	y18	y18	0,91	-
	-	-	-	y20	y20	0,96	-
	-	-	-	12	-	0,988	-
De 1 a 2	y16	Compreensão sobre a Propagação do Calor - [Descritor 11]	⇒ Inferir sobre a «sensação térmica entre corpos com diferentes propriedades térmicas estando na mesma temperatura», no caso, «entrando em contato físico com uma régua de metal e uma de madeira, considera-se que a temperatura independe da sensação térmica»	-	-	1,05	y9; y10; y14
	y14	Compreensão do Comportamento Térmico de acordo com as Propriedades Térmicas - [Descritor 17]	⇒ Inferir uma «igualdade de temperatura entre corpos com diferentes propriedades com base na Lei Zero da Termodinâmica», «no caso, entrando em contato físico com uma cadeira de metal e uma de plástico, considerando uma independência da sensação térmica, portanto, as cadeiras estão à mesma temperatura e não devem ser avaliadas com base na sensação térmica».	-	-	1,20	y9; y10
	y26	Compreensão do Comportamento Térmico de acordo com as Propriedades Térmicas - [Descritores 19, 20 e 21]	⇒ Inferir que «isolantes térmicos não aquecem corpos, apenas dificultam eventuais trocas de calor, sendo eficientes em reter a energia térmica caso o corpo coberto seja um emissor desse tipo de energia em potencial», dessa forma «o uso de cobertores em bonecas não irá aquece-las, pois geralmente elas já se encontram em equilíbrio térmico com o ambiente».	-	-	1,43	y23

	y24	Compreensão do Comportamento Térmico de acordo com as Propriedades Térmicas - [Descritor 17]	⇒ Inferir uma « igualdade de temperatura entre corpos com diferentes propriedades com base na Lei Zero da Termodinâmica », «no caso, entrando em contato físico [sensação térmica] com os palitos de madeira do picolé e o sumo gelado».	-	-	1,45	y9; y10; y14*; y16
	-	-	-	y17	y17	1,55	-
	y9	Compreensão do Comportamento Térmico de acordo com as Propriedades Térmicas - [Descritor 17]	⇒ Inferir uma « igualdade de temperatura entre corpos com diferentes propriedades térmicas com base na Lei Zero da Termodinâmica », «no caso, entrando em contato físico (sensação térmica) com o alumínio (lata), medindo a temperatura (termômetro) do líquido em seu interior (refrigerante) sendo 7 °C, e com isso concluindo sobre a temperatura de um recipiente plástico (garrafa) e do respectivo líquido em seu interior (refrigerante), na qual estavam sob as mesmas condições climáticas ambientais da vizinhança (interior da geladeira), portanto, também igual a 7 °C».	-	-	1,65	-
	-	-	-	y25	y25	1,98	y1
De 2 a 3	y8	<ul style="list-style-type: none"> • Conhecer a Temperatura de um Corpo ou de um Ambiente em Situações Típicas - [Descritor 3]; • Aplicação de Processos de Fervura de acordo com a Variação da Pressão - [Descritor 6] 	⇒ Aplicar um processo de aquecimento considerando que o ponto de ebulição da água varia conforme a variação da pressão, com isso, poder inferir que a « temperatura da água fervente » «varia para mais ou menos do valor de 100 °C» « dependendo da altitude em relação ao nível do mar pelo fato de variar a pressão atmosférica local. No caso de estar numa montanha alta, o ponto de ebulição fica mais baixo, com isso, torna-se mais fácil alcançar o processo de fervura, pois se atinge mais rapidamente o ponto de ebulição».	-	-	2,37	y4
	y19	Aplicação de Processos de Fervura de acordo com a Variação da Pressão - [Descritor 7]	⇒ Inferir que a «temperatura da água fervente» «varia para mais ou menos do valor de 100 °C» «dependendo da variação da pressão em que fica submetida. No caso, estando em um recipiente fechado (panela de pressão) o aquecimento aumenta o movimento e a força de repulsão entre as partículas, e isso resulta em uma maior incidência do choque entre elas, ou seja, a pressão interna fica maior que a pressão atmosférica externa e o ponto de ebulição da água contida na sopa fica mais baixo tornando mais fácil alcançar o processo de fervura e de cozinhamento».	-	-	2,52	y4; y8
	-	-	-	y15	y15	3,00	y9; y10; y14

Legenda: os itens estão representados por: y1, y2, y3...; os grupos normativos referentes as escolas brasileiras por: 1BR, 2BR,...; os grupos normativos brasileiros organizados por turnos, sendo: 1M (1º ano matutino), 2M (2º ano matutino), 3M (3º ano matutino), 1V (1º ano vespertino), ...; os grupos normativos aglomerados por nível de escolaridade (levando em consideração todas as escolas): 1 (1º ano), 2 (2º ano) e 3 (3º ano); os grupos normativos de respondentes organizados por Livro de Física adotado no contexto brasileiro (levando em consideração todas as escolas), como: LDF_2br; LDF_4br, Em Portugal segue o mesmo padrão de leitura e interpretação para a tabela.

4.2.9 Condições de contorno para uma aprendizagem significativa de acordo com as evidências

A análise taxonômica sugere que o *traço latente* se refere a um nível de compreensão conceitual com base em um *conhecimento prévio* dentro do domínio cognitivo (se distingue do psicomotor e do socioafetivo), sendo no caso necessário para responder de acordo com o conhecimento científico. Esse conhecimento pode ser interpretado na concepção ausubeliana como um subsunçor a ser evidenciado.

Na intenção de explorar algumas implicações da psicologia de Ausubel, no que diz respeito às implicações de ensino diante das interpretações psicométricas obtidas, será explorado apenas um item como exemplo, supondo a possibilidade do leitor extrapolar para uma interpretação semelhante ou mais apropriada diante para os outros itens.

No item y1, por exemplo¹⁰², entende-se o subsunçor corresponda a um conhecimento intuitivo de que **a temperatura do gelo dentro do congelador está supostamente abaixo de 0 °C**. Entende-se que essa **habilidade cognitiva** esteja associada a uma **competência científica** voltada para *conhecer a temperatura de um corpo ou de um ambiente em situações típicas e diferenciadas*, de acordo com os descritores 1 e 2, que por sua vez, acredita-se que possa influenciar a resposta do item y2 e y25, apesar de que essa dependência não seja considerada no processo de estimação dos parâmetros da TRI. Esses descritores dizem que é necessário:

- Lembrar o valor da temperatura de fusão do gelo em situações mais comuns;
- Lembrar que em um processo de resfriamento a temperatura dos corpos pode ficar abaixo de 0 °C até um limite de -273,15 °C.

Entende-se que esses descritores se caracterizem como traços latentes subjacentes relacionados ao subsunçor principal, além de sinalizar a necessidade de

¹⁰² O critério para a escolha desse item não se baseou no parâmetro de dificuldade, pois a intenção é de ilustrar uma das formas em que se acredita ser possível extrair inferências pedagógicas para fundamentar diagnósticos instrucionais para o ensino. Entende-se que esse mesmo procedimento deva ser replicado para os outros itens, caso haja interesse e/ou necessidades.

outros conhecimentos associados, que acabam envolvendo também novas concepções associadas ao contexto interpretativo, além das concepções alternativas que são ressaltadas entre os distratores, em especial aquelas que representam o modelo mental incorreto em destaque, caso exista. Evidentemente, é possível ainda ampliar os aspectos que caracterizem de forma pertinente a mesma família de situações-problemas, tornando a competência ainda mais rica, porém, a análise feita se restringe aos itens apresentados no teste TCE, e isso se dá não apenas explorando os descritores, mas também identificando itens que se complementam na realização ou compreensão de uma determinada tarefa.

Os resultados mostram que o subsunçor (presente na letra A) foi mais evidenciado entre alunos entre os alunos portugueses. Entretanto, para os brasileiros predominam dois modelos mentais, de acordo com o *Fator de Concentração de Análise*, sendo a letra B, o modelo incorreto, na qual acredita-se que **o gelo encontra-se sempre a 0°C e/ou não é possível alterar a temperatura** (equívoco [D6], no inventário de concepções), sabe-se que modelo mental incorreto é predominante para esses respondentes, encontra-se entre os distratores. Tendo acesso as evidências e análises empíricas, é possível o professor tomar duas decisões, de modo geral:

1. Identificar em média quais os modelos incorretos predominantes para cada grupo, e buscar instruções gerais no ensino específicas para **todos os indivíduos** de cada grupo, ou seja, que seja pertinente de modo geral;
2. Identificar **os alunos específicos** que possuam uma habilidade ou probabilidade de acerto muito baixa, e buscar instruções diferenciadas para os sujeitos diante dos modelos incorretos, ou seja, explorar atividades que possibilitem os alunos superarem suas dificuldades, e os alunos mais proficientes explorarem atividades em níveis mais avançados e diferenciados.

As análises mostram um coeficiente bisserial igual a -0,15 para o distrator, e isso sugere que tanto alunos bons e ruins assinalaram essa alternativa, portanto, é recomendável que professor desenvolva instruções para o grupo todo. É importante sempre, que durante as aulas, sejam exploradas diversas formas de abordagens que possibilitem corroborar ou não as evidências no teste, como no caso do favorecimento

da argumentação científica, estimulação a formulação de questionamentos, intercâmbio de ideias entre diferentes alunos, etc. Na intenção de explorar possibilidades que sejam úteis e pertinentes na análise de outros itens, decidiu-se explorar diversas situações didáticas de sala de aula.

Assim, pode-se dizer que um perfil de proficiências informa ao professor uma média relacionado as atividades e conteúdos envolvidos, na qual poderá ser conveniente ao professor tomar uma decisão em **avançar diante de novos conteúdos** (os alunos em média dominam o conteúdo) ou **focar nas dificuldades** (os alunos em média não dominam o conteúdo), caso seja pertinente, porém, demonstra ser mais prático e conveniente a primeira situação, em que ele diferencia a instrução de ensino baseada na média ou no perfil de proficiência do grupo e não nas habilidades individuais especificamente, que por sua vez remeteria a um ensino com instrução diferenciada para cada sujeito, sendo algo pouco viável.

Considerando hipoteticamente que esse seja um subsunçor necessário a estar disponível na estrutura cognitiva de todos os alunos na sequência da aprendizagem, como por exemplo, em uma atividade experimental no laboratório, isso poderia remeter à necessidade de ressaltar uma **instrução diferenciada para esses sujeitos**, ou ainda buscar identificar alunos que possuam uma probabilidade maior de acerto nesse item, e assim formar grupo heterogêneos nas atividades de acordo com as habilidades estimadas. Como último recurso, o professor pode apresentar de forma mecânica ou por memorização os conceitos necessários, ou ainda dar um *feedback* aos alunos identificados com dificuldades, recomendando-os a reforçarem esses conceitos, lembrando que o instrumento não deva ser utilizado para elaborar notas.

Para a seleção dos respondentes o professor pode determinar a escola e o turno, por exemplo, considera-se a escola 10BR, na qual utiliza o livro didático 9br. Observando o resultado, tem-se uma grande parte de alunos pré-instruídos ou que ainda receberam as aulas de termodinâmica, que no caso estão no 1º ano do ensino médio. Os escores padronizados (θ) no *modelo geral* (Portugal e Brasil) tornam possíveis uma comparação direta na escala de proficiências, no multigrupo, além de uma comparação dentro do próprio grupo, possibilita uma comparação de probabilidades

de acertos para alunos com a mesma habilidade, enquanto o escore bruto fornece informações exclusivamente relacionada a quantidade de acertos. Obviamente que os modelos da TRI são mais robustos e apresentam informações mais consistentes para as interpretações, porém, deve-se lembrar a recomendação de que o coeficiente de consistência do teste esteja acima de 0,80, para garantir uma avaliação com maior precisão para os indivíduos.

Diante disso, torna-se interessante, dentro de um único contexto, avaliar os alunos que supostamente dominam ou não o conhecimento pela consistência dada pela probabilidade da evidência de existência do subsunçor nos modelos da TRI. O perfil de proficiências pode assim ser avaliado pela CCI para um item em específico, assim como de concepções específicas, ou pela CCT, quando o interesse é o domínio do traço latente predominante, formado pelo resultados dos subconstrutos inerentes.

Interessante saber que indivíduos com a mesmo escore bruto no modelo geral, e que erraram a mesma questão, possuem probabilidades distintas diante da evidência do subsunçor em questão, como é o caso do aluno 1 e o 21, na qual um está praticamente na média ($\theta_1 = 0,10$), e o outro está abaixo ($\theta_{21} = -1,01$), com um intervalo acima de 1 desvio-padrão, ou seja, pode-se dizer que estão em níveis diferentes na escala de proficiências, apesar de possuírem a mesma pontuação. Também pode-se observar que o aluno 14, apesar de ter errado o item y1, possui uma elevada probabilidade de acertar, isso pressupõe que o modelo tenha estimado que esse respondente tenha acertado de forma consistente itens mais difíceis, e talvez, por algum descuido ou falta de atenção tenha errado.

Isso ocorre provavelmente pelo fato do *aluno 21* ter acertado questões difíceis e errado aquelas que são fáceis, pois os modelos interpretam que isso tenha sido ocorrido devido ao acerto aleatório (ou “chute”).

Tabela 45: Exemplo de subamostra de alunos da escola 10BR no contexto brasileiro que erraram o item y1.

Alunos	Nível	Geral (θ)	Multigrupo Não-Restringido (θ)	Escore Bruto (X_i)
1	1	0,10	0,34	4
2	1	-1,18	-0,82	5
3	1	-0,48	0,15	5
4	1	-0,02	0,55	8
5	1	-0,86	-0,43	7
6	1	-1,12	-0,81	6
7	1	-0,01	0,72	8
8	1	-0,89	-0,61	7
9	1	-0,89	-0,61	7
10	1	-0,79	-0,14	9
11	1	0,16	0,72	10
12	1	-1,23	-0,9	6
13	1	-0,95	-0,32	5
14	1	1,31	1,96	12
15	1	0,82	1,08	11
16	1	-0,80	-0,53	10
17	1	-0,02	0,41	9
18	1	-0,54	0,41	8
19	1	-0,13	0,51	8
20	3	-1,15	-0,82	6
21	3	-1,01	-0,84	4
22	3	-1,25	-0,94	2
23	3	-1,18	-0,94	7
24	3	-0,82	-0,2	7
25	3	-0,95	-0,77	6
26	3	-0,73	0,01	7
27	3	-0,87	-0,15	7

O item y1 possui dificuldade estimada em $b = -0,84$, e a probabilidade de acerto para o aluno 1 é de $P(\theta_1) = 75\%$, enquanto do aluno 21 é de $P(\theta_{21}) = 19\%$, apesar de terem o mesmo número de acertos e estarem no grupo inferior de baixo desempenho (ABAIX) segundo a TCT, pode-se dizer que eles possuem na realidade proficiências diferentes, supostamente habilidades distintas ($\theta_1 > \theta_{21}$), sendo o aluno 1 ($\theta_1 = 0,10$) com mais domínio que o 21 ($\theta_{21} = -1,01$), estando o aluno 21 na zona de “guessing”.

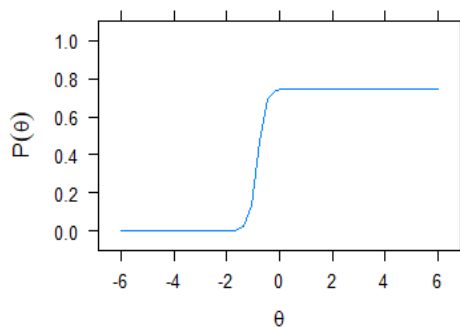


Figura 87: Curva Característica do Item y1 para o modelo geral com 20 itens, ou seja, estando inclusos os dois contextos e as respectivas informações na escala de proficiências técnicas.

NÍVEIS DA ESCALA DE PROFICIÊNCIA (θ)	-1,000	-0,84	0,000	1,000	2,000	3,000
PROPORÇÕES DE ACERTOS EMPÍRICOS (%)	0,193	0,375	0,747	0,750	0,750	0,750

No modelo não-restringido de DIF (Figura 87) é possível comparar a probabilidade de acerto de alunos com a mesma habilidade. Também é possível perceber que o *aluno 1* é mais proficiente que o *aluno 21* ($\theta_1 > \theta_{21}$) por ter maior probabilidade de acerto, sendo o *aluno 1* ($\theta_1 = 0,34$) com mais domínio que o *21* ($\theta_{21} = -0,84$), estando o *aluno 21* na zona de “guessing”. O modelo estimou uma baixa discriminação para o contexto português, ao contrário do Brasil. No tracejado pode-se comparar alunos nos dois contextos com a mesma habilidade, na qual possuem probabilidade de acertos distintos. A sigmoide no contexto brasileiro indica uma melhor discriminação para comparar alunos proficientes e não, enquanto para os portugueses não sendo muito eficiente de fazer essa distinção.

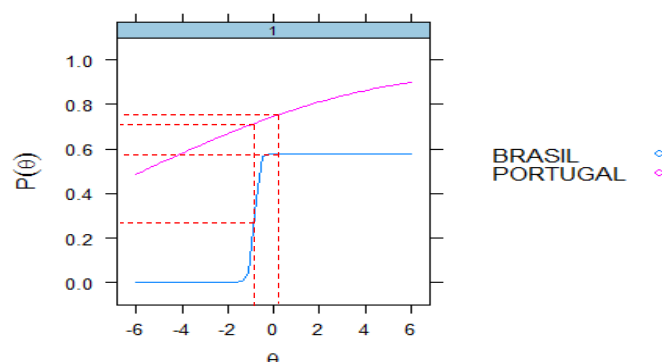


Figura 88: Curva Característica do Item y1 para cada contexto, de acordo com o modelo não-restringido.

Mas porque o item é difícil para alguns e fáceis para outros? Nesse sentido, os resultados apontam para os modelos mentais predominantes, tanto correto como os

incorretos, e sua intensidade depende da força de atração, na qual a AGI sugere bem, e os modelos da TRI discriminam de forma eficiente esses diferentes grupos de alunos. Além dos valores dos índices e parâmetros de dificuldade, é imprescindível uma análise qualitativa e interpretativa do conteúdo do item, buscando identificar conhecimentos potencialmente significativos necessários, além da natureza epistemológica e ontológica das concepções térmicas que caracterizam os modelos metais incorretos. Para um aprofundamento na interpretação sobre perfis conceituais com o foco mais no conteúdo, e não nos desempenhos, pode ser obtido nos trabalhos na literatura (Amaral & Mortimer, 2001; Araújo, 2014; Hitt & Townsend, 2015; Mortimer, 1995; Mortimer, n.d.-a, n.d.-b).

Uma noção sobre as condições térmicas do ambiente de um congelador, assim como uma compreensão mínima de seu funcionamento são questões iniciais a serem verificadas e exploradas. Uma noção de que no funcionamento de um congelador exista um termostato para controlar e manter a temperatura em determinado valor, geralmente a baixo de zero, esta pode ser uma informação ou *conteúdo potencialmente significativo* para ancorar a suposição de que o gelo pode alcançar um equilíbrio abaixo de 0 °C, obviamente o *subsunçor* ou ideia-âncora nesse caso seria o aluno ter um *conhecimento prévio* de que o gelo pode de fato ser resfriado a temperaturas abaixo de 0 °C (é o conhecimento incluso na sentença descritora do item que espera se associar a situação/contexto e alcançar uma compreensão diante do problema e assim tomar uma decisão correta), isso remete a compreender suas propriedades e comportamento, na qual o item y1 procura evidenciar.

Acredita-se na existência de possíveis *subsunçores* relacionados aos equívocos evidenciados nos distratores em que o gelo permanece igual ou acima de 0 °C, essas interpretações eventualmente se ancoram em outros conjuntos de concepções que podem conter, atrair e ancorar novos erros, como por exemplo acreditar que o gelo seja um corpo que possua a temperatura mais baixa possível, talvez pelo valor atribuído na escala Celsius que representa 0 °C. O zero é um valor representado na escala ou plano cartesiano como um valor mínimo para muitas grandezas escalares dentro da Física (área, massa, densidade, tempo, etc.), isso pode influenciar em uma confusão

para o aprendiz levando a acreditar que não exista temperaturas negativas. Uma outra suposição seria o aluno admitir que determinados corpos possuam uma temperatura própria e fixa, no caso da situação-problema a temperatura inalterável para o gelo seria 0 °C. Estas suposições por sua vez podem estar relativamente estáveis e organizadas em suas estruturas cognitivas, exercendo uma função lógica e racional para que o aluno defenda essa opinião, geralmente baseada em situações vivenciadas e transferidas de forma analógica para outros contextos levando consigo esses equívocos. Pode-se dizer que esses conhecimentos são potencialmente “significativos” para ancorar subsunçores preexistentes estruturados em equívocos, dessa forma, um ensino que não esteja atento para diagnosticar essas questões corre o risco de favorecer a construção de um conhecimento falso, em outras palavras seria uma ilusão compreensiva do aprendiz, podendo ficar cada vez mais associado e resistente a mudanças.

Entre sugestões para uma organização hierárquica dos conteúdos e uma abordagem estratégica inicial seria elaborar *organizadores prévios* para explorarem as temperaturas possíveis em que o gelo pode adquirir, sejam em situações controladas em laboratório ou em condições naturais na Terra, assim como aspectos que podem influenciar o seu ponto de solidificação parecem ser situações enriquecedoras e apropriadas, sendo importante ressaltar a recomendação de sempre valorizar e partir de diagnósticos com base na intuição dos alunos. Nessa perspectiva, o conjunto de atividades subsequentes pode buscar apresentá-lo sob diversas condições, esforço esse que corresponderia em identificar uma *família de situações-problemas* que colocasse o gelo em diversos estados térmicos, assim também favorecer o processo de *diferenciação progressiva* ao partir de ideias das mais gerais e inclusivas, depois seguindo para diversas aplicações com incorporação de detalhes, conceitos e conhecimento sobre processos mais elementares.

Entre algumas sugestões temáticas diversificadas a serem desenvolvidas têm o gelo em equilíbrio térmico com outros estados físicos, como no caso do ponto tríplice da água, dentro disso se tornará relevante incorporar e associar o conceito de pressão e sua relação com possíveis temperaturas para o gelo. Uma situação complementar

seria o fenômeno do regelo, descoberto pelo inglês John Tyndall, na qual o ponto de fusão cai à medida em que a pressão aumenta. Explorar como ocorre o processo de solidificação da água seria um aspecto importante para compreender como se forma o gelo, buscando explicar a existência dos cristalinos ou “espaços vazios” juntamente com o seu *comportamento anômalo* (que ocorre entre 4 °C e 0 °C), situação em que se inverte a lógica do aumento de densidade com a diminuição da temperatura, e isso pode ser explorado apresentando (por meio de vídeos ou fotos) situações físicas reais. Um exemplo disso seria mostrar como se forma a superfície de gelo em um lago, assim como das mudanças na sua estrutura molecular, entre outras. Diferenciar o gelo formado em frigoríficos dos flocos de neves poderiam ser conteúdos complementares a serem explorados, em especial quanto a sua estrutura, das condições climáticas e temperaturas em que são submetidos.

A elaboração de um painel sobre as diversas situações do comportamento do gelo poderia ser realizada na função de *organizador posterior*, favorecendo o processo de *reconciliação integrativa* ao possibilitar a realização de tarefas que enfatize “subir e descer” estruturas hierárquicas conceituais na qual fazem parte da explicação de cada situação, ou seja, seguindo lógicas dedutivas e indutivas, indo de conceitos específicos para os mais estruturados e vice-versa, destacando semelhanças, diferenças e discrepâncias. Nisso incluiria os conceitos necessários para estruturar uma argumentação de que o gelo pode ficar abaixo de 0 °C, e numa sequenciação passaria pela compreensão dos processos térmicos da forma como isso ocorre, destacando o *Princípio de Trocas de Calor* até alcançar uma compreensão sobre a *Lei Zero da Termodinâmica*, na qual fundamenta o *equilíbrio térmico* para o gelo ao ser estimado a temperatura dentro das condições climáticas no interior do congelador.

4.3 UMA SÍNTESE DOS RESULTADOS

A síntese visa apresentar breves comentários dos resultados em uma perspectiva comparativa, quanto: (1) aos conteúdos térmicos em destaque em cada contexto; (2) resultados parciais e mais representativos do desempenho, destacando os itens que projetam os perfis de proficiência e de concepções térmicas dos grupos normativos; (3) e uma visão geral de aspectos comuns e diferenciados das concepções de caracteriza o perfil de cada contexto, tendo como base os resultados de desempenho nas diferentes perspectivas de análise (TCT e TRI).

O esquema (Figura 89) destaca de forma ampla e simples os aspectos essenciais adotados no âmbito da investigação comparativa realizada, levando em consideração os diferentes posicionamentos na perspectiva do alinhamento curricular de cada contexto (ver Apêndice D, p. 241), no que se refere aos mais enfatizados conteúdos, das habilidades e competências recomendadas a serem desenvolvidas, e das relações com teste conceitual analisado. Quanto à avaliação de desempenho, têm-se a perspectiva de inferir sobre os grupos e indivíduos, seja por norma e/ou critérios referenciados, e como foi visto, estando inclusas diversas técnicas e procedimentos de análise psicométrico pertinentes a TCT e a TRI.

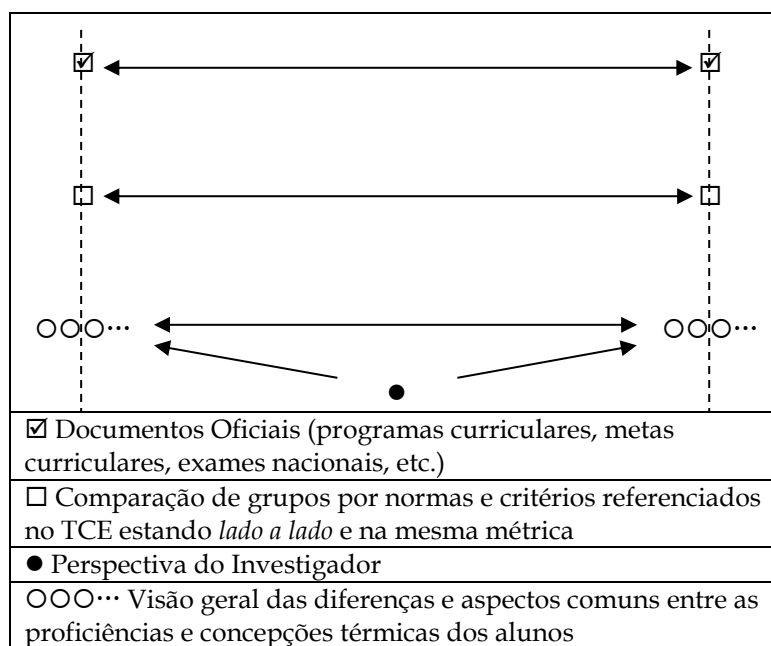


Figura 89: Esquema ilustrativo das principais perspectivas comparativas na investigação.

A perspectiva comparativa quanto as relações entre os paradigmas curriculares em ambos os contextos não possui a pretensão descabida de apontar qual seja o melhor ou pior, pois não faria sentido, seria semelhante a comparar os desempenhos dos estudantes na TCT sem levar em conta que estão em escalas diferentes. Para os resultados na comparação de grupos, serão mantidas a intenção de destacar os itens que se destacam dentre os critérios de validade e consistência, em especial quanto as informações sobre as dificuldades envolvidas e diferenças entre os contextos. Isso irá contribuir na intenção de apresentar uma visão geral e comparativa das diferenças e aspectos comuns nas diferentes perspectivas sobre o perfil de proficiências e concepções térmicas associadas a cada contexto.

Uma verificação de correspondência na relação do conteúdo no TCE e os objetivos dos programas em cada contexto é uma forma de analisar a capacidade de contribuir em um *alinhamento curricular*, e isso diz respeito a validade de construto diante os objetivos gerais dos sistemas educacionais. Diante disso, pode-se dizer que ambos contemplam em grande o conteúdo no teste TCE, como era de se esperar, pois os conceitos básicos e introdutórios devem ser contemplados universalmente na Física, por serem estruturadores, além disso, como se sabe, a temperatura termodinâmica é uma das sete grandezas fundamentais dentro do SIU. As habilidades e competências dentro do conhecimento térmico que se devem desenvolver em Portugal são destacadas pelas atividades prático-laboratoriais obrigatórias, na qual focam em situações-problemas específicos e direcionados a serem explorados. Constatou-se que entre as temáticas mais incidentes nos últimos 10 anos nos exames nacionais para a Física Térmica foi a “*absorção/emissão de radiação*” (45%) e “*rendimento no aquecimento*” (22%), entretanto, é necessário uma verificação do rendimento dos alunos diante esses conteúdos, e da eficiência do modelos de ensino, levando em consideração também as temáticas menos incidentes, servindo de parâmetro na escolha de atividades e como referência na padronização de competências mínimas a serem alcançadas, de acordo com a sugestão e objetivos dos programas. As descrições e orientações metodológicas dessas atividades são reproduzidas em praticamente todos os “*manuals escolares*”, e

há evidências de relatos sobre dificuldades em efetivamente de serem realizadas essas atividades pelos docentes, incluindo a existência de escolas sem os recursos.

No Brasil, os conteúdos no TCE estão em grande parte contempladas nas recomendações dos PCNs, além de possibilitar uma verificação de habilidades e competências que envolvem conceitos fundamentais sugeridos em muitos programas curriculares estaduais, incluindo relação próxima com disciplinas como a Química, que também reivindica uma compreensão básica do conhecimento térmico, como por exemplo, para o Modelo Cinético-Molecular dentro da físico-química e conteúdos termoquímicos. Também foi vista que a última versão da BNCC traz diversas habilidades formalmente descritas e recomendadas a serem desenvolvidas em diversas disciplinas e níveis que antecedem o ensino médio, isso revela uma preocupação pedagógica diante o processo de formação e aquisição básica de conceitos introdutórios da Física Térmica sejam desenvolvidos nos estudantes, tornando o TCE um instrumento diagnóstico importante a ser trabalhado no início do ensino médio, além um campo promissor de investigação e pesquisa.

Esse tipo de instrumento possui obviamente suas limitações, tanto na abrangência do conteúdo quanto na sua função exercida, portanto, entende-se que seja importante o aprimoramento desse e de outros instrumentos semelhante e mais eficientes, bem como o avanço nas mais diversas formas inovadoras de ensino, com designs didáticos diferenciados e eficientes, buscando apoio em recursos das tecnologias da informação e comunicação, como de atividades experimentais investigativas, e aprendizagens colaborativas e mais ativas para os alunos, preocupando-se em contribuir em macrocompetências, como o autodesenvolvimento e o pensamento crítico em questões que relacionem a ciência com problemas sociais, éticos, ambientais e tecnológicos.

Torna-se relevante destacar que o uso dos testes possibilita favorecer o processo de alinhamento curricular na prática docente, atingindo os objetivos de ensino estabelecidos pelo professor, mas sem perder de vista os objetivos mínimos e essenciais que programa curricular apresenta. Obviamente que para isso o professor

deva ter consciência do papel que esse tipo de instrumento deva oferecer, mas que entretanto, não foi o caso, pois o teste-reteste foi aplicado “às cegas”, sem o professor utilizar as informações advindas do teste. Apesar de se defender que o teste possa ser utilizado dentro dessa perspectiva, como também no auxílio de sequenciações de ensino ou atividades específicas mais pontuais, o *design* “às cegas” possui a vantagem do critério da neutralidade como credibilidade em apresentar um diagnóstico não apenas dos alunos pré e pós-instrução, mas também da qualidade ou eficiência do tipo de ensino que está sendo apresentado e desenvolvido com os estudantes.

Quanto aos resultados obtidos, interpreta-se de modo geral que os *itens mais fáceis* permitem discriminar bem entre os sujeitos com posição mais desfavorecida e *itens mais difíceis* aquelas mais capacitadas, com isso, **o teste sendo difícil possui uma forte tendência em destacar uma discriminação em um dos extremos da distribuição, no caso, aquelas mais capacitadas.** E como se constatou, o teste foi considerado difícil para ambos os contextos. Os resultados da TCT e TRI seguem perspectivas de análise diferenciadas na escolha de itens mais fiáveis, válidos e consistentes, porém sabe-se que os resultados devam ser complementares e não controversos, ou que anulem um ao outro. Na TCT destacam-se os ganhos percentuais, a identificação dos modelos mentais (corretos e incorretos) em destaque, sendo útil para observar e atender aos aspectos de legibilidade na investigação, ou seja, observando os diferentes níveis de escolaridade e aspectos peculiares da amostra, controlando vieses e delimitando com prudência as interpretações da amostra geral. Esses índices servem de referência na comparação e consistência diante as análises da TRI, auxiliando de forma mais adequada na interpretação dos resultados, entretanto, serão destacados os resultados da TRI, na qual é mais pertinente diante dos objetivos gerais propostos.

A primeira fase das análises foi fundamental para estabelecer: uma estratégia investigativa; descrever preliminarmente as etapas; elaborar materiais para a coleta de informações; formas de abordagens e tratamento dos dados; definir e testar diferentes tipos de procedimentos de análise; redesenhar um *design* apropriado diante aos objetivos propostos, e realizar um *check-list* para avaliar os riscos de validade; traduzir com 3 juízes; compreender melhor o instrumento realizando uma avaliação da chave-

de-correção com 5 juízes especialistas no EF; analisar a coerência interna e validade de construto; identificar fraquezas, correspondência e dependência entre itens; melhorar a precisão ampliando e detalhando o quadro de tipicidade entre as concepções, os itens e as alternativas; e por fim, realizar as análises comparativas de médias, ganhos normatizados e identificar modelos mentais através do fator de concentração nas análises entre o pré-teste e o pós-teste. A análise da TRI completou as análises, indicando aspectos complementares de desempenho.

O teste-piloto foi aplicado no contexto português em 4 escolas secundárias para aproximadamente 200 alunos, e revelou ganhos de percentual normalizados de baixo impacto ($\langle g \rangle < 0,30$), com um ganhos de $0,14 \pm 0,11$. Partindo para o âmbito geral da pesquisa, no contexto português ($n = 2585$), foi constatado ganhos crescentes entre os níveis de escolaridade ($\langle g \rangle$ 11º ano igual a $0,14 \pm 0,36$, e $\langle g \rangle$ 12º ano igual a $0,24 \pm 0,40$) comparativamente com 10º ano, com a identificação de alguns itens de médio impacto, porém, em proporções de modo geral equivalente ao teste-piloto, ou seja, de **baixo impacto**. Entre os brasileiros praticamente **não houve ganhos** entre os alunos do 2º e 3º ano. Esse tipo de análise, é por vezes comparado entre vantagens e desvantagens diante os modelos psicométricos da TRI, seja pela sua praticidade e utilidade aos professores da educação básica (Hake, 2008), porém, considera-se que sejam complementares.

Também foi realizado uma tentativa de identificação de modelos mentais para ambos os contextos através do fator de concentração de análise de Lei Bao e Redish. O resultado foi que os respondentes brasileiros se concentraram na região BB (dificuldade Baixa ou menor que 40% e concentração Baixa ou menor que 20%) de acerto aleatório, isso significa que os alunos mantiveram os equívocos ao longo das séries para quase todos os itens, porém foi constatado os itens y1, y7 e y12 na região MM (dificuldade Média e concentração Média) de dois modelos (correto e incorreto), muito próximos a região aleatória, revelando uma baixa consistência dos itens. Na região MM indica que alguns alunos obtiveram êxito quando sua resposta foi associada ao modelo correto, mas pelo fato dos três níveis de escolaridade, independentemente da existência de instrução, e mesmo sendo diferentes alunos,

mantiveram os mesmos itens nessa região, isso mostra uma ineficiência do modelo pedagógico exercido para as aulas que envolvem o conhecimento térmico.

O *Fator de Concentração* no contexto português apontou ganhos representativos na curva *S-plot*, inicialmente com os alunos pré-instrução do:

- 10º ano com 1 item (y1) na região AA, e 4 itens (y2, y7, y12, y23) na região MM;
- 11º ano com 1 item (y1) na região AA, e 11 itens (y2, y3, y5, y7, y10, y12, y13, y14, y22, y23, y26) – com progressão de 7 itens com relação ao 10º ano;
- 12º ano com 3 itens (y1, y13, y23) na região AA, e 12 itens (y2, y3, y5, y7, y10, y12, y14, y16, y18, y20, y22, y26) – com a progressão dos itens y13 e y23 para a região AA, mais a inclusão 3 novos itens (y16, y18, y20) na região MM.

Os resultados sugerem uma progressão de proficiências ao longo dos níveis de escolaridade, porém os ganhos não chegaram a ser de médio impacto, conforme mostram os valores para o *Fator de Hake*. Além disso, a permanência de itens em determinadas regiões do gráfico indicam a resistência de modelos incorretos relacionados a esses itens, apesar das aulas terem surtido de modo geral um efeito positivo.

As análise geral dos distratores focou no percentual de determinadas concepções específicas para os respectivos subconstrutos no inventário, e destacou como critério o predomínio nas incidência de concepções alternativas por alunos com desempenhos mais elevados (grupo ACIM). Os resultados convergiram para as concepções relacionadas aos *processos termodinâmicos* e sobre as *propriedades térmicas*. Esse diagnóstico deve ser valorizado e utilizado na elaboração de processos de ensino que visem ajudar os alunos a superarem essas dificuldades, bem como melhorar instrumentos para essa finalidade.

A análise realizada identificou as concepções térmicas subjacentes dentro dos modelos incorretos (distratores), entende-se que essas concepções carregam uma causa relacionada a sua natureza epistemológica e ontológica, como a concepção animista e substancialista do calor, na qual se distingue entre “calor vivo” de “calor inerte”. Entre Portugal e Brasil, foram identificadas de forma relativamente consistente três

concepções térmicas predominantes (com mais incidências entre os percentuais de resposta): C6, A3 e D5.

- C6: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido)
- A3: O quente e frio são diferentes, em vez de extremidades opostas de um *continuum*.
- D5: O ponto de ebulição da água é de 100°C (apenas).

Esses resultados mostram dificuldades nos conceitos de **equilíbrio térmico**, em **diferenciar quente de frio** como sensações térmicas e admitir que a **água pode ferver em temperatura distinta de 100 °C**, como por exemplo, em regiões montanhosas ou no interior de uma panela de pressão. As instruções nesse sentido, poderiam reforçar diante a esses aspectos.

Apesar desses equívocos terem sido coincidentes, elas foram **mais evidenciadas entre os brasileiros**, assim como a D2, A1, B3 e C1, comparadas aos portugueses, sendo a **A3 com 83%** e a **C1 com 80%**.

- D2: Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (*concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não uma energia transitória. O frio também uma substância*)
- A1: O calor é uma substância (*concepção substancialista*).
- B3: Percepções de quente e frio não estão relacionadas com a transferência de energia.
- C1: Aquecimento sempre resulta em um aumento na temperatura. (*não compreende o conceito de calor latente*)

Pode-se destacar ainda que em média os **alunos brasileiros** possuem de forma mais intensa uma **concepção substancialista** diante aos conceitos térmicos, com maior dificuldade em **diferenciar o conceito de quente e frio**, e em admitir que um corpo

pode receber calor sem alterar sua temperatura, ou seja, possuem dificuldade na compreensão sobre o **conceito de calor latente**.

Verificou-se uma ausência das concepções B2, C8, D3 e D10, para ambos os contextos, e algumas que estão ausentes exclusivamente para os alunos portugueses, que são, A6, C7. Entretanto, a análise de coerência interna do instrumento mostrou que essas concepções possuem apenas uma única incidência entre os itens, com exceção do C7, que possui 2 incidências, isso mostra que o **instrumento não avalia muito bem essas concepções**, e o simples fato de não terem sido detectadas no grupo ACIM, não é suficiente para afirmar com maior convicção sobre a sua influência.

Vale ressaltar que a análise de coerência interna no instrumento mostrou uma falta de homogeneidade na distribuição dos construtos, isso pode levar a tendências de resultados enviesados quando explorados em perspectivas globais e percentuais devido a influência da má distribuição das concepções entre os itens. Apesar disso, foram tomadas algumas decisões que pudessem minimizar esses vieses, portanto, considera-se minimamente satisfatório a identificação das concepções em destaque para cada contexto, no sentido de realizar uma interpretação sobre as concepções que habitam dentre os modelos incorretos nos distratores mais atrativos para o grupo ACIM.

As análises da TRI, resultaram na eliminação dos itens y4, y6, y7, y11, y12, y21, devido à falta de consistência nos índices, pois forneceram informações de baixa qualidade para uma interpretação sobre os processos mentais. O modelo 4PL foi o que melhor se ajustou para as análises, com isso, os 20 itens foram responsáveis pela explicação de 78,2% da variância dos padrões de resposta. Para ambos os grupos em uma mesma métrica, os itens considerados *muito difíceis* foram (em ordem crescente de dificuldades estimadas): y26, **y24**, y17, **y9**, y25, **y8**, y19, y15, sendo três deles sendo itens-âncoras (em negrito). Os *itens-âncoras* são os responsáveis pela construção da escala de proficiências e podem ser considerados os itens mais consistentes, ou com poder explicativo mais eficiente devido a qualidade no ajuste dos parâmetros psicométricos, que no caso foram 10 itens (em ordem crescente de dificuldade): **y1**,

y13, y23, y10, y16, y14, y26, y24, y9, y19. Os itens quase-âncora foram 3 itens (em ordem crescente de dificuldade): y3, y22, y8. Apesar de terem sido considerados a metade dos itens do teste na construção da escala, em média todos os grupos normativos ficaram em níveis compreendidos apenas por 6 itens: y1, y3, y13, y23, y10, y22. Desses, pode-se afirmar de forma mais consistente que os grupos normativos brasileiros, em média, dominam somente o conteúdo relacionado ao item y1, estando todos no nível de proficiência entre -1 e 0, e se associa a uma competência mínima relacionada a “conhecer a temperatura de um corpo ou de um ambiente em situações típicas”, sendo a proficiência relacionada ao item y1, sendo uma capacidade cognitiva de “prever que a «temperatura do gelo» dentro do congelador de uma geladeira doméstica «após de um longo tempo está abaixo de 0 °C»”. Essa é uma competência interpretada preliminarmente e encontra-se isolada entre os outros itens, entretanto, entende-se que na escala de proficiência cada nível deve ser interpretado pedagogicamente como uma competência, ou seja, é uma perspectiva epistemológica fundamentada na análise fatorial e não na interpretação qualitativa de habilidades relacionadas entre os itens do ponto de vista de conteúdos e soluções de problemas comuns. Apesar disso essas leituras não devem ser consideradas inválidas, mas sim sob um ponto de vista epistemológico distinto, que podem eventualmente serem complementares.

Para os grupos equivalentes no contexto português, na perspectiva fatorial entende-se que seu nível de proficiência seja acumulado entre os 6 primeiros itens em que os grupos estão inseridos, comportando o item y1 do nível entre -1 e 0 dentro da escala de proficiências e os 5 restantes no nível entre 0 e 1.

Além do item y1, os resultados sustentam afirmar que o perfil de proficiência térmica dos alunos portugueses, se refere em média a:

- *Prever* que a «temperatura do gelo em equilíbrio térmico», permanece a 0 °C, ou seja, recordam que **a temperatura não se altera em um processo de mudança de fase** (y3);
- *Inferir* sobre o sentido de «transferência de energia térmica que ocorre durante o processo de resfriamento, ou seja, compreendem que **a propagação do calor segue naturalmente da extremidade com a temperatura mais elevada ou corpo mais quente**

para o mais frio (y13), compreendendo corretamente ainda se for necessário ter que se basear na sensação térmica (y10);

- Inferir que o mecanismo e utilização de uma agasalho na função de isolante térmico, ou seja, compreendem **o comportamento térmico de corpos de acordo com as propriedades térmicas** que lhe são inerentes (y22 e y23);

Como foi visto, poucos são alunos que alcançam habilidades mais complexas dentro da escala de proficiências térmicas, no caso, os itens mais difíceis restantes (ordem crescente de dificuldade): y16, y14, y26, y24, y9, y8, y19. Esses portanto são itens restritos a alunos considerados *experts* dentro do teste TCE, sendo provavelmente alunos pertencentes do grupo ACIM, como foi dito, a escala é capaz de destacar esses indivíduos, e não inferir muito bem sobre aqueles com baixa proficiência..

Com relação as principais diferenças de compreensão entre os brasileiros e portugueses, a análise de DIF com referência ao parâmetro de dificuldade, pode se afirmar que eles existem: se diferenciam:

- Diferença moderada em relação ao item y24 a **favor do contexto brasileiro**;
- Diferença grande entre os itens y8, y9 e y19, a **favor do contexto brasileiro**;
- Diferença grande no item y13, a **favor do contexto português**.

Na interpretação dessa vantagem é preciso levar em consideração as dificuldades para os grupos. Os itens y8 ($\theta_8 = 2,37$), y9 ($\theta_9 = 1,65$), y19 ($\theta_{19} = 2,52$) e y24 ($\theta_{24} = 1,45$) se classificam como *muito difíceis* ($\theta_{24} < \theta_9 < \theta_8 < \theta_{19}$), estando posicionados na extremidade da escala de proficiência, enquanto o item y13 para os portugueses foi mais fácil, estando no nível entre 0 e 1, ou seja, dentro da média dos grupos normativos. Isso sugere que **entre os itens mais difíceis e consistentes dentro da escala de proficiências, os brasileiros levam vantagem comparativamente**, porém são representados por uma parcela pequena de alunos que alcançam esses níveis. Porém, **em média, os portugueses possuem uma proficiência representativamente mais elevada**, dominando de forma mais ampla os conhecimentos explorados entre os itens no teste.

CAPÍTULO 5 - DISCUSSÃO

5.1 Dialogando com Outras Pesquisas: Comparando Resultados e Apontando Limitações

O TCE se apresenta na literatura como um instrumento utilizado a diversos trabalhos que envolvem conceitos básicos da Física Térmica, sendo aplicado em estudos que exploram concepções relacionados à energia (Tastan & Yalçinkaya, 2008), em diversos contextos e cursos, como o da Química (Wren & Barbera, 2013), além de ser destacado pela facilidade de criação de atividades exploratórias práticas com base nos itens (Otto et al., 2009), entre outros. Um aspecto diferenciado a ser destacado seria a forma da linguagem nos itens, pois buscam reproduzir diálogos do cotidiano dos alunos, e isso projeta o teste em um nível de familiaridade com os respondentes.

Para servir de parâmetro de referência e comparação na discussão com os resultados obtidos em alguns dos estudos que utilizaram o TCE e serão comentados, tem-se algumas informações em destaque. Como foi afirmado, entre as concepções mais fortes para os **alunos brasileiros e portugueses** estão as dificuldades nos conceitos de *equilíbrio térmico* ([C6]: y1c, y2_{a,c,d}, y3a, y6_{c,d}, y9_{a,c,d}, y17_{b,c,d}, y24_{a,b}), em *diferenciar quente de frio* ([A3]: y10a, y13b, y15c, y18_{a,b}, y23a, y24_{b,c}) como sensações térmicas e admitir que a *água pode ferver em temperatura distinta de 100 °C* ([D5]: y4d, y8_{b,c}, y18_{b,c}).

No caso dos **brasileiros** foram mais intenso (em média) os equívocos relacionados às concepções *substancialistas do calor* ([D2] e [A1]: y10_{a,d}, y14d, y16_{c,d}, y20b, y21_{a,b,c}, y22_{c,d}, em *diferenciar quente de frio* dentro de processos de transferência de calor ([B3]: y10d, y18_{a,b}, y21_{a,b}, y22_{a,d}), assim como uma dificuldade diante do *conceito de calor latente* ([C1]: y3_{c,d}, y4c, y5_{c,d}). Lembrando ainda que a análise do modelo 4PL na TRI mostrou os seguintes itens considerados *muito difíceis* sendo (em ordem crescente de dificuldades estimadas): y26, y24, y17, y9, y25, y8, y19, y15.

Interessante ressaltar que o item y15, o mais difícil no teste, envolve a diferença entre os conceitos de calor e temperatura. Esse mesmo tipo de dificuldade é destacado

no item e alternativa y22b, que também envolve a concepção alternativa de que o *conceito de calor e temperatura são iguais* (A4). Sabe-se que essas dificuldades já são bem conhecidas e comentadas na literatura há décadas. Resultados semelhantes dessas dificuldades já foram obtidos em outras pesquisas (Kesidou & Duit, 1993), assim como identificado que crianças já possuem ideias formadas sobre os conceitos de calor e temperatura (Staudt & Forman, 2014).

5.1.1 Austrália, 2001.

Comparando os resultados do teste com o trabalho original (Yeo, & Zadnik, 2001), observa-se uma semelhança no *design* sendo aplicado transversalmente a 478 alunos do 10º, 11º e 12º ano, incluindo um teste-reteste ao 13º ano a um grupo de universitários australianos.

Através do gráfico estima-se uma média do 10º ano ($\approx 0,35$), 11º ano ($\approx 0,53$), 12º ano ($\approx 0,62$), pré-teste do 13º ano ($\approx 0,68$) e pós-teste ($\approx 0,78$). Para o grupo de 13º ano foi constatado um tamanho de efeito de 0,70 e ganho normalizado de $\langle g \rangle = 0,30$, sendo as diferenças significativas ($p < 0,05$).

As médias de Portugal e Brasil podem ser visualizadas no gráfico e comparadas *lado a lado* com os resultados dos autores originais. Os resultados no Brasil encontram-se muito abaixo em comparação a outros contextos. Em Portugal o percentual médio de certos dos alunos portugueses do 10º ano foram similares aos dos australianos para o mesmo nível, porém os ganhos para as outras séries foram diferentes. Isso sinaliza que a forma de tratamento ou instruções de ensino surtiram efeito positivo, porém ainda sendo necessária uma melhor eficiência.

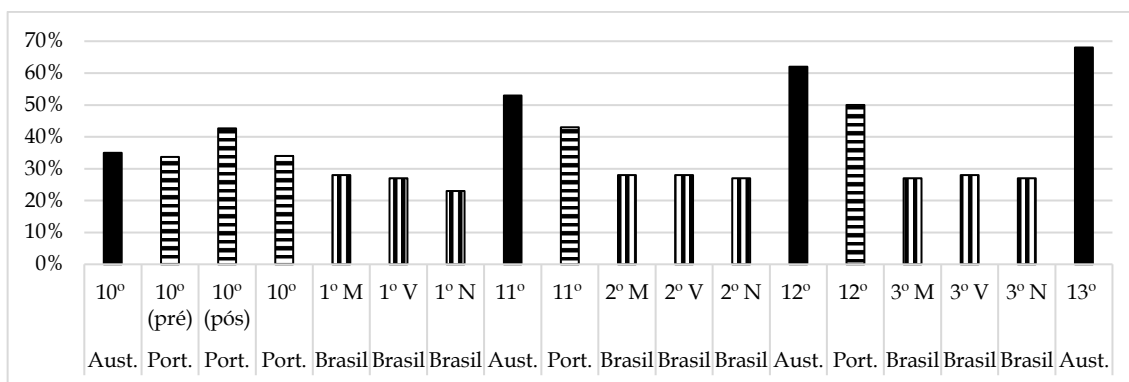


Figura 90: Comparação dos resultados da pesquisa com o trabalho original do TCE em 2001.

Observando os resultados do teste-reteste do 13º ano, verifica-se que foi semelhante ao encontrado no teste-piloto aplicado a alunos do 10º ano em 4 escolas em Portugal, com tamanho de efeito de 0,67, mas com um ganho normalizado de $\langle g \rangle = 0,14$, apesar de que o percentual médio do pré-teste tenha sido de $\langle \text{Pré} \rangle = 0,34$. No trabalho original, se destaca a dificuldade de alunos em admitir a existência da água a 0 °C: "Eu me senti um pouco confuso sobre a água a zero graus então eu pensei que não há nenhuma resposta, porque você não pode ter água a zero graus" (Yeo, & Zadnik, 2001, tradução livre).

O trabalho mostra como exemplo o item y16, na qual apresenta ID = 0,54 e DISCR = 0,80. Nesse estudo realizado, tem-se um ID = 0,20/ DISCR = 0,24 (Brasil) e ID = 0,33/ DISCR = 0,45 (Portugal). Comparativamente, os resultados mostram que houve uma mudança expressiva no contexto dos alunos australianos quanto a crença mais apropriada, que seria acreditar que "o metal retira energia da mão mais rapidamente do que a madeira", ao ser comparar a temperatura de uma régua de metal a outra de madeira. No Brasil essa crença não foi evidenciada predominantemente, sendo um item de acerto aleatório, ou seja, os alunos não conseguiram distinguir a alternativa correta dos distratores. Em Portugal esse também foi um item difícil, porém, apesar do fator de concentração sugerir uma região aleatória, os percentuais brutos sugerem dois modelos mentais (correto e incorreto), sendo o distrator mais atrativo a letra D, que seria a crença de que "Os metais são melhores irradiadores de calor do que a madeira", na qual se associa com uma concepção substancialista do calor [D2] em que o metal possui a capacidade de atrair, manter, intensificar ou absorver calor e frio, e isso reforça um

foco de atenção maior para as instruções diante esse tipo de dificuldade entre os estudantes portugueses.

Com relação à TRI, o estudo realizado, eliminou os itens y4, y6, y7, y11, y12, y21, devido à falta de consistência, estando os itens considerados *muito difíceis* sendo (em ordem crescente de dificuldades estimadas): y26, y24, y17, y9, y25, y8, y19, y15. As CCI para o item y16 mostram que é *muito difícil* para os brasileiros e *difícil* para os portugueses, tendo maior probabilidade de acerto para alunos portugueses com a mesma habilidade (no modelo de DIF comparativo multigrupo não-restringido, considerando-os na mesma métrica). Essa comparação direta na mesma métrica não é possível para as análises da TCT, dessa forma, se tem uma comparação para visualização *lado a lado*, pois os testes possuem funcionalidade distintas, pois estão em escalas supostamente diferentes.

5.1.2 Estados Unidos, 2006.

Os resultados dos ganhos nos EUA em 2006 (Luera et al., 2006) com 24 alunos no pré-teste e 47 no pós-teste mostraram um ganho normalizado de $\langle g \rangle = 0,334$. A limitação da quantidade de estudantes, a idade variando entre 21 e 45 anos, e por serem de nível de escolaridade primária, são aspectos a serem levados em conta, pois se restringem as análises dentro da TCT. Foi argumentado que o TCE poderia ser melhorado para ajudar a diferenciar a falta de conhecimento de equívocos e a identificar melhor os equívocos que os estudantes possuem (Ibid., 2006). Com relação a esse tipo de reivindicação, o atual estudo buscou contribuir detalhando os equívocos itens por itens no quadro de tipicidade, indicando a concepção associada e comentários interpretativos, bem como os equívocos em destaque. O estudo americano destaca algumas implicações na utilização do teste, e assim seleciona cinco objetivos de aprendizagem principais para o conhecimento térmico, na qual os alunos deveriam: (1) *definir energia*; (2) *reconhecer que a energia existe em diferentes formas*; (3) *explicar e sugerir exemplos de conversão ou transformação de energia*; (4) *definir e usar a conservação de energia apropriadamente, e*; (5) *definir e sugerir exemplos de entropia*. Convém lembrar que o conceito de entropia é uma lacuna no TCE.

O resultado obtido no intervalo dos escores foi entre 5 a 15 acertos, com média de 9 acertos (34%). No Brasil, o intervalo variou de 0 a 23 acertos, e a média foi de aproximadamente 7 acertos (27%), dentro de um grupo de 2957 alunos. Em Portugal variou de 0 a 24 acertos, com média aproximadamente de 10 acertos (38%), dentre 2585 alunos. Em todos os cenários o teste sugere ser difícil, tendo em vista que no contexto de Brasil e Portugal, os alunos que acertaram mais itens foram poucos. Já no pós-teste, o estudo apresenta um percentual médio de acertos igual a 56,5%.

O estudo americano considerou que os itens mais consistentes e assinalados corretamente foram os itens y7 (83%), e os itens y13 (74,5%) e y22 (76,6%), podendo dizer que são itens ideais por estarem acima de 62,5%. O item y7 foi considerado *mediano* para o Brasil (54%) e *fácil* para Portugal (60%) dentro das análises clássicas. O y13 foi de 21% e 58% acertos, e y22 de 32% e 51%, estando também acima de 50% para os portugueses.

Para os itens assinalados incorretamente tem-se o y8 (10%), y9 (9%) e y19 (10%), porém eles obtiveram um ganho considerável no pós-teste alcançando 50%, 84,8% e 30,4%. Isso sugere eficiência para as instruções relacionadas aos itens y8 e y9, e uma certa resistência no item y19. O item y19 também demonstrou certa confusão, sendo aleatório entre brasileiros e portugueses. O item explora saber se o aluno compreende o funcionamento de uma panela de pressão com o fato dos alimentos ferverem mais rapidamente, e as concepções associadas foram 3 equívocos, na qual se tem a crença que a temperatura de ebulição é limite (B5), que é sempre constante e igual a 100°C (D5), ou que o vapor se encontra acima de 100 °C (D8). Chama atenção que o item y1 para os americanos mantiveram dificuldade, enquanto para os brasileiros e portugueses foram de modo geral mais fáceis.

Dentre as limitações apontadas:

Os autores do TCE afirmam que a medida *se concentra apenas em conceitos de energia térmica*. Isso limita a utilidade do instrumento para muitos cursos, [...], os próprios conceitos de energia térmica *não estão incluídos uniformemente*; O TCE concentra-se fortemente em transições de fase em água e percepções de calor e frio, enquanto *evita a conservação de energia, entropia e o conceito de fluxo de energia*. (Luera, Otto, &

Zitzewitz, 2006, *tradução livre, grifo nosso*)

Apesar disso, esse estudo considera-o como um instrumento útil, sendo um esforço positivo para tentar avaliar de forma mais precisa a compreensão dos alunos em conceitos importantes na Física, porém ressaltam a necessidade de eventual modificações e adequação ao contexto.

5.1.3 Tailândia, 2004.

Na elaboração de um instrumento contendo itens de diversos trabalhos originais sobre testes em Física Térmica, foram utilizados 3 itens do TCE e aplicados a 212 alunos do secundário em 3 escolas de Bangkok, na Tailândia, após as instruções sobre calor e Termodinâmica: um envolvendo dificuldades na compreensão de *equilíbrio térmico* (y2 e y17); a *invariância da temperatura* durante o processo de mudança de fase (y5) (Kruatong et al., 2006).

Verificando as evidências, foi identificado dificuldades em compreender a influência do tamanho dos cubos de gelo em relação as suas temperaturas, que estaria relacionado na realidade com os itens y1, y9 e y14. Observou-se que 28% dos alunos deram a resposta incorreta de que a temperatura do grande cubo de gelo era maior do que a pequena porque poderia conter mais frio que um pequeno. No Brasil foi de 11% e em Portugal 6%, percentual baixo.

A questão do agasalho se refere aos itens y23 e y26. Os resultados mostraram que cerca de 57% dos alunos identificaram que as bonecas envoltas em cobertores nunca aqueceram porque os cobertores eram feitos de material que não mantinha bem o calor. Esse foi o modelo incorreto predominante entre os brasileiros (37%) e portugueses (27%) diante a alternativa correta, isso reforça uma atenção diante desse equívoco.

De forma correspondente ao item y7, foi constatado que cerca de 70% dos alunos não conseguiu prever a temperatura final da mistura, porém, a maioria entendeu que a temperatura final não pode ser superior à temperatura de duas amostras antes da mistura, que seria a letra "D". Essa foi o modelo incorreto mais

atrativo para os brasileiros, enquanto para os portugueses foi a alternativa correspondente a um valor médio entre as temperaturas, letra “B”.

5.1.4 México, 2009 e Líbia, 2009/2010.

O estudo no México aplicou um pré-teste a 74 alunos calouros universitários do *Centro Universitario de Ciencias Exactas e Ingenierías* (CUCEI) da *Universidad de Guadalajara* (UdG), no México, utilizando itens do TCE (Gómez, & Hernández, 2010). Os resultados apresentaram os seguintes valores para as médias do pré-teste, pós-teste e o *Fator de Hake*:

GC: 9,87; 11,86; $\langle g \rangle = 0,13$ (n=30; *Aula tradicional; Departamento de Física*);

GE1: 9,17; 13,75; $\langle g \rangle_{GE1} = 0,29$ (n=20; *Tratamento*);

GE2: 8,90; 15,74; $\langle g \rangle_{GE2} = 0,42$ (n=24; *Tratamento*).

Pode-se observar que são valores das médias do GC são próximos aos obtidos no estudo para Portugal, entretanto, os grupos experimentais obtiveram ganhos que ultrapassaram a média esperada no teste de 13 acertos, sendo o GE2 com um ganho normalizado de *impacto médio* ($0,30 < \langle g \rangle < 0,70$).

Para a análise dos erros conceituais relacionado ao calor no pré-teste (n=74), foram selecionados os seguintes resultados e comparados aos quadros de tipicidade de itens sugeridos. Para isso entende-se que os resultados do estudo mexicano se referem ao percentual das respostas incorretas dos itens, onde se acredita que todas as alternativas incorretas correspondem às concepções alternativas que se apresenta no quadro de tipicidade dos autores originais, de modo semelhante a outro estudo (Alwan, 2011).

Como foi mostrado, o estudo feito fez uma revisão detalhada em todas as alternativas dos itens (k=109 alternativas) do teste e foram constatadas algumas discordâncias e incompletudes, bem como o fato de que as concepções sugeridas para determinados itens não servem para todas as alternativas desses itens. Isso pode ser

deduzido pelo fato dos autores originais apresentarem itens que correspondem a diferentes concepções. Por exemplo, para o item y10, a concepção C4 somente foi identificada para as letras A e D, estando a letra B um pouco confusa, ou o item y13, que somente foi verificada essa concepção na alternativa B, estando ausente na letra A e C. Dessa forma, entende-se que associar o percentual de todos os distratores de cada item como o percentual correspondente da concepção sugerida originalmente seja uma forma de causar vieses interpretativos, tudo por conta de acreditar às cegas na precisão do inventário proposto, que por ventura não detalha as alternativas, induzindo o pesquisador acreditar que todas as alternativas contemplem essas concepções. Uma sugestão de contorno para isso seria considerar as somente porcentagens de resposta de cada alternativa que de fato corresponda ou contemple a concepção referenciada, podendo, se quiser, tirar uma média dessas porcentagens para uma ideia geral, porém não irá representar um quantitativo de respondentes diretamente. Para fazer uma comparação similar foi colocado na sexta e oitava coluna (da esquerda para a direita) o percentual de todas as incorretas e o percentual médio somente das alternativas que de fato correspondem as concepções dos respondentes de Brasil e Portugal.

O estudo na Líbia (Alwan, 2011) reuniu 53 estudantes universitários de Biologia, Química, Matemática e Física, e aplicou um teste que incluía todos os itens do TCE, mais a inclusão de 4 itens de outros trabalhos originais sobre conceitos básicos da Física Térmica. As respostas dos distratores para todos os grupos foi colocado *lado a lado*, entre tanto deve-se levar em conta que no México e na Líbia o teste foi aplicado a alunos universitários, enquanto para Brasil e Portugal, alunos secundários. O percentual de respostas incorretas, levando em consideração todas os distratores, possibilita perceber a força atrativa geral dos equívocos, e obviamente dos acertos brutos. Quanto menor a porcentagem das respostas incorretas, maiores foram os acertos, naturalmente. Os resultados destacam vantagens em muitos casos para Portugal, e a média dos distratores específicos identificados como legítimos correspondentes das concepções entre Brasil e Portugal, nos dá uma ideia das diferenças do padrão de resposta entre os grupos.

Tabela 46: Comparação de resultados entre México, Líbia, Brasil e Portugal.

CA's	Descrição	Itens	México (n=74)	Líbia (n=53)	Brasil (n=2957)	(%Média BR)	Port. (n=2585)	(%Média PT)
C4	Calor e frio são substâncias (concepção substancialista).	y10	66,2%	77,4%	70%	(26,7%)	58%	(20,6%)
		y13		54,7%	79%		42%	
		y10a			35%		37%	
		y10d			29%		16%	
		y13b			30%		16%	
		y16e			19%		13%	
C5	A temperatura pode ser transferida.	y7	90,5%	50,9%	77%	(25,0%)	69%	(11,0%)
		y13		54,7%	79%		42%	
		y7d			23%		5%	
		y13a			35%		18%	
		y22b			17%		10%	
B7	A temperatura de um corpo depende de seu tamanho.	y1	66,2%	73,6%	54%	(10,3%)	26%	(11,0%)
		y9		62,3%	72%		74%	
		y14		90,6%	88%		64%	
		y1d			11%		6%	
		y9e			10%		21%	
C3	O calor se movimenta de baixo para cima (como no caso da convecção da água).	14c		10%	10%	(31,0%)	6%	(35,0%)
		y20	90,5%	86,8%	79%		64%	
D5	O ponto de ebulição da água é unicamente 100 °C.	y20a			31%	(26,2%)	35%	(28,8%)
		y4	60,8%	49,1%	71%		64%	
		y8		86,8%	71%		71%	
		y19		92,5%	73%		82%	
		y4d			27%		30%	
		y8b			18%		39%	
		y8c			39%		23%	
		y19b			27%		31%	
y19c		20%		18%				
D6	O gelo está a 0°C e não pode mudar de temperatura.	y19d		26%	26%	(38,0%)	32%	(19,0%)
		y1	63,5%	73,6%	54%		26%	
D7	A água não pode estar a 0 °C	y1b			38%	(27,3%)	19%	(22,7%)
		y2	66,2%	62,3%	66%		49%	
		y11		64,2%	80%		73%	
		y2c			38%		37%	
		y2d			19%		8%	
y11e		25%		23%				
D8	O vapor está acima de 100 °C.	y6	71,6%	46,6%	77%	(25,0%)	77%	(27,0%)
		y19		92,5%	73%		82%	
		y6c			31%		46%	
		y6d			24%		17%	
		y19c.			20%		18%	

O atual estudo mostra que as concepções térmicas se manifestam entre distratores específicos dentro de cada item, informação essa que não está precisa no trabalho original de 2001. Observa-se que na concepção C4, há uma divergência por ausência de complementaridade, sendo sugerido os itens y16 (letra “E”) e y21 (letra “B”). Esse é um exemplo que destaca a **importância da precisão do instrumento**, pois o percentual de incidência apontado no estudo do México, provavelmente deva se basear nas sugestões dos autores originais, e com isso, deixando de considerar esses outros itens que também estão incidindo sobre a mesma concepção. Apesar disso, os

elaboradores do TCE em outro artigo (Chu et al., 2012), possibilitam uma maior precisão ao corresponderem as concepções com as alternativas de alguns itens, porém dentro de novas subcategorias. Entretanto, avaliar separadamente as concepções é um aspecto que foi levantado quando se trata de proporção gerais de acerto dentro do próprio teste, pois a má distribuição por acarretar vieses interpretativos.

Pode ser notado que, apesar de alguns distratores se referirem a uma mesma concepção térmica, os percentuais de resposta oscilam em sua correspondência para cada item sugerindo que essas concepções são influenciadas por questões de coerência interna, ou ainda, uma sensibilidade por conta da falta de consistência em relação ao rigor científico do conhecimento envolvido que ocorreu no processo de construção do item, podendo causar uma certa magnitude no *desvio ou distorção conceitual na interpretação do item*. De certo modo, entende-se que quando uma **alternativa for pouco provável ou mais absurda, o percentual tende a ser baixo**, mas se for de certa forma explicada e sustentada por um modelo incorreto dentro de sua lógica interna (equívocos sustentados por teorias alternativas), possa ser mais provável e atrativa, tendendo a um percentual mais elevado. Esses distratores no caso devem ser verificados com mais atenção, e serem comparados com os outros distratores do respectivo item.

Um certo padrão percentual entre distratores que carregam uma mesma concepção térmica pode ser uma evidência que reforça um perfil de concepções térmicas, em que, dentro de um paradigma quantitativo, recomenda-se uma análise fatorial confirmatória e/ou um teste estatístico apropriado que seja possível verificar se existem diferenças significativas entre os padrões de resposta, sendo um campo a ser ainda explorado.

Observa-se, para C4, um padrão médio situa-se em 26,7% entre as respostas dos brasileiros, enquanto para os portugueses seria de 20,6%, sugerindo obviamente uma atração maior para os brasileiros com relação a esse equívoco, entre muitos outros. Considerando o percentual de todas as alternativas incorretas do item y10 e y13 para o Brasil (70%; 79%) e Portugal (58%; 42%), porém, entende-se que o correto para saber

a força atrativa da concepção de um item seria somar as porcentagens dos distratores legítimos identificados no estudo que correspondem a concepção, e não todos os distratores, que no caso, para o item y10_{a,d} resultaria para o Brasil (64%) e Portugal (53%). Para os mexicanos foi apontado 66,2% do total de respondentes para essa concepção. É um pouco desconcertante pelo fato de não ser possível identificar os itens aos quais esses percentuais se refere, no caso, o item y10 ou y13. Também é possível notar uma grande diferença para os percentuais. Os resultados dos portugueses nesses cenários remetem a uma provável vantagem comparativamente devido aos menores percentuais.

5.1.5 Brasil, 2009 e 2010.

O estudo foi aplicado a 433 alunos (n=178 no 1º ano; n=129 no 2º ano; n=123 no 3º ano) do ensino médio em duas escolas estaduais do Rio de Janeiro, no Brasil, em 2009, porém, as análises consideraram 111 alunos pré-instrução (1º ano), 196 alunos pós-instrução (2º ano e alguns do 1º ano) e 126 alunos pós-instrução com reensino dos tópicos de Física Térmica (3º ano) (Gonçalves Júnior, 2012; Gonçalves Júnior & Barroso, 2011b, 2011a, 2014b)

Verificando os resultados (Gonçalves Júnior, 2012) pôde-se constatar que a média desses grupos analisados foram %<pré>=26,9, %<pós>=38,5 e %<pós_com repetição>= 50, com isso, ao comparar com o 1º ano, pode-se estimar em valores absolutos um ganho normalizado de <g_pós-instrução>= 0,16 e <g_pós-instrução_rep>=0,32, ou seja, o 3º ano após um reensino ou reforço nos tópicos de Física Térmica alcançaram um *ganho de médio impacto*, de acordo com os critérios do fator de Hake. Para os grupos correspondentes no Brasil em 2015, em Florianópolis os ganhos foram de <g_2º ano>= 0,003 e <g_3º ano>=0,019, ganho nulo, praticamente. As dificuldades existentes nos dois contextos dentro do Brasil são evidentes, para os resultados de 2014, os efeitos de obliteração podem ser diagnosticados pela estagnação e um relativo retrocesso no rendimento, comparativamente, isso sinaliza uma ineficiência das aulas de Física Térmica e necessidades de melhorias e novas

intervenções, além de atividades de reforço, que parecem influenciar, apesar de que não há detalhes quanto ao tipo de *design* didático-metodológico desenvolvido.

Em Portugal, o teste-reteste para o 10º ano ($n \approx 200$, grupos não-equivalentes) alcançou-se um ganho de $\langle g \rangle = 0,14$, e para grupos correspondentes aos mesmos níveis de escolarização houve ganhos normalizados iguais a $\langle g_{11^\circ \text{ano}} \rangle = 0,14$ e $\langle g_{12^\circ \text{ano}} \rangle = 0,24$, quando comparados ao 10º ano. Os ganhos no Brasil em 2009 para os alunos do último ano foram superiores ao correspondente em Portugal, porém, ressalta-se que houveram aulas de reforço no conteúdo, influenciando inevitavelmente os resultados, enquanto o grupo de pós-instrução naturalmente forma de apenas 0,16, isso mostra que as aulas de Portugal, mesmo sem ensino de reforço obteve relativamente mais efeito.

O estudo afirma que *“grande parte dos alunos respondentes não admitem a existência de água no estado líquido à temperatura de 0 °C, e que o gelo sempre é encontrado à temperatura de 0 °C”* (Gonçalves Júnior & Barroso, 2011a).

Comparando os resultados obtidos nos itens (Gonçalves Júnior & Barroso, 2011b) com o atual, verifica-se que os alunos de modo geral possuem dificuldades em dominar os conceitos básico da Física Térmica. Comparativamente, os resultados em Portugal demonstram um ensino mais eficiente, e mais similar contexto do estudo nas escolas do Rio de Janeiro no Brasil em 2009. Alguns itens para o Brasil em 2009 (Rio de Janeiro) e 2015 (Florianópolis), observa-se alguns aspectos comuns.

Para o item y1, o percentual segue baixo, com uma atração similar do distrator “B”, na qual o gelo estará à 0 °C no congelador após certo tempo, supondo na crença de que o *“gelo sempre é encontrado à temperatura de 0 °C”* (concepção D6 no inventário). Esse também é o distrator mais atrativo para os portugueses, sinalizando um modelo mental de concepção relativamente predominante entre os alunos, independentemente do contexto.

Tabela 47: Comparação dos resultados da pesquisa com o estudo no Brasil em 2009.

Item /Alt	Descrição	Brasil - 2009			Brasil - 2015			Portugal - 2014		
		%1º ano (n=111)	%2º ano (n=196)	%3º ano (n=126)	%1º ano (n=835)	%2º ano (n=1222)	%3º ano (n=900)	%10º ano (n=1313)	%11º ano (n=876)	%12º ano (n=396)
y1	<i>Qual é a temperatura mais provável dos cubos de gelo que se encontram armazenados no congelador de uma geladeira?</i>									
y1a	-10 °C	63,1	49,5	57,9	48,1	43,7	47,0	72,7	75,4	78,5
y1b	0 °C	27,9	45,9	37,3	34,5	39,4	38,0	18,6	19,7	17,5
y1c	5 °C	3,6	1,5	0,0	7,5	5,1	4,0	0,9	1,1	0,8
y1d	<i>A temperatura depende do tamanho dos cubos de gelo</i>	5,4	2,6	4,8	10,0	11,8	10,9	7,9	3,9	3,3
y2	<i>Francisco pega seis cubos de gelo no congelador e coloca quatro deles dentro de um copo com água, deixando os outros dois sobre a mesa. Agita várias vezes o copo até que os cubos de gelo estejam bem pequenos e tenham parado de derreter. Qual é a temperatura mais provável da água nesta situação?</i>									
y2a	-10 °C	5,4	1,0	2,4	11,8	8,9	8,2	4,0	3,1	2,0
y2b	0 °C	26,1	41,3	54,0	34,4	36,2	31,4	46,7	55,8	57,8
y2c	5 °C	41,4	38,3	28,6	35,4	35,6	42,7	39,1	34,6	35,4
y2d	10 °C	27,0	17,9	15,1	18,4	19,3	17,8	10,3	6,6	4,8
y4	<i>Sobre o fogão está uma panela cheia de água. A água começou a ferver rapidamente. A temperatura mais provável da água é:</i>									
y4a	88 °C	25,2	12,2	15,9	14,0	11,5	13,8	15,3	17,5	11,4
y4b	98 °C	26,1	29,6	32,5	29,3	29,5	28,8	37,5	30,5	41,6
y4c	110 °C	26,1	32,6	24,6	31,3	29,3	32,3	16,2	19,5	22,1
y4d	<i>Nenhuma das respostas acima é correta.</i>	22,5	24,5	27,0	25,4	29,7	25,2	30,9	32,2	24,9
y5	<i>Cinco minutos mais tarde, a água na panela ainda ferve. A temperatura mais provável da água agora é:</i>									
y5a	88 °C	9,9	4,1	2,4	7,7	7,2	6,0	6,1	4,4	5,1
y5b	98 °C	27,0	35,7	42,1	21,4	20,5	21,2	36,6	44,4	45,7
y5c	110 °C	34,2	40,3	43,6	37,7	40,8	38,5	43,0	38,6	36,2
y5d	120 °C	28,8	18,9	11,9	33,2	31,5	34,3	14,3	12,7	13,0
y6	<i>Qual será a temperatura previsível do vapor formado, acima da água na panela?</i>									
y6a	88 °C	44,1	16,8	8,7	19,9	22,1	23,5	17,5	9,0	11,6
y6b	98 °C	21,6	19,9	23,8	24,4	24,1	20,4	23,6	22,4	22,9
y6c	110 °C	19,8	45,4	54,7	30,0	31,1	33,3	41,7	50,9	50,1
y6d	120 °C	14,4	16,8	12,7	25,8	22,7	22,8	17,2	17,9	15,4
y8	<i>Pedro acredita que deva usar água fervente para fazer uma xícara de chá. Ele diz aos seus amigos: "Eu não poderia fazer chá se estivesse acampado numa montanha alta, porque a água não ferve a grandes altitudes".</i>									
y8a	<i>[...] água ferve, mas não a uma temperatura tão elevada".</i>	13,5	23,0	30,2	29,8	30,9	26,6	30,1	23,9	35,1
y8b	<i>[...] A água ferve sempre à mesma temperatura".</i>	55,9	37,2	33,3	18,1	16,7	18,5	33,2	47,0	37,6
y8c	<i>[...] a água ainda ferve a 100 °C".</i>	26,1	34,7	33,3	38,3	39,9	39,9	24,9	22,1	20,1
y8d	<i>[...] A água nunca atingirá o seu ponto de ebulição".</i>	4,5	5,1	3,17	13,9	12,4	15,0	11,8	7,3	7,2

O item y2 mostra que após as instruções dos alunos brasileiros (em 2009) o percentual de acerto se equipara ao percentual de pré-instrução de alunos portugueses, revelando uma compreensão dos alunos mais acentuada quanto ao conceito de *equilíbrio térmico*, porém com um reforço na aprendizagem foi possível ultrapassar a média geral dos portugueses. Isso mostra um certo esforço didático-pedagógico necessário para que os alunos superem as dificuldades. O resultado de brasileiros em 2015, sem instrução, revela um padrão similar entorno de 30%, também

há um padrão no distrator “C” (y2c), com percentual de acertos em cerca de 40% para todos os cenários pré-instrução. A força atrativa desse distrator pode ser compreendida pelo modelo mental de que a água agitando o gelo e fazendo-o derreter, faz ele mudar de fase (líquida) e isso leva a crer que a temperatura se elevou de 0 °C, e que supostamente estaria a 5 °C. Para o item y2, esse indica traços de um modelo conceptual incorreto que necessita ser dado maior atenção. No caso, esse tipo de equívoco está associado a duas concepções alternativas: (1) *Que a água no estado líquido não pode estar a 0°C [D7]; e (2) que os objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura, com isso há uma suposta dificuldade em compreender o conceito de equilíbrio térmico [C6].*

Analisando de forma mais ampla os resultados em todos os contextos, percebe-se que os alunos brasileiros sentem uma maior dificuldade em acreditar que a água possa ferver entre 98 °C (y4b) ou 110 °C (y4c) ao ser aquecida em condições típicas, ao contrário dos portugueses, apesar de que, esse distrator também seja o mais atrativo. A crença de que a água após 5 min alcance 110 °C (y5c) também é um destaque predominante entre os alunos. A ideia de que o vapor d’água acima da superfície da água fervente esteja acima de 98 °C se apresenta como uma confusão muito evidenciada, mostrando grande resistência entre os alunos em todos os contextos ao assinalarem que a temperatura é igual a 110 °C (y6c), sendo uma situação mais aleatória para os alunos da amostra em Santa Catarina, no Brasil em 2015. Isso remete a dificuldades de compreensão no comportamento da matéria em processos transitórios, como no caso da mudança de fase. Obviamente que essas dificuldades ou concepções podem estar associadas a diversos conceitos mais específicos, como, o calor latente de ebulição, quantidade de calor latente, etc. Entende-se que o modelo cinético molecular seja um aporte teórico promissor no sentido de contribuir junto aos alunos para desmitificar esses processos termodinâmicos, fazendo-os perceberem o comportamento das partículas nessas fases. Esse modelo mecânico newtoniano, deve, entretanto, avançar da perspectiva corpuscular para a Ondulatória, ou seja, aprofundar os níveis de compreensibilidade física ao serem introduzidos tópicos de

física moderna e contemporânea, porém de forma adequada ao nível de escolaridade.

Observando os parâmetros psicométricos estimados nesse estudo (Gonçalves Júnior & Barroso, 2011b), percebe-se que foi explorado o modelo de 2PL da TRI (apesar de ter sido apresentado equivocadamente o modelo de 1PL, com fator de discriminação $D=1$), na qual foram estimados para o item y_1 os parâmetros de discriminação ($a = -1,7$) e dificuldade ($b = 0,18$), e para o item y_2 ($a = -2,8$ e $b = 0,24$). Infelizmente essa informação se torna vaga por não ter sido informado para qual dos 3 grupos esses parâmetros foram estimados. Apesar disso, pode-se dizer que sejam itens com **dificuldade normal/mediano** ($-0,52 < b \leq 0,51$). Esses autores afirmam demasiadamente que “*para esses estudantes é impossível encontrar água a 0°C ”*, ao mesmo tempo, ao utilizarem o modelo da TRI, sabe-se que não seria impossível, mais sim, dentro de um campo de probabilidade, ainda que seja baixo.

Em 2010, no Rio de Janeiro, um outro estudo também aplicou o teste TCE para 151 alunos do ensino médio de 5 escolas estaduais (Louzada, 2012; Louzada et al., 2015). Além de se considerar que hajam evidências úteis sobre essa análise realizada, a intenção de alguns comentários e críticas é que possam ser relevantes no sentido de tentar contribuir com a melhoria do instrumento e fomentar maiores discussões sobre boas práticas na administração de testes conceituais. Com isso, destacam-se aspectos de cunho etimológico e epistemológico durante as escolhas nos procedimentos de análise, interpretações e conclusões, pois surgem pontos conflitantes na análise que parecem oferecer alguns riscos e vieses.

As categorias relacionadas as concepções alternativas foram denominadas de dimensões, e possuem conceitos físicos imbricados e fortemente associados nos discursos sobre o ensino da Física Térmica. Geralmente a problemática da dimensionalidade é um processo indispensável quando se trata de instrumentos não padronizados, ou seja, testes livres, que podem ser modificados e não seguem procedimentos padrões de administração e de análise, com isso, considera-se importante ter como marco de referência recomendações mais gerais partindo de instituições reconhecimento internacional (International Test Commission, 2014),

principalmente quando se trata de validade de construto e dimensionalidade. A validade de construto pode ser realizada em uma perspectiva qualitativa, dentro de um enfoque interpretativo-hermenêutico dos itens (podendo cruzar com um programa curricular, entre outros), ou quantitativo, em um enfoque hipotético-dedutivo, realizado por análises fatoriais exploratórias e confirmatórias, juntamente com o *Scree plot*.

O estudo afirma um ganho percentual bruto de 17,5% entre o pré-teste (30,2%), aplicado no início do ano, e o pós-teste (47,7%), aplicado ao final do ano, no momento em que os alunos passaram outro nível de escolaridade. Com isso, pode-se estimar um ganho normalizado de *baixo impacto* $\langle g \rangle = 0,25$. O argumento de ganhos sustentado com base na variação do alfa de Cronbach, é no mínimo inapropriado, pois um está relacionado ao tamanho do efeito e o outro com a consistência dos dados, um está sensível ao tamanho da amostra e outro não. Como se afirma, qualquer efeito, por mais minúsculo que seja, pode produzir um pequeno valor de *p-value* se o **tamanho da amostra ou a precisão da medição for suficientemente alta** (Wasserstein & Lazar, 2016), que não foi o caso. Eles chamam de “desconstrução” a soma das diminuições nas porcentagens das concepções alternativas, e “construção” o inverso.

Diante ao trabalho de 2010, pode-se destacar alguns equívocos, ao afirmarem que:

- Os ganhos foram obtidos pela aplicação do instrumento, ao invés de destacar os fatores relacionados ao longo do ensino-aprendizagem;
- O teste seja denominado como “questionário” (fato observado nos dois trabalhos do contexto brasileiro), entretanto, há diferenças não apenas etimológicas para os termos, mas funcionais para esses instrumentos, pois possuem finalidades distintas (McMillan & Schumacher, 1989) (essa crítica foi levantada anteriormente, ver p. 342);
- Que haja ausência de instrumentos semelhantes em língua portuguesa que avaliassem conceitos de Física Térmica, no entanto foram identificados dois instrumentos (Henrique et al., 2010; Lang & Moreira, 1996), como também uma obra

referenciada no próprio trabalho (Köhnlein & Peduzzi, 2002) que apresenta um teste sobre conceito térmicos, mas que não foi levado em consideração;

- O quantitativo de itens errado por “dimensão” (essa crítica foi levantada anteriormente, ver p. 342);
- Por não haver acesso à chave-de-correção, foi elaborada e proposta uma chave com o auxílio de 3 juízes ou avaliadores contendo erros. O estudo atual também realizou um processo de análise para a obtenção de uma chave partir da opinião “às cegas” de 5 juízes (4 sendo doutores), visando identificar fraquezas e ambiguidades dentro do processo de validade de construto. A **análise obteve 100% de acerto** com a chave obtida do TCE a partir do site da AAPT, e ao compará-la ao estudo de 2010, **foram identificadas 2 falhas na chave de correção**, respectivamente nos itens y20 e y21, além da falta de consistência no critério de eliminar os itens y17 e y18 das análises (essa crítica foi levantada anteriormente, ver p. 338).

Entre outras, a “matriz de referência” (no contexto das pesquisas de avaliação educacional, o termo “matriz de referência” é associado à matriz das habilidades e competências, ou aquela na qual são apresentados os descritores, ao contrário de como se apresenta) apresenta uma incompletude estando omissa a 3ª rubrica relacionada as concepções do grupo C, relacionada a transferência de calor e mudança de temperatura. Comenta-se sobre a validade de construto no contexto original, porém, isso não deve ser entendido como um processo de transferibilidade direta entre os contextos devido à necessidade de validação por meio da adaptação transcultural dos itens que compõem o instrumento.

O procedimento adotado baseou-se na medida de consistência do alfa de Cronbach para justificar ganhos e melhorias. Entretanto, sabe-se que esse não é um procedimento padrão a ser utilizado em testes conceituais, além do que os valores foram baixos e não deveriam ser considerados e nem servir de referência, com exceção dos resultados obtidos para o pós-teste. Torna-se preocupante o estudo de 2010 justificar o percentual das respostas do item y15b de modo fechado, ignorando outras possibilidades que levaram os estudantes escolherem essa alternativa. É preciso ressaltar que esse foi o item mais difícil dentro da análise do atual estudo.

5.1.6 Coréia do Sul, 2012.

Nesse estudo fizeram parte os autores originais do teste TCE. O teste foi aplicado a 515 alunos do 10º (n=178), 11º (n=166) e 12º ano (n=171), sendo utilizado 19 dos 26 itens (Chu et al., 2012). Os critérios de qualidade dos itens adotados baseiam-se no índice de dificuldade entre 30% a 70% de acertos e índice de discriminação igual ou maior que 20%, conforme os critérios de apresentados por Arnold J. Lien, em 1971 (Ibid., 2012, p. 1514).

Os resultados mostraram que os itens y9 e y21 não atingiram aos critérios mínimos de dificuldade e discriminação, e o item y12 excedeu o valor máximo do intervalo pretendido para a dificuldade (acima de 0,70), sendo, portanto, na realidade, **um item muito fácil**, e devido à sua exclusão, provavelmente com baixo índice de discriminação e consistência, ou seja, sem qualidade psicométrica para comparar desempenhos. O estudo afirma que 6 itens a mais não obtiveram índice mínimo de dificuldade no intervalo, mas atingiram os níveis aceitáveis de discriminação, e por isso foram mantidos na sequenciação das análises.

Na análise realizada sugeriu a exclusão de 7 itens: y3, y8, y9, **y12**, y19, y20 e **y21**. Comparando com o atual estudo (Portugal e Brasil), a análise da TRI eliminou 6 itens por falta de consistência dentro dos critérios minimamente aceitáveis dos parâmetros logísticos: y4, y6, y7, y11, **y12** e **y21**. As reincidências dos itens y12 e y21 reforçam as evidências de problemas de consistência relacionados a esses itens. A análise fatorial exploratória conseguiu identificar 4 componentes com cargas fatoriais representativas: (1) *transferência de calor e mudanças de temperatura* (y25, y15, y22, y13, y23, y7, y10); (2) *ebulição* (y5, y4, y6); (3) *condutividade e equilíbrio de calor* (y24, y14, y16, y26, y17, y18); e (4) *congelamento e derretimento* (y2, y1, y11). Essas novas subcategorias podem servir na comparação em análises de identificação e justificativa no processo de validação de construto, comparando o *enfoque interpretativo-hermenêutico* dos itens com o *enfoque baseado na análise fatorial*.

Comparando os resultados entre os alunos da Coréia do Sul com os alunos de Portugal e Brasil, pode-se observar que no item/alternativa y7b os alunos brasileiros

possuem uma baixa incidência nesse distrator, sendo mais atraídos pelo y7d (23%), que se refere na soma das temperaturas, ao serem misturados os copos de água de 40 °C e 10 °C. Apesar disso o acerto foi predominante (54%). Verifica-se também um percentual de acerto predominante para os portugueses (60%), apesar de um padrão similar desse distrator com os coreanos.

De modo geral, o Brasil fica abaixo do percentual de “guessing” para todos os níveis de escolaridade em somente concepções alternativas de 5 itens (y7b, **y22b**, y15a, y4c, y17b), os portugueses em 7 itens (y13a, **y22b**, y13b, y25b, y15a, y17b), e os coreanos em 1 item (**y22b**). Além do y22b, os equívocos y15a e y17b demonstram não serem atraídos por nenhum dos grupos, permanecendo relativamente estáveis. Isso sugere que possam ser mais facilmente superados.

Analisando comparativamente a Tabela 48, constata-se que de certa forma, os portugueses levam vantagem por não evidenciarem um número maior de concepções comparativamente com os outros grupos, porém isso não anula o fato de que eles possam estar atraídos por outros distratores (ver Apêndice A, p. 8), entretanto, essa comparação foca apenas nessas concepções especificamente. Relativamente, é possível identificar um padrão para algumas *concepções mais atraídas* pelos três grupos nos três níveis de escolaridade, que são: y4d, y5c, y6c, y14d, y17c, y26c, y2c. Como se sabe, é esperado que o percentual de respostas para esses distratores diminua com o passar dos níveis de escolaridade, porém isso não se observa, sugerindo resistências dessas concepções, na qual incidem sobre o *processo de ebulição* (y4d, y5c, y6c), *condutividade do calor e equilíbrio* (y14d, y17c, y26c), e *congelamento e derretimento* (y2c).

Tabela 48: Tabela comparativa dos Resultados do TCE entre a Coréia do Sul, Portugal e Brasil.

Dimensões e a Descrição das Concepções Relacionadas as Alternativas dos Itens		Itens/ Alt.	Coréia do Sul			Portugal			Brasil		
			10º ano (n=178)	11º ano (n=166)	12º ano (n=171)	10º ano (n=1313)	11º ano (n=876)	12º ano (n=396)	1º ano (n=835)	2º ano (n=1222)	3º ano (n=900)
Transfêrencia de calor e mudanças de temperatura	Duas temperaturas diferentes podem ser adicionadas e calculado a média	y7b	40 (23)	56 (34)	47 (28)	360 (28) n=1308	256 (35) n=873	102 (34) n=396	91 (11) n=830	195 (16) n=1215	91 (16) n=834
	A temperatura pode ser transferida de um corpo para outro	y13a	64 (36)	32 (19)	40 (23)	272 (21) n=1308	184 (15) n=873	93 (13) n=395	250 (31) n=818	413 (34) n=1200	250 (39) n=822
		y22b	30 (17)	23 (14)	34 (20)	163 (13) n=1287	104 (8) n=852	47 (6) n=393	137 (18) n=775	192 (17) n=1154	139 (16) n=779
	Materiais como a lâ têm a capacidade de aquecer as coisas	y23d	43 (23)	52 (31)	48 (28)	425 (33) n=1283	301 (24) n=850	142 (16) n=392	282 (37) n=772	447 (39) n=1157	282 (38) n=776
	'Quente' e 'frio' são diferentes, não nas extremidades opostas de um continuum	y10a	72 (40)	38 (23)	46 (27)	645 (50) n=1297	420 (28) n=862	213 (18) n=391	306 (37) n=826	412 (34) n=1201	306 (35) n=830
		y13b	47 (26)	45 (27)	36 (21)	302 (23) n=1308	196 (10) n=873	94 (6) n=395	254 (31) n=818	353 (29) n=1200	254 (29) n=822
	Não há limite para a temperatura mais baixa que pode ser alcançada	y25b	46 (26)	42 (25)	55 (32)	278 (22) n=1273	168 (20) n=842	79 (19) n=386	208 (27) n=773	284 (25) n=1147	208 (24) n=777
O calor é proporcional à temperatura	y15a	55 (31)	17 (10)	19 (11)	286 (22) n=1300	192 (23) n=867	83 (13) n=394	177 (22) n=795	223 (19) n=1197	179 (22) n=799	
Ebulição	O aquecimento sempre resulta em um aumento de temperatura	y4c	42 (23)	54 (33)	43 (25)	212 (16) n=1306	127 (20) n=872	64 (22) n=396	256 (31) n=819	351 (29) n=1196	256 (32) n=823
	O ponto de ebulição da água é de 100 °C	y4d	80 (45)	76 (46)	76 (44)	404 (31) n=1306	283 (32) n=872	132 (25) n=396	208 (25) n=819	355 (30) n=1196	208 (25) n=823
	A temperatura de ebulição não permanece constante	y5c	68 (38)	82 (50)	69 (40)	561 (43) n=1305	361 (39) n=871	160 (36) n=393	309 (38) n=819	490 (41) n=1200	311 (38) n=823
	O vapor acima da água fervente em uma chaleira fica a uma temperatura superior a 100 °C	y6c	59 (33)	67 (40)	82 (48)	538 (42) n=1291	365 (51) n=860	165 (50) n=389	246 (30) n=821	375 (31) n=1205	246 (33) n=825
Condutividade do calor e equilíbrio	A quantidade de calor em um objeto depende do material que o objeto é composto de	y14d	98 (55)	94 (57)	70 (41)	721 (55) n=1300	489 (30) n=869	211 (30) n=396	395 (49) n=801	650 (54) n=1196	395 (54) n=803
	A régua de metal se sente mais fria do que a de madeira porque o metal é bom radiador	y24c	63 (35)	58 (35)	51 (30)	420 (33) n=1275	288 (28) n=842	136 (22) n=390	231 (30) n=766	349 (30) n=1153	231 (34) n=770
		y16d	39 (22)	29 (18)	38 (22)	398 (30) n=1306	256 (31) n=870	120 (27) n=395	145 (18) n=801	246 (21) n=1197	145 (24) n=805
	Materiais como a lâ têm a capacidade de aquecer as coisas	y17b	46 (26)	39 (24)	38 (22)	205 (16) n=1296	140 (17) n=860	64 (18) n=392	167 (21) n=791	202 (17) n=1194	167 (18) n=795
		81 (46)	60 (36)	59 (35)	665 (51)	441 (43)	212 (37)	301 (38)	482 (40)	303 (41)	

	Os objetos a diferentes temperaturas que estão em contato uns com os outros não se movem necessariamente para a mesma temperatura	y17c				n=1296	n=860	n=392	n=791	n=1194	n=795
	'Quente' e 'Frio' são diferentes, mas não nas extremidades opostas de um continuum	y18a	67 (38)	48 (29)	55 (32)	299 (23) n=1298	208 (10) n=863	105 (5) n=395	254 (32) n=802	331 (28) n=1194	254 (28) n=806
	Alguns materiais são difíceis de aquecer: são mais resistentes ao aquecimento	y26c	61 (34)	60 (36)	58 (34)	390 (30) n=1279	252 (23) n=848	118 (22) n=392	265 (34) n=775	435 (38) n=1152	271 (37) n=803
Congelamento e derretimento	O gelo está sempre a 0 ° C	y1b	88 (50)	70 (42)	57 (33)	243 (19) n=1309	158 (20) n=873	68 (17) n=396	286 (34) n=830	480 (39) n=1217	295 (38) n=859
	O calor é proporcional à temperatura.	y11c	33 (19)	36 (22)	29 (17)	307 (24) n=1296	213 (23) n=862	110 (23) n=394	198 (24) n=820	296 (25) n=1202	200 (23) n=824
	A água não pode estar a 0 ° C	y2c	59 (33)	55 (33)	42 (25)	509 (39) n=1303	335 (35) n=869	143 (35) n=394	293 (35) n=828	429 (36) n=1205	293 (43) n=830

Nota: Os números em negrito indicam as concepções alternativas, de acordo com o critério de porcentagens acima do "Guessing rate" dos itens. As porcentagens das respostas encontram-se entre parênteses. Fonte: Traduzido e adaptado de Chu, Treagust, & Yeo (2012), sendo incorporado os dados da pesquisa.

5.1.7 Turquia, 2006.

Na Turquia foi aplicado a 430 estudantes universitários do Departamento de Educação Básica, na qual Baser (2006) afirma que o teste foi “traduzido e adaptado” para o contexto, e comenta a ausência de itens sobre *isolamento térmico*, com isso, inclui dois itens relacionados. Para o grupo experimental foi adotado como tratamento o *método de Conflito Cognitivo baseado na Instrução* (“*Cognitive Conflict based Instruction – CCI*”) e utilizado a *análise de variância* (ANOVA) como procedimento de análise dos dados obtidos. O estudo revela um ganho considerável fazendo o GE atingir a média de 17,26, enquanto o GC (tradicional) ficar com 11,45. O valor do GC se aproxima dos resultados obtidos em Portugal e Brasil, sinalizando a necessidade de novos métodos de ensino serem desenvolvidos com os alunos. O ganho normalizado (fator de Hake) para o grupo experimental foi $\langle g \rangle_{\text{exp}}=0,427$ (médio impacto), e para o grupo controle $\langle g \rangle_{\text{cont}}=0,147$ (baixo impacto).

5.1.8 Nigéria, 2015.

Na Nigéria, o TCE também foi utilizado para 249 alunos de 2 escolas secundárias para verificar a eficiência do *método de Conflito Cognitivo baseado na Instrução* diante o modelo tradicional (Madu & Orji, 2015). Foi obtido uma consistência com o alfa de Cronbach igual a 0,79.

Algumas justificativas relatadas pelos alunos pelo grupo controle diante o processo de ebulição foram: “a água pode ter impurezas, que aumentam o ponto de ebulição”; “a água pode ferver abaixo do seu ponto de ebulição”; “porque 98 °C é próximo ao ponto de ebulição da água”. As respostas implicaram que os estudantes consideravam o ponto de ebulição da água como consequência de impurezas na água. Na segunda parte, a maioria dos alunos forneceu os seguintes motivos: “o tempo foi aumentado; portanto, o ponto de ebulição deve aumentar”; “a temperatura da água irá diminuir quando o ponto de ebulição for atingido”; “um aumento no calor aumentará o ponto de ebulição”. Os alunos deste grupo não entenderam que o aquecimento de água não depende do tempo, mas da quantidade de calor fornecida (Ibid., 2015).

5.1.9 Singapura, 2017.

Em Singapura, recentemente, foi realizado uma pesquisa-ação para investigar se uma abordagem pedagógica mais interativa e envolvente, como a aprendizagem autêntica e baseada em informações, poderia tornar o pensamento dos alunos mais explícito através de discussões e outras interações sociais (Fernandez, 2017). Dessa vez, o estudo foi realizado para 3 turmas em uma escola secundária (n=89), e comparou o ensino tradicional de Física com o método de *Ensino por Investigação*, na qual é denominado especificamente de *“Authentic Inquiry-Based Instruction (AIBI)”*. No estudo foi constatado e mitigado as diferenças existentes já no pré-teste do TCE entre alunos, e os resultados com o pós-teste atingiram para o grupo N1 uma média de 19,88 ± 4,22.

5.2 Limitações na Investigação

Quanto à função exercida nos testes educacionais em larga escala, sabe-se que adentra em um campo fértil de discussões e controvérsias entre especialistas em avaliação. A perspectiva da psicologia mais geral explora fatores comportamentais a influenciar o mal desempenho de estudantes (depressão, ansiedade, etc.), bem como os processos mentais globais, sendo criticados principalmente na função certificatória, na qual exames selecionam e classificam alunos, profissionais, escolas, etc., considerados testes de alto risco (*“high-stakes test”*) e favoráveis a reducionismo curricular. No que se refere a função avaliativa diagnóstica e regulatória cognitiva dentro da psicologia educacional, quando recaem mais adentro da psicologia experimental, algumas as críticas se sustentam no argumento do excesso de psicologização na educação dentro das escolas ao associar as dificuldades de aprendizagem como distúrbios e doenças sujeitos a intervenção clínica e testes psicológicos (Lajonquière, 2009), outros quanto ao excesso e inadequada utilização de testes de aferição de conhecimento, principalmente quanto a supostas distorções causadas na prática de ensino devido aos exames nacionais (Afonso, 2009, 2011; Alves & De Ketele, 2011). Em contraponto, os psicólogos educacionais não vem demonstrando interesse no paradigma quantitativo psicométrico como campo de

pesquisa educacional a contribuir com as abordagens qualitativas que predominam, sendo raros entre aqueles que criticam um mínimo domínio diante aos procedimentos de análise psicométricos, ignorando que há décadas que o próprio sistema de avaliação educacional é regido por pressupostos dentro desse campo de estudo, e não há perspectivas de mudança nem a médio ou a longo prazo, como no caso do contexto do Brasil.

Um dos aspectos centrais é a crítica de especialistas em educação no uso de testes em larga escala (p. ex.: exames nacionais) em *avaliar competências* e a qualidade de todo um sistema educacional (Alves & De Ketele, 2011) como ocorre no *Sistema Brasileiro de Avaliação da Educação Básica* (SAEB), atualmente sendo reformulado e denominado de *Sistema Nacional de Avaliação da Educação Básica* (SINAEB), mas que de forma semelhante ocorre em outros países. O problema está no fato de que não há um instrumento adequado para avaliar de forma eficaz as competências dentro de todos os domínios do conhecimento (cognitivo, psicomotor e socioafetivo), pois muitas delas estão associadas com a capacidade de resolução de problemas práticos, considerando valores, comportamentos, socialização e vivência em grupos, expondo o aluno em um ambiente ou contexto real de situação.

Defende-se, no entanto sua utilização com uma função mais específica em contextos de sala de aula dentro de tópicos do ensino mais no sentido de orientação por meio de diagnósticos ou regulação dentro do processo formativo, seja avaliando comparativamente (entre grupo experimental e controle) uma forma específica de mecanismo ou proposta de ensino, podendo estar a serviço do aluno e/ou professor, independentemente disso é necessário que seja pertinente diante de um alinhamento curricular quando se assume um compromisso com as finalidades dos documentos reguladores que regem o sistema.

No atual estudo, a compreensão sobre as competências envolvidas foi realizada de forma exploratória através de uma análise interpretativa preliminar, no processo de construção da matriz de habilidades do teste, porém os resultados foram enfatizados na perspectiva advinda das análises fatoriais da TRI, que sustenta

epistemologicamente a construção da escala de proficiências. O fato é que essas diferentes perspectivas ainda podem ser exploradas em análises fatoriais confirmatórias, bem como em estudos complementares de *meta-análise*, possibilitando através dos métodos estatístico de revisão da literatura uma verificação e maior consistência dessas perspectivas para o instrumento.

Quanto aos procedimentos, percebe-se que dentre os trabalhos levantados há um uso indiscriminado da avaliação por meio de testes conceituais que não levam em consideração critérios legítimos de validação, não descrevendo os riscos de validade ou maiores cuidados nos tratamentos e procedimentos de análise, sendo comumente reproduzidos, considerados equivocadamente “adaptados” e reaplicados nos mais diversos contextos, sem ao menos referenciar os procedimentos que visam atender as minimamente as recomendações de boas práticas na administração desses testes, estando ausentes informações, como por exemplo, da existência de correspondência de respondentes entre os grupos de pré-teste e pós-teste.

As análises de DIF também podem ganhar uma nova projeção explicativa, e mais detalhada ou serem considerados uma análise do funcionamento dos itens para os diferentes subgrupos envolvidos (nível de escolaridade, sexo, região, etc.). Isso ampliaria a perspectiva de análise das diferenças no campo da educação comparada, podendo ser ampliada para a análise das causas intrasistemas, como as metodologias exploradas diante os perfis docentes, as condições específicas de recursos físicos de cada escola, dos recursos didáticos adotados, das análises curriculares, ou ainda dos hábitos de aprendizagens que estejam influenciando esse processo de aquisição de conceitos e desenvolvimento de competências. As análises de DIF também ser exploradas por diversos procedimentos, técnicas e modelos estatísticos dentro da Psicometria.

De acordo com os resultados, o teste TCE apresenta limitações para a escala de proficiências, pois o teste remete um nível elevado de dificuldades, necessitando de mais itens fáceis e medianos, para que a escala possa se tornar mais rica quanto as interpretações entre os níveis de proficiência. Isso sugere a necessidade de melhorias no instrumento, apesar de que, esses aspectos também sejam influenciados por

características próprias das populações avaliadas, não podendo ser desconsiderado o fato de que em outros contextos o teste deva estar em um nível mais adequado e equilibrado para a construção da escala. Também pode-se acrescentar problemas de ambiguidade em alguns itens que foram apontados na análise da chave-de-correção, necessidade de melhor coerência interna quanto a distribuição das concepções. Também se afirma que enquanto a distinção de calor e temperatura é destacada, há uma limitação do TCE com a ausência do conceito de entropia (Anna, Laffranchi, & Lubini, 2011, p. 143), sendo um conceito fundamental para a 2ª Lei da Termodinâmica, ainda que aparentemente o instrumento não demonstre se propõe a avaliar esse campo delimitado do conhecimento térmico.

Os indícios de dependência local entre os itens que foram identificados durante a fase exploratória, sendo em grande parte dentro de um enfoque interpretativo-hermenêutico, revelou diversas evidências de dependência que não foram exploradas devido à tomada de decisão em identificar e adotar um modelo logístico que se ajustasse os dados de forma a atender de forma minimamente satisfatória o pressuposto de independência local. Diante disso, uma nova perspectiva também pode ser explorada através de modelos da TRI denominada *testlets*, na qual possibilita admitir de forma direta a dependência local entre os itens (Ostini & Nering, 2006; Wainer, Bradlow, & Wang, 2007; Wainer & Lewis, 1989), podendo ser utilizado o modelo Rasch (Wang & Wilson, 2005), ou explorando DIF e aspectos de multidimensionalidade (Fukuhara & Kamata, 2011; Wainer, 1995).

Vale ressaltar algumas limitações para os testes conceituais apontadas na literatura (Rollnick & Mahooana, 1999), em que muitas vezes, os alunos fornecem ou associam motivos incorretos para suas respostas corretas, com isso os testes de múltipla escolha acabam por não fornecerem uma profunda compreensão das ideias dos alunos sobre o conteúdo envolvido. Sabe-se que os relatórios e entrevistas são métodos de coleta de informações com finalidades afins nesse âmbito, mas que exigem treinamento e dispêndio de tempo por serem mais trabalhosos, com isso, é natural que na comparação com os testes de múltiplas-escolhas para formar diagnósticos cada um possua suas próprias vantagens e desvantagens. Diante disso, a

combinação de métodos tem sido explorada visando atender as fraquezas de cada método. A exploração de um teste de múltipla escolha exigindo ao mesmo tempo um **raciocínio explicativo para a resposta** levou ao desenvolvimento aos *testes de dois níveis* ou *testes de duas camadas* (“*two-tier tests*”), porém uma limitação encontrada para esse processo foi que a *falta de conhecimento* não foi diferenciada de *equivocos*. Para superar essa desvantagem foi elaborado (Hasan, Bagayoko, & Kelley, 1999) um *índice de certeza de resposta* (ICR) diante um teste diagnóstico clássico sobre conceitos de mecânica. O ICR consiste em anexar a cada item um padrão moderador sobre as suas convicções em uma escala *tipo-Likert*, na qual os extremos seria de que a tomada de decisão se baseava em *conhecimento aprendido em sala de aula ou em livros* (classificação 1 a 5) ou *não* (classificação de 0 para uma suposição total, totalmente aleatório, “chute” ou acerto ao acaso) (Ibid., 1999, p. 297, *tradução livre*):

Totalmente uma adivinhação	Quase uma adivinhação	Não tenho certeza	Certeza	Quase convicto	Totalmente convicto
0	1	2	3	4	5

Dessa forma, este índice pode ser usado em conjunto com testes diagnósticos visando para diferenciar respostas erradas com base na **falta de conhecimento** ou de **equivocos**. Os autores afirmam que os alunos que obtiveram uma *baixa pontuação do ICR* por uma resposta incorreta **indica falta do conhecimento**, enquanto uma *alta pontuação do ICR* para uma resposta incorreta **indica um equivoco**. Da mesma forma, para um resposta correta com *alta pontuação do ICR* indica uma **convicção e domínio do conhecimento**. Assim, cada camada de um *teste conceitual de três camadas* (“*three-tier test*”) podem ser descritas como:

- 1ª camada: opção de múltipla escolha;
- 2ª camada: explicação/justificativa da escolha;
- 3ª camada: nível de confiança na escolha através do ICR.

Alguns estudos vem explorando esse tipo de perspectiva (Eryilmaz, 2010), e pode-se dizer que a vantagem de um teste conceitual de três camadas é que a verificação de baixa probabilidade para estimar *falsos positivos* e *falsos negativos*, podem ser utilizados para determinar a validade do teste (Peşman & Eryilmaz, 2010). No caso, falsos positivos ocorrem quando são fornecidas explicações incorretas para respostas

corretas, enquanto que falsos negativos ocorrem quando são fornecidas explicações corretas para respostas incorretas, assim, justificam-se a existência de falsos negativos devido à má elaboração do item, ou seja, ao processo de construção de itens (D Hestenes & Halloun, 1995). Uma revisão básica sobre testes em duas e três camadas, além da proposição de uma **quarta camada** pode ser vista em Caleon & Subramaniam (2010), na qual se inclui a cada item um subitem de múltiplas-escolhas denominado "*razão científica da minha resposta*", porém recomenda-se que o teste de duas camadas possa ser mais vantajoso em alunos graduados.

Contudo, essa é uma limitação presente no estudo, já que foi aplicado e analisado sendo um teste de uma camada. Os riscos de falsos positivos e negativos devem ser considerados, porém, alguns deles foram detectados e sugeridos como rastreá-los através de um **método original** para análise complementar denominada preliminarmente como *níveis de distorção conceptual do item*, na qual foi indicada a possibilidade de vieses interpretativos devido a existência de concepções ou equívocos associados a alternativa correta, ou seja, o fato de ser identificado um modelo correto dentro de um item (um único pico) não representa de forma direta uma infalível certeza de domínio sobre o conteúdo ou de verificação do subsunçor, mas sim como sendo uma possibilidade de evidência.

5.3 Perspectivas Futuras e Complementares de Análise

5.3.1 Contribuições na Função de Desenvolvimento de Competências em Livros Didáticos

Alguns estudos tem utilizado o uso de testes aplicados a professores para identificar conteúdos essenciais e fundamentar uma pedagogia para o EC, como mostram Cobern, Schuster, & Adams (2014). Nesse âmbito, em especial, acredita-se que uma incorporação de conteúdos relevantes e complementares baseados em evidências possam ser incluídos em Livros Didáticos de Física – LDFs (ou áreas afins), possivelmente viáveis nos projetos editoriais que disponibilizam recursos interativos digitais, e com isso tornando possível a disponibilização de conteúdos adequados culturalmente e regionalizados contextualmente aos alunos.

O Livros Didáticos, de modo geral, são recursos regulatórios e orientadores para o currículo, sendo um representante legítimo sobre os conteúdos essenciais a serem ensinados. Diante disso se defende que os critérios na seleção e escolha de conteúdos relevantes e complementares sejam priorizados com referência as dificuldades conceituais identificadas em avaliações diagnósticas, na prática, porém, é preciso garantir minimamente a disponibilização do conhecimento prévio relevante e de tarefas propostas de forma lógica e significativa, entre outros aspectos que visem favorecer uma aprendizagem significativa, conforme propõe Ausubel. Dessa forma, esse processo remete a não somente a encargos pedagógicos textualizados, mas também propostas de intervenção de ensino que visem favorecer a função *Desenvolvimento de Capacidades e Competências* (DCC) nesses recursos instrucionais (Gérard & Roegiers, 1998).

Nessa perspectiva investigativa de complementariedade, considera-se entre os encargos pedagógicos relevantes para potencializar a função de DCC a **elaboração de situações-problemas**. Para isso se torna relevante uma aliança com estudos e análises sobre a qualidade científico-pedagógica dos LDFs, e assim, considerar de forma estratégica as dificuldades dos alunos, diferentes níveis de dificuldade, técnicas específicas para grupos compostos por alunos de baixo e maior desempenho, diferença de gênero, raça, cultura, nível socioeconômico, aspectos intrapsicológicos, entre muitos outros que podem emergir das contribuições de estudos psicométricos diagnósticos em testes, levando em consideração a relevância de que aspectos éticos que possam ser prejudiciais diante de graves problemas no cenário das desigualdades sociais que eventualmente possam existir. Para isso, entretanto, também se torna necessário uma verificação por meio de outros instrumentos e perspectivas sobre os encargos pedagógicos prioritários para os LDFs dentro de uma pedagogia integradora, assim como aliar a métodos qualitativos de acompanhamento quanto ao perfil de utilização dos destinatários.

Uma análise mais apropriada sobre as competências em ação (saber-actuar) podem ser desenvolvidas com o suporte da **Teoria dos Campos Conceituais de Vergnaud**, auxiliando na compreensão sobre os processos operacionais de como os

alunos constroem seu conhecimento, como também em aspectos a serem considerados das situações de aprendizagem numa perspectiva intradisciplinar e interdisciplinar. O avanço dentro das diferentes análises nos mais diversos níveis das competências devem ser potencializadas.

Argumentos quanto à forma de organização de grupos de alunos em *designs* em salas de aula teóricas ou de atividades práticas laboratoriais podem extraídos de índices de discriminação psicométricos, de acordo com o perfil de turmas, desempenho de grupos ou de sujeitos. Além disso, o posicionamento dos itens na escala de proficiência pode remeter a atividades específicas a serem priorizadas no ensino para determinados grupos e indivíduos. Informações pedagógicas sobre os conteúdos podem ainda auxiliar na aplicação em resultados de investigações educacionais, incluindo na construção e utilização de recursos midiáticos digitalizados, semelhantes aos objetos virtuais de aprendizagem disponibilizados pelo governo brasileiro (Ministério da Educação, n.d.), considerados obrigatórios em projetos editoriais de LDFs desde 2015, de acordo com o *Plano Nacional do Livro Didático* (Brasil, 2014).

A elaboração de novos instrumentos para avaliar a qualidade científico-pedagógica dos LDFs pode ser feita com base na construção de escalas que envolvam os encargos pedagógicos dos mesmos, podem ser captado e valorizado opiniões na perspectiva docente e discente (com certa restrição aos alunos), e assim podendo serem comparados aspectos necessários por aqueles que utilizam estes recursos, contribuindo na avaliação de especialistas por meio de critérios mais detalhados e específicos, além da possibilidade de comparação de diferentes LDFs. Obviamente, que a consistência das evidências precisa ser reforçada e explorada por indicadores de outros estudos semelhantes ao serem cruzadas as informações, além das perspectivas teóricas e metodológicas.

5.3.2 Análise Complementar de Aprofundamento: Níveis de Distorção Conceptual do Item e Rede de Dependência Local de Concepções

Um conjunto de eventos relacionados a um mesmo contexto é uma das condições para que haja uma dependência local de itens. As situações-problemas seguem em uma linha sucessiva de eventos, com isso, a natureza do problema varia na temporalidade da situação física que sofre desdobramentos. Nessa cadeia de eventos, a cada vez que se assinala uma alternativa se assume um compromisso de verdade interna e de coerência diante de suas convicções. Isso representa que as informações apresentadas e acreditadas são somadas nos argumentos pessoais e assim levadas em consideração na sequência das situações-problemas que se responde. Isso também pode sugerir que as informações dos itens, consciente ou inconscientemente podem ligar ideias, e se não fazem, não foi intencionado pelos elaboradores dos itens, ainda assim, o respondente pode o fazer, e essas conjecturas precisam ser levantadas, pois podem fazer parte da estrutura lógica preexistente.

Entende-se que as informações fornecidas no enunciado e nas alternativas são dados explícitos, mas remetem indiretamente a uma compreensão do contexto que muitas vezes envolvem conhecimentos implícitos, e participam no processo subjetivo da construção e representação da realidade. No caso, as informações explícitas e/ou implícitas podem carregar ideias de contextos paralelos na temporalidade dos eventos da situação, com indícios sobre uma outra situação física que provavelmente tenha acontecido, está acontecendo, ou vai acontecer. Essas conjecturas são possibilidades que podem estar estruturando as interpretações e as representações na sequência dos eventos dentro do contexto atual que se apresenta. Isso remete a implicações epistemológicas conceituais no processo de compreensão, representação, análise e nas etapas de resolução do problema, e assim, promover estímulos para que se recorra às concepções alternativas disponíveis diante dessas conjecturas.

Para uma análise nessa perspectiva, deve-se identificar na situação-problema uma linha de temporalidade e um conjunto de situações que aconteceram ou podem vir a acontecer, com isso, realizar uma descontextualização, fazendo recortes criteriosos cirúrgicos na didatização do encargo pedagógico, separando dimensões

temporais do evento para assim analisar separadamente as hipotéticas situações em que o contexto sugere que tenha acontecido ou que irá acontecer. Entende-se que a importância disso é que essas interrelações entre os contextos podem trazer informações que coincidem com aquelas construídas no processo de interpretação pelos respondentes, na qual eles as utilizam para acessar as concepções disponíveis diante dos conceitos que não se percebe em uma análise interpretativa superficial e direta da situação-problema.

Esses recortes podem ser considerados partes de uma engenharia de desconstrução e reconstrução didática, e essas informações podem ajudar a compreender não apenas problemas mais internos na elaboração dos itens, como identificar os traços latentes primários e secundários subjacentes ao ligar as ideias dos enunciados com as alternativas, e com isso, possibilitar ainda supostas explicações para as atrações entre os distratores.

O primeiro recorte da situação-problema é sobre o passado, buscando identificar se a situação atual é resultado de algo de uma cadeia de eventos ou de uma única possibilidade de evento. O recorte atual é uma perspectiva direta da situação-problema, ou seja, do presente, enquanto o do futuro remete a uma previsibilidade.

Uma ligação de eventos imbricados implícitos ou explícitos podem exigir uma série de habilidades cognitivas para a compreensão de uma problemática como um todo, entende-se que essa perspectiva interpretativa pode caracterizar na identificação de uma família de situações-problema, e a razão para que se compreenda essa família é o que se pode chamar de competência científica restrita dentro de um campo disciplinar, dentro de uma unidade temática. A interpretação dos conhecimentos envolvidos nessa família interliga e integra um conhecimento principal mais estruturado, e no caso disso estar interno dentro do quadro semântico do enunciado do problema, invoca aquilo que se quer medir e avaliar, é o traço latente predominante do construto.

Em uma perspectiva externa das situações-problemas, item após item, a identificação de traços latentes secundários pode apresentar certo grau de interferência com outros eventuais traços latentes inerentes e indissociáveis a natureza da família

de situações-problemas, e isso pode refletir influenciando negativamente numa distorção interpretativa dos conceitos e conhecimentos do item, e fazendo com que o aluno construa uma rede de concepções que se afaste do traço latente que se pretende avaliar dentro de sua coerência interna entre eles. Por sua vez, poderá fornecer informações sobre a qualidade explicativa, informacional, e grau de confiança nas interpretações das análises psicométricas dos itens, que praticamente não leva em consideração essas conjecturas. Esses conhecimentos podem ser imprescindíveis para uma compreensão mais alargada, que justifique e interprete de forma mais clara quais foram as supostas intenções e influências que o instrumento sugere aos respondentes em sua tomada de decisão. Para identificar e classificar esses conhecimentos envolvidos no item como um todo (enunciado + alternativa), tem-se:

1. **Interpretação *praeteritum*** (do latim: “passado” ou de pretérito): Se caracteriza como um processo de interpretação que recorre a uma *análise investigativa de vestígios*, ou seja, o enunciado apresenta dados de efeito ou consequência de um evento/fenômeno já ocorrido. Assim, a situação-física atual (do contexto do problema) remete a necessidade de conhecer as condições sobre o que já aconteceu em uma linha sequencial temporal de eventos interligados e consecutivos, ou traz implicações de causa que possam influenciar a interpretação pelo respondente. Por exemplo, o trabalho de perícia forense trabalha em especial com esse tipo de problema (homicídios, acidentes, etc.).

- **Conhecimento Implícito-Interpretativo *a priori***: Algumas informações do contexto do problema são aceitas de modo imediato, sem uma verificação ou sem levar em consideração na resolução do problema por não serem consideradas relevantes, pois a tendência de leitor geralmente é filtrar e selecionar de forma pragmática os dados, buscando selecionar as informações minimamente necessárias. Uma análise interpretativa-hermenêutica das condições atuais e que antecedem a situação-física em seu contexto possuem informações que possam ter influência sobre a compreensão do fenômeno, porém são negligenciadas ou desconsideradas geralmente por não serem imprescindíveis para a resolução do problema, mas se torna fundamental para uma compreensão mais profunda do problema e do reconhecimento de elementos

específicos que caracterizam a família de situações-problema na qual ela pertence, sendo informações para compreender os conceitos relacionados às competências envolvidas. O conhecimento implícito pode se estabelecer a partir de dados diretos e/ou indiretos, explícitos e/ou implícitos presentes no enunciado do item.

2. **Interpretação *praesenti*** (do latim: “presente”, atual ou de *status quo*): Se caracteriza como um processo de interpretação e resolução que recorre a uma *análise investigativa de um evento contínuo*, ou seja, a situação-física atual descrita no enunciado apresenta dados que não mudam aparentemente com o tempo (*status quo*, continua como antes), podendo ser um sistema em equilíbrio estático, dinâmico, termodinâmico, ou eletromagnético, independentemente do tipo de equilíbrio (estável, instável ou indiferente). O conhecimento é explícito e direto, possuindo aspectos que caracterizam um processo baseado em previsibilidade nas condições necessárias para que não haja mudanças, na permanência do equilíbrio. Exemplos: as forças que envolvem e mantêm um edifício (equilíbrio estático); a queda de um corpo tendo alcançado sua velocidade limite (equilíbrio dinâmico); as condições ambientais de uma sala de aula sendo controlado por um termostato (equilíbrio termodinâmico); compreensão sobre o comportamento de corpos eletricamente neutros, imãs, e da luz (equilíbrio eletromagnéticos). Existem condições dinâmicas que apresentam características e propriedades constantes (explícitas e conhecidas), como no caso de fluxos de fluidos na hidrodinâmica.

3. **Interpretação *futurum*** (do latim: “futuro” ou daquilo que vai acontecer): Se caracteriza como um processo de interpretação e resolução que recorre a uma *análise investigativa de um evento em processo de mudança*, ou seja, a situação-física descrita no enunciado apresenta dados variando com o tempo. A solução está na compreensão da previsibilidade e se concentra nas condições que causam as mudanças na situação-física onde se buscam os seus efeitos, sendo, portanto, um sistema dinâmico baseado em interações (por contato ou à distância) causadas por forças (gravitacionais ou eletromagnéticas), reações químicas e processos termodinâmicos que envolvam trocas de calor. Exemplos: compreensão de um carro em processo de frenagem (forças de atrito); queima de um corpo por combustão (reação química); funcionamento de um

motor (processo termodinâmico).

- **Conhecimento Implícito-Interpretativo *a posteriori*:** A interpretação sobre a solução do problema é um processo de tomada de decisão entre as alternativas de um teste com respostas fechadas. Os argumentos e justificativas que corroboram as inconsistências das alternativas descartadas fazem parte do processo de domínio e compreensão sobre o problema, e são carregadas geralmente de conhecimento implícito-subjetivo, e podem advir diretamente das informações das alternativas de resposta do item. É natural que a alternativa mais plausível deva, de preferência, apresentar informações de forma altamente satisfatória (consistente, adequada, rigorosa, pertinente,...) em relação ao traço latente que se pretende medir (e avaliar). Nisso, uma verificação minuciosa desse processo analítico interpretativo-hermenêutico alternativa por alternativa é fundamental para aprofundar uma compreensão relacionada ao grau de distorção conceitual ou das concepções diante os conceitos que envolvem o item. Entende-se que esse processo de análise deva ser interno (dentro do próprio enunciado da alternativa, mas com um caráter de pertinência com o enunciado do problema) e comparativo com as outras alternativas, tendo como referência a alternativa correta na chave-de-correção, com isso está condicionada a esse nível interpretativo e com tendência de serem classificadas como unidades ordinais, não necessariamente dentro de uma única dimensionalidade ou contínuo, pois é possível que haja alternativas totalmente fora de contexto em relação as outras.

O conhecimento implícito-interpretativo na perspectiva da *sequenciação temporal das situações-físicas envolvidas* e suas correspondências à situação-física atual dentro do enunciado de um problema, é uma perspectiva de aprofundamento para uma compreensão adequada no processo de resolução de problemas, em especial, para os estímulos de leitura e interpretação do respondente. Os mais experientes podem, entretanto se concentrar nos dois últimos níveis, mas o excesso de confiança é um risco. A profundidade interpretativa na explicação das causas e efeitos fenomenológicos envolvidos no campo da Física podem ser rigorosas, podendo ser em nível macroscópico, mesoscópico ou microscópico, no que diz respeito ao *nível de observação*

e recorte epistemológico na complexidade do conteúdo em questão. Se forem baseados em elementos de observação direta na qual envolvem os sentidos é mesoscópico, mas se forem por observação indireta com fatores que evidenciam e justificam aquilo que se observa, p. ex., se forem em nível de comportamento das partículas constituintes da matéria utilizando o auxílio do MCM é microscópico, e se forem utilizados o comportamento dos corpos celestes seriam macroscópicos, e na perspectiva dos sentidos perceptíveis humanos seria mesoscópico.

5.3.2.1 Níveis de Distorção Conceptual do Item

Um entendimento sobre “*nível de distorção conceptual do item*” é uma perspectiva original e em desenvolvimento dentro de uma perspectiva complementar do atual estudo, na qual se tem a intenção de agregar valor nas análises interpretativas dos itens, comportamento do teste no processo de resolução, construção de itens, análises de distratores, sobretudo, buscando contribuir na validade e qualidade do teste, além possibilitar uma comparação e análise de ranqueamento com as evidências empíricas.

Parte-se de um entendimento relacionado ao afastamento relativo quanto ao nível de compreensibilidade, coerência e adequação em comparação a alternativa mais correta ou da solução mais plausível para a situação-problema proposta nos itens, devendo esta última estar fundamentada com o rigor do conhecimento científico relacionado.

A discriminação das alternativas pelo grau de distorção conceptual (ou conceitual, em determinados aspectos) do item pode ser hierarquizada ou não, dependendo da interpretação qualitativa para a resposta mais correta, na qual pode se recorrer ao critério de maior afastamento da resposta mais provável. Para uma descrição mais clara, em ordem crescente, os *graus de distorção conceitual* na resposta dos itens dentro da Física seriam:

Quadro 71: Descrição dos níveis de distorção conceptual entre os itens.

<p>Nível: 0 – opção plausível</p>	<p>Considerada a com maior probabilidade que ocorra diante das outras alternativas, e sendo por isso considerada a correta. Características: não há aparentemente uma distorção em relação ao conhecimento científico, obedece aos princípios e leis vigentes que envolvem o fenômeno, adequação as normas do SIU, onde a alternativa estaria baseada em condições ideais ou mais prováveis (de acordo com as CNTP). Apesar disso, não há garantias da ausência de uma concepção alternativa envolvida, pois existe a possibilidade de que o indivíduo estabeleça um raciocínio equivocado desconhecido e que seja responsável pela tomada de decisão que por ventura coincida na alternativa correta ou a mais plausível. Entre os possíveis fatores explícitos que possam causar isso, estão o próprio enunciado do item ou na forma de como a resposta se apresenta. Entre os fatores implícitos estão os modelos de aprendizagem subjetivos de cada indivíduo, que por sua vez somente uma investigação centrada na observação da operacionalidade na resolução de problemas possibilitaria melhor identificar.</p>
<p>Nível: 1 – opção plausível</p>	<p>Situação possível, mas pouco comum diante do cotidiano da maioria da população ou das circunstâncias da situação física, por isso deve ser descartada apesar de ser uma alternativa relativamente próxima daquela mais provável. Poderá estar associada a um fenômeno que ocorra em certos ambientes de forma natural, ou seja, existe uma condição para que a situação seja possível obedecendo aos princípios e leis naturais que o regem (relativamente próxima as CNTP). Entre as eventuais características estão: natureza transitória, um fenômeno sazonal, pontual temporalmente, específico regionalmente, ou sendo pouco provável que permaneça nesse estado de acordo com o contexto da situação-problema no enunciado. A uma alternativa que pode causar dúvida. A coerência existente poderá estar associada a teorias intuitivas, sustentada por explicações relativamente estáveis ao ser comparado a situações semelhantes, ou seja, que supostamente possa estar confundindo ou causando uma má interpretação, um erro conceitual por situações análogas, porém inconsistentes. Pouco afastamento da resposta correta.</p>
<p>Nível: 2 – opção não-plausível</p>	<p>Resposta errada, raciocínio equivocado, onde a situação dificilmente ocorreria em condições naturais na Terra (distante da CNTP). Características: relativamente coerente ao corresponder às grandezas físicas e unidades de medidas, obedece aos princípios ou lei que regem o fenômeno somente em situações controladas e laboratoriais, estando à resposta com uma maior distância daquela que seria a mais correta. Médio afastamento da resposta correta.</p>
<p>Nível: 3 – opção não-plausível</p>	<p>Resposta absurda, sem sentido para os conhecimentos científicos envolvidos, viola as leis e princípios físicos, mais de uma evidência em contrassenso, uma situação que jamais ocorreria em situações naturais, distante da resposta correta, sem situações semelhantes, ou seja, sem uma situação análoga comparativa coerente. Grande afastamento da resposta correta.</p>

5.3.2.2 Relação Condicional de Concepções entre os Itens

Rosenblatt & Heckler (2011) em seu artigo “*Systematic study of student understanding of the relationships between the directions of force, velocity, and acceleration in one dimension*”, ressaltam que a compreensão conceitual dos alunos envolve relações condicionais do tipo “dado x, y que é? (notação usada: $x \rightarrow y$), p. ex., dado uma força

resultante sobre um objeto, qual é a sua velocidade?”. Segundo os autores, estudos a partir de Viennot (1979) começaram a levar em consideração *relações condicionais*, porém destaca que até então **não houve nenhum estudo sistemático** sobre a compreensão de estudantes de todos as seis possíveis relações condicionais emparelhados entre os conceitos de força, velocidade e aceleração.

Um estudo sistemático de todos os pares possíveis de relações condicionais [...] permitirá uma visão mais holística sobre a compreensão dos alunos e na possibilidade de determinar o entendimento da forma como uma relação pode afetar (ou prever) o entendimento da outra relação condicional. (Rosenblatt & Heckler, 2011, p. 020112-2, *tradução livre*)

De modo semelhante, apesar da compreensão de que as concepções alternativas se encontram em um emaranhado de relações e associações entre conceitos, relativamente consistentes e com argumentações lógicas para o indivíduo, o levantamento realizado sugere ser desconhecido da mesma forma um estudo sistemático da rede de concepções alternativas para além de pares de relações condicionais.

A preocupação está em considerar todas as possibilidades dentro de uma rede de concepções que possa sustentar e justificar uma alternativa plausível para determinado item, entretanto, esse estudo somente é possível se for detectado uma “dependência local de itens”, que pode ser feito:

- De **modo direto e qualitativo** identificando as implicações de uma afirmação de uma questão (ou item) diante de outra questão (ou item) subsequente, em que esteja evidenciado de modo correspondente os mesmos traços latentes ou uma relação condicionada entre variáveis. O exemplo clássico disso está na proposição de um enunciado válido para duas ou mais questões (ou item). Este tipo de relação condicional é identificado *a priori* no texto dos itens, independentemente da forma como o aluno compreende;
- De **modo indireto e quantitativo** sendo identificado estatisticamente por um dos pressupostos da TRI em que o teste aplicado seja (1) unidimensional e que respeite o (2) teste de independência local dos itens. Este último poderá revelar indícios da existência de relações condicionais existentes entre variáveis. Este tipo de relação

condicional é identificado *a posteriori* no texto dos itens, levando em considerações o resultado de processos mentais dos alunos na sua compreensão.

Em uma perspectiva aproximada com as relações condicionais entre variáveis é o fato de que um teste diagnóstico sobre concepções pode apresentar numa rede de dependência ou condicionamento entre as concepções de diferentes itens, podendo assim revelar as mais comuns incidências de combinações ou relações condicionadas entre as concepções, trazendo, portanto, relevantes informações quanto a estrutura e organização de conceitos na cognição dos estudantes.

Obviamente poderão ser identificadas de relações condicionais pouco consistentes em que dificilmente poderá ser sustentado argumentos que os justifiquem para se manter, essa rede de concepções pouco plausíveis pode ser considerada um extremo absurdo por se configurar uma estrutura ilógica. Para avançar na compreensão dessas estruturas mentais de compreensão o desenvolvimento relacionado ao nível de distorção conceptual do item numa perspectiva de rede de dependência com outros itens, poderá mostrar as relações condicionais entre concepções alternativas em uma nova projeção, mais profunda e complexa, e assim identificar a relevância de conteúdos ou rede concepções tendo critério o nível de dificuldade dentro de um campo de proficiência para assim subsidiar melhores instruções para o ensino e aprendizagem

A opção pelo termo “conceptual” ao invés de “conceitual” é propositada no sentido de considerar não apenas a influência de um conceito, mas uma rede de conceitos que por sua vez caracteriza como uma concepção de valores e crenças sobre o conhecimento científico. A ideia de distorção conceitual se torna conveniente no sentido de que os conceitos podem ser interpretados de forma diferente da formalizada pelo conhecimento científico, e um dos mecanismos de ocorrência está em possíveis concepções textualizadas (no enunciado) e/ou mentalizadas (na mente do sujeito) diante os conceitos científicos, isso possui uma relação de causa de forma enfática com os distratores, que se fossem bem avaliados e identificados aqueles pouco plausíveis, o teste seria muito mais proeminente. Apesar disso, não é exclusivo de distratores, essas concepções também podem coexistir em alternativas corretas. No

caso, os distratores fornecem informações de possíveis *misconceptions* preexistentes, e uma associação entre elas pode revelar possíveis consistências de redes conceituais de erros, equívocos ou total falta de compreensão do conhecimento. Assim como uma competência científica pode indicar certo domínio diante de uma família de situações-problemas, um conjunto consistente de distratores para um conjunto de assertividades poderá indicar uma rede de condicionamento entre concepções, sendo um conjunto de equívocos familiares dentro de um contexto de situações-problemas concretos, específicos e bem definidos.

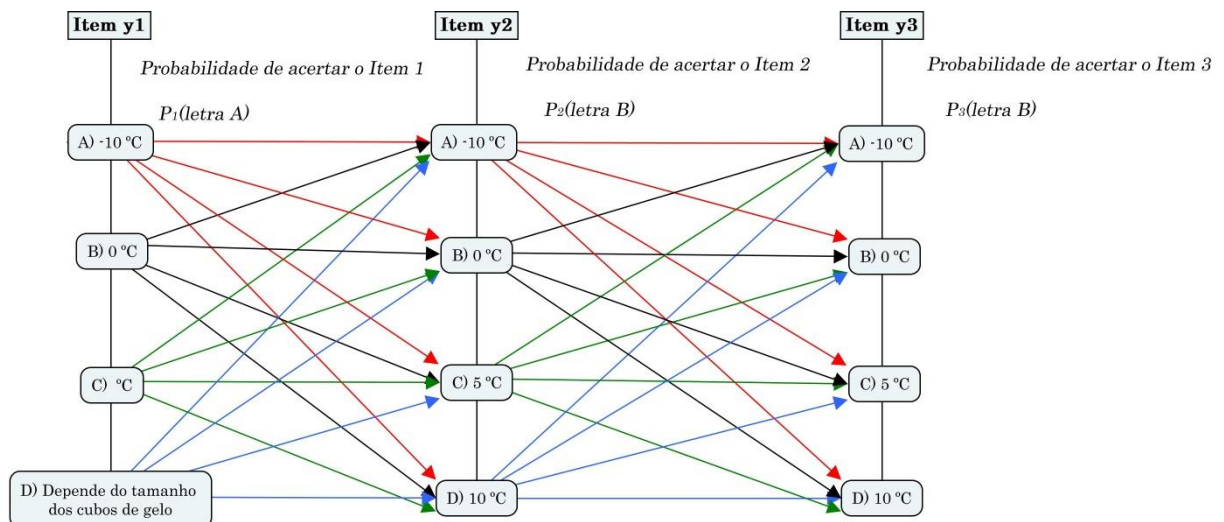
5.3.2.3 Explorando Novas Perspectivas na Interpretação de Dependência Local de Itens

A dependência local dos itens é analisada pela análise fatorial das possibilidades de acerto conjunto entre os itens. No caso da dependência entre os itens y_1, y_2 e y_3 no TCE, consideram-se os seguintes aspectos e definições:

Com base nos indícios de dependência local, pode-se dizer que a probabilidade de que haja essa dependência se define como:

$$p(\text{acerto do item 1}) * p(\text{acerto do item 2}) * p(\text{acerto do item 3}) = p(\text{letra A}) * p(\text{letra B}) * p(\text{letra B})$$

Sendo A a resposta do item y_1 , e B dos itens y_2 e y_3 . As combinações na sequência de respostas com base na lógica da análise fatorial seriam:



Na tentativa de compreender a existência de coerência e consistência na sequenciação dessas respostas, foi analisado e desenvolvido alguns termos que possibilitam uma interpretação clara e precisa, podendo ser útil no avanço de novas perspectivas investigativas sobre o instrumento e diante as evidências das respostas.

5.3.2.3.1 Coerência e Consistência na Interpretação das Alternativas dos Itens

A **coerência** é uma observação de alto nível relacionado a correspondência entre as respostas corretas dos itens, dessa forma não se preocupa com a compreensibilidade envolvida. Portanto, do ponto de vista epistemológico é fundamental investigar e observar em baixo nível a **consistência** existente de cada item, analisando se as possíveis explicações associadas as respostas estão de acordo com o conhecimento científico, assim como do alinhamento na relação de coerência entre as respostas dos itens dependentes.

Em casos extremos quanto as interpretações de dependência, pode-se dizer que **não há** nenhuma possibilidade de haver equívocos, ou a existência do erro ou equívoco é de ordem que dificilmente se encontra uma lógica, com isso, uma conexão na rede de concepções seria próximo a um **absurdo lógico** pela falta de consistência nas informações e do “conhecimento” envolvido, sendo para a grande maioria identificado como erro inconcebível. Para isso se estabeleceu preliminarmente as seguintes definições:

- **Não há** (dependência local): Quando a explicação supostamente não esteja associada (direta ou indiretamente por dependência local) a uma concepção alternativa, com isso, é um pensamento mais adequado ao conhecimento científico.
- **Absurdo** (lógica improvável): Quando a busca por uma justificativa se torna insustentável até para uma explicação baseada em concepções alternativas, ou seja, é um grande contrassenso.

Apesar de haver coerência e plausibilidade na resposta, um item com dependência local não garante que haja consistência de um pensamento adequado ao conhecimento científico, podendo haver compreensões inadequadas ocultas do sujeito devido à não complementariedade ou falta de detalhes no enunciado do item. Nesse

caso pode haver aspectos *Não Possíveis de Verificar* (NPV), onde o item não traz argumentos capazes de avaliar na íntegra o conhecimento científico envolvido no processo, com isso não garante uma **coerência consistente** no pensamento diante da resolução do problema para a tomada de decisão na escolha de uma alternativa mais plausível. A inconsistência não necessariamente advém de falhas ou não complementariedade no enunciado do item em questão, mas sim das explicações possíveis para justificá-lo levando em consideração a dependência local existente, ou seja, pode ser que haja uma correspondência relativamente plausível entre as respostas, com possibilidades da existência da ocorrência em uma situação física próxima, mas não equivalente. O fato é que ao assinalar uma resposta correta, a **dependência local** poderá ser:

- **Coerente** – *com dependência local a uma resposta anterior correta* (Relação entre alternativas corretas): Quando os indícios de dependência local do item possuem uma relação à resposta “correta” do item anterior (ou da cadeia de respostas corretas anteriores). Neste caso a alternativa independe de estar correta ou não, entretanto busca-se apresentar uma explicação associada plausível que não viole a logicidade e adequação ao conhecimento científico das respostas “corretas” anteriores. Para ser coerente é necessário que a resposta dada ao item (estando certa ou errada) esteja fundamentada com uma explicação adequada as questões anteriores. É preciso que haja coerência não apenas com o item anterior, mas também com o «anteanterior». Para analisar os itens errados tem-se a considerar sua consistência na compreensibilidade, na sua adequação ao conhecimento científico;
- **Incoerente** – *com dependência local a uma resposta anterior incorreta* (Relação entre distratores correlatos): Quando os indícios de dependência local do item não possuem uma relação à resposta “correta” do item anterior, ou seja, apresenta uma explicação plausível, mas que viola ou não corresponde à resposta “correta” anterior e sua dependência local é com uma alternativa anterior menos plausível;
- **Consistente** – *sem distorção conceitual causado seja pelo item, entre os itens e/ou pelo sujeito* (Enunciado correto, correspondência anterior sem ruídos, e/ou suposta decisão do sujeito de forma convicta e sustentada pelo conhecimento prévio cientificamente

correta ou modelo mental correto): Quando as explicações correspondentes entre os itens forem plausíveis, não sustentadas por concepções alternativas, não necessitando/possibilitando viabilizar argumentos equivocados para justificar a tomada de decisão pela alternativa correta, e, portanto, não violando as leis e princípios físicos ou distorcendo os conhecimentos científicos envolvidos;

- **Inconsistente** – *distorção conceitual seja pelo item, entre os itens e/ou pelo sujeito* (Enunciado construído com erros, equívocos, com correspondência equívocos anteriores, e/o suposta decisão do sujeito de forma convicta, porém, sustentada por um conhecimento prévio incorreto ou modelo mental incorreto): Quando o distrator não for plausível por envolver concepções alternativas textualizadas, ou carregar equívocos advindos da dependência local dos itens, ou ainda remeter a compreensão equivocada ou mal compreendidas pelo respondentes. Também podem ocorrer devido a existência de argumentos equivocados que *Não Podem ser Verificados* (NPV). A inconsistência pode ainda não envolver concepções alternativas, mas estar caracterizada como linha de pensamento sem lógica e absurda do sujeito, provavelmente pela falta de conhecimento, má interpretação do item, pelo “acerto aleatório”. A inconsistência pelo “erro ao acaso” envolve alunos proficientes que por descuido optam pelos distratores.

De modo geral, o condicionamento de dependência local entre os itens segue de uma perspectiva interpretativa plausível para as informações elaboradas entre os alternativas, partindo da alternativa correta para as distorções conceituais entre os distratores se tornando cada vez mais absurda a resposta.

Relação de Condicionamento na Dependência local			
Coerente		Incoerente	
Consistente	Inconsistente	Consistente	Inconsistente
CC	CI	IC	II
Plausível		Absurdo	

Para a realização das análises verificativas da relação de condicionamento na dependência local entre os itens, foram elaboradas algumas recomendações específicas:

- O fato de responder corretamente um item anterior não garante que opte pela

resposta mais plausível do item posterior, com isso, a incidência ou probabilidade de escolha posterior no que se refere a uma alternativa com maior nível de distorção conceitual é supostamente menor;

- Ainda que a resposta esteja errada, os argumentos e a coerência que a sustentam podem não estar diretamente relacionada a uma concepção alternativa.
- Entende-se que causas dos níveis de distorção conceitual do item podem estar associados a um: (1) *Caráter semântico* (causado pelo **item**) no enunciado e respostas apresentadas, e/ou; (2) *Caráter subjetivo* (causado pelo **sujeito**) ao analisar as possibilidades de compreensibilidade interpretativa (enunciado) e explicativa (respostas) supostas pelo sujeito, com isso pode-se identificar as redes conceituais que envolvem as concepções alternativas interligadas numa sequência não linear, com isso, para cada cadeia de combinações pode se caracterizar um determinado coletivo de pensamento do conhecimento intuitivo.
- O nível de distorção conceptual do item não está necessariamente associado única e exclusivamente a uma concepção equivocada preexistente na alternativa, ainda que saiba que a alternativa do item seja incorreta.
- Como **critérios avaliativos** para determinar os *níveis de distorção conceptual do item*, consideram-se fatores internos e externos que condicionam a relação de dependência local por incoerência e inconsistência, e se justificam nas seguintes condições:
 - (1) *Distorção simples por distanciamento* da resposta mais plausível na sua intensidade/magnitude (grandeza escalar), ou ainda pela direção e sentido quando mensurado por uma grandeza vetorial;
 - (2) *Distorção por incoerência* na dependência local, e;
 - (3) *Distorção por inconsistência* qualitativa na compreensibilidade que se sustenta diante do conhecimento científico (causado pelo item e/ou suposições sujeito).

A determinação do nível de distorção deve levar em consideração esses critérios, porém essa plausibilidade depende das interpretações na avaliação desses

critérios, pois quanto mais erros, equívocos e absurdos lógicos existem, mais afastado ou maior deve ser o nível de distorção, sendo o *nível zero*, para representar a alternativa mais plausível, porém isso não garante que não existam equívocos envolvidos. Essa tomada de decisão deve levar em conta uma comparação na quantidade das incidências de equívocos e erros entre as alternativas do item avaliado, sendo a quantidade de equívocos e a falta de lógica e rigor científico os aspectos determinantes. Assim, as combinações para o condicionamento de dependência local podem levar em consideração às três situações supracitadas. As combinações são:

CC	CI	IC	II
CC _{1,2}	CI _{1,2}	IC _{1,2}	II _{1,2}
CC _{1,3}	CI _{1,3}	IC _{1,3}	II _{1,3}
CC _{1,2,3}	CI _{1,2,3}	IC _{1,2,3}	II _{1,2,3}
CC ₂	CI ₂	IC ₂	II ₂
CC _{2,3}	CI _{2,3}	IC _{2,3}	II _{2,3}
CC ₃	CI ₃	IC ₃	II ₃

Diante disso, pode-se dizer ainda que:

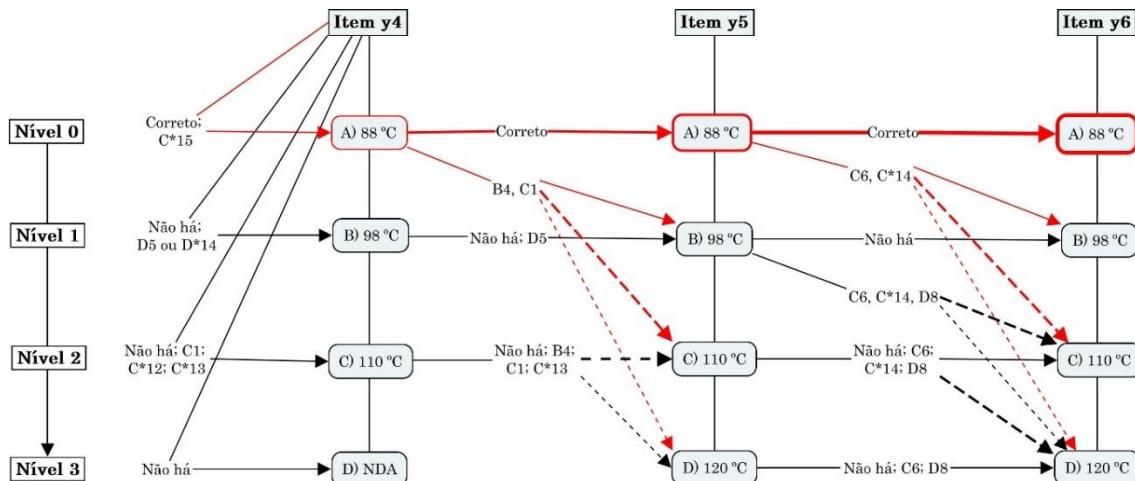
- Uma análise da rede de dependência local nas concepções alternativas envolvidas poderá explicitar e ajudar a compreender uma transferência de erros e conseqüentemente uma ampliação no nível de distorção conceitual para os itens posteriores dentro do grupo de dependência local, tendo em vista seu caráter acumulativo na compreensibilidade para a família de situações-problemas;
- Cada alternativa não necessariamente pode ser considerada uma concepção alternativa por uma análise direta, ou seja, sem levar em consideração as possibilidades de construção do conhecimento implícito, pois não se dá isoladamente por uma única situação-problema, mas sim através de uma rede de concepções multifacetada de combinações associadas. Dessa forma não se pode afirmar que está isenta de concepções equivocadas simplesmente por assinalar a alternativa correta, pois seria ignorar as possibilidades de carregar equívocos anteriores na qual o item está interligado;
- Cada alternativa do item pode ter mais de uma concepção alternativa, incluído

uma explicação que não se caracterize como concepção alternativa;

- A **rede de concepções** com base no condicionamento de dependência local se constrói realizando uma *análise fatorial qualitativa e interpretativa da lógica e consistência existente entre as respostas*, com isso, os níveis de plausibilidade dessas conexões podem ser diferenciados e definidos como:

————>	Provável (P)
----->	Pouco provável (PP)
----->	Muito pouco provável (MPP)

Por exemplo, foi considerado indícios de dependência local para os itens y4, y5 e y6, com isso a análise de rede de concepções resultou na elaboração do seguinte esquema:



Pode-se observar que os níveis de distorção foram fixados na vertical e que algumas conexões são menos plausíveis de serem seguidas em um processo sequencial de respostas do que outras. Na prática, essa sequência nunca é garantida, apesar de ser esperada (o aluno segue respondendo na ordem sequencial dos itens), a não ser que o procedimento seja o de testes adaptativos computadorizados (*“Computerized Adaptive Testing - CAT”*).

Na análise qualitativa de rede concepções surgiram algumas que foram identificados na análise do teste TCE que não estão previstas no inventário do teste. Esse tipo de análise ainda se encontra em processo de investigação, e não está exaustiva. Apesar disso, as concepções identificadas foram destacadas com asterisco, e momentaneamente, podem-se ser destacadas:

Quadro 72: Identificação complementar para o inventário de concepções do TCE.

C - Concepções dos alunos sobre a transferência de calor e mudança de temperatura	
C*10	• Um corpo “frio” tem mais influência que um corpo “quente” quando postos em contato, ou seja, sua temperatura tende a se equilibrar com a temperatura do corpo “frio”.
C*11	• A temperatura final de equilíbrio é igual à média aritmética das temperaturas dos corpos
C*12	• Quanto maior é o ponto de ebulição, mais rápida a água ferve.
C*13	• O aquecimento eleva o ponto de ebulição, da mesma forma o resfriamento abaixa o ponto de solidificação, ou seja, não há pontos fixos para a mudança de fase.
C*14	• No aquecimento, após cada mudança de fase a substância alcança uma temperatura mais elevada. (Substitui as concepções alternativas D7 e D8)
C*15	• A água ferve rapidamente quando em contato com uma superfície superaquecida, ou seja, com temperatura muito superior ao seu ponto de ebulição. (equivoco quanto ao processo de calefação)
C*16	• O aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura. (equivoco sobre o ponto de fusão e complementar a C1 e coincide em parte com D6)
C*17	• O processo de mudança de fase ocorre de forma instantânea e não paulatinamente em fluidos, desconsiderando a forma de condução de calor e de transferência de massa (equivoco sobre o processo de convecção térmica). A variação de temperatura ocorre de forma imediata.
D - Concepções dos alunos sobre "propriedades térmicas de materiais".	
D*13	• As CNTP são condições de validade geral para o comportamento da água, em especial (Substitui e generaliza a concepção alternativa D5).
D*14	• A água pode ser abaixo de 0 °C.
D*15	• O gelo pode assumir uma temperatura acima de 0 °C.
D*16	• O gelo sempre esfria outro corpo, ele jamais aquece.
D*17	• Ao ser aquecido o corpo fica mais frio.
E* - Concepções dos alunos sobre o "funcionamento de artefatos da produção moderna tecnológica".	
E*1	• O tempo na qual um corpo está submetido às máquinas térmicas (ou frigoríficas) não influencia sobre o seu comportamento térmico.
E*2	• Desconhece-se o processo de funcionamento das máquinas caloríficas (ou frigoríficas) e dos dispositivos que possam influenciar sobre o comportamento térmico de um corpo em suas proximidades e/ou em seu interior.

Essas são concepções que envolvem o teste mas que não estão sendo consideradas na análise. Isso mostra a importância de realizar esse tipo de análise, revelando aspectos ocultos, sendo uma das vantagens ao revelar aspectos necessários a contribuir quanto a validade de construto. A precisão do instrumento tende a ser favorecido quando aos critérios referenciados são de interesse ao avaliador.

No sentido de ilustrar o processo de análise de rede conceptual nos critérios de condicionamento de dependência local entre itens, considera-se o fator comum entre os itens y1, y2 e y3 sendo uma sequência de situações interligadas. Sabe-se que os três envolvem o conceito de *equilíbrio térmico* e tratam de uma mesma substância (“água”). O item y2 pede a temperatura de equilíbrio da “água” e o item 3 a do “gelo” em uma

situação semelhante. A diferença é que no item y1 a situação física no *status quo* não envolve trocas de calor nem mudança de fase, enquanto no item y2 e y3 envolve.

Quadro 73: Análise parcial de interpretação implícita dos níveis de distorção conceptual e na relação de condicionamento na dependência local dos itens y2 e y3.

Item y1:

1A - Quem acredita que os cubos de gelo se encontram a -10 °C (letra A) no congelador da geladeira, provavelmente detêm o conhecimento sobre:

(concepção alternativa: Não há - OPÇÃO CORRETA; E*1; E*2; C*17 - NPV) - (nível: 0)

- O processo de resfriamento, na qual a substância permanece na mesma fase com a variando temperatura, ou seja, neste caso o gelo iria atingir temperaturas inferiores a 0 °C. O gelo somente irá parar de esfriar quando atingir a temperatura final de equilíbrio térmico, onde assim serão cessadas as trocas de calor entre os corpos. É preciso reconhecer dentre as propriedades térmicas da água e do gelo como: o ponto de solidificação da água, ou condição para a existência de gelo próximo ao CNTP, no caso 0 °C; capacidade de atingir temperaturas inferiores a 0 °C até o equilíbrio térmico na situação física atual, em condições próximas ao CNTP. Apesar de admitir que o mais provável fosse gelo com temperatura abaixo de 0 °C, ou seja, a única alternativa plausível seria -10 °C (letra A).

(concepção alternativa: Não há - OPÇÃO CORRETA)

- Mas é preciso levar em conta sobre um conhecimento mínimo sobre o funcionamento da máquina frigorífica, ou seja, o processo de trocas de calor nas condições ambientais típicas de um congelador de geladeira sobre os cubos de gelo no decorrer do tempo. Como o equilíbrio térmico depende de uma compreensão sobre o funcionamento de uma máquina frigorífica (ou térmica), é preciso que saiba que um congelador dispõe de um *termostato* para controlar a temperatura do ambiente em seu interior. Dessa forma, o *termostato* poderá ser determinante para compreendê-la numa temperatura final de equilíbrio térmico em *determinado momento*. Esses obstáculos para a compreensão fazem parte de *concepções não de natureza cientificamente conceitual, mas de natureza tecnológica*. Marcel Thouin (2004, pp. 112-115, 120) apresenta concepções de cunho tecnológico para as *ciências físicas, ciências da terra e do espaço, e ciências biológicas*. Neste caso envolve uma máquina frigorífica, com isso, o conhecimento sobre seu típico funcionamento pode ser um fator a influenciar e/ou interagir com outras concepções conceituais equivocadas.

Dessa forma podem-se considerar *concepções tecnológicas* associadas ao congelador ou sobre o **"funcionamento de artefatos da produção moderna tecnológica"**. Neste caso temos:

1. *Tempo na qual um corpo está submetido às máquinas térmicas (ou frigoríficas) não influencia sobre o seu comportamento térmico.* **(concepção alternativa: E*1)**

2. *Desconhece-se o processo de funcionamento das máquinas caloríficas (ou frigoríficas) e dos dispositivos que possam influenciar sobre o comportamento térmico de um corpo em suas proximidades e/ou em seu interior.* **(concepção alternativa: E*2)**

(concepção alternativa: E*1; E*2)

- É implícita a possibilidade errônea de acreditar que o processo de mudança de fase ocorra de forma imediata, instantânea e não paulatinamente ao atingir seus pontos de mudança (solidificação/fusão; condensação/vaporização). Sabe-se que tanto no aquecimento quanto no resfriamento os fluidos conduzem calor e transferem massa por meios de células de convecção ou de Bénard, onde o processo não se dá de forma imediata. Com isso, pode estar implícita a ideia de que o gelo congelou de forma praticamente instantânea e por completo iniciando a variação de temperatura imediatamente, diminuindo até cerca de -10 °C, e se mantendo. Esta é uma suposição que *Não Pode ser Verificada* (NPV) de forma direta, sendo equivalente e bem menos provável para o processo de condensação do vapor d'água, mas que também não podem ser desconsideradas. **(concepção alternativa: C*17 - NPV)**

1B - Quem acredita que os cubos de gelo se encontram a 0 °C (letra B) no congelador da geladeira, provavelmente poderá estar sustentando essa concepção baseado em que:

(concepção alternativa: Não há; D6; C6; E*1; E*2) - (nível: 1)

- A água acabou de solidificar ou o congelador mantém o gelo a uma temperatura a 0°C ou sultimente abaixo. Esta seria uma alternativa plausível. **(concepção alternativa: Não há)**
- O gelo em contato com o ambiente do congelador não tende a uma temperatura de equilíbrio, ou seja, por mais que o ambiente esteja a uma temperatura inferior a 0°C , a temperatura do gelo irá se manter a 0°C . Com isso, se nega duas leis físicas: como não há equilíbrio térmico, *se nega a lei zero da termodinâmica*; e não havendo trocas de calor entre corpos com diferentes temperaturas *se nega a lei geral das trocas de calor*. **(concepção alternativa: D6; C6);**
- Isso também pode implicar concepções tecnológicas sobre o funcionamento das máquinas frigoríficas e de seus efeitos sobre corpos em seu interior ou em suas proximidades.
 1. *Tempo na qual um corpo está submetido às máquinas térmicas (ou frigoríficas) não influencia sobre o seu comportamento térmico.* **(concepção alternativa: E*1)**
 2. *Desconhece-se o processo de funcionamento das máquinas caloríficas (ou frigoríficas) e dos dispositivos que possam influenciar sobre o comportamento térmico de um corpo em suas proximidades e/ou em seu interior.* **(concepção alternativa: E*2)**

1C - Quem acredita que os cubos de gelo se encontram a 5°C (letra C) no congelador da geladeira, provavelmente poderá estar sustentando essa concepção baseado em que:

(concepção alternativa: D*15) - (nível: 3)

- O ponto de solidificação da água é acima de 0°C nas condições ambientais de um congelador de geladeira. É um absurdo nessas condições, tanto do ponto de vista sobre as propriedades térmicas da água quanto do funcionamento de um congelador. Apesar de pouco provável, não se pode ignorar a possibilidade na crença que *o gelo pode assumir em condições próximas ao CNTP uma temperatura acima de 0°C* , levando em consideração o gelo em condições especiais onde o gelo fique sujeito a um *aumento da pressão*. Dificilmente um aluno iria escolher essa opção conscientemente nestas condições, e muito menos em situação próxima as CNTP. **(concepção alternativa: D*15) - (nível: 3)**

1D - Quem acredita que a temperatura mais provável dos cubos de gelo se encontra no congelador da geladeira depende de seu tamanho (letra D), provavelmente poderá estar sustentando essa concepção baseado em que:

(concepção alternativa: B7; E*1; E*2) - (nível: 2)

- A massa do gelo é um fator que influencia no processo de trocas de calor e como consequência a variação da temperatura do corpo no decorrer do tempo, *mas não é responsável para estabilizar e determinar a temperatura em si do corpo*. Levando em consideração que o tamanho (dimensões) geralmente está relacionado com a massa (não necessariamente), apesar do comportamento anômalo da água em seu resfriamento a partir de 4°C , pode-se dizer que a massa é uma variável a influenciar no processo das trocas de calor sendo inversamente proporcional a temperatura final de equilíbrio, mas a determinação da temperatura nesse caso somente seria válida durante o resfriamento antes de alcançar o equilíbrio, e *dependeria da potência da fonte na retirada de calor e do tempo*, caso contrário não faria sentido. Analisando do ponto de vista científico, o conceito de *capacidade térmica* também se pode demonstrar que a massa influencia diretamente na variação de temperatura durante as trocas de calor, mas de um jeito ou de outro, o “tamanho dos cubos de gelo” ou a “massa” somente seria determinante na determinação da temperatura se fosse considerado a potência e o tempo durante o processo de resfriamento, onde um sensor de temperatura teria a finalidade de estabilizar a temperatura ambiente em determinado momento. Como esta alternativa não expressa uma temperatura específica do gelo, como consequência disso, pode se admitir outras concepções implícitas que estejam coexistindo, não apenas de natureza científica para o corpo, mas também *tecnológicas* relacionadas ao funcionamento do congelador:
 - *O tempo na qual um corpo está submetido às máquinas térmicas (ou frigoríficas) não influencia sobre o seu comportamento térmico.*
 - *Desconhece-se o processo de funcionamento das máquinas caloríficas (ou frigoríficas) e dos dispositivos que possam influenciar sobre o comportamento térmico de um corpo em suas proximidades ou em seu interior.* **(concepção alternativa: B7; E*1; E*2)**

Item y2:

2A – Quem acredita que os cubos de gelo se encontram a $-10\text{ }^{\circ}\text{C}$ (letra A) no congelador da geladeira:

(dependência da letra (A) da questão 1 - COERENTE):

a) Após se misturar e se agitar a água dentro de um copo com 4 cubos de gelo a $-10\text{ }^{\circ}\text{C}$, os cubos param de derreter e restam pequenos pedaços, nesse momento a temperatura da água é de $-10\text{ }^{\circ}\text{C}$ (letra A). Supostamente a água preexistente dentro do copo está numa temperatura ambiente de $27\text{ }^{\circ}\text{C}$, pressão de 1 atm, ao nível do mar, latitude de 45° (em CNTP), temperatura essa que irá possibilitar aquecer o gelo e derreter, juntamente com o atrito produzido no copo com a agitação. Pode-se admitir erroneamente que *o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equívoco sobre o ponto de fusão)*, isso justifica a ideia de que o gelo derreta a $-10\text{ }^{\circ}\text{C}$.

Admitir que *o corpo “frio” influencia mais que um corpo “quente” quando postos em contato* justifica a ideia que se tenha mais a influência de esfriar a água do que a água aquecer o gelo, com isso a água iria tender a equilibrar com a temperatura do gelo.

Ainda que seja possível gelo a $-10\text{ }^{\circ}\text{C}$, essa afirmação não respeitaria as leis e princípios físicos, pois apresentam propriedades térmicas irreais tanto para a água líquida (obtendo temperatura inferior a $0\text{ }^{\circ}\text{C}$) quanto ao gelo (derretendo a $-10\text{ }^{\circ}\text{C}$), não correspondendo nem a condições próximas a CNTP, portanto é contrassenso, um absurdo. Apesar disso se enquadra em uma concepção de que *a água não se encontra a $0\text{ }^{\circ}\text{C}$* , o que torna praticamente improvável que este item avalie esse tipo de concepção alternativa ao considerar $-10\text{ }^{\circ}\text{C}$ para a água, pois é bem mais plausível que o aluno imagine a água entre $0\text{ }^{\circ}\text{C}$ e $100\text{ }^{\circ}\text{C}$ e não abaixo de $0\text{ }^{\circ}\text{C}$, ou seja, gelo: $\leq 0\text{ }^{\circ}\text{C}$ e $0\text{ }^{\circ}\text{C} \leq \text{água} \leq 100\text{ }^{\circ}\text{C}$. Apesar de parecer pouco provável, pode não se pode ignorar a possibilidade da existência de um concepção em que se acredita na *existência de água a uma temperatura inferior a $0\text{ }^{\circ}\text{C}$* , e isso somente seria possível se a água estivesse a baixa pressão em condições laboratoriais.

O maior nível de distorção conceitual do item estaria associado ao: (1) **distanciamento da resposta mais coerente**; (2) pela sua **incoerência na dependência local**; (3) e pela sua **inconsistência na compreensibilidade diante do conhecimento científico** (explicação mais detalhada abaixo).

Distorções: Neste caso, (1) se distancia da alternativa mais plausível que seria $0\text{ }^{\circ}\text{C}$ e seria (3) inconsistente teoricamente por admitir a água numa temperatura de $-10\text{ }^{\circ}\text{C}$ juntamente com o gelo em parte derretido, ou estar sendo sustentada implicitamente por concepções alternativas. É coerente na dependência local porque parte de uma concepção correta relacionada à questão anterior, na qual é sua continuação.

(concepção alternativa: C*16; C*10; D*14) - (DLCI_{1,3}) - (nível: 3)

b) Após se misturar e se agitar a água dentro de um copo com 4 cubos de gelo a $-10\text{ }^{\circ}\text{C}$, os cubos param de derreter e restam pequenos pedaços, nesse momento a temperatura da água é de $0\text{ }^{\circ}\text{C}$ (letra B).

O gelo irá variar de temperatura no processo de aquecimento de $-10\text{ }^{\circ}\text{C}$ a $0\text{ }^{\circ}\text{C}$, e que em seguida irá começar a derreter em seu ponto de fusão a $0\text{ }^{\circ}\text{C}$, enquanto isso a água fornece calor e segue esfriando até $0\text{ }^{\circ}\text{C}$ (letra B) permanecendo juntamente com o gelo em parte derretido. Portanto, esse processo seria *totalmente admissível para uma provável temperatura da água a $0\text{ }^{\circ}\text{C}$* (letra B).

O gelo irá derreter no processo de aquecimento de $-10\text{ }^{\circ}\text{C}$ a $0\text{ }^{\circ}\text{C}$, seria admitir o equívoco que ocorre variação de temperatura durante a mudança de fase (fusão do gelo) até atingir água a $0\text{ }^{\circ}\text{C}$ com pequenos pedaços de gelo não derretidos, ou seja, *o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura*, independente se há ou não mudança de fase. Este tipo de afirmação seria uma interpretação implícita, uma suposição *Não Verificável de forma direta (NPV)*. Outra concepção alternativa associada seria acreditar que *o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equívoco sobre o ponto de fusão, onde a configuração do fenômeno seria gelo derretendo a $-10\text{ }^{\circ}\text{C}$ resultando em água a $0\text{ }^{\circ}\text{C}$* .

Relacionada ao equívoco sobre o ponto de fusão, poderá haver implicações a outra concepção errônea na qual a *coexistência de água e gelo a temperaturas diferentes não tendem a estar numa mesma temperatura, seja por contato entre diferentes corpos ou por um dos corpos ser resultado pela mudança de fase do outro.*

Distorção: (3); **(concepção alternativa: Não há - OPÇÃO CORRETA; C1 - NPV; C6; C*16) - (DLCI₃) - (nível: 0)**

c) Após se misturar e se agitar a água dentro de um copo com 4 cubos de gelo a -10 °C, os cubos param de derreter e restam pequenos pedaços, nesse momento a temperatura da água é de 5 °C (letra C). Acreditar que o gelo derreta em parte e a temperatura da água esteja a 5 °C seria um contrassenso, pois não poderia nessas condições existir ainda pedaços de gelo, muito menos a 10 °C (letra D). A única possibilidade de justificar “em parte” esse absurdo seria através de uma concepção alternativa na qual se consideraria válido o *gelo sendo aquecido pela água variando de temperatura até 5°C*, e dessa forma coexistir água também a 5 °C, ainda assim seria um grande contrassenso, um conhecimento implícito supostamente pouquíssimo provável de sua existência e mais improvável se considerar que aconteça isso aos 10 °C.

A temperatura do gelo supostamente varia de -10 °C a 5 °C mudando de fase, nesse caso, o *aquecimento* (ou resfriamento) *sempre resulta num(a) aumento* (ou diminuição) *de temperatura*, e seria possível no caso do gelo estar sobre uma pressão acima de 1 atm, ou seja, em condições não convencionais a situação proposta.

Distorções: (1); (3). **(concepção alternativa: D*15; C1) - (DLCI_{1,3}) - (nível: 1) e (nível: 2)**

2B - Quem acredita que os cubos de gelo se encontram a 0 °C (letra B) no congelador da geladeira:

(dependência da letra (B) da questão 1 - INCOERENTE)

a) Após se misturar e se agitar a água dentro de um copo com 4 cubos de gelo a 0 °C, os cubos param de derreter e restam pequenos pedaços, nesse momento a temperatura da água é de -10 °C (letra A).

Isto seria admitir que o *gelo esfria incessantemente a água diminuindo continuamente sua temperatura* até abaixo da sua própria, como se a esfriasse normalmente até 0 °C. **Concepção alternativa: se admite que a água obtenha temperatura inferior a 0 °C**

A continuação da diminuição da temperatura até -10 °C mesmo estando mais quente que a água. **Concepção alternativa: o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura.**

Associada a concepção anterior, acredita-se que *o gelo sempre esfria outro corpo, ele jamais aquece.*

Distorções: (1); (2); (3). **(concepção alternativa: D*14; C1; D*16) - (DLII_{1,2,3}) - (nível: 3)**

b) Após se misturar e se agitar a água dentro de um copo com 4 cubos de gelo a 0 °C, os cubos param de derreter e restam pequenos pedaços, nesse momento a temperatura da água é de 0 °C (letra B). Isto seria um processo admissível, portanto *não há diretamente uma concepção errônea*. O fato é que o congelador atinge temperaturas inferiores a 0 °C, com isso essa alternativa seria plausível considerando-se que a água tenha acabado de solidificar ou o congelador mantenha o gelo a uma temperatura a 0 °C ou sutilmente abaixo, dessa forma *não teria concepção alternativa relacionada, a explicação estaria plenamente correta.*

Entretanto, pode haver uma concepção na qual o *gelo somente pode estar a 0 °C*, ou que ele em contato com o ambiente interno do congelador não tenda a um equilíbrio térmico.

Admitir que o corpo “frio” influencia mais que um corpo “quente” quando postos em contato justifica a ideia que se tenha mais a influência de esfriar a água do que a água aquecer o gelo, com isso a água iria tender a equilibrar com a temperatura do gelo.

A distorção estaria ligada a (2) incoerência na dependência local com a resposta correta, e na (3) inconsistência teórica por evidências de concepções alternativas associadas. **(concepção alternativa: Não há; D6; C*10) - (DLII_{2,3}) (nível: 0)**

c) Após se misturar e se agitar dentro de um copo com 4 cubos de gelo a 0 °C, os cubos param de derreter e restam pequenos pedaços, nesse momento a temperatura da água é de 5 °C (letra A). Acreditando que a água estaria a 5 °C (letra C) ou 10 °C (letra D) seria incoerente com a presença de pedaços de gelo restantes a 0 °C. Essa ideia pode se sustentar equivocadamente com o fato de que a *água não pode existir a 0 °C.*

Associada a concepção anterior de que os *corpos não tendem a um equilíbrio térmico*.
Supondo que o gelo aqueça e aumente sua temperatura atinja 5 °C ou 10 °C se admite que o *aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura*.
Associado com a concepção anterior, admite-se que o gelo pode assumir uma temperatura acima de 0 °C.
A distorção estaria relacionada ao (1) distanciamento do valor da alternativa correta, na (2) incoerência da correspondência com a dependência local com a resposta correta e (3) inconsistência teórica. **(concepção alternativa: C6; D7; C1; D*15) - (DLII_{1,2,3}) - (nível: 1) e (nível: 2)**

2C - Quem acredita que os cubos de gelo se encontram a 5 °C (letra C) no congelador da geladeira:

(dependência da letra (C) da questão 1 - INCOERENTE)

Já parte de uma concepção pressuposta pouco provável, portanto, todas as alternativas são ligações com pouco sentido ligadas a uma rede de conhecimentos inconsistentes.

a) Seria praticamente inadmissível imaginar nessas condições a água a -10 °C (letra A) estando em contato inicialmente gelo a 5 °C. O absurdo é tanto pelas propriedades da água quanto do gelo, pois necessitaria da combinação quatro concepções pouco improváveis: *que a água pode estar em temperatura inferior a 0 °C; que o gelo pode estar acima de 0 °C; que ambos não tendem um equilíbrio térmico; e que o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão)*. Distorções: (1); (2); (3). **(concepção alternativa: D*14; D*15; C6; C*16) - (DLII_{1,2,3}) - (nível: 3)**

b) Resultando água a 0 °C (letra B) é algo admissível, mas, não para uma mistura com gelo a 5 °C inicialmente. A não ser que se considere a combinação pouco provável de três concepções alternativas: *a de que o gelo possa estar a 5 °C; que ambos não tendem a um equilíbrio térmico; e que o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão)*. Distorções: (1); (2); (3). **(concepção alternativa: D*15; C6; C*16) - (DLII_{1,2,3}) - (nível: 0)**

c) Estando a água a 5 °C (letra C) tem-se somente na combinação das duas concepções: *de que o gelo possa estar a 5 °C juntamente com a água; e que o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão)*. Distorções: (1); (2); (3). **(concepção alternativa: D*15; C*16) - (DLII_{1,2,3}) - (nível: 1)**

d) Estando a água a 10 °C (letra D) em última análise, bem menos provável, na combinação de três concepções: *que o gelo pode estar acima de 0 °C; que ambos não tendem um equilíbrio térmico; e que o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão)*. Distorções: (1); (2); (3). **(concepção alternativa: C6; C*16; D*15) - (DLII_{1,2,3}) - (nível: 2)**

2D - Quem acredita que os cubos de gelo se encontram a 10 °C (letra D) no congelador da geladeira:

(dependência da letra (D) da questão 1 - INCOERENTE)

a) Seria praticamente inadmissível imaginar nessas condições a água a -10 °C (letra A) estando em contato inicialmente gelo a 10 °C. O absurdo é tanto pelas propriedades da água quanto do gelo, pois necessitaria da combinação quatro concepções pouco improváveis: *que a água pode estar em temperatura inferior a 0 °C; que o gelo pode estar acima de 0 °C; que ambos não tendem um equilíbrio térmico; e que o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão)*. Distorções: (1); (2); (3). **(concepção alternativa: D*14; D*15; C6; C*16) - (DLII_{1,2,3}) - (nível: 3)**

b) Resultando água a 0 °C (letra B) é algo admissível, mas, não para uma mistura com gelo a 10 °C inicialmente. A não ser que se considere a combinação pouco provável de três concepções alternativas: *a de que o gelo possa estar acima de 0 °C; que ambos não tendem a um equilíbrio térmico; e que o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão)*. Distorções: (1); (2); (3). **(concepção alternativa: D*15; C6; C*16) - (DLII_{1,2,3}) - (nível: 0)**

c) Estando a água a 5 °C (letra C) somente na combinação das três concepções: a *de que o gelo possa estar acima de 0 °C; que ambos não tendem a um equilíbrio térmico; e que o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão)*. Distorções: (1); (2); (3). **(concepção alternativa: D*15; C6; C*16) - (DLII 1,2,3) - (nível: 1)**

d) Estando a água a 10 °C (letra D) em última análise, bem menos provável, na combinação de duas concepções: *que o gelo pode estar acima de 0 °C; e que o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão)*. Distorções: (1); (2); (3). **(concepção alternativa: D*15; C*16) - (DLII 1,2,3) - (nível: 2)**

Item y3 (Parcial - 1A):

1A - Quem acredita que os cubos de gelo se encontram a -10 °C (letra A) no congelador da geladeira:

(dependência da letra (A) da questão 1):

(concepção alternativa: Não há - OPÇÃO CORRETA; E*1; E*2; C*17 - NPV) - (nível: 0)

2a - E que a temperatura da água é -10 °C (letra A) numa poça junto com pequenos cubos de gelo derretidos:

(dependência da letra (A) da questão 2)

(concepção alternativa: C*16; C*10; D*14) - (DLCI_{1,3}) - (nível: 3)

a) Ao admitir que os cubos de gelos restantes estejam a -10 °C (letra A) após derreter estando inicialmente a -10 °C e formado água a -10 °C se acredita que eles derretam com a temperatura ambiente em cima do balcão, mas não altera sua temperatura. Duas concepções: *o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão); se admite que a água obtenha temperatura inferior a 0 °C*. O absurdo é tanto pelas propriedades térmicas da água quanto do gelo, portanto essa combinação é pouquíssimo provável. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: C*16; D*14) - (DLII 1,2,3) - (nível: 3)**

b) Ao admitir que os cubos de gelos restantes estejam a 0 °C (letra B) após derreter estando inicialmente a -10 °C e formado água a -10 °C seriam admitir o absurdo que quando o gelo derrete ele forma água mais fria. Duas concepções: *a água pode estar em temperatura inferior a 0 °C; que ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (2); (3). **(concepção alternativa: D*14; C*16) - (DLII 2,3) - (nível: 0)**

c) Ao admitir que os cubos de gelos restantes estejam a 5 °C (letra C) após derreter estando inicialmente a -10 °C e formando água a -10 °C. O gelo variou de -10 °C a 5 °C mudando de fase. Cinco concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; a água pode estar em temperatura inferior a 0 °C; que ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; D*14; C6) - (DLII 1,2,3) - (nível: 1)**

d) Ao admitir que os cubos de gelos restantes estejam a 10 °C (letra D) após derreter estando inicialmente a -10 °C e formando água a -10 °C. O gelo variou de -10 °C a 10 °C mudando de fase. Quatro concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; a água pode estar em temperatura inferior a 0 °C; que ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; D*14; C6) - (DLII 1,2,3) - (nível: 2)**

1A - Quem acredita que os cubos de gelo se encontram a -10 °C (letra A) no congelador da geladeira:

(dependência da letra (A) da questão 1)

(concepção alternativa: Não há - OPÇÃO CORRETA; E*1; E*2; C*17 - NPV) - (nível: 0)

2b – E que a temperatura da água é 0 °C (letra B) numa poça junto com pequenos cubos de gelo derretidos:

(dependência da letra (B) da questão 2)

(concepção alternativa: Não há – OPÇÃO CORRETA; C1 – NPV; C6; C*16) - (DLCI3) - (nível: 0)

a) Ao admitir que os cubos de gelos restantes estejam a -10 °C (letra A) após derreter estando inicialmente a -10 °C e formado água a 0 °C. Como o gelo estava a -10 °C e não muda de temperatura. Duas concepções: *o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão); que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: C*16; C6) - (DLII 1,2,3) - (nível: 3)**

b) Ao admitir que os cubos de gelos restantes estejam a 0 °C (letra B) após derreter estando inicialmente a -10 °C e formado água a 0 °C. Sem distorções (questão 2). **(concepção alternativa: Não há) - (DLCC) - (nível: 0)**

c) Ao admitir que os cubos de gelos restantes estejam a 5 °C (letra C) após derreter estando inicialmente a -10 °C e formado água a 0 °C. O gelo variou de -10 °C a 5 °C mudando de fase. Três concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII 1,2,3) - (nível: 1)**

d) Ao admitir que os cubos de gelos restantes estejam a 10 °C (letra D) após derreter estando inicialmente a -10 °C e formado água a 0 °C. O gelo variou de -10 °C a 10 °C mudando de fase. Três concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII 1,2,3) - (nível: 2)**

1A – Quem acredita que os cubos de gelo se encontram a -10 °C (letra A) no congelador da geladeira:

(dependência da letra (A) da questão 1)

(concepção alternativa: Não há – OPÇÃO CORRETA; E*1; E*2; C*17 – NPV) - (nível: 0)

2c – E que a temperatura da água é 5 °C (letra C) numa poça junto com pequenos cubos de gelo derretidos:

(dependência da letra (C) da questão 2)

(concepção alternativa: D*15; C1) - (DLCI_{1,3}) - (nível: 1)

a) Ao admitir que os cubos de gelos restantes estejam a -10 °C (letra A) após derreter estando inicialmente a -10 °C e formado água a 5 °C. Como o gelo estava a -10 °C e não muda de temperatura. Duas concepções: *o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão); que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: C*16; C6) - (DLII 1,2,3) - (nível: 3)**

b) Ao admitir que os cubos de gelos restantes estejam a 0 °C (letra B) após derreter estando inicialmente a -10 °C e formado água a 5 °C. Uma concepção: *que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (2); (3). **(concepção alternativa: C6) - (DLII 2,3) - (nível: 0)**

c) Ao admitir que os cubos de gelos restantes estejam a 5 °C (letra C) após derreter estando inicialmente a -10 °C e formado água a 5 °C. O gelo variou de -10 °C a 5 °C mudando de fase. Três concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII 1,2,3) - (nível: 1)**

d) Ao admitir que os cubos de gelos restantes estejam a 10 °C (letra D) após derreter estando inicialmente a -10 °C e formado água a 5 °C. O gelo variou de -10 °C a 10 °C mudando de fase. Três concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII 1,2,3) - (nível: 2)**

1A – Quem acredita que os cubos de gelo se encontram a -10 °C (letra A) no congelador da geladeira:

(dependência da letra (A) da questão 1)

(concepção alternativa: Não há – OPÇÃO CORRETA; E*1; E*2; C*17 – NPV) - (nível: 0)

2d – E que a temperatura da água é 10 °C (letra D) numa poça junto com pequenos cubos de gelo derretidos:

(dependência da letra (D) da questão 2)

(concepção alternativa: D*15; C1) - (DLCl_{1,3}) - (nível: 2)

a) Ao admitir que os cubos de gelos restantes estejam a -10 °C (letra A) após derreter estando inicialmente a -10 °C e formado água a 10 °C. Como o gelo estava a -10 °C e não muda de temperatura. Concepções: *o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão); que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: C*16; C6) - (DLII_{1,2,3}) - (nível: 3)**

b) Ao admitir que os cubos de gelos restantes estejam a 0 °C (letra B) após derreter estando inicialmente a -10 °C e formado água a 10 °C. Uma concepção: *que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (2); (3). **(concepção alternativa: C6) - (DLII_{2,3}) - (nível: 0)**

c) Ao admitir que os cubos de gelos restantes estejam a 5 °C (letra C) após derreter estando inicialmente a -10 °C estando inicialmente a -10 °C e formado água a 10 °C. O gelo variou de -10 °C a 5 °C mudando de fase. Três concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII_{1,2,3}) - (nível: 1)**

d) Ao admitir que os cubos de gelos restantes estejam a 10 °C (letra D) após derreter estando inicialmente a -10 °C estando inicialmente a -10 °C e formado água a 10 °C. O gelo variou de -10 °C a 10 °C mudando de fase. Três concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII_{1,2,3}) - (nível: 2)**

Item y3 (Parcial – 1B):

1B – Quem acredita que os cubos de gelo se encontram a 0 °C (letra B) no congelador da geladeira:

(dependência da letra (B) da questão 1)

(concepção alternativa: Não há; D6; C6; E*1; E*2) - (nível: 1)

2a – E que a temperatura da água é -10 °C (letra A) numa poça junto com pequenos cubos de gelo derretidos:

(dependência da letra (A) da questão 2)

(concepção alternativa: D*14; C1; D*16) - (DLII_{1,2,3}) - (nível: 3)

a) Ao admitir que os cubos de gelos restantes estejam a -10 °C (letra A) após derreter estando inicialmente a 0 °C e formado água a -10 °C. O gelo varia de 0 °C a -10 °C ao ser aquecido pelo ambiente em cima do balcão, ou seja, *ao ser aquecido diminui sua temperatura e derrete formando água a -10 °C.* Duas concepções: *o aquecimento sobre o gelo sempre resulta em mudança de fase e ele sempre derrete mantendo a temperatura constante, independente de sua temperatura (equivoco sobre o ponto de fusão); se admite que a água obtenha temperatura inferior a 0 °C.* O absurdo é tanto pelas propriedades térmicas da água quanto do gelo, portanto essa combinação é pouquíssimo provável. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: C*16; D*14) - (DLII_{1,2,3}) - (nível: 3)**

b) Ao admitir que os cubos de gelos restantes estejam a 0 °C (letra B) após derreter estando inicialmente a 0 °C e formado água a -10 °C seriam admitir o absurdo que quando o gelo derrete ele forma água mais fria. Duas concepções: *a água pode estar em temperatura inferior a 0 °C; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (2); (3). **(concepção alternativa: D*14; C*16) - (DLII_{2,3}) - (nível: 0)**

c) Ao admitir que os cubos de gelos restantes estejam a 5 °C (letra C) após derreter estando inicialmente a 0 °C e formando água a -10 °C. O gelo variou de 0 °C a 5 °C mudando de fase. Cinco concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; a água pode estar em temperatura inferior a 0 °C; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; D*14; C6; C1) - (DLII_{1,2,3}) - (nível: 1)**

d) Ao admitir que os cubos de gelos restantes estejam a 10 °C (letra D) após derreter estando inicialmente a 0 °C e formando água a -10 °C. O gelo variou de 0 °C a 10 °C mudando de fase. Quatro concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; a água pode estar em temperatura inferior a 0 °C; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; D*14; C6) - (DLII_{1,2,3}) - (nível: 2)**

e) Resultado: **(concepção alternativa: C*16; D*14) - (DLII_{1,2,3}) - (nível: 3)**

1B – Quem acredita que os cubos de gelo se encontram a 0 °C (letra B) no congelador da geladeira:

(dependência da letra (B) da questão 1)

(concepção alternativa: Não há; D6; C6; E*1; E*2) - (nível: 1)

2b – E que a temperatura da água é 0 °C (letra B) numa poça junto com pequenos cubos de gelo derretidos:

(dependência da letra (B) da questão 2)

(concepção alternativa: Não há; D6; C*10) - (DLII_{2,3}) (nível: 0)

a) Ao admitir que os cubos de gelos restantes estejam a -10 °C (letra A) após derreter estando inicialmente a 0 °C e formado água a 0 °C. Como o gelo varia de 0 °C a -10 °C mudando de fase, ou seja, admite-se que *ele é aquecido e fica mais frio*, um contrassenso, considerado uma concepção alternativa pouco provável pelo absurdo de se observar esse tipo de comportamento, sendo difícil justificar como um equívoco. Nesse processo se forma água de acordo com sua temperatura inicial. Duas concepções: *Um corpo ao ser aquecido se esfria; que ambos não tendem a um equilíbrio térmico.*

Distorções (questão 2): (1); (3). **(concepção alternativa: D*17; C6) - (DLII_{1,2,3}) - (nível: 3)**

b) Ao admitir que os cubos de gelos restantes estejam a 0 °C (letra B) após derreter estando inicialmente a 0 °C e formado água a 0 °C. Distorções (questão 2): (2). **(concepção alternativa: Não há) - (DLIC) - (nível: 0)**

c) Ao admitir que os cubos de gelos restantes estejam a 5 °C (letra C) após derreter estando inicialmente a 0 °C e formado água a 0 °C. O gelo variou de 0 °C a 5 °C mudando de fase. Três concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII_{1,2,3}) - (nível: 1)**

d) Ao admitir que os cubos de gelos restantes estejam a 10 °C (letra D) após derreter estando inicialmente a 0 °C e formado água a 0 °C. O gelo variou de 0 °C a 10 °C mudando de fase. Três concepções: *o gelo pode estar acima de 0 °C; o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura; que ambos não tendem a um equilíbrio térmico.* Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII_{1,2,3}) - (nível: 2)**

1B – Quem acredita que os cubos de gelo se encontram a 0 °C (letra B) no congelador da geladeira:

(dependência da letra (B) da questão 1)

(concepção alternativa: Não há; D6; C6; E*1; E*2) - (nível: 1)

2c – E que a temperatura da água é 5 °C (letra C) numa poça junto com pequenos cubos de gelo derretidos:

(dependência da letra (C) da questão 2)

(concepção alternativa: C6; D7; C1; D*15) - (DLII_{1,2,3}) - (nível: 1)

- a) Ao admitir que os cubos de gelos restantes estejam a -10°C (letra A) após derreter estando inicialmente a 0°C e formado água a 5°C . Como o gelo estava a 0°C e varia até -10°C mudando de fase. Duas concepções: *Um corpo ao ser aquecido se esfria*; que *ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*17; C6) - (DLII_{1,2,3}) - (nível: 3)**
- b) Ao admitir que os cubos de gelos restantes estejam a 0°C (letra B) após derreter estando inicialmente a 0°C e formado água a 5°C . Uma concepção: que *ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (2); (3). **(concepção alternativa: C6) - (DLII_{2,3}) - (nível: 0)**
- c) Ao admitir que os cubos de gelos restantes estejam a 5°C (letra C) após derreter estando inicialmente a 0°C e formado água a 5°C . O gelo variou de 0°C a 5°C mudando de fase. Três concepções: *o gelo pode estar acima de 0°C* ; *o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura*; que *ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII_{1,2,3}) - (nível: 1)**
- d) Ao admitir que os cubos de gelos restantes estejam a 10°C (letra D) após derreter estando inicialmente a 0°C e formado água a 5°C . O gelo variou de 0°C a 10°C mudando de fase. Três concepções: *o gelo pode estar acima de 0°C* ; *o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura*; que *ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII_{1,2,3}) - (nível: 2)**

1B - Quem acredita que os cubos de gelo se encontram a 0°C (letra B) no congelador da geladeira:

(dependência da letra (B) da questão 1):

(concepção alternativa: Não há; D6; C6; E*1; E*2) - (nível: 1)

2d - E que a temperatura da água é 10°C (letra D) numa poça junto com pequenos cubos de gelo derretidos:

(dependência da letra (D) da questão 2):

(concepção alternativa: C6; D7; C1; D*15) - (DLII_{1,2,3}) - (nível: 1)

- a) Ao admitir que os cubos de gelos restantes estejam a -10°C (letra A) após derreter estando inicialmente a 0°C e formado água a 10°C . Como o gelo variando de 0°C a -10°C mudando de fase. Concepções: *Um corpo ao ser aquecido se esfria*; que *ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*17; C6) - (DLII_{1,2,3}) - (nível: 3)**
- b) Ao admitir que os cubos de gelos restantes estejam a 0°C (letra B) após derreter estando inicialmente a 0°C e formado água a 10°C . Uma concepção: que *ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (2); (3). **(concepção alternativa: C6) - (DLII_{2,3}) - (nível: 0)**
- c) Ao admitir que os cubos de gelos restantes estejam a 5°C (letra C) após derreter estando inicialmente a 0°C e formado água a 10°C . O gelo variou de 0°C a 5°C mudando de fase. Três concepções: *o gelo pode estar acima de 0°C* ; *o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura*; que *ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII_{1,2,3}) - (nível: 1)**
- d) Ao admitir que os cubos de gelos restantes estejam a 10°C (letra D) após derreter estando inicialmente a 0°C e formado água a 10°C . O gelo variou de 0°C a 10°C mudando de fase. Três concepções: *o gelo pode estar acima de 0°C* ; *o aquecimento(*resfriamento) sempre resulta num(*a) aumento(diminuição) de temperatura*; que *ambos não tendem a um equilíbrio térmico*. Distorções (questão 2): (1); (2); (3). **(concepção alternativa: D*15; C1; C6) - (DLII_{1,2,3}) - (nível: 2)**

Os resultados parciais dessa análise de rede de concepções pode ser observada no quadro colocando *lado a lado* as diversas concepções combinadas, além de identificar os níveis de distorção para cada item (estão ausentes as análises parciais do item y3 que advêm do item y1C e y1D, passando pelas combinações das alternativas do item y2) . Esse tipo de análise acaba revelando evidências que justifiquem uma conversão de teste dicotômico para politômico, tornando assim as respostas dentro de

uma escala ordinal. É possível alterar a chave de correção no modelo dicotômico para os níveis classificados possibilitando uma interpretação focada para os distratores, porém, torna-se mais interessante e talvez mais adequado ampliar essas possibilidades de análise quanto as relações verificativas dessas relações, ao utilizar modelos politômicos, como o de Samejima (Samejima, 1968).

Vale ressaltar que a rede de concepções de dependência é uma perspectiva hipotética, pois os alunos podem responder de forma aleatória tornando sem sentido a interpretação dessa rede de concepções, pois, como se sabe, não corresponde a estrutura cognitiva do respondente, porém não descarta que a partir desse tipo de perspectiva possa emergir evidências secundárias que de certa forma pode estar relacionadas com os complexos modelos mentais preexistentes.

Para um aprofundamento e melhor julgamento é natural que haja uma necessidade de verificação através de outras evidências e métodos de investigação, como por exemplo, entrevistas, observação participante, diferentes formas de registros das atividades que exploram a argumentação científica dos alunos em sala de aula. Ainda assim, essa é uma perspectiva que amplia o conhecimento do traço latente avaliado quando ocorre dependência local, pois se torna explícito uma descrição de influências de concepções subjacente diante de uma mesma situação-problema.

Quadro 74: Resultados parciais da rede de concepções e análise de dependência local entre os itens y1, y2 e y3, segundo os critérios de coerência e consistência para os níveis de distorção conceitual.

Item y1	Item y2	Item y3			
		A	B	C	D
A	A	C*16; D*14 - (DLII 1,2,3) - (nível: 3)	D*14; C*16 - (DLII 2,3) - (nível: 0)	D*15; C1; D*14; C6 - (DLII 1,2,3) - (nível: 1)	D*15; C1; D*14; C6 - (DLII 1,2,3) - (nível: 2)
	B	C*16; C6 - (DLII 1,2,3) - (nível: 3)	Não há - (DLCC) - (nível: 0)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 1)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 2)
	C	C*16; C6 - (DLII 1,2,3) - (nível: 3)	C6 - (DLII 2,3) - (nível: 0)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 1)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 2)
	D	C*16; C6 - (DLII 1,2,3) - (nível: 3)	C6 - (DLII 2,3) - (nível: 0)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 1)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 2)
B	A	C*16; D*14 - (DLII 1,2,3) - (nível: 3)	D*14; C*16 - (DLII 2,3) - (nível: 0)	D*15; C1; D*14; C6; C1 - (DLII 1,2,3) - (nível: 1)	D*15; C1; D*14; C6 - (DLII 1,2,3) - (nível: 2)
	B	D*17; C6 - (DLII 1,2,3) - (nível: 3)	Não há - (DLIC) - (nível: 0)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 1)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 2)
	C	D*17; C6 - (DLII 1,2,3) - (nível: 3)	C6 - (DLII 2,3) - (nível: 0)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 1)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 2)
	D	D*17; C6 - (DLII 1,2,3) - (nível: 3)	C6 - (DLII 2,3) - (nível: 0)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 1)	D*15; C1; C6 - (DLII 1,2,3) - (nível: 2)
C	A	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>
	B	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>

	C	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>
	D	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>
D	A	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>
	B	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>
	C	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>
	D	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>	<i>Em processo de análise</i>

A complexidade na análise fatorial no enfoque qualitativo revela um processo de certa forma tedioso e exaustivo, mas que podem ser observadas influências de concepções advindas de diferentes conexões lógicas de itens anteriores. Apesar do nível de distorção ter se baseado exclusivamente nas informações do próprio item, sem levar em consideração a rede de concepções que eventualmente possa estar influenciando, isso sugere novas perspectivas diante do desenvolvimento de análises relacionadas a interpretações dos traços latentes que estão sendo avaliados, obviamente, no caso de haverem indícios de ocorrência de dependência local. O conjunto dos resultados das concepções envolvidas dentro as combinações analisadas podem mostrar aspectos de consistência na influência de concepções específicas. Destacando os resultados parciais para a sequência de respostas que advêm das alternativas A e B do item y1, nomeadamente quanto as supostas influências da alternativa A (maior distorção) e B (mais plausível) no item y3, tem-se:

1A/2a/3a: (concepção alternativa: C*16; D*14) - (DLII_{1,2,3}) - (nível: 3)

*1A/2a/3b: (concepção alternativa: D*14; C*16) - (DLII_{2,3}) - (nível: 0)*

1A/2b/3a: (concepção alternativa: C*16; C6) - (DLII_{1,2,3}) - (nível: 3)

1A/2b/3b: (concepção alternativa: Não há) - (DLCC) - (nível: 0)

1A/2c/3a: (concepção alternativa: C*16; C6) - (DLII_{1,2,3}) - (nível: 3)

1A/2c/3b: (concepção alternativa: C6) - (DLII_{2,3}) - (nível: 0)

1A/2d/3a: (concepção alternativa: C*16; C6) - (DLII_{1,2,3}) - (nível: 3)

1A/2d/3b: (concepção alternativa: C6) - (DLII_{2,3}) - (nível: 0)

1B/2a/3a: (concepção alternativa: C*16; D*14) - (DLII_{1,2,3}) - (nível: 3)

*1B/2a/3b: (concepção alternativa: D*14; C*16) - (DLII_{2,3}) - (nível: 0)*

1B/2b/3a: (concepção alternativa: D*17; C6) - (DLII_{1,2,3}) - (nível: 3)

1B/2b/3b: (concepção alternativa: Não há) - (DLIC) - (nível: 0)

1B/2c/3a: (concepção alternativa: D*17; C6) - (DLII_{1,2,3}) - (nível: 3)

1B/2c/3b: (concepção alternativa: C6) - (DLII_{2,3}) - (nível: 0)

Analisando a rede de concepções interpretadas segundo os critério de dependência, é possível por exemplo verificar as influências para a alternativa “a” do item y3 (em negrito). Observa-se influências de concepções C e D, no caso,

relacionadas aos *processos termodinâmicos e propriedades térmicas*, sendo uma alternativa pouco plausível, com uma lógica um tanto “absurda” na interpretação física diante das outras alternativas, portanto, com pouca possibilidade de ser um opção admissível.

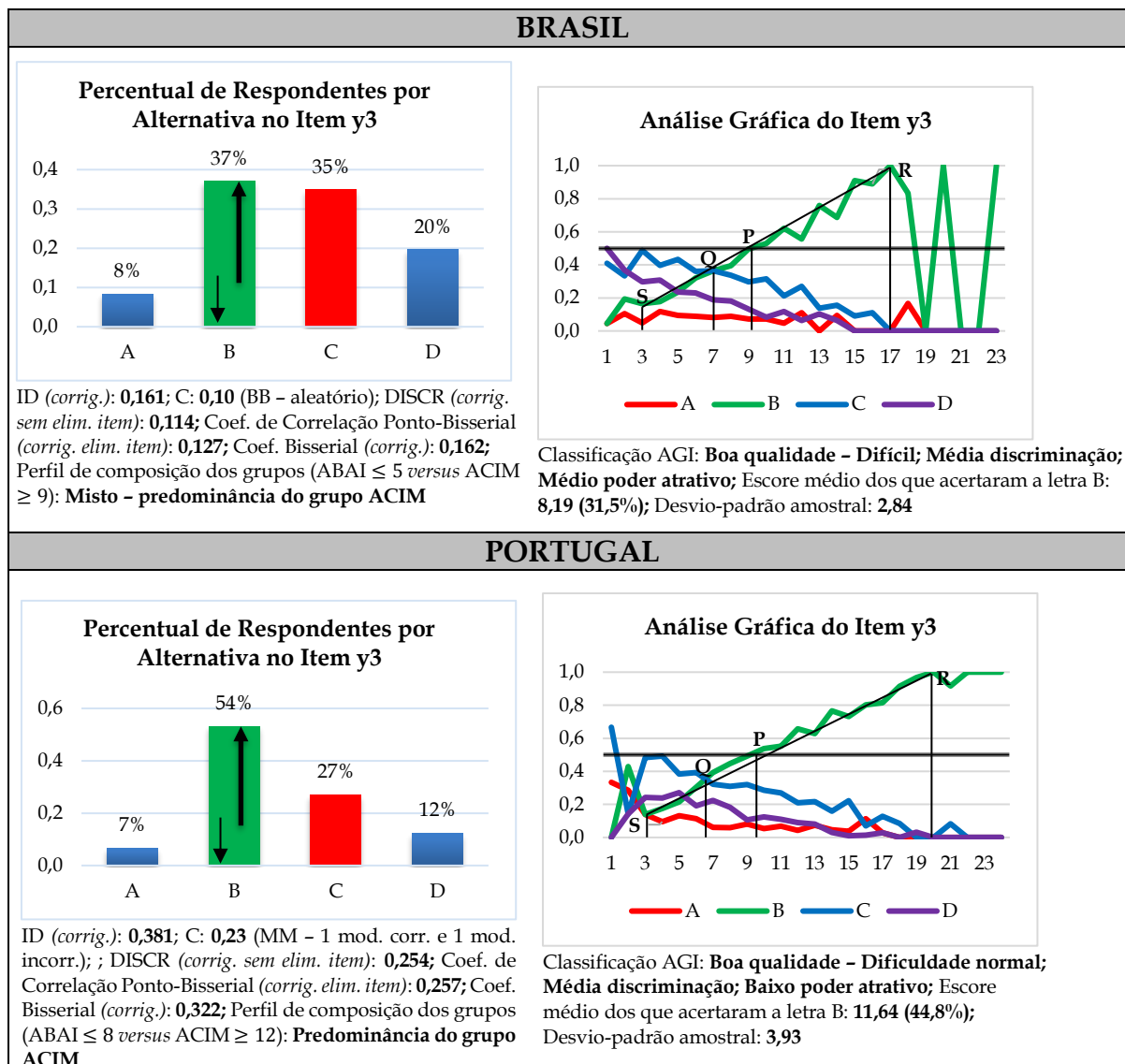


Figura 91: Exemplo comparativo de resultados clássicos do item y3 entre Brasil e Portugal.

Observando os escores brutos do item y3 que encontra-se em detalhes no Apêndice A3 (p. 27), pode-se constatar que os resultados empíricos dos respondentes no Brasil e em Portugal, sugerem os modelos mentais pelo fator de concentração das análises, e esses coincidem com os *níveis de distorções* interpretados preliminarmente, ou seja, acaba por revelar um processo de fazer correspondência com os valores esperados, de acordo com a consistência do item, sendo acrescentadas de informações relacionadas a coerência pode dependência local advindas das análises das redes de

concepções que contribuem na interpretação da carga cognitiva do traço latente avaliado, e que supostamente possa fazer parte da estrutura cognitiva dos respondentes:

- Alternativa A (nível: 3): Com 8% (Brasil) e 7% (Portugal);
- Alternativa B: (nível: 0): Com 37% (Brasil) e 54% (Portugal);
- Alternativa C (nível 1): Com 35% (Brasil) e 27% (Portugal);
- Alternativa D (nível: 2): Com 20% (Brasil) e 12% (Portugal).

Também chama atenção que a resposta correta (letra “b” do item y3, em itálico entre as concepções destacadas) possa carregar concepções alternativas por influência de incoerência na dependência local, ou seja, o modelo mental considerado correto pode eventualmente estar associado aos equívocos C6, D*14 e C*16, porém, os modelos logísticos de análises mais robustos dentro da Psicometria podem fornecer de forma mais consistente a probabilidade de ocorrência desse evento, conforme mostra a *curva característica* e a *curva de informação* do item y3.

Utilizando esses critérios avaliativos foi possível identificar um conjunto de informações mais detalhados sobre o enunciado dos itens, dos estímulos secundários que o aspecto semântico supostamente pode provocar na formulação de ideias durante o processo de leitura e interpretação da situação física, que em parte podem ser argumentos explicativos ou conjecturas sobre a existência de traços latentes secundários diante os traça latente principal e intencional dos itens. A leitura de diferentes níveis para cada item também possibilidades para novas perspectivas de análises sobre as respostas dos alunos e quanto a validade de construto do instrumento. Em síntese, as informações descritivas da análise podem ser lidas dentro dos quadros elaborados.

Quadro 75: Níveis de Distorção Conceptual do Item y1.

y1 – Níveis de Distorção Conceptual do Item y1				
<ul style="list-style-type: none"> • Traços Latentes Semânticos Identificados (secundário): “Qual é a temperatura <i>mais provável...</i>” é baseada numa especulação, numa estimativa de maior probabilidade que ocorra ou não algo preciso. • Traços Latentes Tecnológicos Identificados (secundário): Conhecimento sobre as condições ambientais típicas de um congelador de geladeira. (interpretativo <i>praeteritum</i> – como era antes?); Conhecimento sobre o processo de trocas de calor nas condições ambientais típicas de um congelador de geladeira sobre os cubos de gelo no decorrer do tempo, tendo uma noção sobre o estado atual dentro de um processo de evolução. (interpretativo <i>praesenti</i> – como é agora?) • Traços Latentes Conceituais Identificados (secundário): Conhecimento sobre as propriedades térmicas do gelo: o ponto de solidificação da água, ou condição para a existência de gelo próximo ao CNTP, no caso 0 °C. (interpretativo <i>praeteritum</i> – como vai ser depois?); Conhecimento sobre o processo de resfriamento, na qual a substância permanece na mesma fase variando a temperatura, ou seja, neste caso o gelo iria atingir temperaturas inferiores a 0 °C. O gelo somente irá parar de esfriar quando atingir a temperatura final de <i>equilíbrio térmico</i>, onde assim serão cessadas as trocas de calor entre os corpos. (interpretativo <i>status quo</i>) • Traço Latente Conceitual Primário: Conhecimento sobre as propriedades térmicas do gelo: capacidade de atingir temperaturas inferiores a 0 °C até o equilíbrio térmico na situação física atual, em condições próximas ao CNTP. *Principal grandeza física envolvida: Temperatura. (interpretativo <i>praesenti</i>). 				
Níveis de distorção	0	1	2	3
Alternativas	A	B	D	C
Redes de concepções	Não há; E*1; E*2; C*17 – NPV	Não há; D6; C6; E*1; E*2	B7; E*1; E*2	D*15
Dependência local	<i>Não há indícios na relação de dependência com itens anteriores</i>			

NPV= Esta é uma suposição que *Não Pode ser Verificada* de forma direta.

*Novas concepções identificadas na análise.

Quadro 76: Níveis de Distorção Conceptual do Item y2.

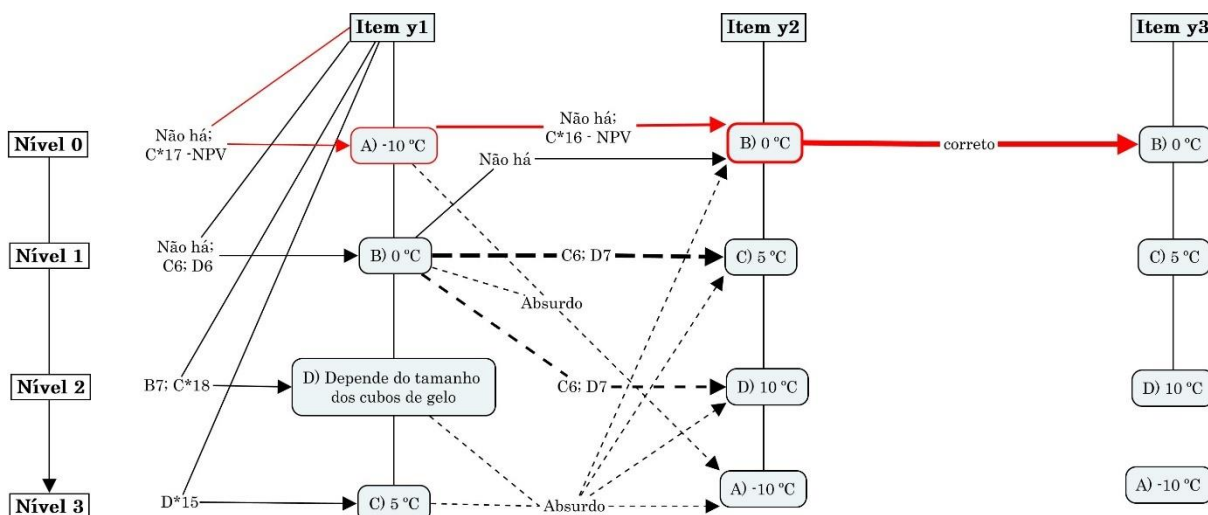
y2 – Níveis de Distorção Conceptual do Item y2				
<ul style="list-style-type: none"> • Traços Latentes Identificados (secundário): Conhecimento sobre as condições ambientais típicas de um congelador de geladeira. (interpretativo <i>praeteritum</i>); Conhecimento sobre as propriedades térmicas do gelo: o ponto de solidificação da água, ou condição para a existência de gelo próximo ao CNTP, no caso 0 °C. (interpretativo <i>praeteritum</i>); Conhecimento sobre o processo de trocas de calor nas condições ambientais típicas de um congelador de geladeira sobre os cubos de gelo no decorrer do tempo. (interpretativo <i>praeteritum</i>); Conhecimento sobre a temperatura provável dos cubos de gelo no congelador, ou seja, um pouco abaixo de 0 °C. (interpretativo <i>praeteritum</i>); Conhecimento da provável temperatura da água contida no copo destinado a serem inseridos quatro cubos de gelo, estando à água supostamente em equilíbrio com a temperatura ambiente, próximo de 27 °C em condições de CNTP. (interpretativo <i>praeteritum</i>); Conhecimento que a agitação dos cubos de gelo em água e nas paredes do recipiente produz fricção (força de atrito) onde conseqüentemente gera <i>calor latente</i>, fazendo com que o gelo derreta, além das trocas de calor envolvidas. (interpretativo <i>praesenti</i>); Conhecimento sobre a temperatura de fusão do gelo, no caso, 0 °C. (interpretativo <i>praesenti</i>) • Traço Latente Conceitual Primário: Conhecimento sobre o processo de mudança de fase, na qual a substância coexiste nas duas fases com a mesma temperatura durante a transição, ou seja, neste caso seria gelo e água a 0 °C. O gelo somente irá parar de derreter quando atingir a temperatura final de <i>equilíbrio térmico</i>, quando serão cessadas as trocas de calor ente os corpos. *Principal grandeza física envolvida: Temperatura. (interpretativo <i>futurum</i>) 				
Níveis de distorção	0	1	2	3
Alternativas	B	C	D	A

Redes de concepções	Não há; C1 – NPV; C6; C*16	Absurdo; C6; D7; D*15	Absurdo; C6; D7; D*15	Absurdo; C*10; C*16; D*14
Dependência local	CC e IC; CI	CI; II	CI; II	CI; II

Quadro 77: Níveis de Distorção Conceptual do Item y3.

y3 – Níveis de Distorção Conceptual do Item y3				
<ul style="list-style-type: none"> • Traços Latentes Identificados (secundário): Conhecimento sobre as condições ambientais típicas de um congelador de geladeira. (interpretativo <i>praeteritum</i>); Conhecimento sobre as propriedades térmicas do gelo: o ponto de solidificação da água, ou condição para a existência de gelo próximo ao CNTP, no caso 0 °C. (interpretativo <i>praeteritum</i>); Conhecimento sobre a temperatura provável dos cubos de gelo no congelador, ou seja, um pouco abaixo de 0 °C. (interpretativo <i>praeteritum</i>); Conhecimento sobre a temperatura provável dos cubos de gelo no congelador, ou seja, um pouco abaixo de 0 °C. (interpretativo <i>praeteritum</i>); Conhecimento sobre a temperatura provável do balcão, ou seja, em torno de 27 °C, a mesma que a temperatura ambiente em condições próximas a CNTP. (interpretativo <i>praeteritum</i>); Conhecimento sobre o processo de trocas de calor entre o balcão e os cubos de gelo no decorrer do tempo. (interpretativo <i>praeteritum</i>); Conhecimento sobre a temperatura de fusão do gelo, no caso, 0 °C. (interpretativo <i>praesenti</i>) • Traço Latente Conceitual Primário: Conhecimento sobre o processo de mudança de fase, na qual a substância coexiste nas duas fases com a mesma temperatura durante a transição, ou seja, neste caso seria gelo e água a 0 °C. O gelo somente irá parar de derreter quando atingir a temperatura final de equilíbrio térmico, quando serão cessadas as trocas de calor ente os corpos. *Principal grandeza física envolvida: Temperatura. (interpretativo <i>futurum</i>) 				
Níveis de distorção	0	1	2	3
Alternativas	B	C	D	A
Redes de concepções	Em processo de análise	Em processo de análise	Em processo de análise	Em processo de análise
Dependência local	Em processo de análise	Em processo de análise	Em processo de análise	Em processo de análise

Entende-se que as redes de concepções do item y3 devam ser discriminadas pelo caminho lógico possíveis entre as combinações no padrão sequencial de respostas. Uma perspectiva ilustrativa dessa rede de concepções seguindo esse padrão de resposta, e para algumas perspectivas da análise de distorção envolvidas, tem-se:



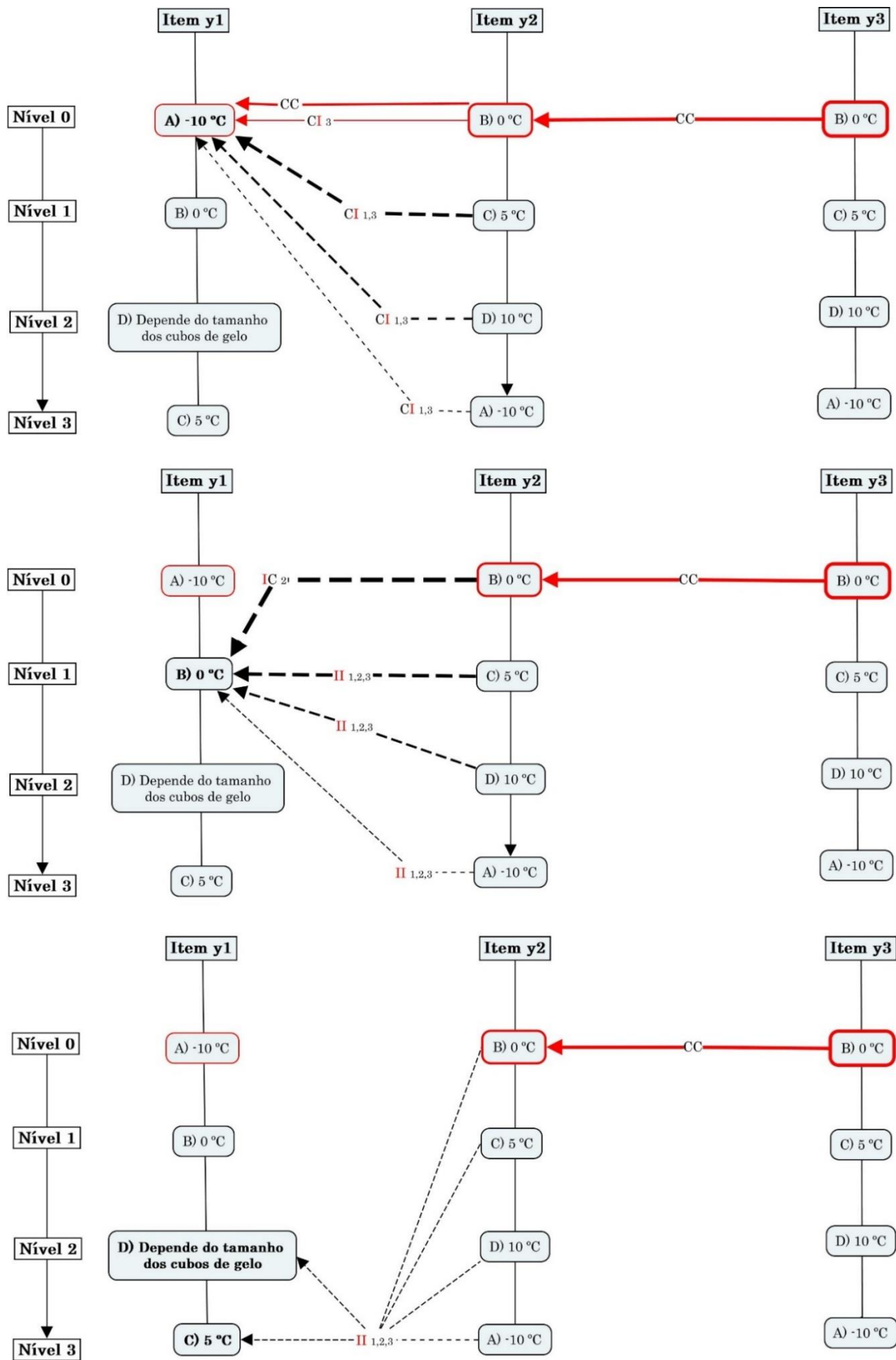


Figura 92: Análises de distorção conceitual diante aos critérios de plausibilidade, no que diz respeito a coerência e consistência dos indícios de dependência local entre os itens y1, y2 e y3.

CONSIDERAÇÕES FINAIS

As análises clássicas reforçaram as evidências empíricas da TRI mostrando ganhos normatizados mais elevados nas escolas portuguesas. Contudo, houve de modo geral uma grande dificuldade diante os conceitos básicos da Física Térmica para ambos os contextos, estando os ganhos ainda estão longe de alcançarem um grande impacto. Comparativamente, é natural admitir que os resultados estão sendo influenciados pelas diferenças nas formas de tratamento existentes entre os sistemas educacionais, como o tempo diferenciado em que os alunos ficam na escola, entre outros diversos fatores intervenientes. Em Portugal a educação em tempo integral é uma realidade, enquanto no Brasil existe a expectativa das escolas atingirem cerca de 25% nos primeiros anos da década de 2020, segundo o *Plano Nacional de Educação* que passa por medidas econômicas de estagnação, podendo comprometer diretamente o cumprimento das metas educacionais previstas.

Em visita as escolas, pode-se observar e contatar através de diálogos com cerca de 30 professores entre 20 escolas secundárias, incluindo um visita técnica no projeto sala de aula do futuro, em Setúbal, que Portugal apresenta ainda algumas escolas deficitárias com relação as condições mínimas de infraestruturas para a realização de atividades prático-laboratoriais. Analisando o programa curricular de 2001, e acompanhando as mudanças mais recentes em 2014, é notório um programa de Física mais reduzido comparativamente com o Brasil, buscando exigir o cumprimento de atividades e abordagem de conteúdos. Além disso, constatou-se o relato de insatisfação dentro da classe de professores com relação as condições de trabalho. No Brasil, os problemas de infraestrutura e de insatisfação da classe docente em média são aparentemente mais graves, porém não é generalizado, com diferenças salariais e de condições de trabalho entre os estados. Atualmente o sistema educacional passa por uma nova formulação estrutural, com o advento da *Base Nacional Comum Curricular* (BNCC) e a reestruturação do *Sistema Nacional de Avaliação da Educação Básica* (SINAEB), e sendo acompanhado da *Reforma do Ensino Médio*, e de novos exames nacionais para os professores e para a educação infantil. Sobretudo, em meio a complexidade, os dois países se apresentam desafiados a superar as dificuldades

existentes, e acredita-se que seja promissor buscar se possível fortalecer o papel das evidências empíricas na busca de melhorias de qualidade pedagógica e rendimento nas aprendizagens, considerando válidos e relevantes as mais diversas formas de métodos de ensino, conforme as tendências contemporâneas são sugeridas dentro do campo investigativo da Didática das Ciências.

Defende-se que o papel avaliativo é exercido através do diagnóstico de testes conceituais nesse processo possam servir como grande aliado, reforçando uma postura *accountable* entre gestores e docentes, e apoiando na regulação do ensino, identificação e valorização das necessidades dos alunos, porém, deve-se evitar os efeitos perversivos apontados na má administração do uso de testes conceituais em larga escala para o sistema educacional, além de funções e contextos adequados. É preciso, sobretudo, garantir um ambiente saudável no cenário pedagógico, valorizando uma formação cidadã e humana, negando a atribuição de punições advindas de perspectivas distorcidas e reducionistas sobre a padronização curricular, entretanto, não convém abrir mão de identificar necessidades gerais para aprendizagem, focar no alcance de metas e no desenvolvimento de competências básicas e específicas, visando favorecer uma progressão da aprendizagem e uma alfabetização científica e tecnológica que torne o aluno preparado para exercer sua cidadania através de um pensamento crítico e postura autônoma no contexto da vida real.

Há uma diferença significativa na atribuição das funções exercidas entre os testes conceituais e exames nacionais, apesar de ambos serem de larga escala. No entanto, algumas considerações gerais quanto às funções exercidas por testes no âmbito dos exames nacionais, podem ser reelaboradas no sentido de evitar os efeitos perversivos e criticados por grupo de avaliadores educacionais, na qual muitos demonstram uma certa aversão ao termo “larga escala”, entre outros que remetam a uma perspectiva “positivista”. Entende-se que hajam argumentos pejorativos e equivocados devido em parte a falta de domínio dentro desse campo de pesquisa, como no caso daqueles que se concentram exclusivamente em abordagens qualitativas de pesquisa, porém muitos dos argumentos levantados são de fato valiosos e pertinentes, principalmente quanto ao uso e abuso desses testes. Na década de 70,

diante das falhas em eficiência e produtividade dos sistemas educacionais, as pesquisas educacionais já recebiam críticas por se concentrar demasiadamente em estatísticas e Psicometria (Pereyra, Kotthoff, & Cowen, 2011). Entre as considerações gerais e relevantes para o uso de testes em larga escala, alguns são bem mais pertinentes aos exames nacionais do que para testes conceituais, e com isso tornam-se alvo de críticas. No caso, tem-se:

- A função diagnóstica deve ser a mais importante para a melhoria da qualidade da educação nas escolas, e não a função de classificação no sentido de promover uma competição em detrimento da menos qualificada pelo indicador.
- Diagnostica as condições da aprendizagem de base no 1º ciclo do nível de escolarização, assim como aquelas ligadas ao final de cada ciclo da educação básica.
- Compromete-se em atender a universalização da educação dando oportunidade de acesso a todos fornecendo indicadores a serem atrelados às necessidades de melhoria e qualidade de ensino.
- Atende as necessidades do interesse do mercado de trabalho e oportuniza cidadãos com idade fora do tempo regular de escolarização quanto a certificação e reconhecimento de competências básicas de acesso ao emprego.
- As escolas com melhor desempenho são premiadas e recebem financiamento para garantir sua qualidade e impulsionar o seu processo de melhoria, assim como dar tutoria as escolas com desempenho inferior através de troca de experiências de gestão e de políticas educacionais implementadas sem puni-las ou compromete-las com recursos as que escolas e alunos que mais necessitam.
- Através do cruzamento entre desempenho e questões socioculturais e econômicas possibilita identificar fatores correlacionados as necessidades de melhoria.
- Possibilita uma comparação entre diferentes regiões e focar os maiores investimentos em determinados locais.
- Contribui para a isonomia dos estudantes, ou seja, estabelece parâmetros para garantir as mesmas condições de direito governamental regido pelas leis.
- Fundamenta e articula o exercício do princípio da equidade no sistema educacional, ou seja, busca desenvolver políticas públicas educacionais que visem

tratar todos de maneira igualitária considerando as diferenças étnicas, físicas, de gênero, raciais, sociais e econômicas.

- Possibilita uma comparação ao longo do tempo quanto aos desempenhos (o que a TRI se denomina séries históricas de desempenho) dos estudantes, estabelecendo com isso parâmetros de observação e acompanhamento de fatores relacionados à qualidade da educação.
- Identificam as potencialidades e necessidades dos estudantes comparativamente quanto aos diferentes fatores ligados as suas características étnicas, físicas, de gênero, raciais, sociais e econômicas diante de seus desempenhos.
- Promove o estabelecimento de metas de melhoria de forma sistemática ao longo dos anos.
- Verifica a necessidade de investimentos quanto à qualificação e valorização docente e de infraestrutura das escolas.
- Elaboração dos itens dos exames possibilita parcerias com as *Instituições de Ensino Superior* (IES) interessadas em elaborar e revisar itens aumentando assim a participação da comunidade acadêmica de todo o Brasil nos processos de avaliação educacional, agregando experiência, conhecimento e diversidade.

Como foi dito, os pesquisadores no EF nos Estados Unidos já vinham desde o início do século XXI liderando o caminho no desenvolvimento e uso de testes objetivos para comparar os ganhos de aprendizagem dos alunos em diferentes tipos de níveis de escolaridade e cursos, e que os pesquisadores educacionais em Química e Biologia, entre outros, também demonstram interesse no desenvolvimento de instrumentos similares (Wood & Gentile, 2003, p. 1510). Nisso, apesar de afirmar que esses testes fornecem evidências convincentes de que os alunos assimilam novos conhecimentos de forma mais eficaz em cursos que incluem atividades ativas, baseadas em questionários/testes, aprendizagem colaborativa, sendo auxiliada pela tecnologia da informação, do que nos cursos com métodos tradicionais, esse tipo de perspectiva metodológica e o uso desses instrumentos como auxílio ao professor continua em expansão no campo de investigação educacional, entretanto, os levantamentos não demonstram interesse nas produções no Brasil, assim como em Portugal. Com isso, as potencialidades de avançar e preencher essas lacunas são um desafio assumido, no

sentido de expandir e contribuir no cruzamento com as abordagens desenvolvidas no âmbito das pesquisas com abordagens qualitativas em educação científica e tecnológica.

Recentemente o levantamento de Schram (2014) em três revistas de alto impacto no EC, duas delas encontram-se na lista do critério Qualis da CAPES como A1, que são as:

- *International Journal of Science Education* – IJSE (Conceito Qualis A1. Fator de Impacto em 2015: 1,000);
- *Research in Science Education* – RSE (o jornal oficial da *Australasian Science Education Research Association*. Conceito Qualis A1. Fator de impacto em 2015: 0,895);
- *Journal of Research in Science Teaching* – JRST (o jornal oficial da *National Association for Research in Science Teaching*. Fator de impacto: 3,052)

O estudo mostrou que os **métodos mistos** encontram-se em **pleno processo de expansão dentro do campo das pesquisas em EC em nível mundial**, pois demonstra um esforço internacional pela diversidade de países buscando melhorias na combinação das abordagens qualitativa e quantitativa para o enfrentamento dos problemas da educação científica. O Brasil não está dentro da lista dos países que somam esforços de forma mais expressiva e nem tão pouco a comunidade demonstra interesse a esses aspectos, na realidade este é provavelmente o país que representa a maior “fatia” dentro da “lacuna do Atlântico”, como se afirma (Creswell & Clark, 2015).

Diversos trabalhos dentro de Psicometria vem sendo desenvolvidos no campo das pesquisas mistas considerando promissor o uso de testes conceituais na educação científica, alguns concentrados na avaliação de inventários de conceitos (Stone, 2006), na validação de escalas psicométricas em educação científica voltada para a saúde na Europa (Jorde & Dillon, 2012), perspectivas de competências educacionais pelo uso de modelos psicométricos inovadores (Fleischer, 2017), entre outros, o próprio teste TCE foi explorado cruzando resultados de procedimentos de estatística inferencial com aspectos de instrução de ensino pelos autores originais (Chu, Treagust, Yeo, & Zadnik, 540

2012).

Enfim, é preciso uma mudança e uma discussão mais profunda sobre a qualidade e praticidade dos trabalhos que vem sendo desenvolvidos tendo como base evidências empíricas e o rigor científico, assim como da necessidade de serem exploradas novas perspectivas para os procedimentos de análise na comunidade de pesquisadores em educação científica, nisso, as contribuições dos inventários de conceitos nesse contexto é um campo ainda a ser explorado. É preciso acompanhar as novas tendências pois a educação da América Latina, em especial a brasileira, está sedenta com indicadores educacionais em retrocesso e novos desafios a serem enfrentados. É preciso que ressoe o chamado de Johnson e Onwuegbuzie de cerca 14 anos atrás (Johnson & Onwuegbuzie, 2004): “*Chegou o momento das pesquisas de métodos mistos!!!*”, ou será que ainda não?

REFERÊNCIAS

Abad, F., Garrido, J., Olea, J., & Ponsoda, V. (2006). Introducción a la psicometría. *Clásica de Los Test y Teoría de La Respuesta*.

Abraham, G. Y. (2012). Book Review: Finnish Lessons, What can the world learn from educational change in Finland? *Drustvena Istrazivanja*, 21(2 (116)), 599–602. <https://doi.org/10.5559/di.21.2.18>

Adams, W., & Wieman, C. (2011). Development and validation of instruments to measure learning of expert- like thinking. *International Journal of Science*. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/09500693.2010.512369>

Adoção de Manuais Escolares | Direção-Geral da Educação. (n.d.). Retrieved from <http://www.dge.mec.pt/adocao-de-manuais-escolares%0A>

Afonso, A. (2009). Nem tudo o que conta em educação é mensurável ou comparável. Crítica à accountability baseada em testes estandardizados e rankings escolares. *Revista Lusófona de Educação*, 13, 13–29. Retrieved from http://www.oei.es/historico/pdf2/revista_lusofona_educacao_13a02.pdf

Afonso, A. (2011). Questões polémicas no debate sobre políticas educativas contemporâneas: o caso da accountability baseada em teste estandardizados e rankings. In *Do currículo à avaliação, da avaliação ao currículo - Coleção Educação e Formação* (pp. 83–101). Retrieved from <https://repositorium.sdum.uminho.pt/handle/1822/15881>

Afonso, M. (2013). *Mensuração da predisposição ao comportamento sustentável por meio da teoria da resposta ao item*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/123004>

Alfaya-Santos, J. V. (2013). *Concepções de progresso biológico em livros didáticos de*

- biologia aprovados pelo PNL D 2012*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/107450>
- Almeida, L. M. P. de. (2007). Competências: um caminho educativo para novos desafios. *Revista Portuguesa de Pedagogia*, 41(3). Retrieved from https://digitalis.uc.pt/pt-pt/artigo/competências_um_caminho_educativo_para_novos_desafios
- Almeida, V. (2009). *Avaliação do desempenho ambiental de estabelecimentos de saúde, por meio da Teoria da Resposta ao Item, como incremento da criação do conhecimento*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/92267>
- Alves, L. (2011). *Desenvolvimento de uma escala para medir o potencial empreendedor utilizando a Teoria da Resposta ao Item (TRI)*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina.
- Alves, M., & De Ketele (orgs.), J. (2011). As abordagens por competências (APC) analisadas do ponto de vista das políticas em educação. In *Do currículo à avaliação, da avaliação ao currículo - Coleção Educação e Formação* (pp. 11–40). Porto: Porto Editora. Retrieved from <https://scholar.google.pt/scholar?hl=pt-PT&q=do+currículo+a+avaliação+de+ketele&btnG=&lr=>
- Alves, M., & De Ketele, J. (2011). Do currículo à avaliação, da avaliação ao currículo. *Coleção Educação e Formação Temática Ciências Da*. Retrieved from https://scholar.google.pt/scholar?q=Do+currículo+à+avaliação%2C+da+avaliação+ao+currículo+-+Coleção+Educação+e+Formação&btnG=&hl=pt-PT&as_sdt=0%2C5
- Alwan, A. A. A. (2011). Misconception of heat and temperature among physics students. *Procedia - Social and Behavioral Sciences*, 12, 600–614. <https://doi.org/10.1016/j.sbspro.2011.02.074>

Amado, J. (2013). *Manual de investigação qualitativa em educação, 2ª Edição (1ª edição)*.

Coimbra: Imprensa da Universidade de Coimbra. Retrieved from

<https://www.google.com/books?hl=pt->

[PT&lr=&id=OnaMCwAAQBAJ&oi=fnd&pg=PA5&ots=-](https://www.google.com/books?hl=pt-PT&lr=&id=OnaMCwAAQBAJ&oi=fnd&pg=PA5&ots=-9rQR1v0q1&sig=ZyMvU8t3xtWpSo67BilBuLUtErU)

[9rQR1v0q1&sig=ZyMvU8t3xtWpSo67BilBuLUtErU](https://www.google.com/books?hl=pt-9rQR1v0q1&sig=ZyMvU8t3xtWpSo67BilBuLUtErU)

Amaral, E. M. R. do, & Mortimer, E. F. (2001). Uma proposta de perfil conceitual para

o conceito de calor. *Revista Brasileira de Pesquisa Em Educação Em Ciências*, 1(3), 05-

18. Retrieved from

<https://seer.lcc.ufmg.br/index.php/rbpec/article/view/2357>

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN

PSYCHOLOGICAL ASSOCIATION, NATIONAL COUNCIL ON

MEASUREMENT IN EDUCATION. (1999). *Standards for Educational and*

Psychological Testing.

American Educational Research Association, American Psychological Association, &

National Council on Measurement in Education. (2014). *Standards for Educational*

and Psychological Testing. Washington.

Ana Maria Lopez Calvo de Feijoo. (1986). *Compreensão X Memorização automática no processo ensino-aprendizagem da disciplina estatística*. (Dissertação de mestrado).

Fundação Getúlio Vargas.

Anderson, L. W. (1999). Rethinking Bloom's Taxonomy: Implications for testing and

assessment. *The Educational Resources Information Center*, 1-25.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E.,

Pintrich, P. R., Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and*

Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. (L. W.

Anderson & D. R. Krathwohl, Eds.). Longman.

Andersson, B. (1980). Some aspects of children's understanding of boiling point.

Cognitive Development Research in Science and Mathematics, 252, 260. Retrieved

from <https://scholar.google.pt/scholar?hl=pt->

PT&as_sdt=0%2C5&q=Some+aspects+of+children's+understanding+of+boiling
+point.&btnG=

Andrade, D. F. De, Tavares, H. R., & Valle, R. D. C. (2000). Teoria da Resposta ao Item: Conceitos e Aplicações. *Simpósio Nacional de Probabilidade e Estatística (SINAPE)*, 3-131.

ANFOPE. (2016). *Carta de Vitória*. Vitória.

Anna, M. D., Laffranchi, G., & Lubini, P. (2011). *Strumenti per l'insegnamento interdisciplinare della termodinamica nelle scienze sperimentali* (Vol. I).

ANPEd. (2017). *A Associação Nacional de Pós-Graduação e Pesquisa em Educação (ANPEd) e a Base Nacional Comum Curricular (BNCC)*.

Araújo-Jorge, T. C. de, Borba, M. de C., & Sovierzoski, H. H. (2017). *Relatório De Avaliação 2013-2016 Quadrienal 2017*.

Araújo-Jorge, T. C. de, & Fonseca, G. R. de S. *Avaliação Trienal da CAPES-2013* (2013).

Araújo, A. O. de. (2014). *O Perfil Conceitual de Calor e Sua Utilização por Comunidades Situadas*.

Araujo e Oliveira, J. B. (2013). *Educação baseada em evidências: o DNA do IAB*. Retrieved June 12, 2017, from <https://www.youtube.com/watch?v=fjvdHhdGoig>

Araujo, J. (2014). *Teoria da resposta ao item em processo de decisão*. (Dissertação de Mestrado em Estatística). Universidade de Brasília. Retrieved from <http://repositorio.unb.br/handle/10482/17243>

Arias, R. M. (2005). *Psicometría: Teoría de los tests psicológicos y educativos* (2ª edição). Madrid: Editorial Síntesis.

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the State of the Art for

Measurement in Chemistry Education Research: Examining the Psychometric Evidence. *Journal of Chemical Education*, 90(5), 536–545.

<https://doi.org/10.1021/ed3002013>

Arruda, J. P. de. (2004). *Cidadania e matemática no livro didático para as séries iniciais do ensino fundamental*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/87510>

Ausubel, D. (2003). Aquisição e retenção de conhecimentos: uma perspectiva cognitiva. *Lisboa: Plátano*. Retrieved from <http://files.mestrado-em-ensino-de-ciencias.webnode.com/200000007-610f46208a/ausebel.pdf>

Ausubel, D., Novak, J., & Hanesian, H. (1980). *Psicologia educacional*. Retrieved from https://scholar.google.pt/scholar?q=psicologia+educacional+dauid+ausubel&btnG=&hl=pt-PT&as_sdt=0%2C5

Ausubel, D. P. (2003). *Aquisição e Retenção de Conhecimentos: Uma Perspectiva Cognitiva*. *Plátano Edições Técnicas* (1ª Edição). Lisboa. Retrieved from <https://www.passeidireto.com/arquivo/19784481/livro-david-ausubel---aquisicao-e-retencao-de-conhecimentos---uma-perspectiva-co>

Azevedo, C. L. N. (2008, March 11). *Modelos longitudinais de grupos múltiplos multiníveis na teoria da resposta ao item: métodos de estimação e seleção estrutural sob uma perspectiva bayesiana*. Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, São Paulo. <https://doi.org/10.11606/T.45.2008.tde-15042008-165256>

Azevedo, M. M. de, & Galhardi, A. C. (2013). O ganho de hake aplicado para replanejamento de disciplina de pós-graduação. In *XLI Congresso Brasileiro de Educação em Engenharia*.

Bachelard, G. (1996). *A Formação do Espírito Científico: Contribuição para uma Psicanálise do Conhecimento*. (E. dos S. (trad.) Abreu, Ed.) (5ª reimpre). Rio de Janeiro: Contraponto.

- Baerisch, S., & Hasselbring, W. (2010). *Domain-Specific Model-Driven Testing*. Retrieved from <http://link.springer.com/content/pdf/10.1007/978-3-8348-9624-7.pdf>
- Balbin Junior, A. (2010). *Avaliação da satisfação de clientes de bancos comerciais utilizando a teoria da resposta do item*. *repositorio.ufsc.br*. Retrieved from <https://repositorio.ufsc.br/handle/123456789/94387>
- Bandeira, M. (n.d.). Texto 9: Definição das variáveis e métodos de coleta de dados.
- Bao, L. (2006). Theoretical comparisons of average normalized gain calculations. *American Journal of Physics*, 74(10), 917. <https://doi.org/10.1119/1.2213632>
- Bao, L., & Redish, E. F. (2001). Concentration analysis: A quantitative assessment of student states. *American Journal of Physics*, 69(1), 45–53. Retrieved from <http://scitation.aip.org/content/aapt/journal/ajp/69/S1/10.1119/1.1371253>
- Bar, V., & Travis, A. S. (1991). Children's views concerning phase changes. *Journal of Research in Science Teaching*, 28(4), 363–382. <https://doi.org/10.1002/tea.3660280409>
- Barbero García, M. I., Vila Abad, E., & Holgado Tello, F. P. (2015). *Psicometría*, 489. <https://doi.org/10.1017/CBO9781107415324.004>
- Barbosa, W. de A., & Miki, P. da S. R. (2007). *Metodologia da Pesquisa*. Manaus: UEA Edições.
- Barros, S. de S. (1994). Pontas de prova para o diagnostico da aprendizagem de fisica na escola: um desafio para o professor. In *IV EPEF - Grupos de Trabalhos*.
- Barton, M. A., & Lord, F. M. (1981). An Upper Asymptote For The Three- Parameter Logistic Item- Response Model. *Educational Testing Service Research Report Series*, (1), 1–8. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1981.tb01255.x/full>
- Baser, M. (1996). *Effect of conceptual change instruction on understanding of heat and*

temperature concepts and science attitude. Unpublished MS Thesis, METU, Ankara, Turkey. Retrieved from <https://scholar.google.pt/scholar?hl=pt-PT&q=Effect+of+Conceptual+Change+Instruction+on+Understanding+of+Heat+and+Temperature+Concepts+and+Science+Attitude&btnG=&lr=>

Baser, M. (2006). Fostering conceptual change by cognitive conflict based instruction on students' understanding of heat and temperature concepts. *Eurasia Journal of Mathematics, Science And*, 2(2), 96–114. Retrieved from <http://www.ejmste.com/022006/d6.pdf>

Başer, M. (2006). Effect of conceptual change oriented instruction on remediation of students' misconceptions related to heat and temperature concepts. *Journal of Maltese Education Research*, 4(1), 64–79. [https://doi.org/10.1061/\(ASCE\)0733-9410\(1991\)117](https://doi.org/10.1061/(ASCE)0733-9410(1991)117)

Batenburg, T. Van, & Laros, J. (2002). Graphical Analysis of Test Items. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 8(3), 319–333. Retrieved from https://www.researchgate.net/profile/Jacob_Laros/publication/232703390_Graphical_Analysis_of_Test_Items/links/09e41508afd209bf9e000000.pdf

Batista, R. (2007). *História da ciência: investigação do tema em livros didáticos do ensino fundamental*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/90347>

Bazarian, J. (1994). *O problema da verdade: teoria do conhecimento* (4ª edição). Alfa e ômega.

Beaton, A., & Allen, N. (1992). Chapter 6: Interpreting Scales Through Scale Anchoring. *Journal of Educational and Behavioral Statistics*, 17(2), 191–204. <https://doi.org/10.3102/10769986017002191>

Beaujean, A. A. (2014). *Latent Variable Modeling Using R. Latent variable modeling using R: A step-by-step guide*. New York: Routledge.

<https://doi.org/10.4324/9781315869780>

Becker, L. A. (2000). *Effect Size (ES)*.

Bernardo, F. (2015). *Análise praxeológica de tópicos de física moderna em livros didáticos do Programa Nacional do Livro Didático*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/160748>

Biggs, J., & Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy. A Handbook for Teaching and Learning in Higher Education: Enhancing Academic Practice*. <https://doi.org/10.1177/089202068700100412>

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1988). *Taxionomia de Objetivos Educacionais. Flávia Maria Sant'Anna (trad.)* (9ª). Rio de Janeiro: Globo.

Borgatto, A., & Andrade, D. de. (2012). Análise Clássica de Testes com diferentes graus de dificuldade. *Estudos Em Avaliação*, 23(52), 146–156. Retrieved from <http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/1934>

Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Adaptação e Validação de Instrumentos Psicológicos entre Culturas: Algumas Considerações. *Paidéia*, 22(53), 423–432. <https://doi.org/http://dx.doi.org/10.1590/1982-43272253201314> Cross-Cultural

Bortolotti, S. (2003). Aplicação de um modelo de desdobramento graduado generalizado da teoria da resposta ao ítem-TRI. Retrieved from <https://repositorio.ufsc.br/handle/123456789/85634>

Braga, M. (2010). *Proposta Metodológica Experimental Demonstrativa Por Investigação: Contribuições para o Ensino da Física na Termologia*. Universidade do estado do Amazonas. Retrieved from <http://www.pos.uea.edu.br/data/area/titulado/download/16-9.pdf>

- Braga, M. B. P., & Kalhil, J. B. (2010). Conceitos microscópicos como necessidade do conhecimento macroscópico na terminologia. *Revista Cubana de Física*, 27(2A), 184-187.
- Braga, M., & Kalhil, J. (2011). As demonstrações investigativas como processo válido na aquisição e retenção significativa de conceitos na física, 1-4.
- Brasil. (1999). *Parecer CNE Nº 16/99 – CEB*. Brasília.
- Brasil. (2000). *Parâmetros curriculares nacionais: Ensino Médio (Bases Legais)*. MEC (Vol. 1). <https://doi.org/10.1017/CBO9781107415324.004>
- Brasil. (2012a). *Entenda a sua nota no ENEM: Guia do participante*. Brasília.
- Brasil. (2012b). *Nota Técnica: Teoria de Resposta ao Item*. Brasília.
- Brasil. (2014). *Plano Nacional de Educação 2014-2024*. Brasília. Retrieved from <http://www.observatoriodopne.org.br/uploads/reference/file/439/documento-referencia.pdf>
- Brasil. (2015). *Deolutivas Pedagógicas das Avaliações de Larga Escala: Fundamentação Teórica e Metodológica*. Brasília.
- Brasil. (2014). *Guia de livros Didáticos: PNLD 2015: Física*. 2014. Brasília. Retrieved from <http://www.fnde.gov.br/programas/livro-didatico/guias-do-pnld/item/5940-guia-pnld-2015>
- Brigo, J. (2010). *As Figuras geométricas no ensino de matemática: uma análise histórica nos livros didáticos*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/94507>
- Brotti, M. (2004). Modelo de avaliação do desempenho da administração da escola como organização sob os prismas dos critérios de eficiência, eficácia, efetividade e relevância. Retrieved from <https://repositorio.ufsc.br/handle/123456789/87437>

- Brown, B. R. (2015). *Developing and assessing research-based tools for teaching quantum mechanics and thermodynamics*.
- Cachapuz, A., Gil-Pérez, D., Carvalho, A., & Praia, J. (2005). *A necessária renovação do ensino das ciências*. São Paulo: Cortez. Retrieved from <https://scholar.google.pt/scholar?hl=pt-PT&q=a+necessária+renovação+do+ensino+das+ciências+-+ResearchGate&btnG=&lr=>
- Caldeira, M. H., & Martins, D. R. (1990). Calor e temperatura: Que noção têm os alunos universitários destes conceitos? *Gazeta de Física*, 13(2), 85-94.
- Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313-337. <https://doi.org/10.1007/s11165-009-9122-4>
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. *Handbook of research on teaching*. Boston, Dallas, Geneva, Hopewell, Palo Alto, London: Houghton Mifflin Company.
- Capelo, A., & Pedrosa, M. A. (2012). Educação para desenvolvimento sustentável e práticas interdisciplinares. In *Para aprender com a Terra: memórias e notícias de Geociências no espaço lusófono*. Coimbra: Imprensa da Universidade de Coimbra. https://doi.org/http://dx.doi.org/10.14195/978-989-26-0533-3_5
- Carrascosa, J. (1983). Errores conceptuales en la enseñanza de la ciencias: selección bibliográfica. *Enseñanza de Las Ciencias* \, 1(1), 63-65.
- Carrascosa, J. (1985). Errores conceptuales en la enseñanza de la física y la química: una revisión bibliográfica. *Enseñanza de Las Ciencias*.
- Carvalho, A. M. P. de, Nardi, R., Vianna, D. M., Almeida, M. J. P. M. de, & Ferracioli, L. (2011). Pesquisa em Ensino de Física. In S. Nogueira & T. Romero (Eds.), *Física 2011 - Estado da arte, desafios e perspectivas para os próximos cinco anos* (1st ed., pp.

115–126). São Paulo: McHilliard Editora.

- Carvalho, O. C. (2006). *Avaliação do desenvolvimento de competências e habilidades: uma proposta de arquitetura de sistema de controle acadêmico para feedback de aprendizagem, em curso de graduação em administração*. Florianópolis, SC. Retrieved from <https://repositorio.ufsc.br/handle/123456789/88893>
- Carvalho, S., Pinto-Gouveia, J., & Pimentel, P. (2011). Propriedades Psicométricas da Escala de Cuidado do Questionário de Experiências de Cuidado e Abuso na Infância (Childhood Experiences of Care and Abuse–CECA. Q). *Psychologica*, 54, 359–383. Retrieved from <http://impactum-journals.uc.pt/index.php/psychologica/article/view/1112>
- Casagrande, G. de L. (2006). *A genética humana no livro didático de biologia*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/88524>
- Castro, S. de J. (2008). *Teoria da resposta ao item: aplicação na avaliação da intensidade de sintomas depressivos*. (Tese de Doutorado). Universidade Federal do Rio Grande do Sul. Retrieved from <http://www.lume.ufrgs.br/handle/10183/17457>
- Caum, C. (2013). *Um livro didático na circulação de sentidos sobre energia*. Dissertação de Mestrado. Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/107407>
- Cerdá, E. (1984). *Psicometría General*. Barcelona.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29.
- Champagne, A. (1992). Cognitive research on thinking in academic science and mathematics: Implications for practice and policy. In *Enhancing thinking skills in the sciences and*. Retrieved from <https://www.google.com/books?hl=pt-PT&lr=&id=ml655fj3rv8C&oi=fnd&pg=PA117&dq=champagne+1992+cognitive+research&ots=i8cO92Vtu3&sig=oHy-Gm8pMBk5OEZD-TSsD7ZufpI>

- Chan, C. C., Tsui, M. S., Chan, M. Y. C., & Hong, J. H. (2002). Applying the Structure of the Observed Learning Outcomes (SOLO) Taxonomy on Student's Learning Outcomes: An empirical study. *Assessment & Evaluation in Higher Education*, 27(6), 511–527. <https://doi.org/10.1080/0260293022000020282>
- Cheung, F., Iliescu, D., & Bartram, D. (2014a). *Declaração do ITC sobre o uso de testes e outros instrumentos de avaliação para fins de investigação científica*. Campinas, São Paulo.
- Cheung, F., Iliescu, D., & Bartram, D. (2014b). *ITC Statement On the Use of Tests and Other Assessment Instruments for Research Purposes*. Campinas. Retrieved from https://www.intestcom.org/files/statement_using_tests_for_research.pdf
- Cheung, F., Iliescu, D., & Bartram, D. (2014c). *ITC Statement On the Use of Tests and Other Assessment Instruments for Research Purposes (Final Version)*. INTERNATIONAL TEST COMMISSION. Campinas. Retrieved from https://www.intestcom.org/files/statement_using_tests_for_research.pdf
- Choppin, B. H., & Postlethwaite, T. N. (1979). *Evaluation in Education* (Vol. 1). Pergamon Press.
- Christophe, M., Elacqua, G., Martinez, M., & Araujo e Oliveira, J. B. (2015). *Educação Baseada em Evidências: Como saber o que funciona em educação*. Brasília: Instituto Alfa e Be.
- Chu, H. E., Treagust, D. F., Yeo, S., & Zadnik, M. (2012). Evaluation of students' understanding of thermal concepts in everyday contexts. *International Journal of Science Education*, 34(10), 1509–1534. <https://doi.org/10.1080/09500693.2012.657714>
- Cigdemoglu, C., & Geban, O. (2015). Improving students' chemical literacy levels on thermochemical and thermodynamics concepts through a context-based approach. *Chemistry Education Research and Practice*. Retrieved from <http://pubs.rsc.org/-/content/articlehtml/2015/rp/c5rp00007f>

- Cindra, J., & Teixeira, O. (2004). Discussão conceitual para o equilíbrio térmico. *Caderno Brasileiro de Ensino de Física*, 21(2), 176–193. Retrieved from [http://files.juliocesarmota.webnode.pt/200000027-a92e3aa286/Discussão Conceitual para o Equilíbrio Térmico.pdf](http://files.juliocesarmota.webnode.pt/200000027-a92e3aa286/Discussão%20Conceitual%20para%20o%20Equilíbrio%20Térmico.pdf)
- Clark Midkiff. (2000). Thermodynamics Concept Inventory (TCI). Retrieved September 16, 2017, from <https://www.physport.org/assessments/assessment.cfm?I=18&A=TCI>
- Clough, E., & Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/sce.3730700412/full>
- CNTE. (2017). *BNCC do MEC golpista, não!*
- Cobern, W., Schuster, D., & Adams, B. (2014). Pedagogy of Science Teaching Tests: Formative assessments of science teaching orientations. *International Journal of Science Education*, 36(13), 2265–2288. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/09500693.2014.918672>
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences, Revised Edition*. *Journal of the American Statistical Association* (First edit, Vol. 73). New York. <https://doi.org/10.2307/2286629>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second Edi). New York: Lawrence Erlbaum Associates.
- Cohen, R., Swerdlik, M., & Sturman, E. (2014). *Testagem e Avaliação Psicológica: Introdução a Testes e Medidas*. Porto Alegre: AMGH. Retrieved from https://scholar.google.pt/scholar?q=Cohen%2C+Swerdlik+e+Sturman+%282014%29&btnG=&hl=pt-PT&as_sdt=0%2C5
- Coletta, V. P., & Phillips, J. a. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics*,

73(12), 1172–1182. <https://doi.org/10.1119/1.2117109>

Condé, F. N. (2008). *RELAÇÃO ENTRE CARACTERÍSTICAS DO TESTE EDUCACIONAL E ESTIMATIVA DE HABILIDADE DO ESTUDANTE*. Tese de doutorado). Universidade de Brasília.

Congress of the United States of America. (2017). *S. 1177 – 114th Congress: Every student succeeds act*. Retrieved from. Washington.

Costa, C. (2015). *Análise da dimensionalidade e modelagem multidimensional pela TRI no Enem (1998-2008)*. Retrieved from <https://repositorio.ufsc.br/handle/123456789/169527>

Creswell, J., & Clark, V. (2007). *Designing and conducting mixed methods research*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1753-6405.2007.00097.x/full>

Creswell, J., & Clark, V. (2015). *Pesquisa de Métodos Mistos: Série Métodos de Pesquisa*. Penso. Retrieved from https://scholar.google.pt/scholar?hl=pt-PT&as_sdt=0%2C5&q=creswell+clark+Pesquisa+de+Métodos+Mistos&btnG=

Creswell, J. W. (2010). *Projeto de pesquisa métodos qualitativo, quantitativo e misto*. (Magda Lopes (trad.), Ed.), *Projeto de pesquisa métodos qualitativo, quantitativo e misto* (3ª). Porto Alegre: Artmed. Retrieved from <http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=LILACS&lang=p&nextAction=lnk&exprSearch=695513&indexSearch=ID>

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*. Retrieved from <http://psycnet.apa.org/journals/bul/52/4/281/>

Crossley, A. (2010). THE IMPACT OF PHYSICAL EXPLANATIONS ON SECONDARY STUDENTS ' KNOWLEDGE ACQUISITION IN THERMODYNAMICS.

- Cubbillos, S. P. Z. (1981). *Influência do conhecimento prévio sobre o desempenho do aluno em um curso de física geral individualizado*. (Dissertação de mestrado). Universidade Federal do Rio Grande do Sul.
- Cumberland, W. N., Fong, Y., Yu, X., Defawe, O., Frahm, N., & De Rosa, S. (2015). Nonlinear Calibration Model Choice between the Four and Five-Parameter Logistic Models. *Journal of Biopharmaceutical Statistics*, 25(5), 972–983.
<https://doi.org/10.1080/10543406.2014.920345>
- Cunha, A. I., Major, S., & Relvas, A. P. (2016). Coping Health Inventory for Parents (CHIP). In *Avaliação familiar: vulnerabilidade, stress e adaptação vol. II*. Coimbra: Imprensa da Universidade de Coimbra.
https://doi.org/https://doi.org/10.14195/978-989-26-1268-3_7
- Custódio, F. L. (2012). A utilização de testes conceituais em Física Básica, 110.
- Custódio, F. L., & Barroso, M. F. (2012). Testes Conceituais em Física Básica: Apresentação e análise dos itens.
- Da Silveira, F. L., Moreira, M. A., & Axt, R. (1989). Validação de um teste para verificar se o aluno possui concepções científicas sobre corrente elétrica em circuito simples. *Ciência & Cultura*, 41(11), 1129–1133.
- Damião, M. H. (2015). Desenvolvimento de competências ou transmissão de conhecimento: acerca da necessidade de superar uma antinomia curricular no Ensino Universitário. *Revista Portuguesa de Pedagogia*, 49(1).
https://doi.org/http://dx.doi.org/10.14195/1647-8614_49-1_3
- Darling-Hammond, L., & Rothman, R. (2011). *Developing effective teachers and school leaders: The case of Finland*. Washington: Alliance for Excellent Education and the Stanford Center for Opportunity Policy in Education. Retrieved from http://admin.kasa.org/Professional_Development/KLA/documents/TeacherLeaderEffectivenessReport.pdf#page=19
- David, M., & Irribarra, T. (2016). Package 'WrightMap.'

- David, R., Paixão, M. P., & Silva, J. T. da. (2009). Interesses e competências percebidas em crianças do ensino básico. *Psychologica*, 51. Retrieved from [https://digitalis.uc.pt/pt-pt/artigo/interesses_e_competências_percebidas_em_crianças_do_ensino_básico](https://digitalis.uc.pt/pt-pt/artigo/interesses_e_competencias_percebidas_em_crianças_do_ensino_básico)
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: The Guilford Press.
- de Carvalho, A. (2004). *Ensino de Ciências: unindo a pesquisa e a prática*. Cengage Learning Editores. Retrieved from <https://www.google.com/books?hl=pt-PT&lr=&id=VI4DGUzL0j0C&oi=fnd&pg=PA2&dq=Ensino+de+Ciências:+unindo+a+pesquisa+e+a+prática.&ots=ib6s20n6Xk&sig=u2jcieEyMbXtoUPcf69bFi8SNpU>
- De Ketele, J.-M. (2006). Caminhos para a avaliação de competências. *Revista Portuguesa de Pedagogia*, 40(3). Retrieved from <http://impactum-journals.uc.pt/index.php/rppedagogia/article/view/1172>
- De Ketele, J.-M. (2008). Enfoque socio-histórico de las competencias en la enseñanza. *Profesorado. Revista de Currículum y Formación de Profesorado*.
- De Ketele, J.-M. (2004). Évaluer des Compétences, ou ne pas se tromper de cible. *Forum American Bar Association*, 7-9.
- De Ketele, J.-M. (2010). La recherche scientifique en éducation : Quels critères de qualité ? *Education, Sciences & Society*.
- De Ketele, J.-M. (1986). *L'évaluation: approche descriptive ou prescriptive?* Retrieved from <http://eduq.info/xmlui/handle/11515/4576>
- De Ketele, J.-M., & Roegiers, X. (1993). *Metodologia da recolha de dados: fundamentos dos métodos de observações, de questionários, de entrevistas, e de estudo de documentos*. (trad. C. Brito), Ed. Epistemologia e Sociedade.

- De Vecchi, G. (1992). *Aider les élèves à apprendre. Pédagogie Pour Demain, Hachette, Education, Paris.*
- Deluiz, N. (2001). O modelo das competências profissionais no mundo do trabalho e na educação: implicações para o currículo. *Boletim Técnico Do SENAC, 27(3)*, 13–25. Retrieved from <http://www.senac.br/INFORMATIVO/BTS/273/boltec273b.htm>
- DeMars, C. (2010). *Item response theory. Understanding statistics measurement.* New York, NY: Oxford University Press.
- Dentz, V. Von, & Truccolo, F. (2013). Mapeamento de pesquisas (teses e dissertações) sobre o Ensino de Ciências da Natureza (Física, Química e Biologia) nos níveis fundamental e médio. *Revista Técnico Científica Do IFSC, 1(2)*, 91. Retrieved from <http://periodicos.ifsc.edu.br/index.php/rtc/article/view/973>
- Dias, I. S. (2010). Competências em educação: conceito e significado pedagógico. *Psicologia Escolar e Educacional (Impresso), 14(1)*, 73–78. <https://doi.org/10.1590/S1413-85572010000100008>
- Domínguez, M. (1985). *Detecção de alguns conceitos intuitivos em eletricidade através de entrevistas clínicas.* Dissertação de Mestrado. Instituto de Física da UFRGS, Porto Alegre, 1985. Retrieved from https://scholar.google.pt/scholar?q=domínguez+1985+detecção+de+alguns+conceitos+intuitivos&btnG=&hl=pt-PT&as_sdt=0%2C5
- Dominguini, L., & Silva, I. B. da. (2010). Obstáculos a Construção do Espírito Científico: Reflexões sobre o Livro Didático. In *V CINFE* (pp. 1–15).
- Dreyfuss, H., & Dreyfuss, S. (1986). *Mind over machine. The Power of Human Intuition and Expertise in the Era.* Oxford: Free Press.
- Dreyfuss, H., Dreyfuss, S., Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The Power of Human Intuition and Expertise in the Era of the Computer. The Power of Human Intuition and Expertise in the Era* (New York). Oxford: The Free Press.

<https://doi.org/10.1017/CBO9781107415324.004>

- Driver, R. (1973). *The representation of conceptual frameworks in young adolescent science students*. (PhD Thesis). University Illinois, Urbana, . Retrieved from <http://psycnet.apa.org/psycinfo/1976-13289-001>
- Driver, R. (1983). *The Pupil as Scientist?* Philadelphia, USA: Open University Press.
- Driver, R. (1985). Children's ideas in science. Retrieved from https://www.google.com/books?hl=pt-PT&lr=&id=GS_1AAAAQBAJ&oi=fnd&pg=PP1&dq=driver+guesne+tiberghien&ots=OOMasKIepB&sig=HIUole0PYTB3eA_y6SSDd1AfskY
- Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education*, 11(5), 481–490.
<https://doi.org/10.1080/0950069890110501>
- Duarte, L. P. A. de A., Gonçalves Júnior, W. P., & Barroso, M. F. (2013). Estudo de um Item Quantitativo na Prova de Física do ENEM 2009. In *XX Simpósio Nacional de Ensino de Física* (pp. 1–9). São Paulo.
- Duit, R. (1991). Students' conceptual frameworks: Consequences for learning science. *The Psychology of Learning Science*. Retrieved from https://www.google.com/books?hl=pt-PT&lr=&id=NZ5lceOc9g0C&oi=fnd&pg=PA65&dq=duit+1991&ots=Oxt7vde_ZV&sig=tDb0EOKy-2aeMF_OZVTV0HZ2u6Q
- Duit, R. (1993). Third Misconceptions Seminar Proceedings. In *In Proceedings of the Second International Seminar: Misconceptions and Educational Strategies in Science and Mathematics* (Vol. III, pp. 111–115). Retrieved from http://www.mlrg.org/proc3pdfs/Echeverria_Solution.pdf
- Duit, R., & Treagust, D. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*. Retrieved from

<http://www.tandfonline.com/doi/pdf/10.1080/09500690305016>

Dworkin, R. (1986). *Law's empire*. Retrieved from

https://www.google.com/books?hl=pt-PT&lr=&id=8TCx_UbbI3AC&oi=fnd&pg=PA1&ots=CNGPCT5j1Q&sig=dzGCBZRdJ5NjKzO723xCLrgpoKY

Ebel, R. (1962). Content standard test scores. *Educational and Psychological*

Measurement. Retrieved from

<http://journals.sagepub.com/doi/pdf/10.1177/001316446202200103>

Ebel, R. L. (1965). *Measuring educational achievement*. Englewood Cliffs.

Einhaus, E. (2007). Schülerkompetenzen im Bereich Wärmelehre: Entwicklung eines

Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells. Retrieved from

https://scholar.google.pt/scholar?q=+Schülerkompetenzen+im+Bereich+Wärmelehre+-+Entwicklung+eines+Testinstruments+zur+Überprüfung+und+Weiterentwicklung+eines+normativen+Modells+fachbezogener+Kompetenzen&btnG=&hl=pt-PT&as_sdt=0%2C5

Embretson, S. (1994). Applications of cognitive design systems to test development.

In *Cognitive Assessment*. Retrieved from

http://link.springer.com/chapter/10.1007/978-1-4757-9730-5_6

Engelhardt, V. P. (1997). *Examining students' understanding of electrical circuits through*

multiple-choice testing and interviews. (Thesis of PhD). North Carolina State University.

Engelhardt, P. V. (2009). An Introduction to Classical Test Theory as Applied to

Conceptual Multiple-choice Tests. *Getting Started in PER*, 1-40. Retrieved from

<http://www.per-central.org/items/detail.cfm?ID=8807>

Erickson, G. L. (1979). Children's conceptions of heat and temperature. *Science*

- Education*, 63(2), 221–230. <https://doi.org/10.1002/sce.3730630210>
- Erickson, G. L. (1980). Children's viewpoints of heat: A second look. *Science Education*, 64(3), 323–336. <https://doi.org/10.1002/sce.3730640307>
- Erthal, T. C. (1987). *Manual de Psicometria*. Rio de Janeiro: Zahar.
- Eryilmaz, A. (2010). Development and application of three-tier heat and temperature test: Sample of bachelor and students graduate. *Egitim Arastirmalari - Eurasian Journal of Educational Research*, (40), 53–76.
- Espírito-Santo, H., & Daniel, F. (2015). Calcular e apresentar tamanhos do efeito em trabalhos científicos: As limitações do $p < 0,05$ na análise de diferenças de médias de dois grupos. *Revista Portuguesa de Investigação Comportamental e Social*, 1(1), 3–16. <https://doi.org/10.5965/1984723816352016153>
- Farias, A. J. O. (1982). *Mapeamento cognitivo em um curso individualizado: um estudo sobre o efeito da abordagem ao conteúdo*. (Dissertação de mestrado). Universidade Federal do Rio Grande do Sul.
- Farias, T. (2013). *Atividades experimentais: contribuições para o aprendizado dos conceitos de calor e de temperatura*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/107370>
- Fávero, M., & Sousa, C. de. (2001). A Resolução de Problemas em Física: Revisão de Pesquisa, Análise e Proposta Metodológica (Problem solving in physics: research review, analysis, and a methodological proposal). *Investigações Em Ensino de Ciências*, 6(2), 143–196. Retrieved from http://www.if.ufrgs.br/public/ienci/artigos/Artigo_ID73/v6_n2_a2001.pdf
- Fechner, G. T. (1860). *Elemente der Psychophysik*. <https://doi.org/10.15713/ins.mmj.3>
- Fernandes, F. (2008). *Proposta de um modelo logístico de três parâmetros a partir da teoria da resposta ao item: uma aplicação em dados do vestibular da universidade federal de lavras*. (Dissertação de Mestrado). Universidade Federal de Lavras. Retrieved

from <http://repositorio.ufla.br/handle/1/3292>

- Fernandez, F. (2017). Action research in the physics classroom: the impact of authentic, inquiry based learning or instruction on the learning of thermal physics. *Asia-Pacific Science Education*. Retrieved from <https://apse-journal.springeropen.com/articles/10.1186/s41029-017-0014-z>
- Ferreira, A. (2001). Elementos fundamentais para a compreensão do estudo da Educação Comparada. *Diálogos Em Educação*. Retrieved from https://scholar.google.pt/scholar?q=Elementos+fundamentais+para+a+compreensao+do+estudo+da+Educacao+Comparad&btnG=&hl=pt-PT&as_sdt=0%2C5
- Ferreira, F. C., Caíres, a. R. L., Silva, a. a., & Oliveira, S. L. (2009). Diagnóstico de dificuldades conceituais em física apresentadas por acadêmicos ingressantes em cursos da UFGD. *VII Encontro Nacional de Pesquisa Em Educação Em Ciências (ENPEC)*.
- Finch, W. H., & French, B. F. (2015). *Latent Variable Modeling with R*. New York, NY: Routledge.
- Findley, W. G. (1956). A rationale for Evaluation of Item Discrimination Statistics. *Educational Testing Service Research Report Series*, 175–180.
- Flanagan, J. (1951). Units, scores, and norms. *Educational Measurement*.
- Fleischer, J. (2017). *Competence Assessment in Education*. <https://doi.org/10.1007/978-3-319-50030-0>
- Fonseca, A. (2010). Consumo de álcool e seus efeitos no desempenho escolar. *Revista Portuguesa de Pedagogia*, 44(1).
https://doi.org/http://dx.doi.org/10.14195/1647-8614_44-1_11
- Fonte, A., & Coutinho, B. (2015). Questionário de Avaliação do Perfil Sazonal: características psicométricas da versão portuguesa do Seasonal Pattern Assessment Questionnaire (SPAQ). *Psychologica*, (58, vol. 1).

https://doi.org/http://dx.doi.org/10.14195/1647-8606_58-1_5

Fukuhara, H., & Kamata, A. (2011). A bifactor multidimensional item response theory model for differential item functioning analysis on testlet-based items. *Applied Psychological Measurement, 35*(8), 604–622.

<https://doi.org/10.1177/0146621611428447>

Gabriel, C. G. (2008). *Alimentos e bebidas comercializados em escolas de oito municípios de Santa Catarina após a instituição da lei estadual das cantinas*. Retrieved from <https://repositorio.ufsc.br/handle/123456789/91998>

García Díaz, J. E. (1998). Hacia una Teoría Alternativa sobre Los Contenidos Escolares. In *Colección Investigación y Enseñanza (Serie Fund)*. Sevilla: Díada.

Garcia, M. I. B., Abad, E. V., & Tello, F. P. H. (2000). *Formulario de psicometría y tablas estadísticas*. Madrid: Sanz y Torres.

Georgiou, H. (2014). *Doing Positive Work : On student understanding of thermodynamics*. (Thesis of Doctor of Philosophy). University of Sydney.

Georgiou, H., Sharma, M., Byrne, J. O., Sefton, I., & Mcinnes, B. (2009). University students' conceptions about familiar thermodynamic processes and the implications for instruction, (1991), 1–7.

Georgiou, H., & Sharma, M. D. (2010). A Report on a Preliminary Diagnostic for Identifying Thermal Physics Conceptions of Tertiary Students. *International Journal of Innovation in Science and Mathematics Education (Formerly CAL-Laborate International), 18*(2), 32–51. Retrieved from <http://openjournals.library.usyd.edu.au/index.php/CAL/article/view/4700>

Georgiou, H., & Sharma, M. D. (2012). University Students' Understanding of Thermal Physics in Everyday Contexts. *International Journal of Science and Mathematics Education, 10*(5), 1119–1142. <https://doi.org/10.1007/s10763-011-9320-1>

- Gérard, F.-M., & Roegiers, X. (1998). *Conceber e avaliar manuais escolares. Ciências da educação* 30.
- Gérard, F. (2010). Le manuel scolaire, un outil efficace, mais décrié. *Education & Formation, e-292*, 13–24. Retrieved from http://www.istres.ien.13.ac-aix-marseille.fr/spip/IMG/pdf/Manuels_scolaires_et_materiel_didactique.pdf#page=13
- Gerard, F., Peyser, A., & Roegiers, X. (2006). Implementing pedagogy of integration: Some thoughts based on a textbook elaboration experience in Vietnam. *Planning and Changing, 37*(1), 37–55.
- Giordan, A., ASTOLFI, J., GOHAU, G., & HOST, V. (1978). *Quelle éducation scientifique pour quelle société. Paris: PUF.*
- Giordan, A., & Vecchi, G. De. (1987). *Les origines du savoir.* Retrieved from <http://www.ldes.unige.ch/publi/livres/couvOriginesSavoir.pdf>
- Giraldi, P. (2005). *Linguagem em textos didáticos de citologia: investigando o uso de analogias.* (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/102533>
- Glas, C. A. W. (2016). Frequentist Model-Fit Tests. In W. J. van der Linden (Ed.), *Handbook of Item Response Theory, Volume Two: Statistical Tools* (pp. 313–361). Boca Raton, FL: CRC Press.
- Glaser, R. (1963). Instructional Technology and the Measurement of Learning Outcomes: Some Questions. *American Psychologist, 18*(8), 519–522. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.1994.tb00561.x/full>
- Gomes, C. A., & Borges, O. (2009). Qualidades psicométricas do conjunto de testes de inteligência fluida. *Avaliação Psicológica.* Retrieved from <http://www.redalyc.org/html/3350/335027279003/>
- Gomes, D. (2014). *Criação de uma escala para avaliar a percepção dos funcionários de*

hospitais universitários no Brasil quanto ao ambiente organizacional. UNIVERSIDADE FEDERAL DE SANTA CATARINA. Retrieved from <https://repositorio.ufsc.br/handle/123456789/129056>

Gomes, H. J. P., & Oliveira, O. B. De. (2007). Obstáculos epistemológicos no ensino de ciências: um estudo sobre suas influências nas concepções de átomo. *Ciências & Cognição*, 12, 96–109.

Gómez, A. L.-B., & Hernández, A. S. (2010). Detección y Clasificación de Errores Conceptuales en Calor y Temperatura. *Lat. Am. J. Phys. Educ.*, 4(2), 399–407. Retrieved from <https://dialnet.unirioja.es/descarga/articulo/3696425.pdf>

Gonçalves, E. (1979). *Laboratório Estruturado versus não estruturado: estudo comparativo em um curso de Física Geral*. Porto Alegre: UFRGS–Instituto de Física. (Dissertação de mestrado). Universidade Federal do Rio Grande do Sul. Retrieved from https://scholar.google.pt/scholar?q=Laboratório+estruturado+versus+não-estruturado%3A+estudo+comparativo+em+um+curso+de+Física+Geral&btnG=&hl=pt-PT&as_sdt=0%2C5

Gonçalves, J. M. L. (2011). O uso do manual escolar enquanto recurso promotor do desenvolvimento de competências históricas relatório final o manual escolar enquanto recurso promotor do desenvolvimento de competências históricas, (3), 110. Retrieved from <http://www.google.pt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=12&ved=0CCYQFjABOAO&url=http://repositorio-aberto.up.pt/bitstream/10216/57058/2/TESEMESJOANAGONALVES000141874.pdf&ei=19J1VZjDI4TAmwXEx4HoAg&usg=AFQjCNE6l3VfZLpGgXIxWwsdhaJsbD7Ltw>

Gonçalves Júnior, W. P. (2012). *A Física no ENEM: Primeiras análises*. Rio de Janeiro.

Gonçalves Júnior, W. P. (2012). *Avaliações em larga escala e o professor de Física*. Universidade Federal do Rio de Janeiro.

- Gonçalves Júnior, W. P., & Barroso, M. F. (2011a). Conceitos de Física térmica: Estudo por meio de um questionário, 2009–2011.
- Gonçalves Júnior, W. P., & Barroso, M. F. (2011b). O que alunos entendem a respeito de conceitos básicos da Física térmica, 1–8.
- Gonçalves Júnior, W. P., & Barroso, M. F. (2012). ENEM: Os itens e o desempenho dos estudantes em 2009 (pp. 1–11).
- Gonçalves Júnior, W. P., & Barroso, M. F. (2014a). As questões de Física e o desempenho dos estudantes no ENEM. *Revista Brasileira de Ensino de Física*, 36, 1402-1-16. <https://doi.org/10.1590/S1806-11172014000100017>
- Gonçalves Júnior, W. P., & Barroso, M. F. (2014b). Identificando as concepções não científicas que dificultam a compreensão do fenômeno de mudança de fase. In *XV Encontro de Pesquisa em Ensino de Física* (pp. 1–6). Maresias, São Paulo.
- Gonçalves, W., & Barroso, M. F. (2011). O que alunos entendem a respeito de conceitos básicos da Física térmica. In *XIX Simpósio Nacional de Ensino de Física*. Manaus.
- Gonnelli Netto, O. (2014). *Análise da didatização do tema radiação de corpo negro sob a luz da teoria antropológica do didático*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/123351>
- Gorri, A. (2014). *Análise semiótica de representações moleculares na comunicação de conhecimentos sobre ácidos e bases em livros-texto de química orgânica: primeira e segunda*. Dissertação de Mestrado. Universidade Federal de Santa Catarina. <https://doi.org/Dissertação de Mestrado>
- Gottschalk, P. G., & Dunn, J. R. (2005). The five-parameter logistic: A characterization and comparison with the four-parameter logistic. *Analytical Biochemistry*, 343(1), 54–65. <https://doi.org/10.1016/j.ab.2005.04.035>

- Graeff, J. (2011). *Sensemaking e sensegiving: a construção da lógica institucional do campo da agricultura familiar na Região Metropolitana de Curitiba/PR no período de 1990-2010*. (Tese de doutorado). Universidade Federal do Paraná. Retrieved from <http://acervodigital.ufpr.br/handle/1884/27191>
- Gramowski, V. (2014). *O livro didático de ciências: a persistência da fragmentação dos conteúdos*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/129522>
- Grings, E. de O., Caballero, C., & Moreira, M. (2006). Possíveis indicadores de invariantes operatórios apresentados por estudantes em conceitos da termodinâmica. *Revista Brasileira de Ensino de Física*, 28(4), 463–471. Retrieved from <http://www.scielo.br/pdf/%0D/rbef/v28n4/a09v28n4.pdf>
- Grings, E. T. de O., Caballero, C., & Moreira, M. A. (2008). Avanços e retrocessos dos alunos no campo conceitual da Termodinâmica. *Revista Electrónica de Enseñanza de Las Ciencias*, 7(1), 23–46.
- Guerra, W. (1983). *Mapas conceituais como instrumentos para investigar a estrutura cognitiva em Física. Mapas conceituais como instrumentos para investigar*. (Dissertação de mestrado). Universidade Federal do Rio Grande do Sul.
- Guewehr, K. (2007). *Teoria da resposta ao item na avaliação de qualidade de vida de idosos*. (Dissertação de Mestrado Epidemiologia). Universidade Federal do Rio Grande do Sul.
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology*. Retrieved from <http://www.sciencedirect.com/science/article/pii/089543569390142N>
- Guisbond, L., Neill, M., & Schaeffer, B. (2012). A década de progresso educativo perdida sob a NCLB: que lições tirar desse fracasso político? *Educação & Sociedade*, 33(119). Retrieved from

<http://www.redalyc.org/html/873/87323122005/>

Gulliksen, H. (1950). *Theory of mental tests*. Retrieved from

<https://www.google.com/books?hl=pt-PT&lr=&id=HAu-B6vgd38C&oi=fnd&pg=PP2&dq=Theory+of+MENTAL+TESTS&ots=Uy8xZvMEDL&sig=7VI461GuINKQ2q3eOKISb0E4fMk>

Gunstone, R. F., & White, R. T. (1981). Understanding of gravity. *Science Education*, 65(3), 291–299. <https://doi.org/10.1002/sce.3730650308>

Guy Le Boterf. (2000). La gestión por competencias. *Ideia*, 110–113.

Hake, R. (2008). Design-based research in physics education: a review. In & B. Kelly, Lesh (Ed.), *Handbook of Design Research Methods in Mathematics, Science, and Technology Education*. Erlbaum. Retrieved from https://www.researchgate.net/profile/Richard_Hake/publication/250270804_DESIGN-BASED_RESEARCH_IN_PHYSICS_EDUCATION_A_REVIEW/links/54886b1d0cf289302e30acb7.pdf

Hake, R. R. (1998a). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>

Hake, R. R. (1998b). *Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses*.

Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items* (Edition, T). Mahwah, New Jersey.

Haladyna, T. M., & Rogríguez, M. C. (2004). Developing and validating test items.

Halloun, I., & Hestenes, D. (1985a). Common sense concepts about motion. *American Journal of Physics*. Retrieved from

<http://aapt.scitation.org/doi/abs/10.1119/1.14031>

Halloun, I., & Hestenes, D. (1985b). The initial knowledge state of college physics students. *American Journal of Physics*. Retrieved from <http://aapt.scitation.org/doi/abs/10.1119/1.14030>

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410611758>

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. <https://doi.org/10.1007/978-94-017-1988-9>

Harrison, A. G., Grayson, D. J., & Treagust, D. F. (1999). Investigating a grade 11 student's evolving conceptions of heat and temperature. *Journal of Research in Science Teaching*, 36(1), 55–87. [https://doi.org/10.1002/\(SICI\)1098-2736\(199901\)36:1<55::AID-TEA5>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1098-2736(199901)36:1<55::AID-TEA5>3.0.CO;2-P)

Hasan, S., Bagayoko, D., & Kelley, E. (1999). Misconceptions and the Certainty of Response Index (CRI). *Physics Education*, 34(5), 294–299. Retrieved from https://www.researchgate.net/profile/Diola_Bagayoko/publication/241530804_Misconceptions_and_the_Certainty_of_Response_Index_CRI/links/53d2e74d0cf220632f3cc30a.pdf

Henrique, O., Nardi, R., & Laburú, C. E. (2010). Um estudo dos avanços conceituais dos estudantes sobre calor e temperatura decorrentes da aplicação de uma estratégia de ensino inspirada na teoria de Lakatos A study about students' conceptual development on heat and temperature due to the application, 1–18.

Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller. *The Physics Teacher*. Retrieved from <http://aapt.scitation.org/doi/pdf/10.1119/1.2344278>

Hestenes, D., & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*. Retrieved from <http://aapt.scitation.org/doi/pdf/10.1119/1.2343498>

- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141. <https://doi.org/10.1119/1.2343497>
- Hitt, A. M., & Townsend, S. J. (2015). The Heat Is On! Using Particle Models to Change Students' Conceptions of Heat and Temperature. *Science Activities: Classroom Projects and Curriculum Ideas*, 52(2), 45-52. <https://doi.org/10.1080/00368121.2015.1049580>
- Hively, W. (1973). Domain-Referenced Curriculum Evaluation: A Technical Handbook and a Case Study from the MINNEMAST Project. CSE Monograph Series in Evaluation, 1. Retrieved from <https://eric.ed.gov/?id=ED177226>
- Hively, W. (1974a). Domain-referenced testing. Retrieved from <https://www.google.com/books?hl=pt-PT&lr=&id=DTVxUonow5gC&oi=fnd&pg=PA3&dq=domain-referenced+testing&ots=lc6JsuTVM9&sig=UmDaauKaKQuT6GARgEDIO5EBfrc>
- Hively, W. (1974b). Introduction to domain-referenced testing. *Educational Technology*. Retrieved from <https://www.google.com/books?hl=pt-PT&lr=&id=DTVxUonow5gC&oi=fnd&pg=PA5&dq=domain-referenced+testing&ots=lc6JsuTVM9&sig=kMFm8XomWtLCZEwpLuK4bPNvwY8>
- Hively, W., & Patterson, H. (1968). A "universe- defined" system of arithmetic achievement tests. *Journal of Educational*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3984.1968.tb00639.x/full>
- Hively, W., & Reynolds, M. (1975). Domain-Referenced Testing in Special Education. Retrieved from <https://eric.ed.gov/?id=ED108410>
- Hoffmann, B. (1961). Testing. *Physics Today*, 14(10), 38-42. Retrieved from <http://adsabs.harvard.edu/abs/1961PhT....14j..38H>
- Holland, P., & Thayer, D. (1988). Differential item performance and the Mantel-Haenszel procedure. *Test Validity*, 129-145. Retrieved from

<https://www.google.com/books?hl=pt-PT&lr=&id=1i98Bl6EEZ0C&oi=fnd&pg=PA129&dq=Holland+%26+Thayer,+1988&ots=Ymr7XHSshh&sig=QzV6VH9lB6grXOo47uk0basMKCE>

Holland, P. W., & Thayer, D. T. (1985). *An Alternate Definition of the statistics*.

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation*. Retrieved from <https://eric.ed.gov/?id=EJ933682>

Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (2015). *Psicometria*. Porto Alegre: Artmed.

Inácio, M. (2009). *Aprender a aprender no processo de reconhecimento, validação e certificação de competências (RVCC): estudo de caso duplo*. (Dissertação de mestrado). Universidade de Lisboa. Retrieved from <https://repositorio.ul.pt/handle/10451/3434>

International Test Commission (ITC). (2014). *ITC Statement On the Use of Tests and Other Assessment Instruments for Research Purposes (Final Version)*. Campinas.

Izquierdo, M. (2005). Hacia una Teoría de los Contenidos Escolares. *Enseñanza de Las Ciencias*, 23(1), 111–122. Retrieved from <https://ddd.uab.cat/record/1681>

Jacobi, A., Martin, J., & Mitchell, J. (2003). A concept inventory for heat transfer. *Frontiers in Education*,. Retrieved from <http://ieeexplore.ieee.org/abstract/document/1263338/>

Jacobi, A., Martin, J., & Mitchell, J. (2004). Work in progress: A concept inventory for heat transfer. *Frontiers in Education*,. Retrieved from <http://ieeexplore.ieee.org/abstract/document/1408538/>

Jacomelli, K. (2006). *A linguagem natural e a linguagem algébrica: nos livros didáticos e em uma classe de 7a. série do ensino fundamental*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/88270>

- Jacques, V. (2008). *A energia no ensino fundamental: o livro didático e as concepções alternativas*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/91750>
- Johnson, R. B. R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7), 14–26. <https://doi.org/10.3102/0013189X033007014>
- Jorde, D., & Dillon, J. (2012). *Science education research and practice in Europe*. *Science education research and practice in Europe*. Retrieved from http://link.springer.com/chapter/10.1007/978-94-6091-900-8_1
- Joshua, S., & Dupin, J. (1993). Introduction à la didactique des sciences et des mathématiques.(p. 61-74). *France: Presses Univ. de France*.
- Jusoh, N. M., & Nor, K. (2014). Adoption of digital Games in Malaysia : Rasch Difference Item Functioning. *International Journal of Interactive Digital Media*, 2(1), 22–26.
- Justino, G. (2007). *Desenvolvimento de ambiente computacional para implementação de métodos da teoria da resposta ao item*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/90760>
- Kearney, M., Treagust, D. F., Yeo, S., & Zadnik, M. G. (2001). Student and teacher perceptions of the use of multimedia supported predict – observe – explain tasks to probe understanding. *Research in Science Education*, 31, 589–615. <https://doi.org/10.1023/A:1013106209449>
- Keevy, J., & Chakroun, B. (2015). *Level-setting and recognition of learning outcomes: The use of level descriptors in the twenty-first century*. United Nations Educational, Scientific and Cultural Organization.
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17–24.

<https://doi.org/10.1037/h0057123>

- Kesidou, S., & Duit, R. (1993). Students' conceptions of the second law of thermodynamics – an interpretive study. *Journal of Research in Science Teaching*, 30(1), 85–106. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/tea.3660300107/full>
- Köhnlein, J. (2001). *Um estudo sobre as concepções alternativas de calor e temperatura*. Trabalho de Conclusão de Curso. Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/105335>
- Köhnlein, J. F. K., & Peduzzi, L. O. Q. (2002). Sobre a concepção empirista-Indutivista no Ensino de Ciências. *VIII EPEF - Encontro de Pesquisa Em Ensino de Física*, 1–18.
- Köhnlein, J., & Peduzzi, S. (2002). Um Estudo a Respeito das Concepções Alternativas sobre Calor e Temperatura. *Revista Brasileira de Investigação Em Educação Em Ciências*, 2(3), 84–96. Retrieved from http://www.educadores.diaadia.pr.gov.br/arquivos/File/2010/artigos_teses/fisica/artigos/um_estudo.pdf
- Köhnlein, J., & Peduzzi, S. (2003). Calor e Temperatura: Uma Intervenção em Sala de Aula. In *IV Encontro Nacional de Pesquisa em Educação em Ciências*. Bauru, SP: fep.if.usp.br. Retrieved from <http://fep.if.usp.br/~profis/arquivos/ivenpec/Arquivos/Painel/PNL091.pdf>
- Kruatong, T., Sung-ong, S., Singh, P., & Jones, A. (2006). Thai high school students' understanding of heat and thermodynamics. *Kasetsart Journal - Social Sciences*, 27(2), 321–330. Retrieved from http://kasetsartjournal.ku.ac.th/kuj_files/2008/A080403160712.pdf
- Lacerda, R., Martins, C., & Reis, E. M. (2009). Mapas conceituais como instrumento de avaliação e aprendizagem de conceitos físicos sobre mecânica do voo
Resumo. *Revista Brasileira de Pesquisa Em Educação Em Ciências*, 9(1).
- Lahera, J., & Forteza, A. (2006). Ciências físicas nos ensinos fundamental e médio:

modelos e exemplos.

- Lajonquière, L. (2009). *Infância e ilusão psicopedagógica: escritos de psicanálise e educação. Psicanálise e Educação* (4ª edição). Vozes. Retrieved from <http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=ADOLEC&lang=p&nextAction=lnk&exprSearch=626055&indexSearch=ID>
- LANDSHEERE, V. DE, & LANDSHEERE, G. DE. (1978). *DEFINIR LES OBJECTIFS DE L'ÉDUCATION*.
- Lang, F. (2013). Comentários críticos sobre cinco questões de Física na prova de Ciências da Natureza – Enem 2013.
- Lang, F., & Moreira, M. A. (1996). Validación de um test para verificar si el alumno posee concepciones científicas sobre calor , temperatura y energia interna
Validação de um teste para verificar se o aluno possui concepções científicas sobre calor , temperatura e energia interna *DE UN TE*, 14(1), 75-86.
- Latour, B. (1998). *Ciência em Ação: Como seguir engenheiros e cientistas sociedade afora*. São Paulo: Editora UNESP.
- Laurence, V. (1979). *Le raisonnement spontané en dynamique élémentaire*. Paris: Hermann. Retrieved from https://scholar.google.pt/scholar?q=viennot+1979+le+raisonnement+spontane++en+dynamique&btnG=&hl=pt-PT&as_sdt=0%2C5
- Laveault, D., & Grégoire, J. (2014). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. Bruxelles: de boeck.
- Lazarsfeld, P. (1950). The logical and mathematical foundation of latent structure analysis. *Studies in Social Psychology in World War II Vol. IV*: Retrieved from <http://ci.nii.ac.jp/naid/10011739862/>
- Lazarsfeld, P. (1959). Latent structure analysis. In *Psychology: A study of a science* (pp.

476–543). New York: McGraw-Hill.

Leite, L. (1999). Heat and Temperature: An Analysis of How These Concepts Are Dealt with in Textbooks. *European Journal of Teacher Education*, 22(1), 75–88.
<https://doi.org/10.1080/0261976990220106>

Leite, L., Costa, C., & Esteves, E. (2008). Os manuais escolares e a aprendizagem baseada na resolução de problemas : um estudo centrado em manuais escolares de Ciências Físico-Químicas do Ensino Básico. *Actas Do XXI CONGRESO DE ENSEÑANTES DE CIÊNCIAS DE GALICIA*.

Lewis, E. L. (1991). The development of understanding in elementary thermodynamics. *Aera*.

Lewis, E. L., & Linn, M. C. (1994). Heat Energy and Temperature Concepts of Adolescents, Adults, and Experts: Implications for Curricular Improvements. *Journal of Research in Science Teaching*, 31(6), 657–677.
<https://doi.org/10.1002/tea.3660310607>

Lewis, E. L., Stern, J. L., & Linn, M. C. (1993). The effect of computer simulations on introductory thermodynamics understanding. *Educational Technology | Educational Technology*, 33(1), 45–58.

Lima, C. A. (2008). *Aproximações entre ciência-tecnologia-sociedade e os temas transversais no livro didático de matemática do ensino fundamental de 5a. a 8a. séries*. Florianópolis, SC. Retrieved from
<https://repositorio.ufsc.br/handle/123456789/91773>

Lima, A. O. G. de. (1981). *Uma abordagem ausubeliana à organização do conteúdo em termodinâmica e teoria cinética dos gases ao nível de Física Geral*. (Dissertação de mestrado). Universidade Federal do Rio Grande do Sul.

Lima, J. M. de A. (2013). Algumas reflexões sobre o segundo princípio da termodinâmica ou princípio da entropia / João Maria de Almeida Lima. Retrieved from <https://digitalis.uc.pt/pt->

pt/fundo_antigo/algumas_reflexões_sobre_o_segundo_princípio_da_termodinâmica_ou_princípio_da_entropia

Lindquist, E. (1942). *A first course in statistics*.

Linsingen, L. (2008). *Literatura infantil no ensino de ciências: articulações a partir da análise de uma coleção de livros*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/91784>

Lobo Antunes, M. P. (2015). Manuais escolares de Ciências Naturais de 8º ano em Portugal e estrutura conceptual do PISA 2006. *Revista Portuguesa de Educação*, 28(1), 139–169.

Lobo, L. Q., & Ferreira, A. G. M. (2005). *Termodinâmica e propriedades termofísicas: vol.2 teoria cinética e propriedades de transporte dos gases*. Coimbra: Imprensa da Universidade de Coimbra. <https://doi.org/http://dx.doi.org/10.14195/978-989-26-0488-6>

Lobo, L. Q., & Ferreira, A. G. M. (2006). *Termodinâmica e propriedades termofísicas: vol.1 termodinâmica das fases*. Coimbra: Imprensa da Universidade de Coimbra. <https://doi.org/http://dx.doi.org/10.14195/978-989-26-0487-9>

Loken, E., & Rulison, K. L. (2010). Estimation of a four-parameter item response theory model. *The British Journal of Mathematical and Statistical Psychology*, 63(3), 509–525. <https://doi.org/10.1348/000711009X474502>

Lopes, P. (2015). Avaliação de competências de literacia mediática: instrumentos de recolha de informação e opções teórico- metodológicas. *Media & Jornalismo: Uma Revista Do Centro de Investigação Media e Jornalismo*, 15(27). https://doi.org/http://dx.doi.org/10.14195/2183-5462_27_2

Lord, F. (1952). A theory of test scores. *Psychometric Monographs*, (7), 84. <https://doi.org/10.1039/tf9524800166>

- Lord, F. M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika*, (2), 181-194.
- Louzada, A., & Elia, M. da F. (2015). Concepções alternativas dos estudantes sobre conceitos térmicos: Um estudo de avaliação diagnóstica e formativa. *Revista Brasileira de Ensino de Física*. Retrieved from https://www.researchgate.net/profile/Marcos_Elia/publication/277964725_Concepcoes_alternativas_dos_estudantes_sobre_conceitos_termicos_Um_estudo_de_avaliacao_diagnostica_e_formativa/links/56fa9b6508ae8239f6dad090.pdf
- Louzada, A. N. (2012). *Uso da Modelagem Computacional Como Proposta de Inovação Curricular para o Ensino de Física no Ensino Médio*. (Dissertação de mestrado). Universidade Federal do Rio de Janeiro.
- Louzada, A. N. (2015). 2015_ARTIGO OURO-Louzada_et_al-2015-Revista_Brasileira_de_Ensino_de_Fsica - EDITÁVEL.
- Louzada, A. N., Elia, F., & Sampaio, F. (2015). Concepcoes alternativas dos estudantes sobre conceitos termicos : Um estudo de avaliacao diagnostica e formativa, 1508.
- Louzada, A. N., Elia, M. da F., Sampaio, F. F., & Vidal, A. L. P. (2014). Validating the ACE Model for Evaluating Student Performance Using a Teaching-Learning Process Based on Computational Modeling Systems. *Informatics in Education*, 13(1), 73-85. Retrieved from http://search.proquest.com.ezp.lib.unimelb.edu.au/docview/1562511872?accountid=12372%5Cnhttp://sfx.unimelb.hosted.exlibrisgroup.com/sfxlcl41?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Aeducation&atitle=Valida
- Luera, G. G., Otto, C. C., & Zitzewitz, P. P. (2006). Use of the Thermal Concept Evaluation to Focus Instruction. *Phys. Teach.*, 44(3), 162-166. <https://doi.org/10.1119/1.2173324>

- Macedo, S. S. (2008). *Estratégias didáticas de professores do ensino fundamental ao lidar com concepções alternativas de alunos*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/91319>
- Machado, A. P., & Morona, V. C. (2007). *Manual de Avaliação Psicológica* (Coletânea).
- Machado, C., Alavarse, O., & Arcas, P. (2015). Sistemas estaduais de avaliação: interfaces com qualidade e gestão da educação. *Revista Brasileira de Política e Administração Da Educação*, 31(3), 667–680. <https://doi.org/10.21573/vol31n32015.63800>
- Madsen, A., McKagan, S. B., & Sayre, E. C. (2014). Best Practices for Administering Concept Inventories, (approximately 4), 1–11. Retrieved from <http://arxiv.org/abs/1404.6500>
- Madsen, A., Sayre, E., & McKagan, S. (2016). Effect size: What is it and when and how should I use it?
- Madu, B. C., & Orji, E. (2015). Effects of cognitive conflict instructional strategy on students' conceptual change in temperature and heat. *SAGE Open*, July-Sept(1), 9. <https://doi.org/10.1177/2158244015594662>
- Madu, B., & Orji, E. (2015). Effects of Cognitive Conflict Instructional Strategy on Students' Conceptual Change in Temperature and Heat. *SAGE Open*. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/2158244015594662>
- Mafra, P. (2011). *Proposta de uma sistemática para a modelagem de risco de crédito sob a perspectiva da teoria da resposta ao item*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/95148>
- Magis, D., Béland, S., Tuerlinckx, F., & Boeck, P. de. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42(3), 847–862. <https://doi.org/10.3758/BRM.42.3.847>

- Maia, C. K. (2008). *A organização praxiológica do objeto triângulo nos livros didáticos da 7ª série do ensino fundamental*. Florianópolis, SC. Retrieved from <https://repositorio.ufsc.br/handle/123456789/90898>
- Mantel-Haenszel DIF and PROX are Equivalent! (n.d.). Retrieved December 8, 2017, from <https://www.rasch.org/rmt/rmt32a.htm>
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies. *Journal of the National Cancer Institute*, 22(4), 719–748. <https://doi.org/10.1093/jnci/22.4.719>
- Marques, J. R. (2016). O que é Rapport | IBC Coaching. Retrieved October 9, 2017, from <http://www.ibccoaching.com.br/portal/coaching-e-psicologia/o-que-e-rapport/>
- Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, 75(2007), 87. <https://doi.org/10.1119/1.2372468>
- Marzano. (2006). The Need for a Revision of Bloom's Taxonomy. *The New Taxonomy of Educational Objectives*, 1–20. <https://doi.org/10.1207/s15430421tip4104>
- Marzano, R., & Kendall, J. (2007). *The new taxonomy of educational objectives* (2nd ed.). California: Corwin Press. Retrieved from https://www.google.com/books?hl=pt-PT&lr=&id=JT4KAgAAQBAJ&oi=fnd&pg=PP1&ots=xjo_KClz7y&sig=9KUAWudQ1WF1NYe1PmMDXrId4s4
- Matthews, M. (2004). Reappraising Positivism and Education: The Arguments of Philipp Frank and Herbert Feigl. *Science & Education*, 13(1–2), 7–39. <https://doi.org/10.1023/B:SCED.0000018544.34295.8b>
- Mattos, C., Drumond, A. V. N., & Paulo, G. S. (2004). Sensação Térmica : Uma Abordagem Interdisciplinar. *Caderno Brasileiro de Ensino Em Física*, 21(1), 7–34.
- Matui, J. (1995). *Construtivismo: teoria construtivista sócio-histórica aplicada ao ensino*. São

Paulo: Editora Moderna.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, 71(4), 713–732.

<https://doi.org/10.1007/s11336-005-1295-9>

Mcdermott, L. C. (1993). Guest Comment : How we teach and how students learn – A mismatch? *American Journal of Physics*, (61), 295.

<https://doi.org/10.1119/1.17258>

McDermott, L. C. (2001). Oersted Medal Lecture 2001: “Physics Education Research – The Key to Student Learning.” *American Journal of Physics*, 69(11), 1127.

<https://doi.org/10.1119/1.1389280>

McDermott, L. C. (2006). Preparing K-12 teachers in physics: Insights from history, experience, and research. *American Journal of Physics*, 74(9), 758.

<https://doi.org/10.1119/1.2209243>

McDermott, L. C. L. C., Shaffer, P. S. P. S., & Constantinou, C. P. (2000). Preparing teachers to teach physics and physical science by inquiry. *Phys. Educ*,

35(November), 71–85. <https://doi.org/10.1088/0031-9120/35/6/306>

McDermott, L. C., & Redish, E. F. (1999). Resource letter: PER-1: Physics education research. *American Journal of Physics*, 67(9), 755–767.

<https://doi.org/10.1119/1.19122>

McKagan, S., Sayre, E., & Madsen, A. (2017). Normalized gain: What is it and when and how should I use it? Retrieved September 15, 2017, from

<https://www.physport.org/recommendations/Entry.cfm?ID=93334>

McMillan, J., & Schumacher, S. (1989). *Research in education: a conceptual introduction*. *Research in Education: A conceptual introduction, Ed.* (2nd ed.). New York: Harper Collins.

MEC/BRASIL. (2015). Base Nacional Comum Curricular. *Base Nacional Comum*

Curricular. <https://doi.org/10.1007/s13398-014-0173-7.2>

Mendonça, C., Huet, I., & Alves, M. G. (2014). Da construção à validação de um referencial de competências para uma Licenciatura em Enfermagem. *Revista Portuguesa de Pedagogia*, 48(2).
https://doi.org/http://dx.doi.org/10.14195/1647-8614_48-2_6

Menegon, L. (2013). *Mensuração de conforto e desconforto em poltrona de aeronave pela teoria da resposta ao item*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/122869>

Midkiff, C., Litzinger, T., & Evans, D. (n.d.). Thermodynamics Concept Inventory Assessment.

Miller, K., Lasry, N., Reshef, O., Dowd, J., Araujo, I., & Mazur, E. (2010). Losing it: The influence of losses on individuals' normalized gains. *AIP Conference Proceedings*, 1289, 229–232. <https://doi.org/10.1063/1.3515208>

Ministério da Educação. (n.d.). SEED - Banco Internacional de Objetos Educacionais - Ministério da Educação. Retrieved August 25, 2017, from <http://portal.mec.gov.br/seed-banco-internacional-de-objetos-educacionais>

Moleiro, M. J., Damião, M. H., & Festas, M. I. F. (2014). Contextualização da aprendizagem: sua representação em manuais escolares de Estudo do Meio. *Revista Estudos Do Século XX*, (14).
https://doi.org/http://dx.doi.org/10.14195/1647-8622_14_18

Montalvão Neto, A. L. (2016). *Discursos de genética em livro didático: implicações para o ensino de biologia*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina.

Monteiro, A. (2010). A (re) valorização de outras fontes históricas: a problemática dos manuais escolares. In *Outros combates pela história*. Coimbra: Imprensa da Universidade de Coimbra. https://doi.org/http://dx.doi.org/10.14195/978-989-26-0199-1_25

- Monteiro, A. J. (2008). "Heróis do lar, nação católica e rural": os livros únicos do ensino primário no Estado Novo. *Revista Estudos Do Século XX*, (8).
https://doi.org/http://dx.doi.org/10.14195/1647-8622_8_14
- Montero, A., & Hierrezuelo, J. (1988). *La ciencia de los alumnos: Su utilización en la didáctica de la física y de la química*. Madrid: Laia/MEC.
- Moreira, M. A.; Silveira, F. L. (1993). Validação de instrumentos de medida aplicados à pesquisa em Ensino de Física. *Instrumentos de Pesquisa Em Ensino e Aprendizagem*, (December), 67-95.
- Moreira Junior, F. (2010). Aplicações da teoria da resposta ao item (TRI) no Brasil. *Rev. Bras. Biom*, 28(4), 137-170. Retrieved from
http://www.sigmees.com.br/files/ARTIGO_Fernando_TRI.pdf
- Moreira Junior, F. (2011). *Sistemática para a implantação de testes adaptativos informatizados baseados na teoria da resposta ao item*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from
<https://repositorio.ufsc.br/handle/123456789/95506>
- Moreira, M. A. (1999). Teorias de aprendizagem. Retrieved from
<http://leticiaawfrancomartins.pbworks.com/w/file/fetch/97972008/Cap9Moreira.pdf>
- Moreira, M. A. (2006). Mapas conceituais e diagramas V. *Porto Alegre: Ed. Do Autor*, 103. Retrieved from
http://www.if.ufrgs.br/~moreira/Livro_Mapas_conceituais_e_Diagramas_V_COMPLETO.pdf
- Moreira, M. A. (2011). Unidades de Ensino Potencialmente Significativas-UEPS. *Porto Alegre: UFRGS*.
- Moreira, M. A. (1998). Mapas Conceituais E Aprendizagem Significativa. *Cadernos de Aplicação*, 2(2), 87-95. Retrieved from
http://www.virtual.ufc.br/cursouca/modulo_4_projetos/conteudo/unidade_3

/mapas conceituais e aprendizagem significativa.pdf

- Moreira, M., & Masini, E. (2006). *Aprendizagem significativa: a teoria de David Ausubel* (2nd ed.). São Paulo: Centauro. Retrieved from https://scholar.google.pt/scholar?q=aprendizagem+significativa+moreira+masini+a+teoria+de+david+ausubel+centauro+2006&btnG=&hl=pt-PT&as_sdt=0%2C5
- Morin, E., & Moigne, J. Le. (2007). *Inteligência da complexidade: epistemologia e pragmática*. (J. (trad. . Duarte, Ed.)). Retrieved from https://scholar.google.pt/scholar?q=Inteligência+da+Complexidade+Epistemologia+e+Pragmática&btnG=&hl=pt-PT&as_sdt=0%2C5
- Morse, J. M. (1991). Approaches to Qualitative-Quantitative Methodological Triangulation. *Nursing Research*, 40(2), 120–123. <https://doi.org/10.1097/00006199-199103000-00014>
- Mortimer, E. (1994). *Evolução do atomismo em sala de aula: mudança de perfis conceituais*. Universidade de São Paulo. (Tese de doutorado). Universidade de São Paulo.
- Mortimer, E. (1995). Conceptual change or conceptual profile change? *Science & Education*, 4(3), 267–285. Retrieved from <http://www.springerlink.com/index/L3026827846056Q5.pdf>
- Mortimer, E. (1996). Construtivismo, mudança conceitual e ensino de ciências: para onde vamos? *Investigações Em Ensino de Ciências*, 1(1), 20–39. Retrieved from <https://www.if.ufrgs.br/cref/ojs/index.php/ienci/article/view/645>
- Mortimer, E. F. (n.d.-a). Bases Teóricas E Epistemológicas Da Abordagem Dos Perfis Conceituais Theoretical and Epistemological Grounds of the Conceptual Profile Approach. *Biologia*.
- Mortimer, E. F. (n.d.-b). *Conceptual Profiles*.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Ediciones

- Pirâmide. Retrieved from
https://scholar.google.pt/scholar?q=josé+muñiz+introducción+a+la&btnG=&hl=pt-PT&as_sdt=0%2C5
- Muñiz, J., Fidalgo, A., García-Cueto, E., Martínez, R., & Moreno, R. (2005). *Análisis de los ítems. Cuadernos de Estadística n° 30*. Madrid: La muralla.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological Testing: Principles and Applications*. Pearson Education.
- Nagem, R. L., Carvalhaes, D. D. O., & Dias, J. (2001). Uma proposta de metodologia de ensino com analogias. *Revista Portuguesa de Educação, 14*(1), 197–213.
- Nardi, R., & de Almeida, M. J. (2007). Investigação em Ensino de Ciências no Brasil segundo pesquisadores da área: alguns fatores que lhe deram origem. *Pro-Posições, 18*(1), 213–226.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*. Retrieved from
<http://journals.sagepub.com/doi/pdf/10.1177/001316445401400101>
- Nedelsky, L. (1961). Science Teaching and science testing. Retrieved from
https://scholar.google.pt/scholar?q=Nedelsky+Science+Teaching+and+science+testing%2C+1961.+&btnG=&hl=pt-PT&as_sdt=0%2C5
- Nelson, M. A., Geist, M. R., Miller, R. L., Streveler, R. A., & Olds, B. M. (2007). How to create a concept inventory: The thermal and transport concept inventory. *Annual Conference of the American Educational Research Association*.
- Neves, C. (2010). *Experimentos de microarrays e teoria da resposta ao item*. (Dissertação de Mestrado). Universidade de São Paulo. Retrieved from
<http://www.teses.usp.br/teses/disponiveis/45/45133/tde-24052010-140944/en.php>
- Neves, P., & Valadares, J. (2004). O contributo dos manuais de física para o

- enriquecimento conceptual dos alunos. *Revista Brasileira de Pesquisa Em Educação Em Ciências*, 4(2), 5–15. Retrieved from <https://repositorioaberto.uab.pt/handle/10400.2/1327>
- Niaz, M. (2000). a Framework To Uunderstand Students' Differentiation Between Heat Energy and Temperature and Its Educational Implication. *Interchange*, 31, 1–20.
- North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.
- Novak, J. D., & Gowin, D. B. (1984). *Aprender a aprender*. (Carla Valadares (trad.), Ed.) (1ª Edição). Lisboa.
- Nunes, P., & Del Pino, J. C. (2008). MAPA CONCEITUAL COMO ESTRATÉGIA PARA A AVALIAÇÃO DA REDE CONCEITUAL ESTABELECIDADA PELOS ESTUDANTES SOBRE O TEMA ÁTOMO (Concept mapping as strategy to assess the conceptual network established by the students about the atom subject). *Experiências Em Ensino de Ciências*, 3(1), 53–63.
- Nunes, T. (2016). *Modelagem e Simulações Computacionais: Uma Abordagem para o Ensino de Gases e Termodinâmica no Ensino Médio*. Retrieved from <https://repositorio.ufsc.br/handle/123456789/174273>
- OCDE. (2005). *La definición y selección de competencias clave: Resumen ejecutivo*. Retrieved from <http://www.deseco.admin.ch/bfs/deseco/en/index/03/02.parsys.78532.downloadList.94248.DownloadFile.tmp/2005.dscexecutivesummary.sp.pdf>
- OCDE. (2015). *Trabalhando com o Brasil*. Paris.
- Olds, B. M., Streveler, R. A., & Miller, R. L. (2004). Preliminary results from the development of a concept inventory in thermal and transport science. *Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition*, (703). Retrieved from

http://www.researchgate.net/publication/228597130_Preliminary_results_from_the_development_of_a_concept_inventory_in_thermal_and_transport_science/file/3deec521f897117e85.pdf%5Cnpapers3://publication/uuid/DD17816A-AC24-45B1-B541-35EBAC1E08C7

- Oliveira, B. (2014). A Qualidade das Questões de Física no Enem e o Problema da Dimensionalidade da Prova de Ciências da Natureza e Suas Tecnologias. In *XV Encontro de Pesquisa em Ensino de Física*. Maresias, São Paulo.
- Oliveira, B. (2015). Interdisciplinaridade e dimensionalidade das provas do Enem. In *VIII Reunião da ABAVE - Avaliação de Larga Escala no Brasil: Ensinaamentos, Aprendizagens e Tendências* (Vol. 8, pp. 423–426). Retrieved from http://www.abave.com.br/ojs/index.php/Reunioes_da_Abave/article/view/336
- Oliveira, K. (2010). *Um Estudo da evolução da GQT por meio do modelo de resposta gradual da teoria da resposta ao item*. (Dissertação de Mestrado). Universidade Federal do Ceará. Retrieved from <http://repositorio.ufc.br/handle/riufc/3029>
- Orlando, M., & Thissen, D. (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, 24(1), 50–64. <https://doi.org/10.1177/01466216000241003>
- Osborne, R. J., & Wittrock, M. C. (1983). Learning science: A generative process. *Science Education*, 67(4), 489–508. <https://doi.org/10.1002/sce.3730670406>
- Osburn, H. (1968). Item sampling for achievement testing. *Educational and Psychological Measurement*. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1177/001316446802800109>
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*.
- Ostini, R., & Nering, M. (2006). Polytomous item response theory models. Retrieved from <https://www.google.com/books?hl=pt->

PT&lr=&id=wS8VEMtJ3UYC&oi=fnd&pg=PR5&dq=testlets+in+item+response+theory&ots=m9rCFnv5gJ&sig=JOkPnjsOf58oT7mICAOQtQAIZao

- Otto, C., Luera, G., & Everett, S. (2009). An innovative course featuring action research integrated with unifying science themes. *Journal of Science Teacher Education*. Retrieved from <http://link.springer.com/article/10.1007/s10972-009-9146-7>
- Pacheco, A. P. R. (2010). *Competências essenciais: proposta de um modelo de concepção*. (Tese de doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/93801>
- Pacheco, J. (2014). *Benchmarking by Item Response Theory (BIRTH): método de benchmarking utilizando a teoria da resposta ao item na construção de escalas de competitividade*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/129310>
- Pacheco, M. A., & Henriques, M. H. (2012). As temáticas do Ano Internacional do Planeta Terra nos manuais escolares de Geologia dos 10º e 11º anos de escolaridade do Ensino Secundário Português. In *Para aprender com a Terra: memórias e notícias de Geociências no espaço lusófono*. Coimbra: Imprensa da Universidade de Coimbra. https://doi.org/http://dx.doi.org/10.14195/978-989-26-0533-3_4
- Pacheco, S. M. V., & Damasio, F. (2009). Mapas conceituais e diagramas V: ferramentas para o ensino, a aprendizagem e a avaliação no ensino técnico. *Ciências & Cognição*, 14(2), 166–193. <https://doi.org/ISSN 1806-5821>
- Palma, L. O. (2012). *Revisión de los conceptos de calor y temperatura y elaboración de una estrategia didáctica en la Educación Secundaria*. Universidad de La Rioja.
- Parceria MEC/SEB/COEDI, & UFMG, U. e U. (2017). *Posicionamento do projeto leitura e escrita na educação infantil em relação à terceira versão da BNCC*.
- Parreiral, R., Pereira, L. G., & Gomes, C. (2011). Representações pictóricas nos

manuais de ensino de Ciências Naturais (7º ano) e de Geologia (10º ano). In *Modelação de sistemas geológicos: livro de homenagem ao Professor Manuel Maria Godinho*. Coimbra: Imprensa da Universidade de Coimbra.
https://doi.org/http://dx.doi.org/10.14195/978-989-26-1009-2_21

Partanen, A. (2011). What Americans keep ignoring about Finland's school success. *The Atlantic*. Retrieved from

<http://harlanfalcons.org/ourpages/auto/2012/2/22/58526503/FINLAND.pdf>

Pasquali, L. (1996). *Teoria e Métodos de Medida em Ciências do Comportamento*. Brasília.

Pasquali, L. (1999). Histórico dos instrumentos psicológicos. *Instrumentos Psicológicos: Manual Prático De*. Retrieved from

<http://www.nnce.org/Arquivos/Aulas/bioestatistica/3-pasquali-instrumentacao.pdf>

Pasquali, L. (2011). *Psicometria: teoria dos testes na psicologia e na educação* (4th ed.). Petrópolis, RJ: Vozes.

Pasquali, L., & Primi, R. (2003). Fundamentos da Teoria da Resposta ao Item – TRI Basic Theory of Item Response Theory – IRT. *Avaliação Psicológica*, 2(2), 99–110.

Peduzzi, L. de Q. (1997). Sobre a resolução de problemas no ensino da física. *Caderno Brasileiro de Ensino de Física*, 14(3), 229–253.

Peduzzi, S. S. (1981). *Uma abordagem ausubeliana ao ensino de eletricidade e magnetismo em nível universitário básico*. (Dissertação de mestrado). Universidade Federal do Rio Grande do Sul.

Peixe, B. (2014). *Mensuração da maturidade do sistema de gestão ambiental de empresas industriais utilizando a teoria da resposta ao item*. (Tese de Doutorado).

Universidade Federal de Santa Catarina. Retrieved from
<https://repositorio.ufsc.br/handle/123456789/128931>

Pelizzari, A., Kriegl, M. de L., Baron, M. P., Finck, N. T. L., & Dorocinski, S. I. (2002).

- Teoria da aprendizagem significativa segundo Ausubel. *Revista PEC*, 2(1), 37-42.
Retrieved from
<http://portaldoprofessor.mec.gov.br/storage/materiais/0000012381.pdf>
- Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Retrieved from
<https://eric.ed.gov/?id=ED458233>
- Pereira, C., Andrade, D., & Tavares, H. (2004). Detection of determinant genes and diagnostic via Item Response Theory. Retrieved from
<http://www.repositorio.ufpa.br/jspui/handle/2011/4057>
- Pereira, M., & Barros, S. de S. (2010). Análise da produção de vídeos por estudantes como uma estratégia alternativa de laboratório de física no Ensino Médio. *Revista Brasileira de Ensino De*. Retrieved from
https://www.researchgate.net/profile/Marcus_Pereira2/publication/262615617_Analysis_of_students'_video_production_as_an_alternative_strategy_for_the_physics_lab/links/57925d3c08aed51475afc957.pdf
- Pereira, V. (2004). *Métodos alternativos no critério Brasil para construção de indicadores socioeconômicos: Teoria da resposta ao item*. (Dissertação de Mestrado). Pontifícia Universidade Católica do Rio de Janeiro. Retrieved from
http://www.lambda.maxwell.ele.puc-rio.br/5253/5253_1.PDF
- Pereyra, M. A., Kotthoff, H.-G., & Cowen, R. (2011). *PISA UNDER EXAMINATION: Changing Knowledge, Changing Tests, and Changing Schools*. Sense Publishers.
- Pérez, D. (1994). Diez años de investigación en didáctica de las ciencias: realizaciones y perspectivas. *Enseñanza de Las Ciencias: Revista de Investigación Y*. Retrieved from
<http://www.raco.cat/index.php/Ensenanza/article/viewFile/21355/93310>
- Pérez, D. G., Alís, J. C., & Terrades, F. M. (1999). El Surgimiento de la Didáctica de las Ciencias como Campo Específico de Conocimientos. *Revista Educación y*

Pedagogía, 25, 15–65.

Peşman, H., & Eryilmaz, A. (2010). Development of a Three-Tier Test to Assess Misconceptions About Simple Electric Circuits. *The Journal of Educational Research*, 103(3), 208–222. <https://doi.org/10.1080/00220670903383002>

Peters, J. R. (2005). *A História da Matemática no Ensino Fundamental Uma análise de livros didáticos e artigos sobre história*. (Dissertação de mestrado). Universidade Federal de Santa Catarina.

Physport. (n.d.). *PhysPort Implementation Guide : Thermal Concept Evaluation (TCE) Purpose of the TCE*.

Piaget, J. (1933). *Psychologie de l'enfant et enseignement de l'histoire*.

Piaget, J. (1947). *La représentation du monde chez l'enfant*. Retrieved from http://www.ufrgs.br/psicoeduc/arquivos/piaget_repres_monde_intro.pdf

Piaget, J. (1952). *The origins of intelligence in the child* (M. Cook, Trans.). *Work Published 1936) Piaget The Origins of Intelligence*

Piaget, J. (1973). *The child's conception of the world*. (Routledge, Ed.). London. Retrieved from <https://archive.org/details/childsconception01piag>

Pines, A., Novak, J., Posner, G., & VanKirk, J. (1978). *The clinical interview: A method for evaluating cognitive structure*. *Department of Education Curriculum Series*. Ithaca, NY. Retrieved from https://scholar.google.pt/scholar?hl=pt-PT&as_sdt=0,5&q=pines+1978+the+clinical

Pinho, J., & Sacramento, A. (2009). Accountability: já podemos traduzi-la para o português? *Revista de Administracao Publica-RAP*, 43(6), 1343–1369. Retrieved from <https://repositorio.ufba.br/ri/handle/ri/1645>

Portal do FNDE - Guias do Livro Didático. (n.d.). Retrieved from <http://www.fnde.gov.br/programas/livro-didatico/guias-do-pnld%0A>

- Posner, G., Strike, A., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a Theory of Conceptual Change. *Science Education*, 66(2), 221-227.
- Pozo, J., & Crespo, M. (2009). A aprendizagem e o ensino de ciências: do conhecimento cotidiano ao conhecimento científico. *Porto Alegre: Artmed*.
- Prado, M. P. (2009). *Um estudo de caracterização e avaliação de critérios de teste estruturais entre os paradigmas procedimental e OO*. (Dissertação de mestrado). Universidade de São Paulo.
- Preece, P. (1976). Mapping cognitive structure: A comparison of methods. *Journal of Educational Psychology*. Retrieved from <http://psycnet.apa.org/journals/edu/68/1/1/>
- Preece, P. (1978). Exploration of semantic space: Review of research on the organization of scientific concepts in semantic memory. *Science Education*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/sce.3730620415/full>
- Primi, R. (2004). Avanços na interpretação de escalas com a aplicação da Teoria de Resposta ao Item. *Avaliação Psicológica*. Retrieved from http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712004000100006
- Primi, R. (2017). *Teoria de Resposta ao Item com R*. São Francisco.
- Primi, R., Muniz, M., & Nunes, C. (2009). Definições contemporâneas de validade de testes psicológicos. In C. S. Hutz (Org.). *Avanços e polêmicas em avaliação psicológica* (pp. 243-265). São Paulo: Casa do Psicólogo.
- Prince, M., Vigeant, M., & Nottis, K. (2012). Development of the heat and energy concept inventory: Preliminary results on the prevalence and persistence of engineering students' misconceptions. *Journal of Engineering Education*, 101(3), 412-438. <https://doi.org/10.1002/j.2168-9830.2012.tb00056.x>

- QMSS e-Lessons | Validity and Reliability. (n.d.). Retrieved August 25, 2017, from http://ccnmtl.columbia.edu/projects/qmss/measurement/validity_and_reliability.html
- Quaresma, E. de S. (2014, May 28). *Modelagem para construção de escalas avaliativas e classificatórias em exames seletivos utilizando teoria da resposta ao item uni e multidimensional*. Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, Piracicaba. <https://doi.org/10.11606/T.11.2014.tde-25062014-103226>
- R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rabelo, M. (2013). *Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro*. Rio de Janeiro: SBM (1ª). Rio de Janeiro.
- Ramos, I. O. (2011). *Escola pública integrada: uma proposta sob análise*. (Dissertação de Mestrado em Educação). Univali. Retrieved from [http://siaibib01.univali.br/pdf/Ines Odorizzi Ramos.pdf](http://siaibib01.univali.br/pdf/Ines%20Odorizzi%20Ramos.pdf)
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA press. Retrieved from <https://eric.ed.gov/?id=ED419814>
- Redish, E. F., & Burciaga, J. R. (2004). *Teaching Physics with the Physics Suite*. *American Journal of Physics* (Vol. 72). <https://doi.org/10.1119/1.1691552>
- Resnick, L. (1989). Convictions ontologiques dans l'apprentissage de la physique. N. Bednarz et C. Gamier, *Construction Des Savoirs*.
- Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*. Evanston, Illinois.
- Rey, B. (1996). *Les compétences transversales en question*. Retrieved from <http://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/69298/Home>

- Rezende, F., Ostermann, F., & Ferraz, G. (2009). Ensino-aprendizagem de física no nível médio: o estado da arte da produção acadêmica no século XXI. *Revista Brasileira de Ensino de Física*, 31(1), 1402. Retrieved from <http://www.scielo.br/pdf/rbef/v31n1/v31n1a08>
- Ricardo, E. C. (2005). *Competências, interdisciplinaridade e contextualização: dos parâmetros curriculares a uma compreensão para o ensino das ciências*. Florianópolis, SC. Retrieved from <https://repositorio.ufsc.br/xmlui/handle/123456789/102668>
- Ricketts, J. H., & Head, G. a. (1999). A five-parameter logistic equation for investigating asymmetry of curvature in baroreflex studies. *The American Journal of Physiology*, 277(2 Pt 2), R441-54. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10444551>
- Rodrigues, M. (1993). *O uso de role-play e dramatização no ensino de física do 2o. grau: quatro casos em estudo*. (Dissertação de mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/bitstream/handle/123456789/111478/279333.pdf?sequence=1>
- Rodrigues, M. (2006). Proposta de análise de itens das provas do SAEB sob a perspectiva pedagógica e a psicométrica. *Estudos Em Avaliação Educacional*. Retrieved from <http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/2117>
- Rodrigues, R. (2016). *Formação e evolução estelar como uma proposta de contextualização para o ensino de termodinâmica no ensino médio*. Retrieved from <https://repositorio.ufsc.br/handle/123456789/174129>
- Roegiers, X. (2000). Saberes, capacidades y competencias en la escuela: una búsqueda de sentido. *Innovación Educativa*.
- Roegiers, X. (2006). *Aprendizagem integrada: situações do cotidiano escolar*. (J. (trad. .

- Wolff, Ed.). Porto Alegre: Artmed.
- Roegiers, X., & De Ketele, J.-M. (2004). *Uma pedagogia da integração: competências e aquisições no ensino*. (C. (trad. . Huang, Ed.) (2ª). POрто Alegre: Artmed.
- Rollnick, M., & Mahooana, P. (1999). A quick and effective way of diagnosing student difficulties: two tier from simple multiple choice questions. *South African Journal of Chemistry*, 52(4), 161-164. Retrieved from https://www.researchgate.net/profile/Marissa_Rollnick/publication/287485956_A_quick_and_effective_way_of_diagnosing_student_difficulties_Two_tier_from_simple_multiple_choice_questions/links/57306a4708aeb1c73d14720c.pdf
- Roque, G. O. B. (2004). *Uma Proposta de um Modelo de Avaliação de Aprendizagem Por Competências para Cursos a Distância baseados na Web*. (Dissertação de mestrado). Universidade Federal do Rio de Janeiro.
- Rosa, L. da. (2009). Sobre a Análise do Conceito de Conhecimento. *Intuitio*, 2(3), 191-202. Retrieved from <http://revistaseletronicas.pucrs.br/ojs/index.php/intuitio/article/view/5992/0>
- Rosa, M. (2013). *Seleção e uso do livro didático na visão de professores de ciências: um estudo na rede municipal de ensino de Florianópolis*. Dissertação de Mestrado. Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/107364>
- Rosenblatt, R., & Heckler, A. F. (2011). Systematic study of student understanding of the relationships between the directions of force, velocity, and acceleration in one dimension. *Physical Review Special Topics - Physics Education Research*, 7(2), 1-20. <https://doi.org/10.1103/PhysRevSTPER.7.020112>
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. McGraw-Hill Humanities Social. Retrieved from <https://scholar.google.com/scholar?q=related:q9FDXUArSKUJ:scholar.google.c>

om/&hl=en&as_sdt=0,5

- Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods, 1*(4), 331–340. <https://doi.org/10.1037/1082-989X.1.4.331>
- Rué, J., Almeida, M. de, & Arantes, V. (2009). *Educação e competências: pontos e contrapontos*. Summus.
- Sahlberg, P. (2011a). Lessons from Finland. *American Educator, 2*(1), 34–38. <https://doi.org/10.1146/annurev.genom.2.1.103>
- Sahlberg, P. (2011b). PISA in Finland: An education miracle or an obstacle to change? *CEPS Journal: Center for Educational Policy*. Retrieved from <http://search.proquest.com/openview/5789896d782ba725bec886c6f1a93843/1?pq-origsite=gscholar&cbl=1336356>
- Sahlberg, P. (2011c). The Professional Educator: Lessons from Finland. *American Educator*. Retrieved from <https://eric.ed.gov/?id=EJ931215>
- Sahlberg, P. (2013). Teachers as Leaders in Finland. *Educational Leadership*. Retrieved from <https://pasisahlberg.com/wp-content/uploads/2012/12/Teachers-as-leaders-20131.pdf>
- Samejima, F. (1968). *Estimation of latent ability using a response pattern of graded scores*. *ETS Research Report Series*. Princeton. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1968.tb00153.x/full>
- Santos, A. A. dos, Neto, J. L. H., & Junqueira, R. D. (2017). *Sistema Nacional de Avaliação da Educação Básica (Sinaeb): Proposta para atender ao disposto no Plano Nacional de Educação*. (E. G. Rodrigues, Ed.), INEP, MEC, Brasil. Brasília: Inep/MEC – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

- Santos, C. dos, & Moreira, M. (1979a). Aplicação da análise de agrupamentos hierárquicos ao mapeamento cognitivo de conceitos físicos. *Revista Brasileira de Física*. Retrieved from <http://www.sbfisica.org.br/bjp/download/v09/v09a57.pdf>
- Santos, C. dos, & Moreira, M. (1979b). Aplicação da Análise Multidimensional ao Mapeamento Cognitivo de Conceitos Físicos* 4. *Revista Brasileira de Física*. Retrieved from <http://www.sbfisica.org.br/bjp/download/v09/v09a56.pdf>
- Santos, C. P. dos C. dos, Araújo, L., & Graça, M. (2006). Resultados das Provas de Aferição e dos Exames de Matemática e de Português do 9º ano: Que conclusões tirar? *Revista Portuguesa de Pedagogia*, 40(3), 199–218. Retrieved from https://digitalis.uc.pt/pt-pt/artigo/resultados_das_provas_de_aferição_e_dos_exames_de_matemática_e_de_português_do_9º_ano_que
- Santos, F. M. T., & Greca, I. M. (2013). Metodologias de pesquisa no ensino de ciências na América Latina: como pesquisamos na década de 2000. *Ciência & Educação*, 19(1), 15–33. Retrieved from <http://riubu.ubu.es/handle/10259.4/2566>
- Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, 2(6), 110–114.
- Schnittka, C., & Bell, R. L. (2011). Engineering Design and Conceptual Change in Science: Addressing thermal energy and heat transfer in eighth grade. *International Journal of Science Education*, 33(13), 1861–1887. <https://doi.org/10.1080/09500693.2010.529177>
- Schram, A. B. (2014). A Mixed Methods Content Analysis of the Research Literature in Science Education. *International Journal of Science Education*, 0693(July), 1–20. <https://doi.org/10.1080/09500693.2014.908328>
- Sciarretta, M. R., Stilli, R., & Missoni, M. V. (1990). On the thermal properties of

materials: Common- sense knowledge of italian students and teachers. *International Journal of Science Education*, 12(4), 369–379.
<https://doi.org/10.1080/0950069900120404>

- Sepulveda, C., & Mortimer, E. (2013). Construção de um Perfil Conceitual de Adaptação: Implicações Metodológicas para o Programa de Pesquisa sobre Perfis Conceituais e o Ensino de Evolução. *Investigações Em Ensino de Ciências*, 18(2), 439–479. Retrieved from <http://search.proquest.com/openview/95b5c73987bb990201971c074212c47d/1?pq-origsite=gscholar&cbl=2032603>
- Serpa, D. (2011). Nigel Brooke: “Avaliação é apenas um ponto de partida.” Retrieved November 17, 2017, from <https://novaescola.org.br/conteudo/978/nigel-brooke-avaliacao-e- apenas-um-ponto-de-partida>
- Serrano, C. I. (2008). Manuais escolares: “a ponte e a porta” da vida e da cultura. *Revista Estudos Do Século XX*, (8).
https://doi.org/http://dx.doi.org/10.14195/1647-8622_8_16
- Silva, D. N. da. (2013, April 5). *Ensino e aprendizagem da termodinâmica: questões didáticas e contribuições da história da ciência*. Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, São Paulo.
<https://doi.org/10.11606/T.81.2013.tde-16052013-123501>
- Silva, J. da, Paixão, M., & Albuquerque, A. (2009). Características psicométricas da versão Portuguesa da Career Decision Self-Efficacy Scale–Short Form (CDSE–SF). *Psychologica*. Retrieved from <http://impactum-journals.uc.pt/index.php/psychologica/article/view/1013>
- Silva, M. da. (2014). A prática educativa baseada em evidências: contribuições na formação de docentes interdisciplinares. *Revista Do Grupo de Estudos e Pesquisa Em Interdisciplinaridade.*, (5), 25–30. Retrieved from <https://revistas.pucsp.br/index.php/interdisciplinaridade/article/view/20744>

- Silveira, F. L. da, Barbosa, M. C. B., & Silva, R. da. (2015). Exame Nacional do Ensino Médio (ENEM): Uma análise crítica. *Revista Brasileira de Ensino de Física*, 37(1).
- Silveira, F. L. da, Moreira, M. A., & Axt, R. (1986). Validação de um teste para detectar se o aluno possui a concepção newtoniana sobre força e movimento. *Ciência e Cultura*. Retrieved from http://www.if.ufrgs.br/~lang/Textos/Teste_Mecanica_1.pdf
- Silveira, F. L. da, Moreira, M. A., & Axt, R. (1992). Estrutura interna de testes de conhecimento em Física: um exemplo em Mecânica. *Enseñanza De Las Ciencias*, 10(2), 187-194.
- Silveira, P. M. da. (2015). *Criação de um índice de satisfação com a vida por meio da teoria da resposta ao item e fatores associados em trabalhadores brasileiros*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/135687>
- Silveira, F., Stilck, J., & Barbosa, M. (2014). Manifesto sobre a qualidade das questões de Física na Prova de Ciências da Natureza no Exame Nacional do Ensino Médio. *Caderno Brasileiro de Ensino de Física*, 31(2), 473-479.
- Simões, M. O., Simões, T. S., Caldeira, H., Bello, A., & Pina, E. P. (2005). *Programa de Física e Química A 10º Ano*.
- Simon Schwartzman. (2015). Simon's Site » O impasse do Ensino médio e o Funil do ENEM. Retrieved August 27, 2017, from <http://www.schwartzman.org.br/sitesimon/?p=5289&lang=pt-br>
- Sireci, S. G., Wainer, H., & Braun, H. (1998). *Psychometrics*.
- Skbkekas. (2012). File:Cohens d 4panel.svg. Retrieved September 16, 2017, from https://translate.google.pt/translate?hl=pt-PT&sl=en&u=https://en.wikipedia.org/wiki/Effect_size&prev=search
- Soares, A. M., Pereira, M. D. de A., & Canavarro, J. M. A. P. (2010). Competências do

século XXI, transição para o ensino superior e sucesso acadêmico: estudo do nível de proficiência linguística em inglês e português dos estudantes do 1º ano da Universidade de Coimbra. *Revista Portuguesa de Pedagogia*, 44(2).
https://doi.org/http://dx.doi.org/10.14195/1647-8614_44-2_7

Soares, M. (2004). *Projetos de trabalho e avaliação por competências: encontros, desencontros e contribuições à educação matemática*. Florianópolis, SC. Retrieved from <https://repositorio.ufsc.br/handle/123456789/87718>

Sobral vira modelo nacional de gestão - Educação - Estadão. (2014). Retrieved from <http://educacao.estadao.com.br/noticias/geral,sobral-vira-modelo-nacional-de-gestao-imp-,1158465>

Staudt, C., & Forman, G. (2014). Sensing Science: Temperature and Heat Readiness for Early Elementary Students – Concord Consortium. Retrieved December 11, 2017, from <https://concord.org/newsletter/2014-spring/sensing-science/>

Stepans, J. (1996). Targeting students' science misconceptions: Physical science concepts using the conceptual change model.

Stewart, J. (1979). Content and cognitive structure: Critique of assessment and representation techniques used by science education researchers. *Science Education*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/sce.3730630311/full>

Stocker, J., Neves, S. P., & Faria, L. (2010). Competência Percebida e Realização Escolar: Que Relações em Alunos do Ensino Secundário? *Psychologica*, 1(52). Retrieved from https://digitalis.uc.pt/pt-pt/artigo/competencia_percebida_e_realizacao_escolar_que_relacoes_em_alunos_do_ensino_secundario

Stone, A. (2006). A Psychometric Analysis of the Statistics Concept Inventory. Retrieved from <http://www.shareok.org/bitstream/handle/11244/1013/3208004.PDF?sequenc>

e=1

- Strike, K., & Posner, G. (1992). A revisionist theory of conceptual change. In *Philosophy of science, cognitive psychology, and educational theory and practice*. Retrieved from <https://www.google.com/books?hl=pt-PT&lr=&id=bJpW2WvfVkcC&oi=fnd&pg=PA147&dq=A+revisionist+theory+of+conceptual+change&ots=1e8lVi8Bti&sig=XrqpRf-q45XmN9-WvVmotamhuSw>
- Stroud, A. H., & Secrest, D. (1966). *Prentice-Hall Series in Automatic Computation George Forsythe , editor*. (G. Forsythe, Ed.). Prentice-Hall International.
- Suñé, L. S., Araújo, P. J. L., & Urquiza, R. de A. (2015). Desenho de currículo para desenvolver competências: uma proposta metodológica. *Aracaju: Editora Unit*. Retrieved from https://scholar.google.pt/scholar?q=sune+desenho+de+curriculo+para+desenvolver+competencias&btnG=&hl=pt-PT&as_sdt=0%2C5
- Swackhammer, G. (2005). Energy concept inventory. <https://www.msu.edu/course/te/407/FS05Sec3/Te802/Files/EnergyConceptInventory.Pdf>.
- Talaeb, P., & Wattanakasiwich, P. (2010). Development of thermodynamic conceptual evaluation. *AIP Conference*. Retrieved from <http://aip.scitation.org/doi/abs/10.1063/1.3479864>
- Talaia, M., & Silva, M. (2014). *Ambiente térmico de sala de aula pode condicionar o desenvolvimento de competências e avaliação de alunos*. Coimbra: Imprensa da Universidade de Coimbra. https://doi.org/http://dx.doi.org/10.14195/978-989-96253-3-4_77
- Talaia, M., & Vigário, C. (2016). Temperatura de ponto de orvalho: um risco ou uma necessidade. In *Geografia, cultura e riscos: livro de homenagem ao Prof. Doutor António Pedrosa*. Coimbra: Imprensa da Universidade de Coimbra. https://doi.org/https://doi.org/10.14195/978-989-26-1237-9_7

- Tanahoung, C., Chitaree, R., Soankwan, C., Sharma, M., & Johnston, I. (2006). Surveying Thai and Sydney introductory physics students' understandings of heat and temperature. *Proceedings of the Assessment in Science Teaching and Learning Symposium*.
- Tankersley, K. (2007). *Tests That Teach: Using Standardized Tests to Improve Instruction*.
- Tarekegn, G. (2009). Can computer simulations substitute real laboratory apparatus? *Lat. Am. J. Phys. Educ. Vol, 3(3)*, 506–517. Retrieved from http://journal.lapen.org.mx/sep09/2_LAJPE_282_Tarekegn.pdf
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Vol. 46). Sage. Retrieved from [https://www.google.com/books?hl=pt-PT&lr=&id=qtW04-pRJZ0C&oi=fnd&pg=PR9&dq=Tashakkori+e+Teddlie+\(1998&ots=6g6qpBc2uR&sig=xLvb2zfDJ7abtFAdeHeiQ7Suv0o](https://www.google.com/books?hl=pt-PT&lr=&id=qtW04-pRJZ0C&oi=fnd&pg=PR9&dq=Tashakkori+e+Teddlie+(1998&ots=6g6qpBc2uR&sig=xLvb2zfDJ7abtFAdeHeiQ7Suv0o)
- Tastan, Ö., & Yalçinkaya, E. (2008). Effectiveness of conceptual change text-oriented instruction on students' understanding of energy in chemical reactions. *Journal of Science Education and Technology, 17(5)*, 444–453. <https://doi.org/10.1007/s10956-008-9113-7>
- Tavares, R. (2007). Construindo mapas conceituais. *Ciências e Cognição, Vol.12(21)*, 72–85. <https://doi.org/10.4181/RNC.2014.22.02.editorial894.2p>
- Terrades, F. M. (1998). La didáctica de las ciencias como campo específico de conocimientos. Génesis, estado actual y perspectivas. Retrieved from https://scholar.google.pt/scholar?q=La+didáctica+de+las+ciencias+como+campo+especifico+de+conocimientos.+Génesis%2C+estado+actual+y+perspectivas.&btnG=&hl=pt-PT&as_sdt=0%2C5
- Testa, W. (2014). *Criação de uma escala da qualidade de vida para profissionais de educação física por meio da teoria da resposta ao item*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from

<https://repositorio.ufsc.br/handle/123456789/123399>

Tezza, R. (2009). *Proposta de um construto para medir usabilidade em sites de e-commerce utilizando a Teoria da Resposta ao Item*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/92424>

Tezza, R. (2012). *Modelagem multidimensional para mensurar qualidade em website de e-commerce utilizando a teoria da resposta ao item*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/103430>

Thacker, B. A. (2003). Recent advances in classroom physics. *Reports on Progress in Physics*, 66(10), 1833–1864. <https://doi.org/10.1088/0034-4885/66/10/R07>

Thomas, G., & Pring, R. (2004). *Evidence-Based Practice. Reading* (First). New York: Open University Press, McGraw-Hill Education.

Thompson, B., & Levitov, J. (1985). Using Microcomputers to Score and Evaluate Items. *Collegiate Microcomputer*. Retrieved from <https://eric.ed.gov/?id=EJ320128>

Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics*, 66(4), 338–352. <https://doi.org/10.1119/1.18863>

Thorton, R., & Sokoloff, D. (2001). Heat and Temperature Conceptual Evaluation. *Physport. Org Didownload Dari Hhttps://Www. Physport*. Retrieved from https://scholar.google.pt/scholar?q=Heat+and+Temperature+Conceptual+Evaluation&btnG=&hl=pt-PT&as_sdt=0%2C5&oq=+Heat+and+Temperature+Conceptual+Evaluation

Thouin, M. (2001). Notions de culture scientifique et technologique: Concepts de base, percées historiques et conceptions fréquentes.

- Thouin, M. (2004). *Ensinar as Ciências e a Tecnologia nos Ensinos Pré-Escolar e Básico 1.º Ciclo*. Coleção Horizontes Pedagógicos, Lisboa, Edições.
- Tirloni, A. (2013). *Avaliação ergonômica de carteiras universitárias: validação de um instrumento utilizando a teoria da resposta ao item (TRI)*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/107247>
- Toquetto, A. R. (2016). *Os temas “vidros e metais” em livros didáticos de química*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/169087>
- Torres, F. (2015). *Prevalência da maloclusão e o impacto na qualidade de vida de pré-escolares no município de Florianópolis-SC*. Retrieved from <https://repositorio.ufsc.br/handle/123456789/169493>
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159–169. <https://doi.org/10.1080/0950069880100204>
- Tremblay, G. (1994). Pedagogía colegial, marzo, material fotocopiado, s/ed.
- Trochim, W. M. K. (2007). *Research Methods: Knowledge Base*. (Shomoita Alam Lopa, Ed.), *Journal of Bussiness & Economic Research - March 2007* (Vol. 5). <https://doi.org/10.1111/j.1740-8784.2007.00058.x>
- Turik, C. (2010). *Análise de atitudes de alunos universitários em relação à estatística por meio da Teoria da Resposta ao Item (TRI)*. (Dissertação de Mestrado). Pontifícia Universidade Católica do Rio Grande do Sul. Retrieved from <http://repositorio.pucrs.br/dspace/handle/10923/3093>
- Tytler, R., & Prain, V. (2010). A framework for re- thinking learning in science from recent cognitive science perspectives. *International Journal of Science Education*. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/09500690903334849>

- UFSC. (2017). *DOCUMENTO SÍNTESE DA UNIVERSIDADE FEDERAL DE SANTA CATARINA SOBRE A 3ª VERSÃO DA BASE NACIONAL COMUM CURRICULAR*. Florianópolis.
- Valadares, J., & Fonseca, F. (2011). Uma Estratégia Construtivista e Investigativa para o Ensino da Óptica. *Revista Brasileira de Pesquisa Em Educação Em Ciências*, 4(3), 74–85. Retrieved from <http://ufpa.br/ensinofts/artigos2/v4n3a5.pdf>
- Vargas, V. (2007). *Medida padronizada para avaliação de intangíveis organizacionais por meio da teoria da resposta ao item*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/90026>
- Vaughan, A. (2002). Standards, accountability, and the determination of school success. In *The Educational Forum* (Vol. 66, pp. 206–213). Taylor & Francis Group. <https://doi.org/10.1080/00131720208984830> To
- Vendramini, C. M. M., & Dias, A. S. (2005). Teoria de Resposta ao Item na análise de uma prova de estatística em universitários. *PsicoUSF*, 10(2), 201–210. <https://doi.org/10.1590/S1413-82712005000200012>
- Vendramini, C., Silva, M., & Canale, M. (2004). Análise de itens de uma prova de raciocínio estatístico. *Psicologia Em Estudo*, 9(3), 487–498. Retrieved from <http://www.scielo.br/pdf/pe/v9n3/v9n3a16>
- Vergara, L. (2005). *Avaliação do Ensino de Ergonomia para o Design aplicando a Teoria da Resposta ao Item (TRI)*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/102916>
- Vernet, O. (2013). *Redigindo Itens de Conteúdo Dinâmico no Sistema AtenaME (Nota Técnica)*.
- Vey, I. (2011). *Avaliação de desempenho logístico no serviço ao cliente baseada na teoria da resposta ao item*. (Tese de Doutorado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/94947>

- Vieira, C. M. C. (n.d.). A validade de uma investigação de natureza quantitativa.
- Vieira, C. M. C. (2008). Género e cidadania nas imagens de História: estudos de manuais escolares e software educativo. *Revista Portuguesa de Pedagogia*, 42(1). Retrieved from https://digitalis.uc.pt/pt-pt/artigo/recensao_nunes_m_t_genero_e_cidadania_nas_imagens_de_historia_estudos_de_manuais_escolares_e
- Vieira, N. (2016). *As provas das quatro áreas do Enem vistas como prova única na ótica de modelos da teoria da resposta ao item uni e multidimensional*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/169668>
- Viennot, L. (1977). *Le raisonnement spontané en dynamique élémentaire*. *Revue française de pédagogie*. <https://doi.org/10.3406/rfp.1978.1688>
- Viennot, L. (1978). *Le raisonnement spontané en dynamique élémentaire*. *Revue Française de Pédagogie*. Retrieved from <http://www.jstor.org/stable/41161770>
- Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education*. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/0140528790010209>
- Von Korff, J., Archibeque, B., Gomez, K. A., McKagan, S. B., Sayre, E. C., Schenk, E. W., ... Sorell, L. (2016). Secondary Analysis of Teaching Methods in Introductory Physics: a 50k-student Study. *American Journal of Physics*, 84(12), 969–974. <https://doi.org/10.1119/1.4964354>
- W. James Popham. (2016). The Fatal Flaw of Educational Assessment. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2016/03/23/the-fatal-flaw-of-educational-assessment.html>
- Wainer, H. (1995). Precision and Differential Item Functioning on a Testlet-Based Test: The 1991 Law School Admissions Test as an Example. *Applied Measurement in Education*, 8(2), 157–186. https://doi.org/10.1207/s15324818ame0802_4

- Wainer, H., Bradlow, E., & Wang, X. (2007). Testlet response theory and its applications. Retrieved from <https://www.google.com/books?hl=pt-PT&lr=&id=9yAceUBHGdcC&oi=fnd&pg=PA3&dq=testlets+in+item+response+theory&ots=CYRPZJn10G&sig=789FTNIr6vjznVZKAAxk5EDEoc>
- Wainer, H., & Lewis, C. (1989). *Toward a psychometrics for testlets*. Educational Testing Service Research Report Series. Princeton. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3984.1990.tb00730.x/full>
- Wandersee, J., Mintzes, J., & Novak, J. (1994). Research on alternative conceptions in science. *Of Research on Science Teaching and Learning*. Retrieved from https://scholar.google.pt/scholar?q=wandersee+j.a++research+on+alternative+conceptions+in+science+1994&btnG=&hl=pt-PT&as_sdt=0%2C5
- Wang, W.-C., & Wilson, M. (2005). The Rasch Testlet Model. *Applied Psychological Measurement*, 29(2), 126–149. <https://doi.org/10.1177/0146621604271053>
- Ward, H., Roden, J., Hewlett, C., & Foreman, J. (2009). Ensino de ciências.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p -Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Wattanawasiwich, P., Taleab, P., Sharma, M. D., & Johnston, I. D. (2013). Development and implementation of a conceptual survey in thermodynamics. *International Journal of Innovation in Science and Mathematics Education*, 21(1), 29–53.
- Weale, S. (2017, March). Teachers must ditch “neuromyth” of learning styles, say scientists. Retrieved from https://www.theguardian.com/education/2017/mar/13/teachers-neuromyth-learning-styles-scientists-neuroscience-education?CMP=share_btn_fb
- Webb, N. L. (1997a). Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. *National Institute for Science Education*

- (NISE) *Publications*, 1(8), 1–39. Retrieved from <http://facstaff.wceruw.org/normw/WEBBMonograph6criteria.pdf>
- Webb, N. L. (1997b). Determining Alignment of Expectations and Assessments in Mathematics and Science Education. *NISE Brief*, 1(2), 10. Retrieved from <http://eric.ed.gov/?id=ED405190>
- Webb, N. L. (1999). *Alignment of Science and Mathematics Standards and Assessments in Four States* (Research Monograph No. N° 18). Washington.
- Webb, N. L. (2002). *An Analysis of the Alignment Between Mathematics Standards and Assessments for Three States*.
- Webb, N. L. (2006). Identifying content for student achievement tests. In *Handbook of test development* (pp. 155–180). Retrieved from https://www.google.com/books?hl=pt-PT&lr=&id=ed-NAgAAQBAJ&oi=fnd&pg=PA155&dq=Identifying+Content+for+Student+Achievement+Tests&ots=FsUnaPS0_r&sig=1xSXdmfNsewaenEAlqNWgL20rY0
- Weiss, D., & Davison, M. (1981). Test theory and methods. *Annual Review of Psychology*. Retrieved from <http://www.annualreviews.org/doi/pdf/10.1146/annurev.ps.32.020181.003213>
- Welch, B. (1947). The generalization of student's' problem when several different population variances are involved. *Biometrika*, 34(1/2), 28–35. Retrieved from <http://www.jstor.org/stable/2332510>
- Wood, W. B., & Gentile, J. M. (2003). Teaching in a research context. *Science*, 302(November), 1510.
- Wormeli, R. (2003). *Day one & Beyond: Practical matters for new middle-level teachers*. (National Middle School Association, Ed.). Portland, Maine: Stenhouse Publishers.
- Wormeli, R. (2006). *Fair Isn't Always Equal: assessing & Grading in the Differentiated*

Classroom. Portland, Maine: Stenhouse Publishers.

- Wormelli, K. (2003). Differentiating instruction: A modified concerto in four movements. *LD Online*. Retrieved from [http://users.manchester.edu/Student/GJTribbett/Webpage/Differentiating Instruction.pdf](http://users.manchester.edu/Student/GJTribbett/Webpage/DifferentiatingInstruction.pdf)
- Wren, D., & Barbera, J. (2013). Gathering evidence for validity during the design, development, and qualitative evaluation of Thermochemistry Concept Inventory items. *Journal of Chemical Education*, 90(12), 1590–1601. <https://doi.org/10.1021/ed400384g>
- Yeo, S., & Zadnik, M. (2001). Introductory thermal concept evaluation: Assessing students' understanding. *The Physics Teacher*. 39(11), 496-504. Retrieved from <http://aapt.scitation.org/doi/abs/10.1119/1.1424603>
- Yurdugül, H., & Batenburg, T. Van. (2006). Item Difficulty From Graphical Item Analysis. *Eurasian Journal Of*. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=1302597X&AN=24124384&h=6aKPfV5bR1nyN4xu8anKpiQuG3mnayjgVHJ37uk8ufFR5nLyIPEzQCoPm12OunbX39mFQeU6MF8F78SFO%2BXIXA%3D%3D&crl=c>
- Zabala, A., & Arnau, L. (2010). *Como aprender e ensinar competências*.
- Zancan, L. R. (2015). *O professor e o uso do livro didático de biologia*. (Dissertação de Mestrado). Universidade Federal de Santa Catarina. Retrieved from <https://repositorio.ufsc.br/handle/123456789/160550>
- Zwick, R., Thayer, D. T., & Lewis, C. (1999). An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. *Journal of Educational Measurement*, 36(1), 1–28. <https://doi.org/10.1111/j.1745-3984.1999.tb00543.x>
- Zwierewicz, M. (2013). *Competências de pesquisadores manifestadas em teses de doutorado em psicologia e educação de universidades do Brasil, Espanha e Suécia*. (Tese de

doutorado). Universidade Federal de Santa Catarina. Retrieved from
<https://repositorio.ufsc.br/handle/123456789/107332>



Marcel Bruno Pereira Braga

Escala de Proficiências em Concepções Térmicas: Diagnóstico Psicométrico de Estudantes em Portugal e Brasil

Vol. II: Apêndices e Anexo – Informações Detalhadas dos Itens

Tese de Doutoramento em Ensino das Ciências, Ramo de Ensino da Física, em regime de cotutela entre a Universidade de Coimbra e a Universidade Federal de Santa Catarina, orientada pelos professores doutores Décio Ruivo Martins, Maria Augusta Vilalobos Filipe Pereira do Nascimento e José André Peres Angotti e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologias da Universidade de Coimbra

Junho de 2018



UNIVERSIDADE DE COIMBRA

Marcel Bruno Pereira Braga

Escala de Proficiências em Concepções Térmicas:
Diagnóstico Psicométrico de Estudantes em Portugal e Brasil

Vol. II: Apêndices e Anexo – Informações Detalhadas dos Itens

Tese de Doutoramento em Ensino das Ciências, Ramo de Ensino da Física, em regime de cotutela entre a Universidade de Coimbra e a Universidade Federal de Santa Catarina, orientada pelos professores doutores Décio Ruivo Martins, Maria Augusta Vilalobos Filipe Pereira do Nascimento e José André Peres Angotti e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologias da Universidade de Coimbra



UNIVERSIDADE DE COIMBRA



**UNIVERSIDADE FEDERAL
DE SANTA CATARINA**
Centro de Ciências Físicas
e Matemáticas - CFM

LISTA DE FIGURAS

Figura 1: Quantidade de itens relacionados à temática em <i>Termodinâmica</i> nos exames nacionais em Portugal.....	250
Figura 2: Incidências nas temáticas em termodinâmica entre 2006 e 2016.....	251
Figura 3: Habilidades consideradas pertinentes ao desenvolvimento de competências científicas dentro do conhecimento térmico, de acordo com a 3ª versão da BNCC.	258
Figura 4: Comparativo do percentual de respostas para o pré-teste e o pós-teste em Portugal.	263
Figura 5: Comparativo do índice de dificuldade e do percentual dos distratores assinalado entre os respondentes de Brasil e Portugal.....	265

LISTA DE TABELAS

Tabela 1: Variação da densidade da água em função da temperatura.	103
Tabela 2: Condutividade térmica de materiais a 27 °C (300 K).	120
Tabela 3: Dados da População (Escolas Secundárias) de acordo com o Distrito, Concelho, Nome da Escola e Livros Didáticos de Física e Química Adotados (vigência 2008-2014).	217
Tabela 4: Identificação de todas as escolas da Região da Grande Florianópolis, de acordo com o quantitativo de alunos (estimado) e modalidade de educação desenvolvida na unidade de ensino (referente a 2014).	221
Tabela 5: Controle logístico dos materiais distribuídos por escola, livros didáticos, sujeitos estimados e participantes.	226
Tabela 6: Mapeamento na abordagem de temáticas relacionadas à termodinâmica no ensino experimental exigidas pelo programa curricular português, que no caso foram cobradas aos alunos nos exames nacionais de 2006 a 2016.	245
Tabela 7: Valores dos Fatores de Concentração nas Análises dos Itens da amostra portuguesa no teste-piloto, nomeadamente as porcentagens de acerto e concentração dos grupos do 10º ano no pré-teste e pós-teste, conforme o <i>design</i> estabelecido.....	264
Tabela 8: Parâmetros clássicos para cada item do conjunto dos alunos do 1º, 2º e 3º ano na amostra do contexto no Brasil/2015.	266
Tabela 9: Parâmetros clássicos para cada item do conjunto dos alunos do 10º, 11º e 12º ano na amostra do contexto em Portugal/2014.	267
Tabela 10: Escala de desempenho (escore bruto) dos alunos brasileiros de acordo com sua distribuição nas unidades escolares, turno e nível de escolarização (em 2015).	268
Tabela 11: Escala de desempenho (escore bruto) dos alunos portugueses de acordo com sua distribuição nas unidades escolares e nível de escolaridade (em 2014).	269
Tabela 12: Teste-t <i>student</i> comparando diferenças significativas entre as médias do 1º, 2º e 3º ano com relação ao turno matutino.	270
Tabela 13: Teste-t <i>student</i> comparando diferenças significativas entre os grupos de respondentes conforme o Livro Didático de Física adotado na escola.....	270
Tabela 14: Teste-t <i>student</i> comparando diferenças significativas entre os grupos de respondentes conforme os níveis de escolaridade e na perspectiva geral entre o contexto Brasil e de Portugal.	271
Tabela 15: Valores dos Fatores de Concentração nas Análises dos Itens da amostra brasileira, nomeadamente as porcentagens de acerto e concentração dos grupos do 1º, 2º e 3º ano, respectivamente.	271
Tabela 16: Valores dos Fatores de Concentração nas Análises dos Itens da amostra portuguesa, nomeadamente as porcentagens de acerto e concentração dos grupos do 10º, 11º e 12º ano, respectivamente.	272
Tabela 17: Classificação do Índice de Dificuldades e do Fator de Concentração nas Análises de Portugal e Brasil, com relação aos níveis de escolaridade e geral em cada contexto.	273
Tabela 18: Distratores mais atrativos entre os contextos, com base na AGI dentro do grupo ACIM.	274

Tabela 19: Matriz de correlações tetracóricas dos itens - Amostra Geral (Brasil e Portugal).	275
Tabela 20: Identificação dos itens-âncoras e quase-âncoras, de acordo com os critérios preestabelecidos relacionados a consistência dos itens.	276

LISTA DE QUADROS

Quadro 1: Tipo de violações na AGI.	3
Quadro 2: Explicações sobre o quadro descritivo e o painel de estatísticas do item.....	4
Quadro 3: Análise e considerações diante dos riscos à validade interna do <i>design</i> experimental do estudo piloto.....	227
Quadro 4: Análise e considerações diante dos riscos à validade externa do <i>design</i> experimental do estudo piloto.....	230
Quadro 5: Análise e considerações diante dos riscos à validade interna dos <i>designs</i> de Portugal e Brasil.....	233
Quadro 6: Análise e considerações diante dos riscos à validade externa dos <i>designs</i> de Portugal e Brasil.....	236
Quadro 7: Capacidades a serem desenvolvidas na unidade temática de termodinâmica, de acordo com o programa de Físico-Química A, versão 2005.....	241
Quadro 8: Capacidades a serem desenvolvidas na unidade temática de termodinâmica, de acordo com o programa de Físico-Química A, versão 2014.....	242
Quadro 9: Sugestões Temáticas do PCN+/2002 para Desenvolver Competências na Física Térmica.	252
Quadro 10: Habilidades da Matriz de Referência de Ciências da Natureza e suas Tecnologias consideradas pertinentes ao TCE.....	252
Quadro 11: Sugestões para Etapas e Procedimentos na Aplicação de Testes ou Exames em Larga Escala.	260
Quadro 12: Tempo estimado para a realização do teste TCE baseados em experiência prévia.	262

SUMÁRIO

LISTA DE FIGURAS	V
LISTA DE TABELAS.....	VII
LISTA DE QUADROS	IX
APÊNDICES.....	1
A - Itens Comentados.....	1
Considerações sobre os Itens Comentados.....	1
A1: ITEM y1	7
A2: ITEM y2	17
A3: ITEM y3	27
A4: ITEM y4	35
A5: ITEM y5	43
A6: ITEM y6	51
A7: ITEM y7	57
A8: ITEM y8	65
A9: ITEM y9	71
A10: ITEM y10	81
A11: ITEM y11	91
A12: ITEM y12	99
A13: ITEM y13	105
A14: ITEM y14	113
A15: ITEM y15	121
A16: ITEM y16	129
A17: ITEM y17	137
A18: ITEM y18	145
A19: ITEM y19	153
A20: ITEM y20	161
A21: ITEM y21	169
A22: ITEM y22	177
A23: ITEM y23	187
A24: ITEM y24	195
A25: ITEM y25	203
A26: ITEM y26	209
B - População e Amostra.....	217

B1: População – Região Centro, Portugal.....	217
B2: População – Região da Grande Florianópolis, Brasil.....	221
B3: Quantitativo de alunos da amostra por escola e turno – Brasil.....	225
B4: Controle no Processo de Recolha de Dados – Portugal.....	226
C – Validade no <i>Design</i> da Pesquisa.....	227
C1: Análise de Riscos sobre a Validade Interna: Estudo Piloto.....	227
C2: Análise de Riscos sobre a Validade Externa: Estudo Piloto.....	230
C3: Análise de Riscos sobre a Validade Interna – Portugal e Brasil.....	233
C4: Análise de Riscos sobre a Validade Externa – Portugal e Brasil.....	236
D – A Física Térmica nos Programas Curriculares.....	241
D1: Portugal.....	241
D2: Brasil.....	252
E – Detalhes nos Resultados Obtidos.....	259
E1: Teste-Piloto (quatro escolas secundárias em Portugal).....	259
E2: Índice de Dificuldade dos Itens nas amostras do Brasil e de Portugal.....	265
E3: Parâmetros Clássicos da amostra no Brasil e em Portugal.....	266
E4: Escala de desempenho (escore bruto) dos alunos brasileiros e portugueses.....	268
E5: Teste-t <i>student</i> comparando diferenças significativas entre as médias do 1º, 2º e 3º ano com relação ao turno matutino.....	270
E6: Teste-t <i>student</i> comparando diferenças significativas entre as médias dos grupos de escolas conforme os Livros Didáticos de Física (ou Manuais Escolares) em comum, tendo como referência o grupo com maior média.....	270
E7: Teste-t <i>student</i> comparando diferenças significativas entre as médias dos diferentes níveis escolares nos dois contextos, assim como entre as médias gerais entre Brasil e Portugal.....	271
E8: Fator de Concentração nas Análises dos Itens – Brasil.....	271
E9: Fator de Concentração nas Análises dos Itens – Portugal.....	272
E10: Fator de Concentração nas Análises dos Itens (Visualização Geral) – Teste-Piloto, Portugal e Brasil.....	273
E11: Análise dos Distratores e as Respectivas Concepções Térmicas em Destaque dentro dos Grupos Superiores (ACIM).....	274
E12: Matriz de Correlações Tetracóricas dos Itens – Amostra Geral (Brasil e Portugal).....	275
E13: Identificação dos Itens-âncoras na Escala de Proficiências Térmicas para o Modelo 4PL (20 itens) – Brasil e Portugal.....	276
F – Sintaxes.....	278
F1: Alguns Resultados da Análise Clássica no BILOG-MG – Brasil/2015.....	278
F2: Alguns Resultados da Análise Clássica no BILOG-MG – Portugal/2014.....	279

G - Documentos Formais	280
G1: Consentimento da Autora do Teste TCE via <i>e-mail</i>	280
G2: Carta aos Diretores (Modelo Brasil).....	281
G3: Carta de Apresentação aos Professores (Modelo Portugal)	282
G4: Carta de Apresentação aos Professores (Modelo Brasil)	283
G5: TCLE - Encarregado de Educação.....	284
G6: TCLE - Professores (Modelo Portugal)	285
G7: TCLE - Professores (Modelo Brasil)	286
G8: TCE - Orientação aos Professores.....	287
H - Instrumentos.....	288
H1: Tradução do TCE - Versão Portugal.....	288
H2: Tradução do TCE - Versão Brasil.....	292
ANEXOS.....	298
1: CNPD - Autorização Formal para a Realização da Pesquisa	298

APÊNDICES

A - Itens Comentados

Considerações sobre os Itens Comentados

Basicamente, o modelo de itens comentados desenvolvido foi inspirado a partir dos exemplos ilustrados no livro de Rabelo (2013) e na plataforma *online Devolutivas Pedagógicas* (INEP, n.d.). No caso, o modelo elaborado pelo INEP visa, entretanto, contribuir nas escolas, em especial aos gestores e professores no sentido de apresentar alguns¹⁰³ resultados do SAEB e sugerir-lhes informações que contribuam na tomada de decisões baseadas em evidências de boas práticas tanto de gestão escolar quanto na prática pedagógica. Essa é uma perspectiva concreta que motiva as expectativas de possíveis contribuições sociais advindas no âmbito dessa investigação, podendo ser ampliada em diversos aspectos.

Por ser um aspecto valorizado nesse estudo, considera-se desafiador e necessário não somente uma interpretação e utilização adequada (e responsável) dessas informações como também a necessidade de estudos complementares mais amplos e aprofundados diante das contribuições que o conhecimento psicométrico possa oferecer para as pesquisas no EC. Dentre algumas limitações no modelo disponibilizado pelo INEP, percebe-se que os erros em destaque que são caracterizados pelos distratores não ganham uma devida importância, pois quando são referenciados, os comentários apresentados se mostram superficiais e sem proposições mais práticas e específicas para o cenário pedagógico no que diz respeito a uma fundamentação dentro da Psicologia da Educação e na Didática das Ciências.

Entende-se que o conhecimento psicométrico que deveria estar a serviço da psicologia, se restringe basicamente no entendimento dos dados estatísticos, na qual fazem referência em interpretações de diferentes perfis de composições de grupos de desempenho e índices de dificuldades, porém estão ausentes contribuições e sugestões

¹⁰³ O SAEB não disponibiliza todas as questões por motivo de validade, tendo em vista que utiliza processos psicométricos que dependem de itens-âncora para possibilitar o desenvolvimento da construção de séries histórias de desempenho.

mais concretas para a melhoria no cenário pedagógico, mesmo contendo normas e critérios referenciados. Diante disso, realiza-se um esforço de complementariedade no âmbito desse estudo nesse sentido.

Os itens comentados mostram de forma detalhada uma correspondência dos resultados estatísticos com suas interpretações pedagógicas, sendo um mecanismo avaliativo útil a contribuir de forma mais específica na tomada de decisões educacionais baseada em evidências de competências científicas (em potencial) e de informações relevantes para intervenções de ensino em escolas, além da possibilidade de melhoria de recursos instrucionais diante das dificuldades conceituais dos alunos.

Essa análise é feita em larga escala e comparativamente entre os alunos brasileiros e portugueses, todos em uma mesma base de dados, possibilitando colocá-los dentro de uma mesma métrica e assim realizar comparações entre os desempenhos de diferentes grupos normativos. A sondagem busca evidenciar uma magnitude de generalização e comparação dentro de cada contexto, e com isso mostrar características quanto as diferentes concepções térmicas existentes entre alunos nos níveis de escolarização dos três últimos anos da educação básica ou pré-universitária.

A construção da escala de proficiências de concepções térmicas possibilita não apenas diferenciar os grupos de acordo com as suas proficiências, como também ajudar na elaboração de um perfil dos alunos diante de suas capacidades para as concepções de modo geral dentro de cada contexto.

Os itens são comentados com base na elaboração de um painel de estatísticas para cada item, sendo constituído de dados e gráficos que contemplam a TCT e a TRI. Esse painel duplo apresenta as seguintes informações:

- TCT: Parâmetros clássicos e a *Análise Gráfica dos Itens - AGI*, de acordo com a proposta de Batenburg & Laros (2002), incluindo o *fator de concentração de análise* de Bao & Redish (2001). Incluem-se os valores corrigidos para os índices de discriminação, coeficiente bisserial, ponto-bisserial, e de dificuldade;

- TRI: Parâmetros do *modelo 4PL*¹⁰⁴ para o contexto geral, e uma representação gráfica de DIF para comparar as curvas em cada contexto separadamente (modelo restrungido *versus* não restrungido), assim como aspectos comparativos no desempenho de grupo normativos, incluindo interpretações gráficas das CCI e CII.

Nas AGIs com a amostra brasileira os níveis de escore 19, 21 e 22 estão ausentes de sujeitos, e o 20 e 23 apenas com um único indivíduo cada um (alunos do turno matutino da escola 22BR), desse modo as análises mais representativas devem se concentrar nos níveis que antecedem, pois apresentam maior quantidade de informação devido a conter maior número de respondentes. Os níveis ausentes na análise dos gráficos descem todos ao nível zero justificado pela ausência de sujeitos, desse modo convêm interpretar a proporção assinalada nas alternativas indo do 18 para o 20 e depois para o 23. Para uma melhor interpretação e comparação quanto ao ângulo de inclinação dos gráficos, é importante verificar o intervalo de discriminação em cada um e o comprimento no eixo X que vai do ponto S ao ponto P (intervalo entre os escores), pois quanto menor for essa distância, maior será o ângulo medido.

Para uma análise da qualidade do item, quanto à discriminação, dificuldade, *guessing*, erro ao acaso, serão contabilizadas a quantidade de violações, que serão numeradas em violações do tipo 1, 2, 3 e 4, conforme as recomendações dos autores para a técnica de AGI.

Quadro 78: Tipo de violações na AGI.

Violações na Análise das Curvas
Tipo 1: Analisar o número de violações contra a suposição de que os aumentos das alternativas corretas devam seguir crescentemente com a pontuação total;
Tipo 2: Analisar o número de violações contra a suposição de que a proporção dos distratores (falsas alternativas) deva diminuir o aumento da pontuação global;
Tipo 3: Analisar o número de intersecções entre a alternativa correta e as falsas após o início do intervalo de discriminação;
Tipo 4: Analisar a ausência de discriminação ou de uma baixa potência/poder de inclinação.

O item comentado possui inicialmente um *quadro de identificação* com o enunciado e suas alternativas, sendo acompanhados com o indicativo das concepções

¹⁰⁴ O processo de escolha para os modelos da TRI adotados nas análises foi realizado por procedimentos estatísticos que encontram-se explicados mais adiante.

envolvidas conforme os códigos do IC, breves justificativas com as concepções sugeridas, destaque para o modelo mental correto (em verde) e incorreto (em vermelho), descrição das competências técnicas associadas e das proficiências dos itens, juntamente com os respectivos descritores da matriz de habilidades do teste e interpretação dos subsunçores inerentes. Esses parâmetros do item são objetos na análise taxonômica e na determinação das proficiências científicas envolvidas. Para uma melhor interpretação, o modelo do quadro de identificação do item segue da seguinte forma:

Quadro 79: Explicações sobre o quadro descritivo e o painel de estatísticas do item.

Item yX

(Enunciado)
<p>Apresenta a situação-problema na qual é possível identificar a habilidade, o conteúdo, o contexto (disciplinar, trabalhista ou relacionado ao cotidiano), e eventuais relações com outros itens. Também é objeto na análise taxonômica ao ser verificado a semântica dos verbos de ação com os conteúdos envolvidos, bem como o contexto e sua relação com o conhecimento necessário para que se resolva a situação-problema, podendo ser identificado o <i>conhecimento potencialmente significativo</i> na concepção ausubeliana, e na alternativa correta o conhecimento associado que supostamente corresponda ao <i>subsunçor</i>. Quando o aluno responde optando por um distrator, considera-se a possibilidade de que o conhecimento prévio seja um equívoco (Concepção térmica), ou seja, um <i>subsunçor distrator</i>.</p>
(Alternativa correta) – Modelo mental correto
<p>Mesmo sendo assinalado a alternativa correta, não se descarta a existência de concepções alternativas associadas e identificadas de forma indireta. O percentual acerto predominante sugere um modelo mental correto, sustentado pelo conhecimento científico, diferentemente do senso comum, ou seja, as evidências sugerem que o aluno adquiriu a proficiência que item exige. Entretanto, ainda assim, possam haver equívocos que influenciam na tomada de decisão, como por exemplo, no caso de uma relação dependência com respostas equivocadas e assinaladas anteriormente.</p>
(Distratores) – Modelo mental incorreto
<ul style="list-style-type: none"> • Concepções alternativas associadas a alternativa “b”, estando codificada de acordo com o inventário de concepções térmicas (verificar no quadro de tipicidade dos grupos de concepções A, B, C e D, que fazem correspondência com os itens/distratores); • Os distratores serão acompanhados de um breve argumento justificador; • O distrator pode estar associado a mais de uma concepção alternativa; • O conjunto de concepções alternativas que o distrator representa sugere um modelo mental do aluno diante o conhecimento térmico envolvido na relação do item com o distrator; • O modelo mental pode reincidir de forma plena ou parcial entre os outros distratores, ainda que o conhecimento envolvido na assertiva sustente um argumento aparentemente distinto. Quando isso acontecer considera-se que o modelo mental está reforçado como um padrão mais consistente dentro do grupo; • Podem haver mais de um modelo mental baseado nos equívocos. Isso pode sugerir problemas no item ou grande dificuldade dos alunos no conhecimento envolvido, ou ainda, forte ligação com as concepções envolvidas entre os modelos mentais; • Podem haver um modelo mental (correto ou incorreto) ou dois modelos mentais (correto e incorreto; incorreto e incorreto); • Pode não demonstrar nenhum modelo mental. Isso ocorre quando não houver uma concentração significativa em uma das alternativas, podendo indicar problema no item ou grande dificuldade pelos alunos no conhecimento envolvido.

(Eventual justificava caso haja uma falta de correspondência com os distratores sugeridos pelos autores originais)
Apresentação das concepções que não foi identificada como uma concepção térmica associada. A concepção térmica é acompanhada por uma breve justificativa.
(A competência científica associada)
Representa uma perspectiva que abriga itens com características semelhantes – situações, contextos, um conhecimento que se exige relativamente próximo, e vem acompanhado de um identificador do descritor envolvido [Descritor x], na qual deve ser verificado os detalhes na matriz de referência do teste.
(Proficiência científica)
Detalhamento e especificidade do conhecimento envolvido que se exige para responder o item corretamente. São destacados com <i>itálico</i> o verbo de ação que caracteriza a capacidade (ou habilidade cognitiva), por “«” e “»” os conteúdos e textos complementares, e em negrito o <i>conhecimento potencialmente significativo</i> e o <i>subsunçor</i> identificado que está envolvido.
Painel de Estatísticas do Item
Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)
Tem-se as informações básicas da amostra: o quantitativo, os níveis de escolaridade e o ano em que o teste foi administrado. Nos índices da Psicometria Clássica , tem-se: Informações do gabarito; índice de dificuldade (ID), percentual de acertos do grupo superior (ACIM) e inferior (ABAIX), índice de discriminação (DISCR), coeficiente bisserial e ponto-bisserial, e índice de fiabilidade (IF). Tem-se as proporções de acerto de cada resposta/alternativa; e os coeficientes bisseriais de todas as respostas/alternativas. Também estão inclusos os respectivos valores corrigidos de todos os índices, juntamente do fator de concentração de análise do item. A interpretação sobre o perfil de composição dos grupos com base no valor da DISCR, por exemplo: ABAI ≤5 versus ACIM ≥9, que seria analisar e classificar segundo a diferença entre os grupos superiores e inferiores, que pode ser totalmente mista, ou com sutil predominância de um dos extremos, ou com predominância absoluta de um dos extremos No gráfico do lado esquerdo se tem os percentuais absolutos de cada alternativa, destacando em verde o modelo mental correto e em vermelho o modelo mental incorreto em destaque. No gráfico da direita se tem a <i>Análise Gráfica dos Itens (AGI)</i> , acompanhado de sua classificação, percentual médio dos respondentes que acertaram, desvio-padrão amostral, o quantitativos das violações das curvas e o valor do ângulo de inclinação, indicando o poder discriminativo
Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)
Tem-se as indicações do modelo logístico da TRI para o DIF (no caso, o não -restringido), o tamanho da amostra, o contexto, o quantitativo e número de itens considerados. Entre os parâmetros da Psicometria Moderna para o modelo de 4PL para cada contexto, tem-se: a discriminação; a dificuldade; o acerto ao acaso; e o erro ao acaso. Os dois primeiros gráficos são a <i>Curva Característica do Item (CCI)</i> e a <i>Curva de Informação do Item (CII)</i> No modelo geral, o tipo de informação é o mesmo que o anterior, sendo acrescentado os níveis da escala posicionando o item e apresentando as proporções de acertos empíricos.

Também são apresentados “indícios” da relação de dependência local para cada item, apesar de serem explorados modelos da TRI que pressupõem uma independência local, esse portanto é um aspecto delimitador no estudo e que se propõe uma complementariedade. Para cada item buscou-se elaborar um *painel de estatísticas* enriquecido de informações com base nas análises da TCT e TRI (apresentação de gráficos), contendo com isso uma perspectiva ampla e detalhada do comportamento do item dentro de cada contexto (no contexto geral, considerado conjuntamente os dois

países e todos os níveis de escolaridade), servindo assim de auxílio nas interpretações ao longo do texto e das análises mais refinadas. Segue para a parte mais qualitativa, ou da transposição quantitativa para a qualitativa (interpretação dos valores estatísticos) na qual formam os *comentários pedagógicos do item*. Eles possuem uma: a descrição dos valores dos índices da TCT e dos parâmetros da TRI (incluindo os valores de índices corrigidos, buscando tornar uma interpretação mais fiável e válida), bem como da escala de proficiências técnicas; análise explicativa sobre os distratores mais atrativos; e algumas sugestões de conteúdos com base no conhecimento prévio dentro da perspectiva ausubeliana.

A1: ITEM y1

1. Qual é a temperatura mais provável dos cubos de gelo que se encontram armazenados no congelador de uma geladeira?

a) $-10\text{ }^{\circ}\text{C}$

b) $0\text{ }^{\circ}\text{C}$

[D6]: O gelo encontra-se sempre a $0\text{ }^{\circ}\text{C}$ e/ou não é possível alterar a temperatura.

⇒ Por acreditar que o gelo somente está a $0\text{ }^{\circ}\text{C}$, em qualquer situação, ou seja, é a temperatura mais baixa que ele pode obter.

c) $5\text{ }^{\circ}\text{C}$

[C6] pouco provável: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido)

⇒ Por acreditar que o gelo encontra-se acima de $0\text{ }^{\circ}\text{C}$, independentemente que hajam trocas de calor com o ambiente interno da geladeira.

d) A temperatura depende do tamanho dos cubos de gelo

[B7]: A temperatura de um objeto depende de seu tamanho.

⇒ Por acreditar que quanto maior o corpo, mais baixa será a sua temperatura, pois é mais difícil resfriar e alcançar o equilíbrio, por isso, mais baixa será a temperatura.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Conhecer* a Temperatura de um Corpo ou de um Ambiente em Situações Típicas - [Descritores 1 e 2].

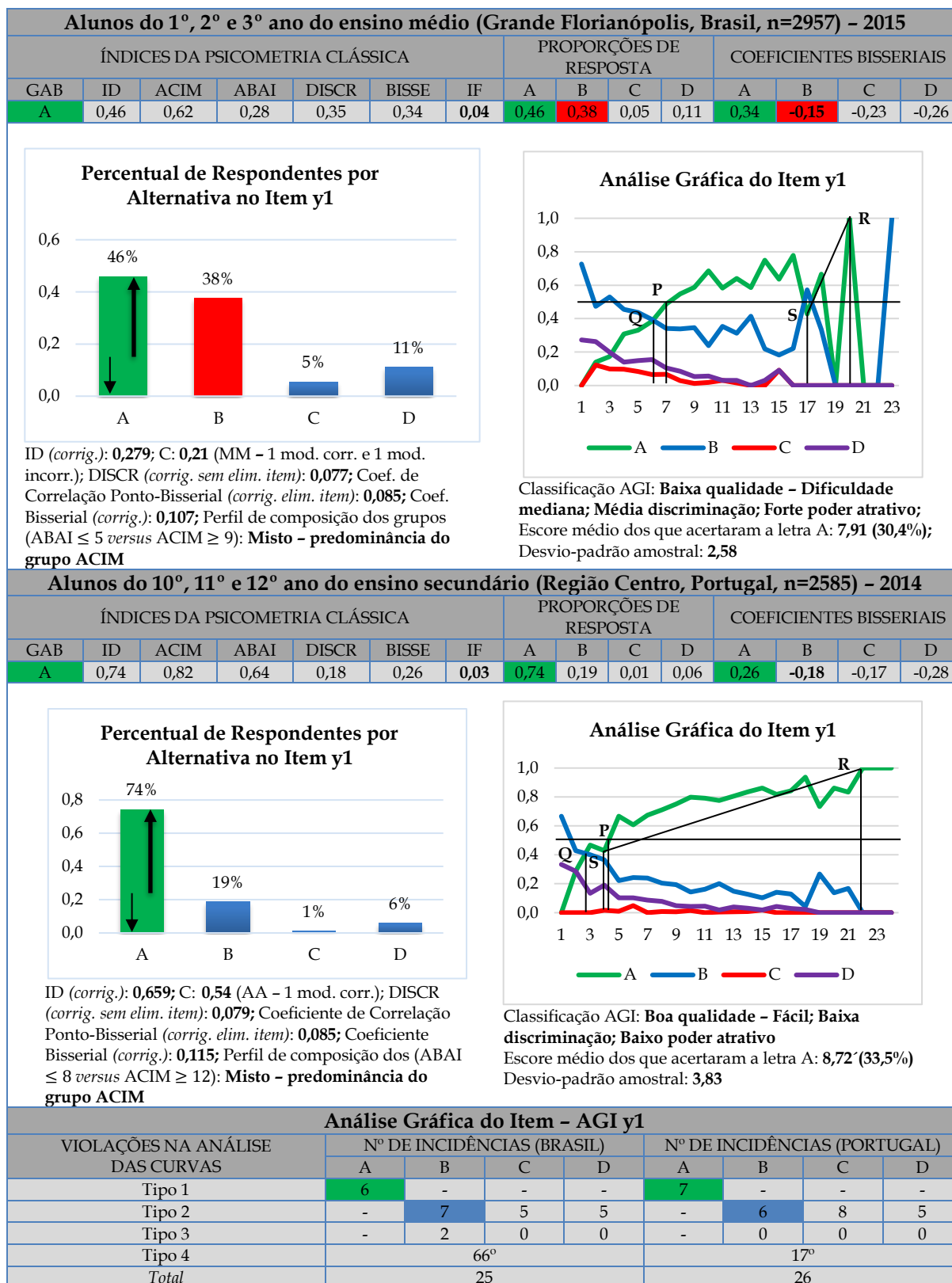
PROFICIÊNCIA*: *Prever* que a «temperatura do gelo» «dentro do congelador de uma geladeira doméstica» «após de um longo tempo está abaixo de $0\text{ }^{\circ}\text{C}$ ».

INDÍCIOS DE DEPENDÊNCIA LOCAL: não há item anterior.

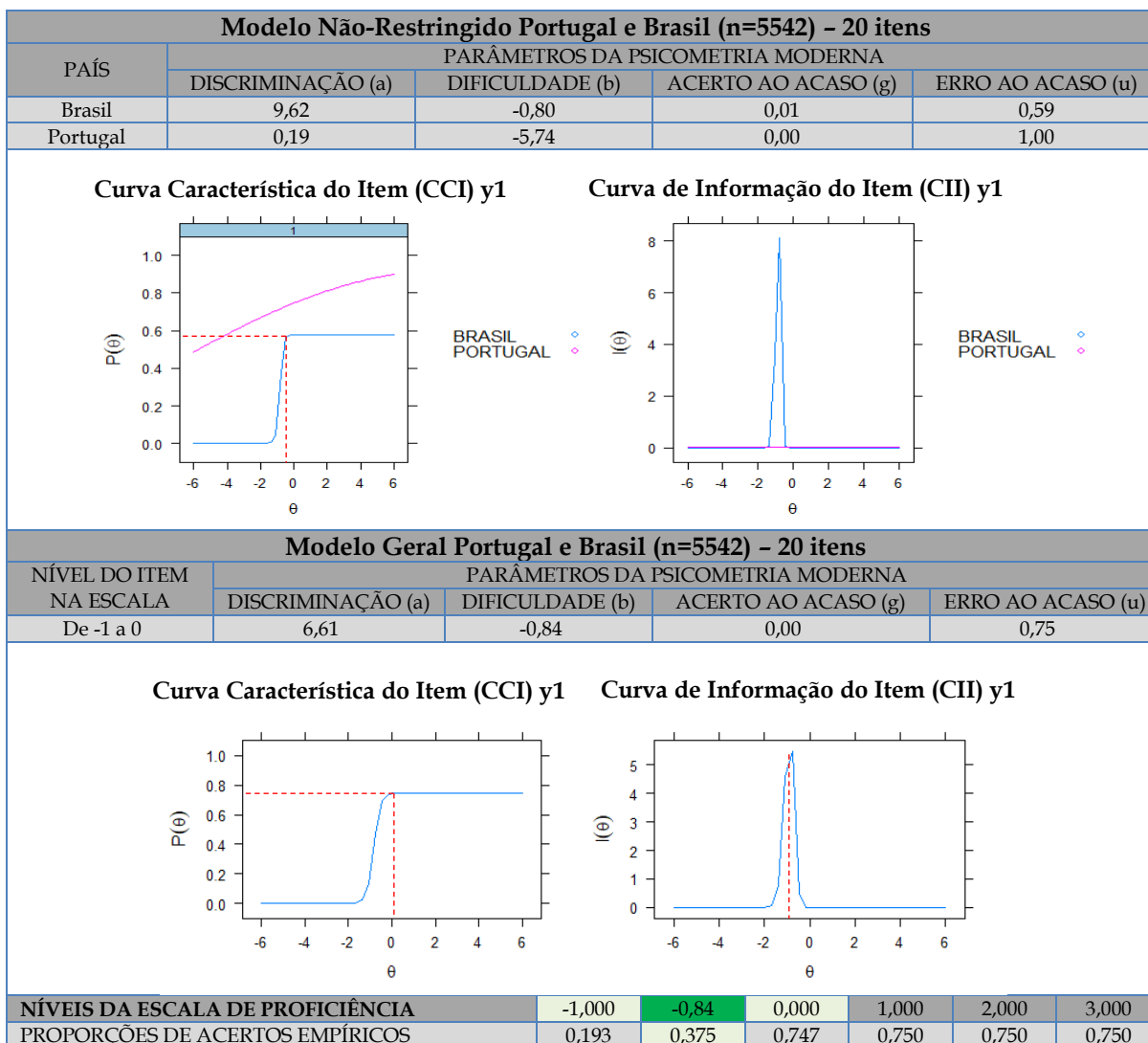
**Considera-se em negrito o subsunçor (ou parte dele) a ser evidenciado pelos respondentes de modo a se associar as informações do contexto do problema entre os outras que serão ou foram abordadas (conhecimentos potencialmente significativos), para assim representar uma assimilação conceitual.*

PAINÉL DE ESTATÍSTICAS DO ITEM y1

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y1

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Normal/mediano** (proporção esperada de respondentes 40%; $0,45 \leq ID < 0,55$) e **bom** ($0,30 \leq DISCR < 0,40$), a ser *aprimorado* (Brasil); **Fácil** (proporção esperada de respondentes 20%; $0,55 \leq ID < 0,75$) e **inadequado** ($DISCR < 0,20$), a ser *rejeitado* (Portugal). É importante ressaltar que os baixos valores dos coeficientes ponto-bisserials nos dois contextos indicam uma baixa correlação do item com o escore bruto, comprometendo a fidedignidade e consistência do teste.

O item teve predomínio de acerto absoluto (>50% do total de respostas) somente em Portugal com 74%, nos Brasil foi um predomínio relativo diante das outras alternativas com 46%, ou seja, o item é mais fácil para os alunos portugueses. O distrator "B" para os brasileiros se apresenta como um modelo mental incorreto de relevância entre os respondentes com uma força atrativa de 38%, esse mesmo distrator para os portugueses encontra-se na zona de *guessing* (18%).

No Brasil, 62% dos respondentes que acertaram pertencem ao grupo de maior desempenho, e 28% os de desempenho inferior, e em no outro contexto foi de 82% e 64% para os grupos correspondentes. Isso mostra que apesar do item ser mais fácil para Portugal, ele não tem um bom índice de discriminação (valor de 0,18), pois ambos os perfis de composição dentro dos seus grupos possuem um índice acima de 50%, enquanto no Brasil sugere discriminar melhor com um valor de 0,35.

O coeficiente bisserial da alternativa A, que corresponde a 46%, foi de 0,34 para os brasileiros, enquanto para os 74% foi de 0,26, ou seja, apesar de ser um índice maior um pouco que Portugal, ele está relacionado a um percentual inferior de indivíduos, isso significa que apesar de ter sido mais discriminativo está relacionado a uma quantidade inferior de indivíduos que acertaram e com isso, de modo geral, pode-se dizer que encontraram maiores dificuldades. Com isso, os perfis de composição de grupo (de alto e baixo desempenho) dentro de cada contexto para essa alternativa sugerem que sejam mistos, com leve predominância de indivíduos com maior desempenho relativamente, porém não muito representativo por não estar tão próximo do valor de 1, que seria o ideal.

A AGI revelou um número equiparado de violações de modo geral nos dois contextos, com maior destaque para a letra B no Brasil, e letra C em Portugal, sinalizado incertezas entre esses distratores para os grupos de diferentes níveis de escore bruto. As incertezas mais intensas proporcionalmente estão com o distrator B, em especial para alunos brasileiros com 18 acertos e portugueses com 21 acertos. Comparando os ângulos de inclinação, o Brasil demonstra um maior poder de discriminação, ou seja, a convicção de sobre a alternativa correta cresce de forma mais abrupta entre os alunos de bom desempenho, porém o gráfico mostra que isso ocorre

dentro de um intervalo de escores muito estreito na faixa de alunos com maior número acertos, e isso não é um bom sinal, ao contrário de como ocorre entre os portugueses.

Entre os distratores chama atenção a alternativa B, nos dois contextos, sendo o poder de atração maior para os respondentes brasileiros, tendo mais de 50% de alunos com desempenho de 17 valores assinalando essa alternativa, enquanto em Portugal alunos com escores iguais a 19, 20 e 21 se sentem atraídos. O bisserial dessa alternativa no Brasil foi de -0,15 e em Portugal de -0,18, isso significa que o perfil dos grupos que assinalaram o distrator B é composto mais por alunos com baixo desempenho, mas como é próximo de zero, se configura mais como um grupo misto, ou seja, quase a metade faziam parte do grupo de melhor desempenho demonstrando a força atrativa dessa alternativa.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido (cada contexto sendo analisado independentemente um do outro, porém dentro de uma mesma métrica, possibilitando assim uma comparação) o item pode ser classificado como: **Fácil** (proporção de respondentes esperados 20%; $-1,27 \leq b \leq -0,52$) e com uma **discriminação muito alta** ($a > 1,70$) (Brasil); **Muito fácil** (proporção de respondentes esperados 10%; $b \leq -1,28$) e com **discriminação muito baixa** ($0,0 < a \leq 0,35$) (Portugal). Isso demonstra problemas para o item, em especial para o contexto português.

A *Curva Característica do Item* (CCI) entre os brasileiros sugere uma elevada discriminação com valor 9,62 e dificuldade abaixo do valor médio na escala de proficiências sendo configurado dentro de um intervalo como fácil, com medida igual a -0,80. Para os portugueses a curva não revela uma forma de sigmoide relativamente uniforme, e foram obtidos os valores de discriminação igual a 0,19 e dificuldade com -5,74, ou seja, um item quase sem discriminação e muito fácil. O modelo não estima acerto ao acaso ou *guessing*, e isso indica coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os respectivos níveis de proficiência, porém um erro ao caso para alunos brasileiros de 0,59, isso representa uma probabilidade de 41% dos alunos com maior desempenho se equivocarem por descuido a partir de $\theta = 0$ (aproximadamente), seja por má interpretação, precipitação em responder, ou talvez

problemas no item. Comparativamente, pode se afirmar que dentre os alunos com mesma proficiência θ , os de Portugal possuem maior probabilidade de acerto.

A *Curva de Informação do Item* (CII) deve seguir um padrão de modelo de ogiva normal, e figurativamente mostra praticamente nenhuma informação para o contexto português, e uma maior informação na região do posicionamento do item por influência dos respondentes brasileiros devido a curva estar mais elevada, estando abaixo da média e representado por um intervalo em cerca de 1 desvio-padrão.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de -1 a 0, sendo um patamar de competência representado unicamente pela proficiência de responder o item y_1 , ou seja, é um nível com pouca informação na qual sugere uma capacidade (habilidade cognitiva) de responder o próprio y_1 (ver a escala de proficiência).

A CCI representa uma sigmoide bem inclinada e os parâmetros de dificuldade e discriminação classificam o item sendo: **Fácil** (proporção de respondentes esperados 20%; $-1,27 \leq b \leq -0,52$) e **adequado**, com uma *discriminação muito alta* ($a > 1,70$) no valor de 6,61 e índice de dificuldade igual a -0,84 na qual representa o posicionamento do item em um percentual de acertos empíricos estimados em 37,5% entre aqueles que alcançam esse nível de habilidade θ . O parâmetro do erro ao acaso sugere que a partir de 75% de respostas corretas, todos os respondentes mantêm a sua probabilidade de acerto mesmo elevando sua proficiência, como consequência se pode afirmar que eles possuam uma probabilidade de 25% de errar o item mesmo tendo boa proficiência (grupo de alto desempenho), supostamente um equívoco provocado por descuido de interpretação ou “pegadinha” do item com efeitos negativos entre os respondentes. Por fim, a CII indica uma boa discriminação, com isso, o y_1 apresenta informações relevantes para a interpretação da escala dentro desse intervalo.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item indaga qual deva ser a temperatura de cubos de gelo que já se encontram a certo tempo dentro de um congelador. Espera-se que os alunos tenham a compreensão que a água se congela a 0 °C, porém o congelador tende a baixar um pouco mais essa temperatura, com isso, a única alternativa que sinaliza essa situação é a alternativa A, com -10 °C.

A forte atração do distrator B entre os alunos brasileiros ocorre supostamente por acreditarem que a água congela a 0°C , e dessa forma o gelo permanece com essa temperatura de modo inalterado. Isso também pode estar sendo influenciado pela crença que essa seja a menor temperatura possível para qualquer corpo, de modo geral. O erro também pode estar relacionado conjuntamente a magnitude zero representar simbolicamente um valor limite e inferior para as medidas de várias outras grandezas físicas. Ou ainda que o gelo supostamente independa de as condições climáticas do congelador ficarem abaixo de 0°C . O distrator D foi a 3ª opção e menos atrativa, porém se acredita que esteja ligada a ideia errônea de que quanto maior o tamanho do gelo, “mais frio se sintam”, ou “menos calor se tenham”, conseqüentemente, a temperatura pode alcançar valores mais baixos de acordo com a suas dimensões.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: Uma opção para ser aprimorado seria a inclusão contextualizada de objetos, termos ou processos que sinalizem a possibilidade de o interior do congelador estar abaixo de 0°C . Uma sugestão seria afirmar a presença de outras substâncias congeladas, como líquidos com álcool, ou de que existem dispositivos de controle de temperatura em seu interior. Um raciocínio comparativo e analógico pode ajudar a reforçar a ideia de que o gelo esteja a uma temperatura inferior a 0°C .

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: A análise taxonômica sugere que o *traço latente* se refere a um nível de compreensão conceitual com base em um *conhecimento prévio* dentro do domínio cognitivo (se distingue do psicomotor e do socioafetivo), sendo no caso necessário para responder de acordo com o conhecimento científico. Esse conhecimento pode ser interpretado na concepção ausubeliana como um subsunçor a ser evidenciado e que corresponde um conhecimento intuitivo de que a temperatura **do gelo dentro do congelador está supostamente abaixo de 0°C** . Entende-se ainda que este traço latente se agregue não somente a subsunçores como de concepções associadas a esse conhecimento que tem em seu eixo determinados conceitos relacionados. Nisso, as

concepções caracterizadas como erro ou equívoco diante do conhecimento científico podem ser denominadas como concepções alternativas na qual são ressaltadas entre os distratores. Esse subsunçor (presente na letra A) foi mais evidenciado entre alunos entre os alunos portugueses. Pegando como exemplo a escola 1 do contexto português, constata-se 27 de um total de 117 alunos (entre o 10º, 11º e 12º ano) que não optaram por essa opção, sendo 4 alunos do 10º ano, com isso poderia ser ressaltado a necessidade de uma intervenção diferenciada a esses sujeitos para disponibilizar esses subsunçores através uma aprendizagem mecânica.

Uma noção sobre as condições térmicas do ambiente de um congelador, assim como uma compreensão mínima de seu funcionamento são questões iniciais a serem verificadas e exploradas. Uma noção de que no funcionamento de um congelador exista um termostato para controlar e manter a temperatura em determinado valor, geralmente a baixo de zero, esta pode ser uma informação ou *conteúdo potencialmente significativo* para ancorar a suposição de que o gelo pode alcançar um equilíbrio abaixo de 0 °C, obviamente o *subsunçor* ou ideia-âncora nesse caso seria o aluno ter um *conhecimento prévio* de que o gelo pode de fato ser resfriado a temperaturas abaixo de 0 °C (é o conhecimento incluso na sentença descritora do item que espera se associar a situação/contexto e alcançar uma compreensão diante do problema e assim tomar uma decisão correta), isso remete a compreender suas propriedades e comportamento, na qual o item y1 procura evidenciar.

Acredita-se na existência de possíveis *subsunçores* relacionados aos equívocos evidenciados nos distratores em que o gelo permanece igual ou acima de 0 °C, essas interpretações eventualmente se ancoram em outros conjuntos de concepções que podem conter, atrair e ancorar novos erros, como por exemplo acreditar que o gelo seja um corpo que possua a temperatura mais baixa possível, talvez pelo valor atribuído na escala Celsius que representa 0 °C. O zero é um valor representado na escala ou plano cartesiano como um valor mínimo para muitas grandezas escalares dentro da Física (área, massa, densidade, tempo, etc.), isso pode influenciar em uma confusão para o aprendiz levando a acreditar que não exista temperaturas negativas. Uma outra suposição seria o aluno admitir que determinados corpos possuam uma temperatura

própria e fixa, no caso da situação-problema a temperatura inalterável para o gelo seria 0 °C. Estas suposições por sua vez podem estar relativamente estáveis e organizadas em suas estruturas cognitivas, exercendo uma função lógica e racional para que o aluno defenda essa opinião, geralmente baseada em situações vivenciadas e transferidas de forma analógica para outros contextos levando consigo esses equívocos. Pode-se dizer que esses conhecimentos são potencialmente “significativos” para ancorar subsunçores preexistentes estruturados em equívocos, dessa forma, um ensino que não esteja atento para diagnosticar essas questões corre o risco de favorecer a construção de um conhecimento falso, em outras palavras seria uma ilusão compreensiva do aprendiz, podendo ficar cada vez mais associado e resistente a mudanças.

Entre sugestões para uma organização hierárquica dos conteúdos e uma abordagem estratégica inicial seria elaborar *organizadores prévios* para explorarem as temperaturas possíveis em que o gelo pode adquirir, sejam em situações controladas em laboratório ou em condições naturais na Terra, assim como aspectos que podem influenciar o seu ponto de solidificação parecem ser situações enriquecedoras e apropriadas, sendo importante ressaltar a recomendação de sempre valorizar e partir de diagnósticos com base na intuição dos alunos. Nessa perspectiva, o conjunto de atividades subsequentes pode buscar apresentá-lo sob diversas condições, esforço esse que corresponderia em identificar uma *família de situações-problemas* que colocasse o gelo em diversos estados térmicos, assim também favorecer o processo de *diferenciação progressiva* ao partir de ideias das mais gerais e inclusivas, pois seguindo para diversas aplicações com incorporação de detalhes, conceitos e conhecimento sobre processos mais elementares.

Entre algumas sugestões temáticas diversificadas a serem desenvolvidas tem o gelo em equilíbrio com outros estados físicos, como no caso do ponto tríplice da água, dentro disso se tornará relevante incorporar e associar o conceito de pressão e sua relação com possíveis temperaturas para o gelo. Uma situação complementar seria o fenômeno do regelo, descoberto pelo inglês John Tyndall, na qual o ponto de fusão cai à medida em que a pressão aumenta. Explorar como ocorre o processo de solidificação

da água seria um aspecto importante para compreender como se forma o gelo, buscando explicar a existência dos cristalinos ou “espaços vazios” juntamente com o seu *comportamento anômalo* (que ocorre entre 4 °C e 0 °C), situação em que se inverte a lógica do aumento de densidade com a diminuição da temperatura, e isso pode ser explorado apresentando (por meio de vídeos ou fotos) situações físicas reais. Um exemplo disso seria mostrar como se forma a superfície de gelo em um lago, assim como das mudanças na sua estrutura molecular, entre outras. Diferenciar o gelo formado em frigoríficos dos flocos de neves poderiam ser conteúdos complementares a serem explorados, em especial quanto a sua estrutura, das condições climáticas e temperaturas em que são submetidos.

A elaboração de um painel sobre as diversas situações do comportamento do gelo poderia ser realizada na função de *organizador posterior*, favorecendo o processo de *reconciliação integrativa* ao possibilitar a realização de tarefas que enfatize “subir e descer” estruturas hierárquicas conceituais na qual fazem parte da explicação de cada situação, ou seja, seguindo lógicas dedutivas e indutivas, indo de conceitos específicos para os mais estruturados e vice-versa, destacando semelhanças, diferenças e discrepâncias. Nisso incluiria os conceitos necessários para estruturar uma argumentação de que o gelo pode ficar abaixo de 0 °C, e numa sequenciação passaria pela compreensão dos processos térmicos da forma como isso ocorre, destacando o *Princípio de Trocas de Calor* até alcançar uma compreensão sobre a *Lei Zero da Termodinâmica*, na qual fundamenta o *equilíbrio térmico* para o gelo ao ser estimado a temperatura dentro das condições climáticas no interior do congelador.

A2: ITEM y2

2. Francisco pega seis cubos de gelo no congelador e coloca quatro deles dentro de um copo com água, deixando os outros dois sobre a mesa. Agita várias vezes o copo até que os cubos de gelo estejam bem pequenos e tenham parado de derreter. Qual é a temperatura mais provável da água nesta situação?

a) $-10\text{ }^{\circ}\text{C}$

[C6] *pouco provável*: **Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura.** (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por supor que a água ao ser agitada com o gelo fica abaixo de $0\text{ }^{\circ}\text{C}$.

b) $0\text{ }^{\circ}\text{C}$

c) $5\text{ }^{\circ}\text{C}$

[D7] *muito provável*: **A água no estado líquido não pode estar a $0\text{ }^{\circ}\text{C}$.**

⇒ Por acreditar que inevitavelmente a $0\text{ }^{\circ}\text{C}$ a água se torna gelo.

[C6] *muito provável*: **Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura.** (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por supor que a água ao ser agitada com o gelo esfria, mas acima de $0\text{ }^{\circ}\text{C}$.

d) $10\text{ }^{\circ}\text{C}$.

[D7] *pouco provável*: **A água no estado líquido não pode estar a $0\text{ }^{\circ}\text{C}$.**

⇒ Por acreditar que inevitavelmente a $0\text{ }^{\circ}\text{C}$ a água se torna gelo.

[C6] *pouco provável*: **Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura.** (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por supor que a água ao ser agitada com o gelo esfria, mas acima de $0\text{ }^{\circ}\text{C}$.

COMPETÊNCIAS CIENTÍFICAS ASSOCIADAS A:

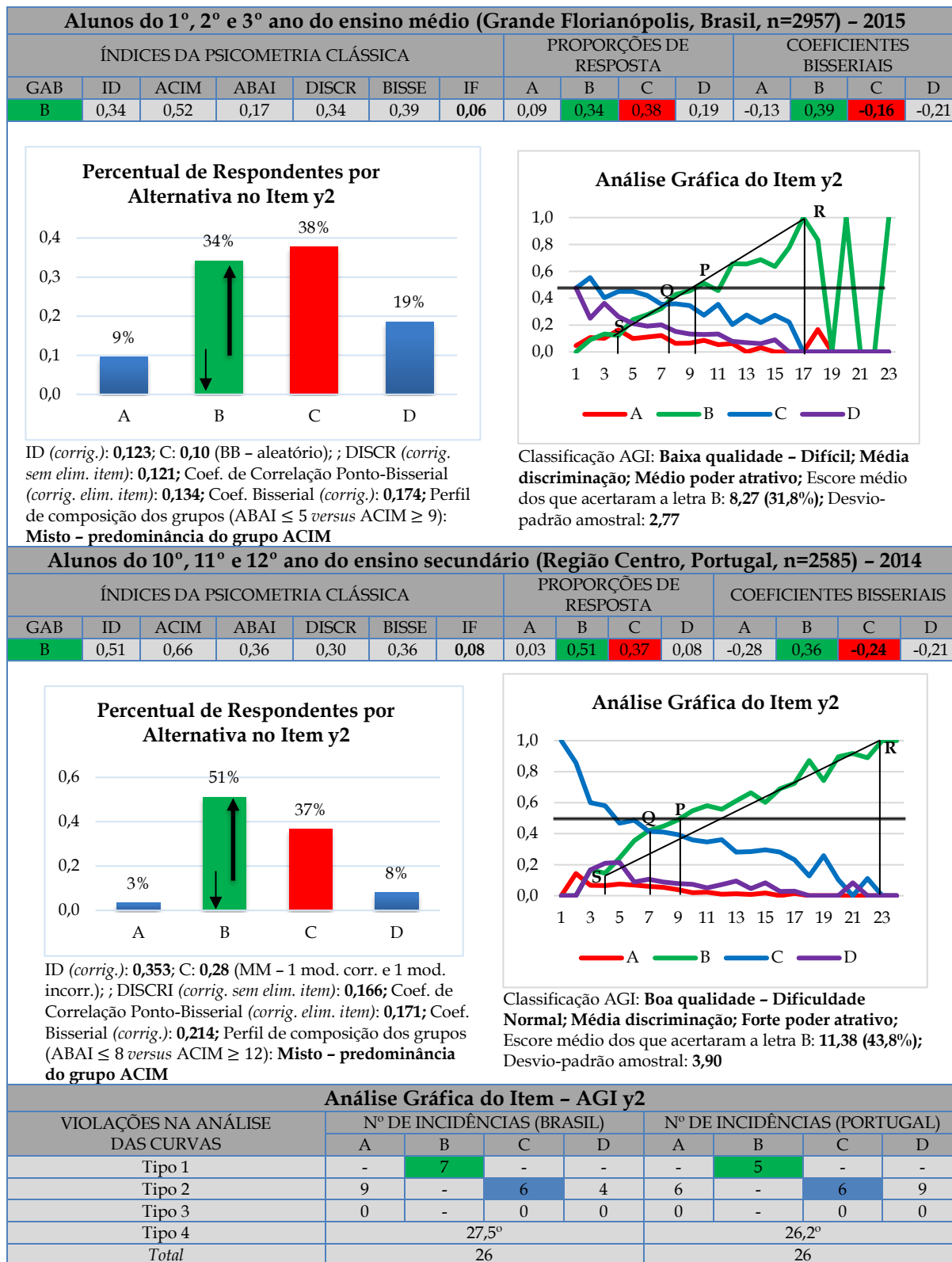
- *Conhecer* a Temperatura de um Corpo ou de um Ambiente em Situações Típicas – [Descritor 1];
- *Recordar* sobre a Invariância da Temperatura em um Processo de Mudança de Fase – [Descritor 5].

PROFICIÊNCIA: *Prever* que a «**temperatura da água em equilíbrio térmico**» «que se forma a partir do derretimento do gelo» «**permanece igual a $0\text{ }^{\circ}\text{C}$, mesmo após cessar o processo de fusão**».

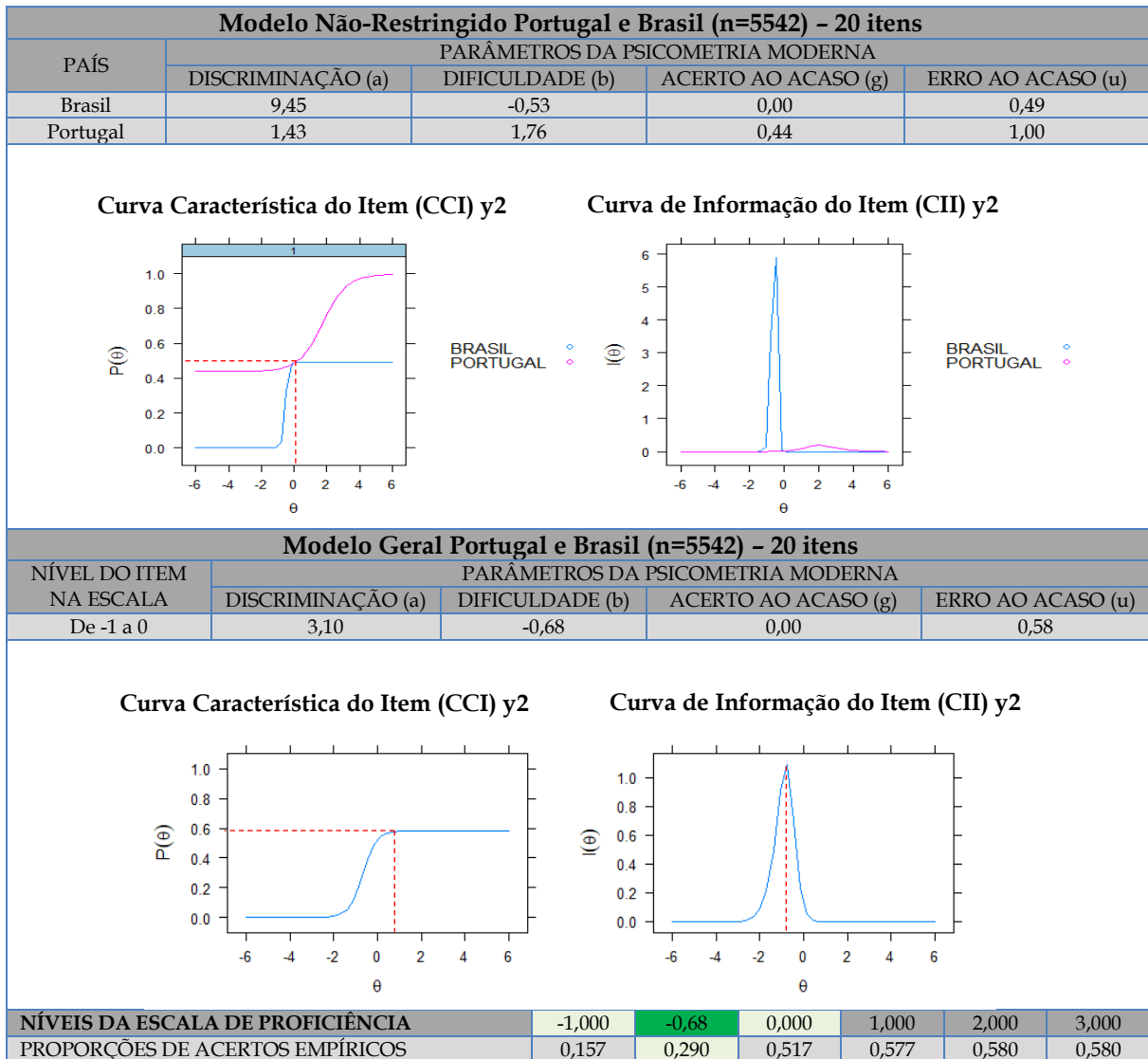
INDÍCIOS DE DEPENDÊNCIA LOCAL: O contexto do item y2 é uma sequência do item y1, na qual envolvem concepções relacionadas ao estado térmico em que o objeto estava anteriormente, no caso, a temperatura do gelo em um congelador. Com isso, se pode considerar que o item y1 e y2 fazem parte da *mesma família de situações-problema* quando ao contexto da situação de vida real. Apesar da eventual dependência da condição inicial, a questão se concentra numa compreensão sobre a temperatura de equilíbrio térmico da mesma substância em dois estados físicos.

PAINÉL DE ESTATÍSTICAS DO ITEM y2

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y2

Com base na estimação dos parâmetros da TCT os indicadores da dificuldade e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **bom** ($0,30 \leq DISCR < 0,40$), mas a *ser aprimorado* (Brasil); **Normal/mediano** (proporção esperada de respondentes 40%; $0,45 \leq ID < 0,55$) e **bom** ($0,30 \leq DISCR < 0,40$), mas *ser aprimorado* (Portugal).

A alternativa correta B para o Brasil ficou em 2ª posição na proporção total com 34%, perdendo para o distrator C com 38%. Em Portugal houve uma predominância

absoluta da alternativa correta (>50% do total de respostas) com 51%, porém, o distrator C demonstrou também força diante dos outros com 37%.

O grupo de melhor desempenho de alunos brasileiros (de 9 a 23 acertos no escore bruto) teve uma proporção de acerto de 52%, enquanto o grupo inferior com 17% (até de 5 acertos no escore bruto). No outro contexto esses valores foram de 66% (de 12 a 23 acertos no escore bruto) e 36% (até de 8 acertos no escore bruto), com uma discriminação sutilmente menor, sendo o grupo inferior acertando um pouco mais do dobro que o grupo correspondente no Brasil.

Os valores dos coeficientes bisseriais foram de 0,39 (Brasil) e 0,36 (Portugal), e isso sugere um perfil de composição de respondentes que acertaram formado por alunos pertencentes aos grupos de alto e baixo desempenho de modo equivalente, porém distante do valor igual a 1, que seria mais discriminativo.

O distrator C se destaca devido força atrativa nos dois contextos, sendo a alternativa com maior proporção entre os alunos brasileiros (38%), e em Portugal, atraindo alunos com desempenho igual a 22 acertos. Os bisseriais dessa alternativa foram iguais a -0,16 e -0,24, como no Brasil é mais próximo de zero tende a ser menos discriminativo, ou seja, mais misto, sendo formado por alunos de maior e menor desempenho.

A AGI mostra um mesmo número de violações entre os contextos, assim como um poder de discriminação equivalente e baixo devido ao ângulo de inclinação ser para ambos em torno de 27°. A alternativa correta teve um pouco mais de oscilações percentuais entre os brasileiros (violações do tipo 1), sinalizando incertezas entre os respondentes com 15 e 18 acertos ao serem atraídos pelos distratores C e A. O distrator A apresentou o maior número de violações do tipo 2 entre os brasileiros, enquanto em Portugal foi a letra D, entretanto são em sua maioria variações de pequena proporção.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Fácil** (proporção esperada de respondentes 20%; $-1,27 \leq b \leq -0,52$) e com uma **discriminação muito alta** ($a > 1,70$) (Brasil); **Muito difícil** (proporção esperada de respondentes 10%; $b > 1,27$) e com

discriminação alta ($1,35 \leq a \leq 1,70$) (Portugal). Isso demonstra problemas para o item, em especial para o contexto português.

A *Curva Característica do Item* (CCI) sugere uma maior dificuldade para os alunos portugueses e os brasileiros com uma melhor discriminação, porém chama a atenção que o ajuste indica um elevado acerto ao acaso com um valor de 0,44 e ausência de erro por descuido entre os alunos de elevada proficiência. O elevado índice de *guessing* indica falta coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os respectivos níveis de proficiência alcançados. Para os brasileiros além de uma proficiência que atingem uma probabilidade de 49% de acertos, a curva sugere uma probabilidade de erro (“upper”) de 51% por descuido a partir de $\theta = 0$ (aproximadamente). Comparando a CCI de cada contexto, se pode afirmar que dentre os alunos com mesma proficiência θ , os de Portugal possuem maior probabilidade de acerto, com exceção do ponto de interseção entre as curvas com proficiência em torno da média, na qual ambos possuem a mesma probabilidade de acerto, sendo próximo de 50%.

A *Curva de Informação do Item* (CII) deve seguir um padrão de modelo de ogiva normal, e figurativamente mostra praticamente nenhuma informação para o contexto português, e uma maior informação na região do posicionamento do item por influência dos respondentes brasileiros devido a curva estar mais elevada e numa região de baixa proficiência, estando abaixo da média e representado por um intervalo em cerca de 1 desvio-padrão.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de -1 a 0, representado por um patamar de competência com proficiência em responder somente o item y_1 , ou seja, de pouca informação.

A CCI representa uma sigmoide bem inclinada e os parâmetros de dificuldade e discriminação classificam o item sendo: **Fácil** (proporção esperada de respondentes 20%; $-1,27 \leq b \leq -0,52$) e **Excelente**, com uma *discriminação muito alta* ($a > 1,70$) no valor de 3,10 e índice de dificuldade igual a -0,68 na qual representa o posicionamento do item em um percentual de acertos empíricos estimados em 29% entre aqueles que

alcançam esse nível de habilidade θ . O parâmetro do erro ao acaso sugere que a partir de 58% de respostas corretas, todos os respondentes mantêm a sua probabilidade de acerto mesmo elevando sua proficiência, como consequência se pode afirmar que eles possuam uma probabilidade de 42% de errar o item mesmo tendo boa proficiência (grupo de alto desempenho), partindo de $\theta = 1$ (aproximadamente) supostamente um equívoco provocado por descuido de interpretação ou “pegadinha” do item com efeitos negativos entre os respondentes. Por fim, a CII indica uma boa discriminação, com isso, o y_2 apresenta informações relevantes para a interpretação da escala dentro desse intervalo, porém ele não foi considerado como um item-âncora.

DANDO SENTIDO AOS ERROS/DIFICULDADES: Os alunos sentem atraídos supostamente pela crença de que ao “agitar” os gelos a temperatura irá aumentar e será obtida água a uma temperatura sutilmente superior à do gelo em fusão (0 °C). A palavra “agitar” pode remeter ideia de que a fricção, e essa de que provoca aquecimento e elevação da temperatura, com isso, se constrói a crença de que a “agitação” do gelo produza calor e se eleve a temperatura, dessa se torna conveniente acreditar que a temperatura da água ultrapasse 0 °C e alcance 5 °C, ou ainda 10 °C. Esses os equívocos se relacionam com uma falta de compreensão sobre as propriedades e comportamentos térmicos do gelo. Apesar do enunciado afirmar que o gelo “para” de derreter, sabe-se que isso não acontece efetivamente, pois o sistema continua a trocar calor com o meio, esse talvez seja ainda um fator que possa estar sendo levado em consideração e se conclua o fim se dá em uma temperatura mais próxima à do ambiente, e muito provavelmente acima de 0 °C. Remotamente, se supõe ainda que haja um erro associado ao aumento da pressão (devido à agitação do gelo no copo) sobre a alteração na temperatura de mudança de fase (fusão/solidificação), de forma correspondente as variações das condições da pressão atmosférica local. De modo geral, os possíveis equívocos demonstram um desconhecimento da constância da temperatura durante a mudança de fase, e de que a água e o gelo podem coexistir com a mesma temperatura em estados diferentes no momento em que cessem trocas de calor entre eles.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: Não se considera a necessidade de aprimoramento no enunciado ou entre os distratores, pois as informações estão claras e minimamente necessárias para uma boa interpretação e tomada de decisão.

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: O enunciado apresenta a ideia sobre o estado de equilíbrio térmico da água com o gelo ao enunciar que o gelo parou de derreter, esse tipo de informação é potencialmente significativo para se associar a ideia de que a **temperatura da água permanece igual a 0 °C, mesmo após cessar o processo de fusão**, que no caso se caracteriza como parte do conhecimento do *subsunçor*.

Os *subsunçores* estão relacionados com uma compreensão de conceitos e processos que justificam a temperatura de um corpo permanecer constante ao receber/absorver energia térmica em uma mudança de fase, destacando a situação na qual o equilíbrio térmico possa ser alcançado levando em conta que partes de uma mesma substância coexistam em fases distintas. Na situação-problema ocorre que parte do gelo derrete e ainda assim se alcance o equilíbrio térmico com a água líquida advinda de seu próprio derretimento. Chama atenção as evidências de *subsunçores* que relacionadas a equívocos na qual se associam a falsos conhecimentos, que caso são os distratores. Acredita-se que o gelo alcance a temperatura de 5 °C e ainda 10 °C, considerando que uma das possíveis causas seja a transferência analógica de que se fricção entre duas mãos (p.ex.) produza calor e aqueça elevando a temperatura, e o mesmo deva acontecer com a água produzida pela fricção do gelo na agitação do copo.

Do ponto de vista dos *conhecimentos potencialmente significativos* e da organização hierárquica dos conteúdos supõe-se necessário explorar os conceitos de calor e temperatura com um olhar sobre o comportamento das partículas partindo de uma noção intuitiva, e assim seguir para um entendimento mais elaborado levando em consideração a situação como ocorre o aquecimento. A abordagem inicial deve seguir conjuntamente com as leis mais gerais da Termodinâmica, e sempre que possível e necessário, recorrer ao Modelo Cinético-Molecular (MCM), seja para auxiliar na

explicação dos conceitos, desmistificar os diversos aspectos fenomenológicos da física térmica, assim como na elaboração e uso de aparatos experimentais, sequenciação de atividades e articulador na argumentação de quem ensina.

Diversas faces da mesma situação, incluindo diferentes níveis de observação, que pode ser direta (mesoscópica), indireta (microscópica ou macroscópica), são caminhos a serem explorados na qual envolvem múltiplas habilidades cognitivas e faces distintas quanto a natureza epistemológica do conteúdo sobre o mesmo fenômeno, esse tipo de perspectiva revela que dentro de uma situação-problema podem caber internamente uma família problemas que fazem parte da mesma situação e que ajudam e enriquecem a compreensão sobre o fenômeno ao possibilitar uma aprendizagem mais aprofundada e um conhecimento mais inclusivo e integrador. Na sequenciação da aprendizagem, o MCM pode vir auxiliar apresentando uma mudança na estrutura e organização das partículas constituintes da matéria entre as fases, cabendo uma intradisciplinaridade com outras situações e fenômenos térmicos correspondentes, que podem ser caracterizadas dentro de uma mesma família de situações-problema quanto a contextualização disciplinar (e não necessariamente de situação real do cotidiano ou trabalhista), como por exemplo a expansão gasosa, a dilatação térmica, o funcionamento do termômetro de mercúrio, entre outros, entretanto depende da estratégia de ensino e dos objetivos estabelecidos. Uma abordagem com aplicações diversificadas que envolva os conceitos mais relevantes e as leis e princípios que regem o mesmo fenômeno térmico é um aspecto que favorece o processo da *diferenciação progressiva* ao tornar o conhecimento cada vez incorporado em detalhes e especificidades, com isso favorecendo uma compreensão mais alargada, processo de aprendizagem na qual pode ser compreendida no âmbito de desenvolvimento e estabelecimento de competências científicas cognitivas pela múltiplas habilidades envolvidas

Na situação-problema apresentada, os conceitos-chave a serem explorados de modo a contribuir numa aprendizagem mais efetiva seriam: calor, temperatura, produção do calor por fricção, equilíbrio térmico e uma noção sobre calor latente. Convém ressaltar que o estudo de Astrid Doppenschmidt e Hans Jurgen Butt em 1998

mostrou que o derretimento do gelo é desencadeado por um processo de relaxamento das ligações químicas da molécula, na qual ocorre bem antes de temperatura propriamente dita de fusão, e existem diversas implicações nesse processo que explicam o seu baixo coeficiente de atrito, sua facilidade de compactação, e a alta aderência de superfície (Butt & Kappl, 2010). Essas questões são *complementares*, mas podem ser exploradas de forma didaticamente adequada aos alunos quando se refere ao processo de derretimento do gelo.

A3: ITEM y3

3. Os cubos de gelo que Francisco deixou sobre o balcão quase derreteram por completo e repousam sobre uma poça de água, ficando ainda pequenos pedaços de gelo. Qual é a temperatura mais provável desses cubos de gelo menores?

a) $-10\text{ }^{\circ}\text{C}$

[C6] *pouco provável*: **Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura.** (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por supor que o gelo ao repousar no balcão derrete ficando abaixo de $0\text{ }^{\circ}\text{C}$.

[C7] *pouco provável*: **Objetos quentes naturalmente se resfriam, objetos frios naturalmente se aquecem.** (tange a concepção animista, como se os objetos tivessem vontade própria de se aquecerem ou se resfriarem espontaneamente, independentemente de qualquer fator externo).

⇒ Por acreditar que o gelo derreta naturalmente, se aquecendo por si só, sem trocas de calor com o balcão.

b) $0\text{ }^{\circ}\text{C}$

c) $5\text{ }^{\circ}\text{C}$

[C1] *pouco provável*: **Aquecimento sempre resulta em um aumento na temperatura.** (não compreende o conceito de calor latente).

⇒ Por acreditar que o gelo permaneça a aquecer e elevar sua temperatura acima de $0\text{ }^{\circ}\text{C}$

d) $10\text{ }^{\circ}\text{C}$

[C1] *pouquíssimo provável*: **Aquecimento sempre resulta em um aumento na temperatura.** (não compreende o conceito de calor latente).

⇒ Por acreditar que o gelo permaneça a aquecer e elevar sua temperatura acima de $0\text{ }^{\circ}\text{C}$

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: Recordar sobre a Invariância da Temperatura em um Processo de Mudança de Fase – [Descritor 5].

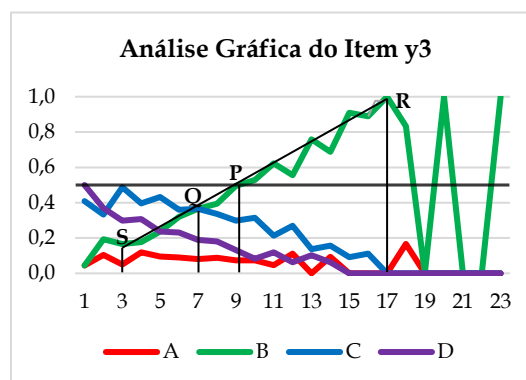
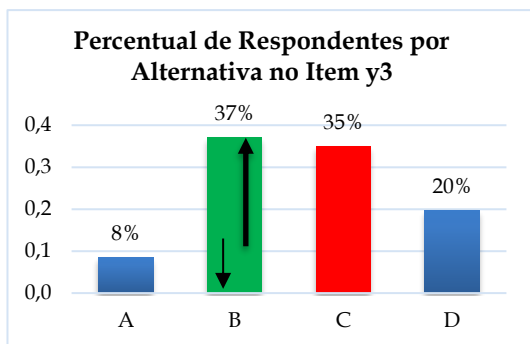
PROFICIÊNCIA*: Prever que a «temperatura do gelo em equilíbrio térmico» «com a água que se forma com seu derretimento» «permanece a $0\text{ }^{\circ}\text{C}$, mesmo após cessar o processo de fusão».

INDÍCIOS DE DEPENDÊNCIA LOCAL: O contexto do item y3 é uma variação dentro do mesmo contexto do item y2 que segue desde o item y1, na qual envolvem concepções relacionadas ao estado térmico em que o objeto estava inicialmente, nomeadamente sobre a sua [gelo] temperatura em um congelador. Os itens y1, y2 e y3 se caracterizam dentro de uma *mesma família de situações-problemas*, pois ainda que a situação foque em um objeto em suas distintas, eles se apresentam interligados dentro um *mesmo contexto de vida real* (suposto) na sequência do enunciado e apresentam um conjunto de conhecimentos associados a serem aferidos. Apesar da eventual dependência da condição inicial, o item y3 se concentra numa compreensão sobre a temperatura de equilíbrio térmico da mesma substância (água) em dois estados físicos (sólido-líquido).

PAINÉL DE ESTATÍSTICAS DO ITEM y3

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

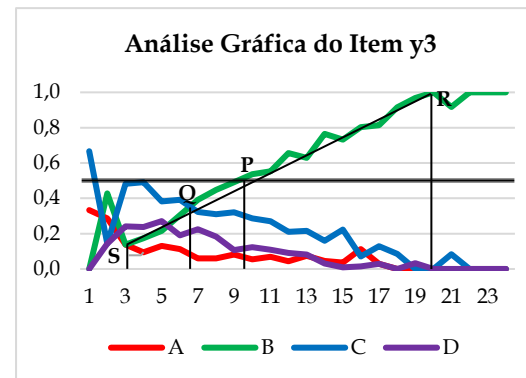
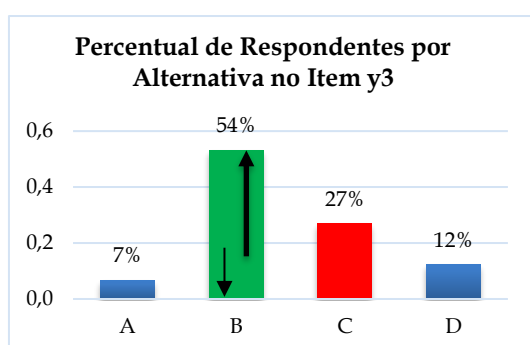
Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
B	0,37	0,56	0,20	0,36	0,39	0,05	0,08	0,37	0,35	0,20	-0,06	0,39	-0,17	-0,26



ID (corr.): **0,161**; C: **0,10** (BB - aleatório); DISCR (corr. sem elim. item): **0,114**; Coef. de Correlação Ponto-Bisserial (corr. elim. item): **0,127**; Coef. Bisserial (corr.): **0,162**; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): **Misto - predominância do grupo ACIM**

Classificação AGI: **Boa qualidade - Difícil; Média discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra B: **8,19 (31,5%)**; Desvio-padrão amostral: **2,84**

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
B	0,54	0,74	0,34	0,40	0,47	0,13	0,07	0,54	0,27	0,12	-0,17	0,47	-0,28	-0,33



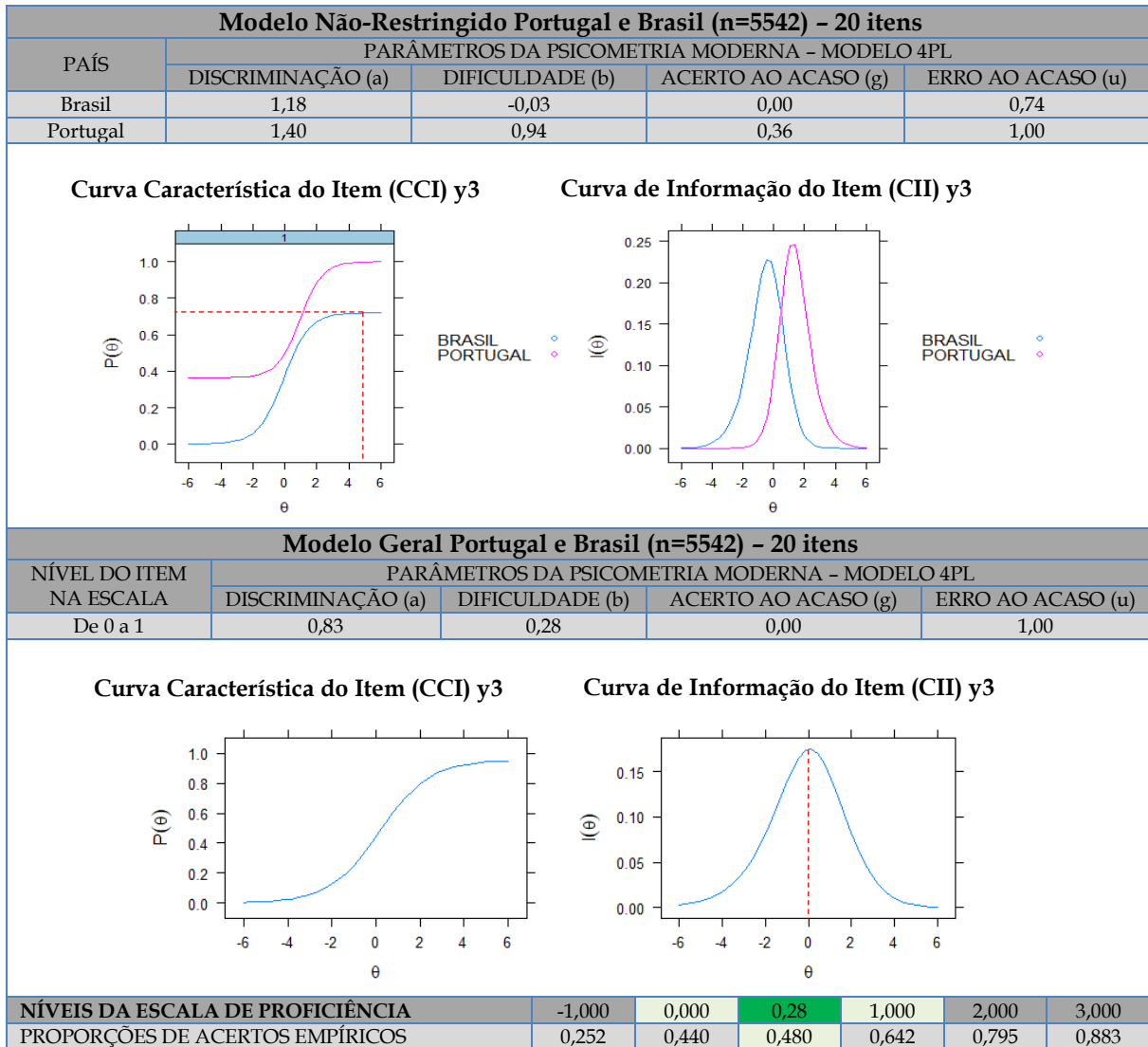
ID (corr.): **0,381**; C: **0,23** (MM - 1 mod. corr. e 1 mod. incorr.); DISCR (corr. sem elim. item): **0,254**; Coef. de Correlação Ponto-Bisserial (corr. elim. item): **0,257**; Coef. Bisserial (corr.): **0,322**; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): **Predominância do grupo ACIM**

Classificação AGI: **Boa qualidade - Dificuldade normal; Média discriminação; Baixo poder atrativo**; Escore médio dos que acertaram a letra B: **11,64 (44,8%)**; Desvio-padrão amostral: **3,93**

Análise Gráfica do Item - AGI y3

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	-	7	-	-	-	4	-	-
Tipo 2	6	-	7	3	5	-	8	8
Tipo 3	0	-	0	0	0	-	0	0
Tipo 4	29º				26,3º			
Total	23				25			

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y3

Com base na estimação dos parâmetros da TCT os indicadores de dificuldades e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **bom** ($0,30 \leq DISCR < 0,40$), mas *sujeito a aprimoramento* (Brasil), e **Normal/mediano** (proporção esperada de respondentes 40%; $0,45 \leq ID < 0,55$) e **adequado** ($DISCR \geq 0,40$) (Portugal).

A porcentagem de acertos (37%) foi quase igual ao distrator C (35%) para o Brasil, e em Portugal há um predomínio absoluto da alternativa correta B (54%), porém o distrator C também revela forte atração (27%) no percentual total. Os alunos

brasileiros de maior desempenho representam 56% dentro dos 27% com melhores escores (a partir dos que alcançaram 9 valores), e 20% para os de desempenho inferior (até 5 valores no escore bruto). Em Portugal esses grupos correspondem a 74% (de 12 até 24 valores no escore bruto) e 34% (todos com escore de 1 até 8 valores), com uma diferença maior entre esses grupos no contexto português.

O valor do bisserial da alternativa correta é igual a 0,39 e 0,47, sendo o valor maior para Portugal, significa que o perfil de composição desse grupo seja composto mais por alunos com bom desempenho que no Brasil. É recomendável verificar o distrator C, pois foi atrativo no Brasil para alunos com desempenho de 18 escores e em Portugal para alunos com 21 acertos. O bisserial da letra C foi de -0,17 e -0,28, ou seja, um grupo relativamente misto, com sutil predominância de alunos com baixo desempenho em ambos os contextos, sendo um pouco mais discriminante nesse sentido entre os alunos portugueses.

A AGI mostrou uma quantidade de violação quase igual entre os contextos, entretanto, a alternativa A e C chamam atenção no Brasil, em especial a letra A para os grupos com 18 acertos, enquanto no outro contexto os distratores C e D, em especial a letra C para os alunos com 21 acertos. O poder de discriminação, ou seja, o arranque desde o valor proporcional mais baixo de acertos (ponto S) até o mais elevado (ponto R, ou seja, 100% de respostas corretas em determinado nível de escore) mostra valores próximos para o ângulo de inclinação entre eles.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Normal/mediano** (proporção esperada de respondentes 40%; $-0,51 \leq b \leq 0,51$) e **bom**, *sujeito a aprimoramento*, com uma *discriminação moderada* ($0,65 \leq a \leq 1,35$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **muito bom**, *sujeito a aprimoramento*, com uma *discriminação alta* ($1,35 \leq a \leq 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma maior dificuldade para os alunos portugueses, porém chama a atenção que o ajuste indica um elevado acerto ao acaso com um valor de 0,36 (valor esperado de $\frac{1}{4}$ ou 0,25) e ausência de erro por descuido entre os alunos de elevada proficiência. Para os brasileiros o ajuste sugere

ausência de acerto ao acaso (*guessing* ou “chute”), indicando assim coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência. Aqueles com proficiência além dos primeiros que atingem uma probabilidade de 74% de acertos, estima-se uma probabilidade de erro de 26% por descuido a partir de $\theta = 5$ (aproximadamente), ou seja, erram ainda que sua proficiência aumente. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto.

A elevação na *Curva de Informação do Item* (CII) para cada contexto indica uma boa informação do item, sendo a escala mais bem representada pelos alunos portugueses pela proximidade com o posicionamento do item y_3 , estando em um nível mais elevado de proficiência sendo representando por um intervalo em cerca de 2 desvios-padrões.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 0 a 1, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y_3 , y_{13} , y_{23} , y_{10} , y_{22} (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI representa uma sigmoide bem inclinada e os parâmetros de dificuldade (0,28) e discriminação (0,83) classificam o item sendo: **Normal/ Mediano** (proporção esperada de respondentes 40%; $-0,51 < b \leq 0,51$) e **Bom, sujeito a aprimoramento**, com uma *discriminação moderada* ($0,65 < a \leq 1,35$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 48% entre aqueles que alcançam esse nível de habilidade θ . Não foram estimados acertos ou erros ao acaso. Por fim, a CII indica uma boa discriminação, com isso, o y_3 apresenta informações relevantes para a interpretação da escala dentro de um intervalo de aproximadamente 3 desvios-padrões em torno da média geral, sendo considerado um item-âncora.

DANDO SENTIDO AOS ERROS/DIFICULDADES: A letra C diz que os pequenos pedaços de gelo derretido sobre a mesa em meio a uma poça de água encontram-se a uma temperatura igual a 5 °C, isso leva a crer que os alunos devam

imaginar que os gelos quando aquecidos e derretidos aumentam de temperatura nesse processo, ou seja, não associam que no processo de mudança de fase a temperatura deva se manter constante. Obviamente que essa opinião seja mais comum entre alunos do 1º ano (Brasil) e 10º ano (Portugal) ao fato de ainda não terem sido discutidos esses assuntos, apesar de que os alunos brasileiros tenham visto de forma superficial na série anterior, na qual apresenta e discute qualitativamente esses fenômenos.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: Não se considera a necessidade de aprimoramento no enunciado ou entre os distratores, pois as informações estão claras e minimamente necessárias para uma boa interpretação e tomada de decisão.

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: O *subsunçor* a ser evidenciado se refere ao conhecimento preexistente de que a **temperatura do gelo permanece a 0 °C, mesmo após cessar o processo de fusão**. Considera-se que ele possa ter sido internalizado de forma arbitrária e literal, ou não, para cada situação mudam-se o seu potencial de ancoragem para uma nova informação. Sendo não-literal, acredita-se que hajam supostos conhecimentos associados previamente a esse subsunçor, e esses se concentram na zona de compreensão sobre o processo de mudança de fase. Esse tipo de conhecimento se configura como capaz de ancorar novas informações advindas de situações diversificadas, incluindo o estado de equilíbrio térmico entre o gelo e a água. Obviamente que novas informações posteriormente abordadas podem se configurar como potencialmente significativas a esse subsunçor, como por exemplo a compreensão da relação dentro de um conjunto de propriedades térmicas, químicas e gerais, assim como situações em que se alteram que possam ser alteradas o ponto de fusão, ou quando a água se mistura a outras substâncias e se torna heterogênea, entre outras novas informações.

Para o processo de seleção e organização hierárquica dos conteúdos é importante destacar que uma mesma substância pode coexistir em diferentes estados físicos, com isso, investigar e evidenciar essas situações empiricamente são caminhos

que podem explorar muita discussão e riqueza no cenário pedagógico, obviamente, sempre partindo das noções que os alunos possuem previamente sobre essas questões. Um caminho que parece promissor estaria no evento do *ponto triplo* ou *ponto tríplice da água*, na qual envolvem questões relacionadas as condições sobre os valores de pressão e temperatura para uma substância em questão, apresentar sob exemplo de diferentes substâncias poderia contribuir bastante nesse processo. Substâncias com poliformismo são exemplos em que os equilíbrios entre estados físicos podem acarretar diferentes estados sólidos para um determinado estado por exemplo, e o quantidade de pontos tríplices para p fases possíveis pode ser dada pela relação:

$$p = \frac{p!}{(p-3)! \cdot 3!}$$

Sugestões do item anterior são válidas para a identificação e o enfretoamento dos mesmos equívocos do item y3.

A4: ITEM y4

4. Sobre o fogão está uma panela cheia de água. A água começou a ferver rapidamente. A temperatura mais provável da água é:

a) 88 °C

⇒ Acredita-se que a água aqueceu rapidamente por atingir o ponto de ebulição a uma temperatura mais baixa que o habitual. (*): *Não houve correspondência direta com nenhuma concepção do inventário).*

b) 98 °C

c) 110 °C

[C1]: Aquecimento sempre resulta em um aumento na temperatura. (*não compreende o conceito de calor latente*)

⇒ Por acreditar que quando se ferve a temperatura continua a crescer e portanto deva estar acima de 100 °C.

d) Nenhuma das respostas acima é correta.

[D5]: O ponto de ebulição da água é de 100°C (apenas).

⇒ Por acreditar que a temperatura de fervura não se altera jamais, mesmo que o corpo continue sendo aquecido, ou haja qualquer tipo de outra influência externa.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Conhecer* a Temperatura de um Corpo ou de um Ambiente em Situações Típicas – [Descritor 3].

PROFICIÊNCIA*: *Recordar* que a «**temperatura mais provável da água fervente**» «**é igual ou próximo a 100 °C**, considerando condições mais comuns, ou seja, próximo ao nível do mar».

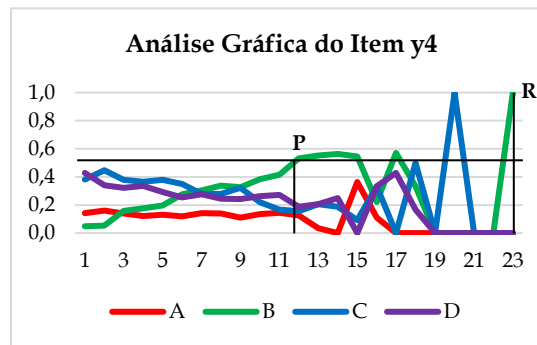
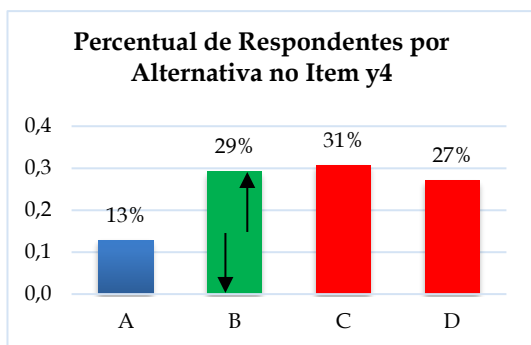
Indícios de Dependência Local: Não foi identificado entre os itens anteriores.

**Considera-se em negrito o subsunçor (ou parte dele) a ser evidenciado pelos respondentes de modo a se associar as informações do contexto do problema entre os outras que serão ou foram abordadas (conhecimentos potencialmente significativos), para assim representar uma assimilação conceitual.*

PAINÉL DE ESTATÍSTICAS DO ITEM y4

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

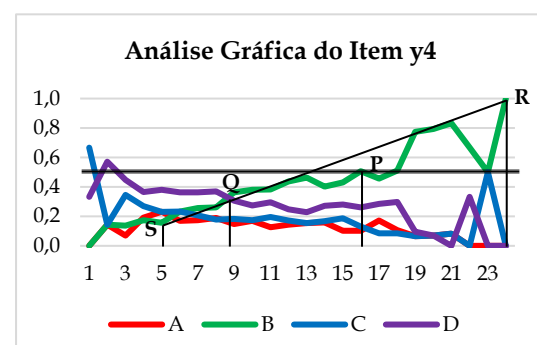
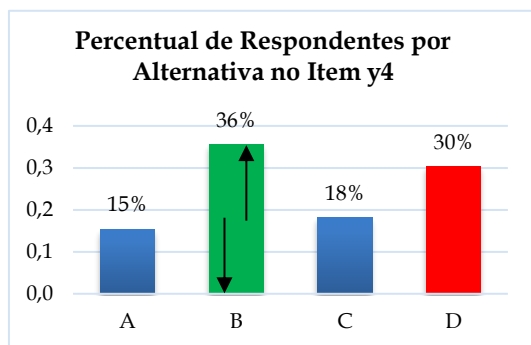
Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSEERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
B	0,29	0,39	0,17	0,22	0,27	0,01	0,13	0,29	0,31	0,27	-0,03	0,27	-0,15	-0,08



ID (corr.): **0,056**; C: **0,04** (BB - aleatório); DISCR (corr. sem elim. item): **0,031**; Coef. de Correlação Ponto-Bisserial (corr. elim. item): **0,033**; Coef. Bisserial (corr.): **0,044**
 Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): **Misto**

Classificação AGI: **Baixa qualidade - Difícil; Baixa discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra B: **7,98 (30,7%)**; Desvio-padrão amostral: **2,71**

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSEERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
B	0,36	0,48	0,23	0,25	0,01	0,06	0,15	0,36	0,18	0,30	-0,06	0,01	-0,11	-0,13



ID (corr.): **0,144**; C: **0,05** (BB - aleatório); DISCR (corr. sem elim. item): **0,115**; Coef. de Correlação Ponto-Bisserial (corr. elim. item): **0,118**; Coef. Bisserial (corr.): **0,152**
 Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): **Misto**

Classificação AGI: **Baixa qualidade - Difícil; Média discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra B: **11,51 (44,3%)**; Desvio-padrão amostral: **4,01**

Análise Gráfica do Item - AGI y4

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	-	5	-	-	-	6	-	-
Tipo 2	6	-	7	8	12	-	9	9
Tipo 3	-	-	-	-	0	-	1	0
Tipo 4	INDETERMINADO				20º			
Total	26				36			

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)

Modelo Geral Portugal e Brasil (n=5542) – 26 itens				
NÍVEL DO ITEM NA ESCALA*	PARÂMETROS DA PSICOMETRIA MODERNA – MODELO 4PL			
	DISCRIMINAÇÃO (a)	DIFICULDADE (b)	ACERTO AO ACASO (g)	ERRO AO ACASO (u)
De 3 a 4	0,50	3,41	0,21	0,91

Curva Característica do Item (CCI) y4

Curva de Informação do Item (CII) y4

NÍVEIS DA ESCALA DE PROFICIÊNCIA	-1,000	0,000	1,000	2,000	3,000	3,410	4,000
PROPORÇÕES DE ACERTOS EMPÍRICOS	0,277	0,315	0,369	0,440	0,524	0,560	0,612

*O item y4 foi excluído da escala de proficiência formada pelos 20 itens, sendo o posicionamento nesta suposta escala de acordo com a estimativa inicial contendo todos os itens. Esta análise serve para uma compreensão de sua caracterização em um contexto geral considerando 26 itens de acordo com o modelo de 4PL, portanto, a análise da TRI não possibilitará comparações.

COMENTÁRIO PEDAGÓGICO DO ITEM y4

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem uma mesma tipologia para o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **marginal**, *sujeito a reelaboração*, $0,20 \leq DISCR < 0,30$) (Brasil e Portugal).

A porcentagem de acertos (letra B) de modo geral não foi predominante no Brasil (29%) sendo equivalentes aos distratores C (31%) e D (27%). Em Portugal há uma predominância relativa da opção correta diante os distratores com 36%, na qual o distrator D (30%) demonstrou força atrativa. Isso demonstra evidência que de fato esse item deva ser reelaborado, pois não foi verificado um predomínio absoluto da alternativa correta em ambos os contextos.

Os grupos superiores (27% dos sujeitos com melhores escores, variando para cada contexto) e inferiores com relação a proporção de acertos dentro de seus grupos foram de 39% e 17% entre respondentes brasileiros e de 48% e 23% para os portugueses, ou seja, com uma discriminação mediana para ambos.

O bisserial da alternativa correta foi de 0,27 e 0,01. Como em Portugal foi praticamente zero, significa que o grupo que assinalou a alternativa correta é composta igualmente por alunos de baixo e bom desempenho. No Brasil é um perfil misto, com leve predominância de alunos com bom desempenho. Os distratores C e D devam ser revistos, pois foram alternativas assinaladas por alunos de maior desempenho em ambos os contextos, e de forma correspondente os seus bisseriais próximos de zero sinalizam uma composição mista para os grupos que foram atraídos e assinalaram esses distratores. É considerado um item confuso, em que as causas podem estar no enunciado da questão.

A AGI revela uma indeterminação para o intervalo de discriminação (Q) e com isso o poder de discriminação no contexto brasileiro, característica de um item de baixa qualidade. Em Portugal o número de violações é maior, porém ainda se consegue discriminar minimamente, indicado por baixo poder de discriminação com um ângulo com aproximadamente 20°. Para o Brasil, os distratores C e D foram os que tiveram maior quantidade de violações gráficas, com destaque para o D, com picos de maiores proporções. Para Portugal foi a alternativa A, porém, as oscilações foram mais sutis, mas chama atenção o fato de alunos que alto desempenho (23 acertos) terem assinalado.

Com base na estimação dos parâmetros da TRI através do modelo geral com 26 itens, o item y4 pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **ruim**, *sujeito a reelaboração* ($0,35 < a \leq 0,65$).

A CCI apresenta uma sigmoide com uma baixa inclinação, revelando pouca discriminação, ou seja, o item não consegue distinguir respondentes alta e baixa proficiência. Estando mais à direita, indica uma posição do item em uma alta habilidade θ , possuindo um percentual de acertos empíricos estimados em 56% entre aqueles que alcançam esse nível de habilidade θ . O índice de acerto ao acaso foi de 0,21, um pouco abaixo do valor esperado que seria de $\frac{1}{4}$ ou 0,25, isso representa uma probabilidade de 21% dos respondentes acertarem o item sem possuírem habilidades suficientes. O índice de *guessing* um pouco abaixo do esperado indica uma boa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os

respectivos níveis de proficiência alcançados. O índice de erro ao acaso foi de 0,91, isso representa 9% de probabilidade de os alunos cometerem equívocos, ainda que possuam uma boa habilidade e esse nível se eleve. Por fim, a CII indica uma baixa discriminação, com isso, o item y4 não apresenta informações relevantes para uma interpretação pedagógica, estando sujeito a rejeição.

DANDO SENTIDO AOS ERROS/DIFICULDADES: Sabe-se que a água deva ferver próximo a 100 °C, dependendo a altitude, na qual influencia a pressão atmosférica local. Quando o enunciado diz que a água na panela ferve rapidamente, supostamente deva fazer com os alunos pensem que a temperatura está muito elevada e assinalam a letra D. Por outro lado, podem haver aqueles que imaginam pôr a água ferver rapidamente é porque alcança a temperatura de ebulição mais rapidamente, para isso, uma das possibilidades seria sua temperatura de ebulição estar mais baixa que o normal, e com isso assinalaram a letra A. Outras ainda podem imaginar que somente a 100 °C a água ferva, e com isso marcaria a letra D, ou seja, nenhuma das alternativas é correta. A falta de informação sobre a altitude no local, de declarar se o evento ocorre em *Condições Normais de Temperatura e Pressão* (CNTP), ou se a rapidez é devido a fonte calorífica mais potente, sejam alguns possíveis agravantes.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/REELABORADOS AO ITEM: Existem quatro aspectos que podem ser considerados necessários para que se mantenha o traço latente de conhecimento cognitivo relacionado ao *subsunçor* e evite vieses interpretativos. O primeiro se refere em deixar claro o tipo de panela, tipo, “panela comum” ou “panela sem a tampa”, assim evita-se a possibilidade de um respondente imaginar uma panela de pressão. O segundo se refere aos termos “ferve rapidamente”, isso pode remeter a ideia errônea de um ponto de ebulição mais baixo. A sugestão seria a retirada do termo “rapidamente”. O terceiro seria informar dentro do contexto se o aquecimento ocorre em regiões próximas ao nível do mar ou em regiões montanhosas, ou ainda deixar explícito a informação sobre o fenômeno ocorrer em condições normais de temperatura e pressão (0 °C e pressão de 1 atm numa latitude de 45° ao nível do mar - CNTP). Um quarto aspecto a ser levado em consideração é o fato da água imaginada não ser uma substância pura e

homogênea, como geralmente se encontra, dessa forma se torna relevante afirmar que seja. Apresentar um valor sutilmente abaixo de 100 °C é ponto positivo por retratar melhor as situações reais e deve se manter, caso seja reelaborada para uma região pouco acima do nível do mar, sendo a situação mais comum.

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: O *subsunçor* envolvido está relacionado no conhecimento do respondente saber que **a temperatura mais provável da água fervente é igual ou próximo a 100 °C**. Este é um saber internalizado que se sustenta com a memorização de uma informação (conhecimento factual), em que sua aquisição na estrutura cognitiva possa ter ocorrido de forma arbitrária e literal, ou não, porém sua existência é o que basta para aferir ou evidenciar um acerto no item y4. Na sequência da aprendizagem, uma condição favorável para a ocorrência de uma *assimilação subsunciva* (ou por subsunção) estaria na associação/interação do subsunçor com um *conhecimento potencialmente significativo* mais inclusivo, que dentro do contexto, seria uma compreensão da variação do ponto de ebulição com a altitude.

Como *organizadores prévios*, podem ser abordados inicialmente diversas técnicas de ensino que explorem sobre o processo de fervura da água, entre outras substâncias. Recomenda-se aquelas de caráter investigativo e que visem o desenvolvimento de competências científicas. Atividades baseadas em projetos também são uma opção para *organizadores avançados* (também podem ser prévios, ou ainda atividades formativas regulatórias, dependendo da estratégia de ensino), possibilitando associar esse fenômeno com temas transversais, articular uma intradisciplinaridade na sequência dos conteúdos a serem abordados, e com isso favorecer o processo de *reconciliação integrativa* caso seja desenvolvido visando analisar similaridades e diferenças de opiniões iniciais e finais dentro de um ciclo de aprendizagem. O uso de outras técnicas também pode ser incorporado a esse processo, como o caso de mapas conceituais como atividade a explorar conceitos hierárquicos e favorecer a *retenção subsunciva*, e com isso enfrentar o processo de *assimilação obliteradora*, na qual se perde parte das informações supostamente incorporadas no

processo de aprendizagem, causando, entretanto, uma modificação ao subsunçor inicial.

Para o processo de *diferenciação progressiva*, parece promissor explorar as diferenças de como as partículas se comportam nos processos de evaporação, ebulição e calefação, assim como as distinções entre vapor e gás. Destacar as grandezas físicas envolvidas em cada caso são implicações importantes para uma melhor compreensão, e o Modelo Cinético-Molecular pode surgir como um aliado facilitador e estratégico para fazer a relação entre uma observação indireta (mesoscópica) desses fenômenos e uma forma de visualizar, de forma simplificada, como ocorrem esses processos. É importante destacar que na apresentação de modelos simplificados se tenha atenção a reducionismos extremos, apresentando limitações sobre o modelo explicativo.

A5: ITEM y5

5. Cinco minutos mais tarde, a água na panela ainda ferve. A temperatura mais provável da água agora é:

a) 88 °C

⇒ Acredita-se que a água aqueceu rapidamente por atingir o ponto de ebulição a uma temperatura mais baixa que o habitual, e deve permanecer na mesma temperatura ainda que continue sendo aquecido. (*): *Não houve correspondência direta com nenhuma concepção do inventário*).

b) 98 °C

c) 110 °C

[B4]: Quando a temperatura em ebulição permanece constante, algo está “errado”.

⇒ Por acreditar que um corpo ao ser aquecido nunca permanece com a mesma temperatura, nem durante a mudança de fase, no caso, a fervura.

[C1] *muito provável*: Aquecimento sempre resulta em um aumento na temperatura. (não compreende o conceito de calor latente).

⇒ Por acreditar que quando se ferve a temperatura continua a crescer e portanto deva estar acima de 100 °C

d) Nenhuma das respostas acima é correta.

[B4]: Quando a temperatura em ebulição permanece constante, algo está “errado”.

⇒ Por acreditar que um corpo ao ser aquecido nunca permanece com a mesma temperatura, nem durante a mudança de fase, no caso, a fervura.

[C1] *pouco provável*: Aquecimento sempre resulta em um aumento na temperatura. (não compreende o conceito de calor latente).

⇒ Por acreditar que quando se ferve a temperatura continua a crescer e, portanto, deva estar acima de 100 °C

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Recordar* sobre a Invariância da Temperatura em um Processo de Mudança de Fase - [Descritor 5].

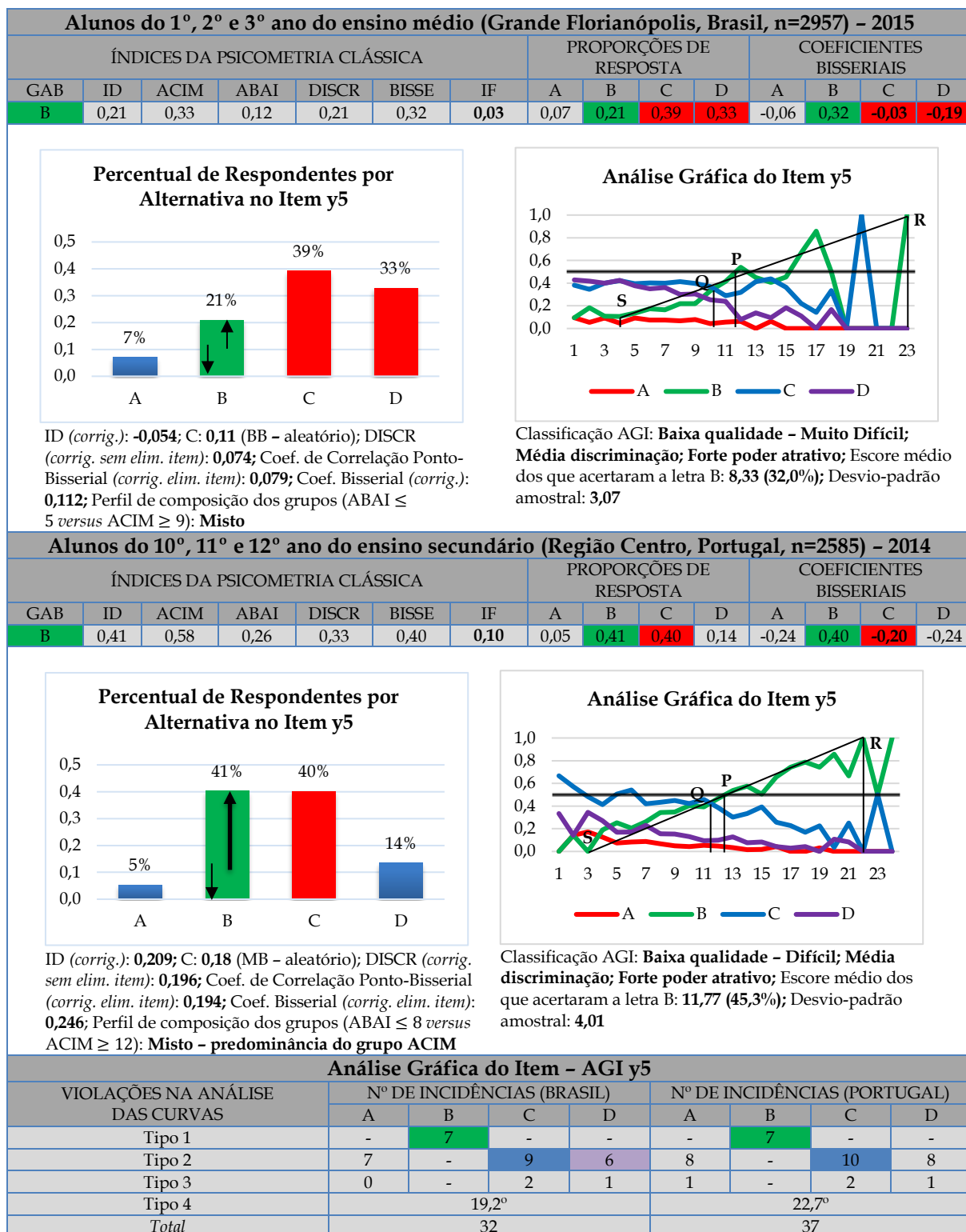
PROFICIÊNCIA*: *Prever* a «temperatura mais provável da água fervente» «durante todo o processo de vaporização» «permanece com a temperatura do ponto de ebulição local, sendo igual ou próximo a 100 °C em regiões próximas ao nível do mar».

INDÍCIOS DE DEPENDÊNCIA LOCAL: É uma sequência do contexto do item y4 e envolve uma intuição de valor de medida sobre a mesma grandeza indagada sobre o objeto em uma situação aparentemente próxima.

**Considera-se em negrito o subsunção (ou parte dele) a ser evidenciado pelos respondentes de modo a se associar as informações do contexto do problema entre os outras que serão ou foram abordadas (conhecimentos potencialmente significativos), para assim representar uma assimilação conceitual.*

PAINÉL DE ESTATÍSTICAS DO ITEM y5

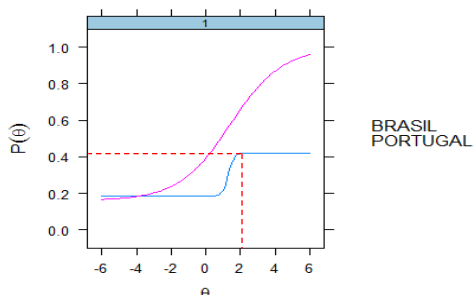
Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



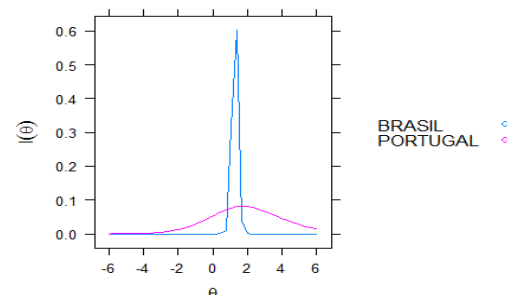
Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)

Modelo Não-Restringido Portugal e Brasil (n=5542) - 20 itens				
PAÍS	PARÂMETROS DA PSICOMETRIA MODERNA - MODELO 4PL			
	DISCRIMINAÇÃO (a)	DIFICULDADE (b)	ACERTO AO ACASO (g)	ERRO AO ACASO (u)
Brasil	7,18	1,29	0,19	0,42
Portugal	0,53	1,10	0,07	0,99

Curva Característica do Item (CCI) y5



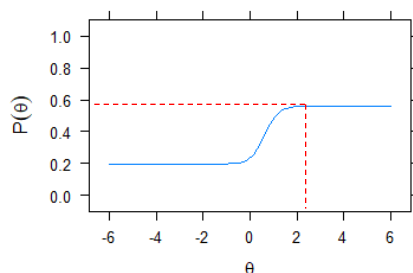
Curva de Informação do Item (CII) y5



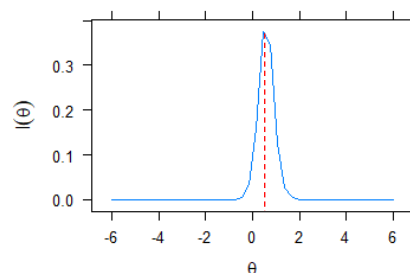
Modelo Geral Portugal e Brasil (n=5542) - 20 itens

NÍVEL DO ITEM NA ESCALA	PARÂMETROS DA PSICOMETRIA MODERNA - MODELO 4PL			
	DISCRIMINAÇÃO (a)	DIFICULDADE (b)	ACERTO AO ACASO (g)	ERRO AO ACASO (u)
De 0 a 1	3,45	0,61	0,20	0,55

Curva Característica do Item (CCI) y5



Curva de Informação do Item (CII) y5



NÍVEIS DA ESCALA DE PROFICIÊNCIA	-1,000	0,000	0,61	1,000	2,000	3,000
PROPORÇÕES DE ACERTOS EMPÍRICOS	0,201	0,238	0,375	0,478	0,547	0,550

COMENTÁRIO PEDAGÓGICO DO ITEM 5

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **Marginal**, sujeito a reelaboração) ($0,20 \leq DISCR < 0,30$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **Bom**, sujeito a aprimoramento ($0,30 \leq DISCR < 0,40$) (Portugal).

A porcentagem absoluta dos que acertaram ficou apenas em 21%, sendo os distratores C (39%) e D (33%) com proporções maiores, ou seja, a alternativa correta foi de modo geral a 3ª opção entre os respondentes brasileiros, e isso é um péssimo sinal de início. Em Portugal a letra B e C ficaram com percentual igual a 41% e 40%, ou

seja, é um item que não apresentou nem um predomínio absoluto percentual sobre a alternativa correta (>50%) e nem parcial sobre os distratores, e ainda revela que o distrator C possui grande poder atrativo, e deve ser averiguado.

Os grupos de desempenho superior e inferior que assinalaram a letra B ficou em 33% e 12% no Brasil, e 58% e 26% em Portugal, com uma maior discriminação entre os alunos portugueses.

Os bisseriais da B foram de 0,32 e 0,40, ou seja, há uma predominância de que os alunos que acertaram fazem parte do grupo de maior desempenho. O distrator C apresenta bisserial igual a -0,03 e -0,20 para Brasil e Portugal, ou seja, apresentam um perfil misto pela proximidade com o zero, com isso, tanto bons como alunos de desempenho baixo assinalaram essa alternativa. Chama atenção do aluno brasileiro com desempenho igual a 20 acertos ter assinalado essa alternativa, assim como do grupo de alunos portugueses com 23 acertos.

A AGI apresenta uma quantidade de violações equivalentes a alternativa correta, e entre os distratores o maior destaque é para a alternativa C, no 9 e 10 violações comparando Brasil e Portugal, apesar de que no geral, Portugal tenho um pouco mais de violações. O ângulo de inclinação demonstra um baixo poder de discriminação, sendo o ângulo de Portugal sutilmente superior, mas sem grande diferença.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **ruim**, *sujeito a reelaboração*, com *discriminação baixa* ($0,35 \leq a \leq 0,65$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade relativamente próxima entre os contextos, com uma superioridade entre os alunos brasileiros que ainda possuem uma maior discriminação. Foi estimado um índice de acerto ao acaso (*guessing*) de 0,19 e 0,07, sendo para os portugueses um índice abaixo do esperado, que seria de $\frac{1}{4}$ ou 0,25, isso indica uma ótima coerência entre responder uma alternativa

(fácil, mediana ou difícil) de acordo com os níveis de proficiência. O erro ao acaso estimado foram de 0,42 e 0,99, praticamente nulo entre os portugueses devido a um valor próximo de 1, mas para os brasileiros uma probabilidade de 58% para a ocorrência de equívocos por descuido entre alunos com proficiência acima de $\theta = 2$ (aproximadamente), correspondente a 42% de acertos empíricos. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto, com exceção dos alunos brasileiros com proficiência entre $-6,00 < \theta < -5,00$ (aproximadamente).

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros, e com isso uma boa informação do item, no caso estando acima da média e em um nível mais elevado de proficiência (entre 1 e 2, aproximadamente), e sendo representando por um intervalo em cerca de 1 desvio-padrão.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 0 a 1, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y3, y13, y23, y10, y22 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (0,61) e discriminação (3,45) classificando o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado** para uma reaplicação, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 37,5% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso foi de 0,20, abaixo do esperado que seria de $\frac{1}{4}$ ou 0,25, isso indica coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência dos sujeitos. O erro ao acaso sugere de modo geral que os alunos com probabilidade de certos acima de 0,55 possuam cerca de 45% de chances de se equivocarem por descuido, mesmo que suas proficiências se tornem mais elevadas a partir de $\theta = 1,9$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação, apesar do y5 não ser considerado um item-âncora.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O enunciado é uma continuação do item y4 e fala de a panela estar com a água fervente por cinco minutos e questiona a mais provável temperatura após esse tempo decorrido. Obviamente que a convicção da resposta anterior supostamente irá influenciar os alunos nessa questão, entretanto, aqueles que imaginam uma temperatura constante na mudança de fase tende a manter a mesma alternativa, porém aqueles que acreditam que o calor fornecido tende a fazer com que a temperatura da água continua a crescer tende de assinalar alternativas com valores superiores ao anterior, e fazem uma estimativa de 5 min de aquecimento. A forte evidencia da letra leva a entender que muitos possam ter noção que a água ferva a 100 °C, porém levar a crer que os 5 min de aquecimento sobre a água fervente façam com que ela fique ainda mais quente e atinja 110 °C, pois é mais razoável do que atingir 120 °C. De modo geral, a sequência das respostas dos alunos leva a acreditar em dificuldades na compreensão do conceito de equilíbrio térmico e que na mudança de fase a temperatura de substâncias homogêneas ela tende a se manter constante. É importante ressaltar que os eventuais problemas e críticas citados no item anterior tendem a refletirem os equívocos nesse.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: As sugestões são as mesmas dadas para o item y4.

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: O subsunçor se refere a compreender que a **temperatura da água fervente durante o processo de vaporização permanece igual ou próximo a 100 °C.**

A identificação preliminar de um subsunçor que deva preexistir cognitivamente de forma não-arbitrária e não-literal se associa na compreensão diante do processo de mudança no estado de agregação da matéria (tanto no aquecimento quanto no resfriamento), levando em consideração que supostamente a sua temperatura nessas condições não deva se alterar com tempo. Como conhecimentos potencialmente significativos estão as causas e efeitos a serem investigados durante esse processo, diferentes níveis de observação e análise do fenômeno, assim como das diferenças da ocorrência desse fenômeno com as diferentes substâncias, sendo puras ou misturas

homogêneas e heterogêneas, de preferência que se tenha como exemplo substâncias diferentes além da água, ou mesmo, misturando substâncias na própria água destilada. Explorar essas diversidades favorece o processo de *diferenciação progressiva* sugerida por Ausubel.

Entre outras possibilidades de abordagem está a mudança na fonte térmica, intensificando/diminuindo o fluxo de calor, assim como o tempo de observação do aquecimento. Dentro disso, atividades experimentais investigativas podem ser *mecanismo de ensino potencialmente significativos* quando buscam colocar o aluno em uma situação familiar a ele, e podem ainda favorecer a uma aprendizagem significativa quando forem priorizadas *tarefas logicamente significativas* envolvendo confronto cognitivo, estímulo a questionamentos e formulação de problemas, descrição do fenômeno observado, valorização de opiniões intuitivas individuais (previsões), incluindo discussões e atividades coletivas, sobretudo considera-se fundamental que haja uma correspondência no andamento da abordagem dos conteúdos organizados hierarquicamente (como p. ex. sugerido nos itens anteriores) e o desenvolvimento das atividades propostas.

A6: ITEM y6

6. Qual será a temperatura previsível do vapor formado, acima da água na panela?

a) 88 °C

⇒ Acredita-se que a água aqueceu rapidamente por atingir o ponto de ebulição a uma temperatura mais baixa que o habitual, permanecendo na mesma temperatura ao continuar a ser aquecido, ou seja, tanto a água fervente quanto o vapor d'água estão a 88 °C. (*): *Não houve correspondência direta com nenhuma concepção do inventário).*

b) 98 °C

c) 110 °C

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por acreditar que o vapor não irá trocar calor com o ar e se resfriar. Ref. [D8].

[D8] muito provável: O vapor d'água está sempre a uma temperatura superior a 100°C.

⇒ Por acreditar que o vapor está supostamente acima de 110 °C, pois quando muda-se de fase, a temperatura necessariamente deve ser superior a do ponto de ebulição.

d) Nenhuma das respostas acima é correta.

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por acreditar que o vapor não irá trocar calor com o ar e se resfriar. Ref. [D8].

[D8] pouco provável: O vapor d'água está sempre a uma temperatura superior a 100°C.

⇒ Por acreditar que o vapor está supostamente acima de 110 °C, pois quando muda-se de fase, a temperatura necessariamente deve ser superior a do ponto de ebulição.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: Recordar sobre a Invariância da Temperatura em um Processo de Mudança de Fase - [Descritor 5].

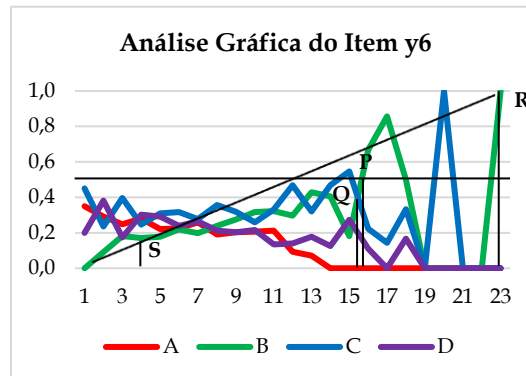
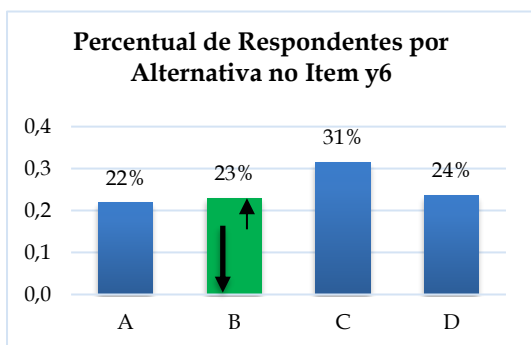
PROFICIÊNCIA*: Prever que a «temperatura mais provável do vapor d'água» «formado logo acima da superfície da água fervente» «é igual ou próximo a da água fervente durante todo o processo de vaporização, no caso, em cerca de 100 °C em regiões ao nível do mar, em outras acima desse nível (maioria das cidades) o ponto de ebulição fica sutilmente abaixo de 100 °C».

INDÍCIOS DE DEPENDÊNCIA LOCAL: Os itens y4 e y5 estão na mesma família de situações, no caso, a temperatura da água e do seu vapor em processo de ebulição. O raciocínio de que a temperatura se mantém constante na mudança de fase é análoga a descrita nos itens y2 e y3, na qual ocorrem o processo de fusão, com isso, pode-se considerar indícios de dependência local na compreensão entre esses itens.

PAINÉL DE ESTATÍSTICAS DO ITEM y6

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

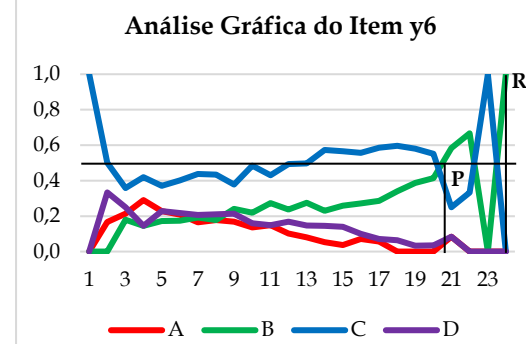
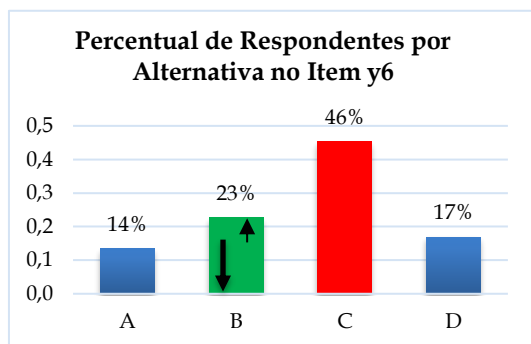
Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
B	0,23	0,31	0,16	0,15	0,22	-0,0002	0,22	0,23	0,31	0,24	-0,13	0,22	0,04	-0,11



ID (corr.): -0,026; C: 0,01 (BB - aleatório); DISCR (corr. sem elim. item): -0,001; Coef. de Correlação Ponto-Bisserial (corr. elim. item): -0,001; Coef. Bisserial (corr.): -0,001; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): Misto - tendência ao grupo ABAI

Classificação AGI: **Baixa qualidade - Muito difícil; Baixa discriminação; Médio poder atrativo;** Escore médio dos que acertaram a letra B: 7,89 (30,3%); Desvio-padrão amostral: 2,92

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
B	0,23	0,28	0,17	0,10	0,18	0,01	0,14	0,23	0,46	0,17	-0,27	0,18	0,13	-0,16

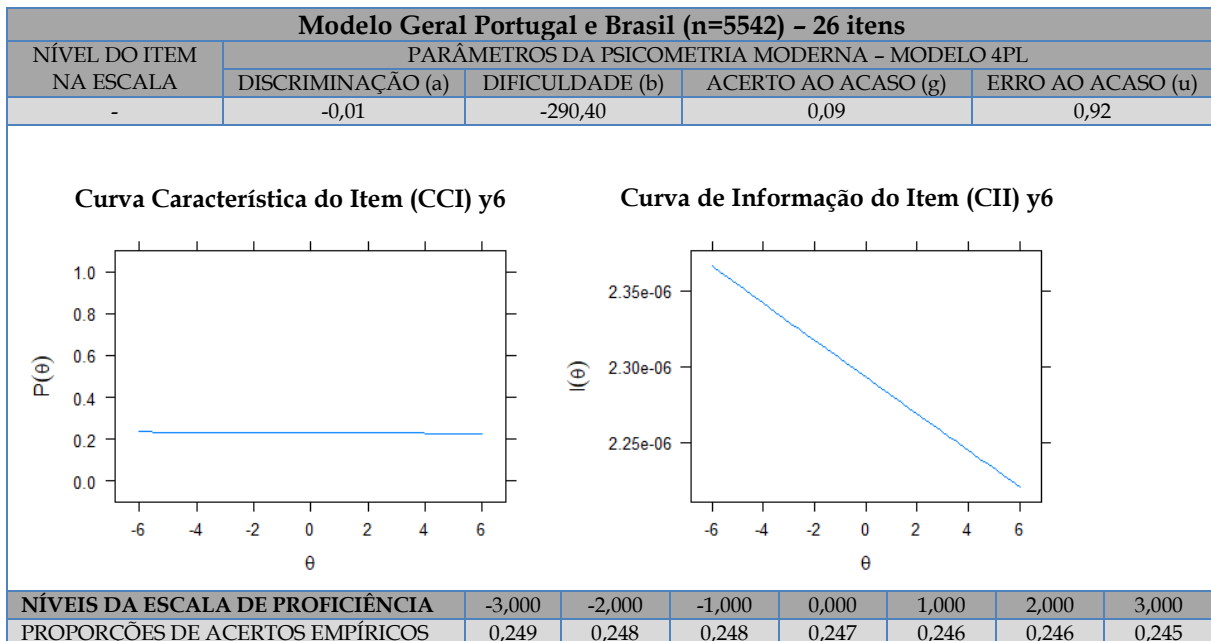


ID (corr.): -0,025; C: 0,12 (BB - aleatório) DISCR (corr. sem elim. item): 0,019; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,016; Coef. Bisserial (corr.): 0,023; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): Misto - tendência ao grupo ABAI

Classificação AGI: **Baixa qualidade - Muito difícil; Baixa discriminação; Forte poder atrativo;** Escore médio dos que acertaram a letra B: 11,19 (43,0%); Desvio-padrão amostral: 4,12

Análise Gráfica do Item - AGI y6								
VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	-	7	-	-	-	6	-	-
Tipo 2	6	-	10	8	7	-	11	7
Tipo 3	0	-	0	0	0	-	2	0
Tipo 4	22,5º				INDETERMINADO			
Total	31				33			

Coeficientes Estatísticas e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y6

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **inadequado**, *sujeito a rejeição* (DISCR < 0,20) (Brasil e Portugal).

A porcentagem absoluta do item mostra que não há uma predominância para nenhuma das populações. No Brasil, a maior porcentagem ficou no distrator C com 31%, sendo a letra B com 23% e D com 24%, considerados na zona de *guessing*. Em Portugal o distrator C também foi a alternativa mais atrativa com 46%, sendo a letra B com apenas 23%, coincidentemente com o Brasil, e isso sinaliza problemas graves no item.

O grupo superior e inferior ficou com 31% e 16% no Brasil e 28% e 17% em Portugal, demonstrando uma baixa discriminação, ou seja, não consegue diferenciar de forma satisfatória os alunos de mal e bom desempenho que responderam corretamente o item. Isso pode ser evidenciado com os valores dos bisseriais, que com 0,22 e 0,18, demonstrando um perfil de grupo misto para ambos os contextos, apesar da leve predominância de alunos com o desempenho superior.

O distrator C foi o mais atrativo para ambos os alunos e teve um coeficiente de bisserial discriminativo de 0,04 e 0,13, ou seja, tanto alunos com bom e mal desempenho assinalaram essa alternativa, tendo aluno brasileiro com 20 acertos e alunos portugueses com 23 acertos sendo atraídos, e isso não deve ser facilmente tolerado.

A AGI mostra uma quantidade de violações equiparáveis entre os grupos na alternativa correta, mas de proporções maiores para alunos brasileiros. A alternativa C sofre muitas oscilações, não tendendo linearmente a um decréscimo como se espera, isso demonstra incertezas entre grupos com diferentes níveis de desempenho no teste.

Com base na estimação dos parâmetros da TRI através do modelo geral com 26 itens, o item y6 pode não possuir uma classificação típica, estando a discriminação negativa e um número muito elevado de dificuldade, demonstrando ser totalmente inadequado e com uma péssima qualidade, sendo assim sem a possibilidade de uma interpretação qualitativa e útil para o campo pedagógico, sujeito a ser reelaborado ou rejeitado. Pode-se observar que o item não atende ao pressuposto da TRI relacionada a monotonicidade, ou seja, que a probabilidade de acertar deva crescer proporcionalmente ao aumento da habilidade dos respondentes, com isso, é um item que não possa ser interpretado através da TRI, e por isso deve ser excluída do processo de construção da escala de proficiência, bem como de outras interpretações. O mesmo serve para outros itens com características semelhantes.

DANDO SENTIDO AOS ERROS/DIFICULDADES: A temperatura do vapor formado acima da panela com água fervente é um processo de mudança de fase (líquido - gasoso) e espera-se que os alunos compreendam que a temperatura se mantém constante nesse processo, ou seja, sendo mais provável com 98 °C. Apesar disso, a atração pela letra C foi evidenciada, e isso leva a acreditar que os alunos tenham uma convicção de que no estado gasoso o vapor esteja numa temperatura superior à da água fervente. Essa concepção possivelmente acompanha equívocos da alternativa anterior e pode estar relacionada ainda a outros fatores, como por exemplo a sensação térmica de vapores em situações do cotidiano, ou ainda de que a medida

em que se aqueça e o estado físico se altere, a temperatura deva se elevar ao mudar de fase, tipo: a água após derretimento do gelo é mais quente, e o vapor após a água ferver é mais quente.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: Apesar de ser sugerido uma rejeição do item, isso deve ser feita na análise, porém na prática isso deve ser evitado buscando uma reelaboração. A situação-problema segue uma sequência lógica dos itens na qual envolve uma compreensão sobre a temperatura dos corpos na transição de fases. O item pode ser aprimorado explicitando que a água está fervente e que a temperatura do vapor formado se refere aqueles próximos a superfície da água fervente, pois o vapor tende a trocar calor com ar e resfriar imediatamente, e uma noção sobre isso pode influenciar negativamente o respondente.

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: Espera-se evidenciar um subsunçor relacionado a compreensão de que **a temperatura do vapor d'água acima da superfície da água fervente é igual ou próximo de 100 °C**. O fornecimento contínuo de energia térmica a água fervente aliado a ideia de sua mudança de fase pode agravar na concepção errônea de aquecimento do vapor, e assim ser obtida uma temperatura acima de 100 °C. Um entendimento sobre a permanência da temperatura em um mesmo valor na mudança de fase segue o mesmo raciocínio do processo de fusão na transição do gelo em água líquida nos itens y2 e y3, com isso, os conhecimentos que envolvem esse tipo de compreensão podem ser considerados uma zona mais direta e imediata entre aqueles supostos potencialmente significativos dentro do *contexto disciplinar*. Falar em significância para o sujeito remete a focar a um contexto mais familiar a ele, dessa forma, aplicações diversificadas e que envolva o seu cotidiano ou a seu campo de interesse profissional são consideradas promissoras.

Para a organização hierárquica dos conteúdos acredita-se que seja estruturador nesse âmbito uma compreensão sobre o conceito de calor latente, pois um entendimento de que a energia térmica recebida por um corpo não produz como efeito

única e exclusivamente a aumento de sua temperatura, seja *potencialmente significativo* para admitir uma mesma temperatura na mudança de fase. Dentro disso, explorar que a energia térmica recebida resulta em grande parte no afastamento das moléculas, mudanças na relação da força de coesão e repulsão entre elas, entre outras modificações na estrutura molecular, fazendo com que a temperatura se mantenha constante é algo que possa ser enriquecedor e favorável para uma assimilação subsunciva significativa. Uma abordagem dentro desse enfoque e sequenciação remete a importância recorrer ao Modelo Cinético-Molecular (MCM) e as propriedades gerais da matéria, assim como das propriedades quânticas. O MCM é considerado um dos pilares para o campo da Termodinâmica, ou seja, é considerado um conhecimento mais estruturado e inclusivo, assim, na sequenciação e organização da abordagem dos conteúdos deve ser um ponto de partida, conjuntamente com diagnósticos sobre o conhecimento prévio dos alunos, e seguindo para a incorporação e associação de um conhecimento cada vez mais estruturado a medida em que são exploradas a compreensão das diferentes faces da situação-problema. A diversificação e as diferentes aplicações fazem parte do processo de *diferenciação progressiva*, e deve sempre que possível transitar entre conceitos mais simples e complexos no andamento da aprendizagem para que ocorra uma *reconciliação integrativa*, como recomenda Ausubel.

A7: ITEM y7

7. Susana pega dois copos com água a 40 °C e mistura-os à água de um copo a 10 °C. Qual é a temperatura mais provável da mistura?

a) 20 °C

⇒ Nas trocas de calor o corpo frio é mais difícil de ser aquecido. (a temperatura de equilíbrio tende de estar mais próxima ao corpo mais frio). (*): *Não houve correspondência direta com nenhuma concepção do inventário).*

b) 25 °C

[A5]: O calor é proporcional à temperatura.

⇒ Por considerar que a igualdade do calor cedido e recebido sejam iguais, a temperatura final será a média aritmética das temperaturas, independente das massas.

c) 30 °C

⇒ Poderá haver um equívoco de que o equilíbrio seja a subtração de 40 °C com 10 °C.

d) 50 °C

[A6]: O calor não é mensurável, conceito quantificável.

⇒ Por não somar as temperaturas e não considerar a quantidade de calor no processo

[C5]: A temperatura pode ser transferida (concepção substancialista, a temperatura é uma substância passível de se deslocar de um corpo para outro).

⇒ Por considerar que seja a soma aritmética.

Não se percebe que:

[B6]: Um corpo frio não possui a capacidade de aquecer um outro corpo (somente corpos quentes “possuem calor” e podem aquecer, ou seja, se admite uma concepção ontológica substancialista).

⇒ Por ter resfriado o corpo de 40 °C nas alternativas “a”, “b” e “c”, e na letra d ele supostamente “aquece” fazendo com que a temperatura saia de 40 °C e vá para 50 °C.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Recordar* sobre a Invariância da Temperatura em um Processo de Mudança de Fase – [Descritor 5].

PROFICIÊNCIA*: *Prever* que a «**temperatura de equilíbrio** mais provável **na mistura de corpos idênticos**», «porém com massas diferentes, a temperatura final **tende a ser mais próxima da temperatura inicial do corpo com maior massa**», no caso, fica mais baixa e próxima ao corpo de massa 2m (mistura dos dois copos com água de massa m) que possui 40 °C, portanto, 30 °C.

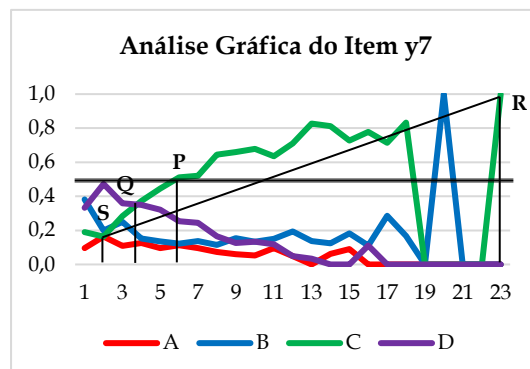
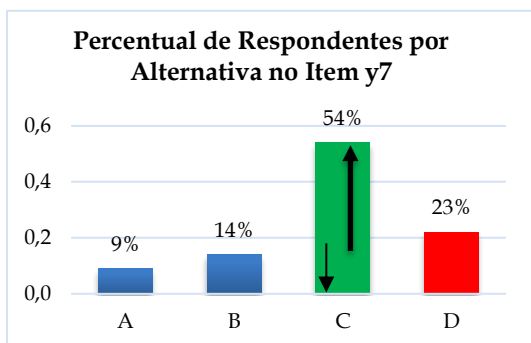
INDÍCIOS DE DEPENDÊNCIA LOCAL: Não foi identificado entre os itens anteriores.

*Considera-se em negrito o subunçor (ou parte dele) a ser evidenciado pelos respondentes de modo a se associar as informações do contexto do problema entre os outras que serão ou foram abordadas (conhecimentos potencialmente significativos), para assim representar uma assimilação conceitual.

PAINÉL DE ESTATÍSTICAS DO ITEM y7

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

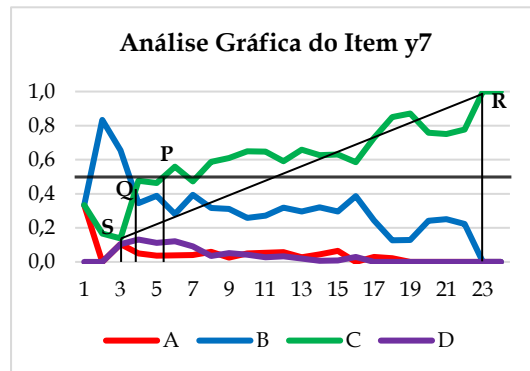
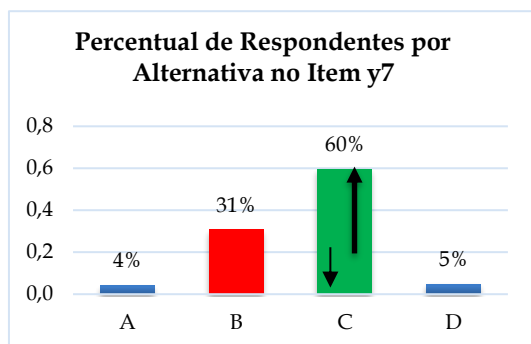
Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
C	0,54	0,68	0,37	0,31	0,31	0,03	0,09	0,14	0,54	0,23	-0,13	-0,03	0,31	-0,30



ID (corr.): 0,386; C: 0,22 (MM - 1 mod. corr. e 1 mod. incorr.); DISCR (corr. sem elim. item): 0,057; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,062; Coef. Bisserial (corr.): 0,078; Perfil de composição dos grupos (ACIM \geq 9 versus ABAI \leq 5): Misto - predominância do grupo ACIM

Classificação AGI: Boa qualidade - Dificuldade normal; Baixa discriminação; Forte poder atrativo; Escore médio dos que acertaram a letra C: 7,74 (29,8%); Desvio-padrão amostral: 2,65

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
C	0,60	0,66	0,51	0,15	0,20	0,02	0,04	0,31	0,60	0,05	-0,08	-0,10	0,20	-0,31



ID (corr.): 0,466; C: 0,36 (MM - 1 mod. corr. e 1 mod. incorr.); DISCR (corr. sem elim. item): 0,031; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,037; Coef. Bisserial (corr.): 0,047; Perfil de composição dos grupos (ACIM \geq 12 versus ABAI \leq 8): Misto - predominância do grupo ACIM

Classificação AGI: Baixa qualidade - Fácil; Baixa discriminação; Forte poder atrativo; Escore médio dos que acertaram a letra C: 10,78 (41,5%); Desvio-padrão amostral: 3,90

Análise Gráfica do Item - AGI y7

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	-	-	6	-	-	-	10	-
Tipo 2	6	8	-	3	10	10	-	7
Tipo 3	0	0	-	0	0	0	-	0
Tipo 4	19,5º				21,8º			
Total	23				37			

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)

Modelo Geral Portugal e Brasil (n=5542) – 26 itens					
NÍVEL DO ITEM NA ESCALA	PARÂMETROS DA PSICOMETRIA MODERNA – MODELO 4PL				
	DISCRIMINAÇÃO (a)	DIFICULDADE (b)	ACERTO AO ACASO (g)	ERRO AO ACASO (u)	
De -4 a -3	0,27	-3,65	0,06	0,76	

Curva Característica do Item (CCI) y7

Curva de Informação do Item (CII) y7

NÍVEIS DA ESCALA DE PROFICIÊNCIA	-4,000	-3,65	-3,000	-2,000	-1,000	0,000
PROPORÇÕES DE ACERTOS EMPÍRICOS	0,395	0,412	0,442	0,487	0,530	0,570

COMENTÁRIO PEDAGÓGICO DO ITEM y7

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Normal/mediano** (proporção esperada de respondentes 40%; $0,45 \leq ID < 0,55$) e **bom, sujeito a aprimoramento** ($0,30 \leq DISCR < 0,40$) (Brasil); **Fácil** (proporção esperada de respondentes 20%; $0,55 \leq ID < 0,75$) e **inadequado, sujeito a rejeição** ($DISCR < 0,20$) (Portugal).

As porcentagens totais foram predominantes na alternativa correta C nos dois contextos, com 54% (Brasil) e 60% (Portugal), sendo o distrator mais atrativo aos brasileiros a letra D com 23% (zona de *guessing*) e para os portugueses a letra B com 31%. O distrator com maior destaque nos dois contextos é o B, pois chega a ser uma opção assinalada por alunos brasileiros e portugueses com alto desempenho.

Os grupos superior e inferior (dos 27% da amostra com melhor e pior desempenho) correspondem a 68% e 37% no Brasil e 66% e 51% em Portugal, entre aqueles que responderam corretamente dentro de seus grupos, isso mostra uma discriminação muito baixa entre alunos portugueses e uma discriminação aceitável para os brasileiros. Os bisseriais de 0,31 e 0,20 para a letra C refletem isso, e sugere um perfil de composição misto com predomínio de alunos com melhor desempenho.

O distrator B possui bisseriais de -0,03 e -0,10. A proximidade desses valores com o zero sugere uma composição de grupo com alto e baixo desempenho admitindo essa alternativa, incluindo aluno brasileiro com 20 acertos e portugueses com 20, 21 e 22 acertos, e isso é um problema.

A AGI revela uma quantidade maior de violações para Portugal, tanto para a alternativa correta quanto para os distratores, e isso reflete as incertezas do sujeito e instabilidade do item, que são mais intensas que para os brasileiros. Apesar disso a letra B é a de maior oscilação para os brasileiros, e para os portugueses as letras A e B possuem a mesma quantidade de violações, porém o gráfico mostra que B possui oscilações mais intensas proporcionalmente. O poder de discriminação associado a acividade são baixas, com ângulos em torno de 20°.

Com base na estimação dos parâmetros da TRI através do modelo geral com 26 itens, o item y7 pode ser classificado como: **Muito fácil** (proporção esperada de respondentes 10%; $b \leq -1,28$) e **muito ruim**, *sujeito a reelaboração* ($0,0 < a \leq 0,35$).

A CCI não apresenta uma sigmoide tendo uma baixa inclinação, revelando pouca discriminação, ou seja, o item não consegue distinguir respondentes alta e baixa proficiência. Estando mais à esquerda, indica uma posição do item em uma baixa habilidade θ , possuindo um percentual de acertos empíricos estimados em 41,2% entre aqueles que alcançam esse nível de habilidade θ . O índice de acerto ao acaso foi de 0,06, um pouco abaixo do valor esperado que seria de $\frac{1}{4}$ ou 0,25, isso representa uma probabilidade de 6% dos respondentes acertarem o item sem possuírem habilidades suficientes. O índice de *guessing* um pouco abaixo do esperado indica uma boa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os respectivos níveis de proficiência alcançados. O índice de erro ao acaso foi de 0,76, isso representa 24% de probabilidade de os alunos cometerem equívocos, ainda que possuam uma boa habilidade e esse nível se eleve. Por fim, a CII indica uma baixa discriminação, com isso, o item y7 não apresenta informações relevantes para uma interpretação pedagógica, estando sujeito a rejeição.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item busca identificar uma temperatura de equilíbrio após misturar dois copos a 40 °C com um copo a 10 °C. A letra C é a correta, mais pode ter sido assinalada apenas pelo raciocínio de que o equilíbrio se alcança subtraindo 10 °C por 40 °C, e isso seria um equívoco que iria passar despercebido. Assinando a alternativa B, se opta por um equilíbrio a 25 °C, ou seja, um volume de 2V a 40 °C com 1V a 10 °C, e pode levar a crer que sendo 1/3 do volume, a temperatura diminuiria 1/3, ou seja, 16,7 °C, e o valor mais próximo seria 25 °C. Uma forma mais simples e equivocada de chegar nesse valor seria admitir que na mistura as temperaturas são somadas (40 °C + 10 °C) e o “equilíbrio” estaria associado à “média” das temperaturas, ou seja, a 25 °C. Essa talvez seja o equívoco mais provável, mas é recomendável garantir a informação de que os copos possuam o mesmo volume.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: O item pode ser aprimorado informando que os copos são idênticos em seus formatos, feitos de um mesmo material, e que possuem a mesma quantidade de água. Deixar uma ressalva para ignorar as trocas de calor com o meio externo pode ser fator positivo para evitar vieses interpretativos.

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: O item remete a verificar um *subsunçor* relacionado a compreensão de que a **temperatura de equilíbrio na mistura de corpos idênticos tende a ser mais próxima da temperatura inicial do corpo com maior massa**. O *conhecimento potencialmente significativo* pode ser construído com base mais racional e abstrato por meio de deduções matemáticas de forma numérica ou literal, mas também pode seguir um caminho que explore um enfoque empírico-concreto através de um ensino experimental. Sugere-se partir de uma abordagem mais concreta seguindo para um contexto mais abstrato, sempre buscando estimular o questionamento e argumentação dos alunos, sobretudo, realizando diagnósticos sobre suas concepções e as utilizando na busca de um enfoque que seja mais significativo aos alunos.

Compreender o *Princípio do Processo de Trocas* e *Lei Zero da Termodinâmica* são conhecimentos mais estruturados propedêuticos que devem servir de ponto de partida na organização hierárquica diante da situação-problema apresentada, e obviamente que também sirva como ponto de partida que convirja para desmistificar a sua família de situações-problema. Acredita-se que uma compreensão sobre a influência do conceito de *capacidade térmica* dentro do *processo de trocas de calor* entre corpos de um mesmo material irá remeter a ideia de que uma maior massa tende a ceder/absorver maior quantidade de energia térmica, pois o *calor específico sensível* é o mesmo. Assim, nessas condições esse conceito exerce uma função explicativa em potencial para que se tenha uma noção de qual corpo possui uma maior capacidade de “atrair” a temperatura final de equilíbrio térmico em um processo de trocas de calor, ou seja, considera-se que seja um *conhecimento potencialmente significativo* a ser ancorado pelo subsunçor. A interpretação matemática do conceito é um nível mais aprofundado que deve ser alcançado, ele ajuda dar um entendimento de proporcionalidade entre as medidas das grandezas, e de identificar a variável explicativa, no caso, sendo corpos de materiais iguais significa que possuem o mesmo calor específico, com isso a quantidade de massa é o fator que mais influencia. A interpretação da unidade de medida é fundamental para dar o sentido físico, assim, quanto maior a capacidade térmica, maior a quantidade de calor para variar de temperatura, ou seja, quem possui a maior capacidade térmica possui a tendência de ficar mais próximo da temperatura de equilíbrio. Durante as trocas de calor, o corpo que

$$C = m \cdot c = \frac{Q}{\Delta\theta} = \frac{\langle 1 \text{ cal} \rangle}{\langle 1 \text{ }^\circ\text{C} \rangle}$$

A *diferenciação progressiva* sugere explorar múltiplas possibilidades educativas, seja para alcançar um mesmo entendimento dentro de um campo de conhecimento específico, como para ampliar o campo de aplicação desse conhecimento em outros contextos, situações e problemas. Como sugestão para uma aprendizagem significativa diante da situação-problema apresentada, os alunos podem ser submetidos a realizarem *tarefas logicamente significativas* que investiguem as relações entre o volume ou quantidade de “água” (podendo ser outra substância) no processo de mistura e obtenção da temperatura de equilíbrio térmico, sendo um caminho para

a construção de argumentos e evidências que ajude os alunos perceberem os fatores que influenciam o alcance desse estado térmico. Entre os recursos, se deve recorrer ao uso do termômetro para aferir as temperaturas e de recipientes com mesmas dimensões para controlar a variável massa (ou utilizar uma balança para garantir a mesma quantidade de massa). Utilizar diferentes medidas de volume de água e temperaturas explorando os dados obtidos no estado térmico inicial e final por meio de gráficos podem contribuir na compreensão da relação existente entre elas em um campo mais abstrato e racional, na qual se encontra a matematização. A verificação dessas relações matemáticas, não deve, entretanto, deixar de focar na compreensão física e conceitual das grandezas envolvidas. Em uma dedução, pode-se chegar a expressão:

$$\theta_{eq} = \frac{M \cdot \theta_1 + m \cdot \theta_2}{M + m}$$

Pode-se interpretar que a temperatura de equilíbrio θ_{eq} seja uma *média ponderada* tendendo seu valor ao corpo de maior massa.

A8: ITEM y8

8. Pedro acredita que deva usar água fervente para fazer uma xícara de chá. Ele diz aos seus amigos: “Eu não poderia fazer chá se estivesse acampado numa montanha alta, porque a água não ferve a grandes altitudes”.

a) João diz: “Sim, a água ferve, mas não a uma temperatura tão elevada como aqui”.

b) Tânia diz: “Não é verdade. A água ferve sempre à mesma temperatura”.

[D5]: O ponto de ebulição da água é de 100°C (apenas).

⇒ Por acreditar que a temperatura de fervura não se altera jamais, mesmo que o corpo continue sendo aquecido, ou haja qualquer tipo de outra influência externa.

c) Luís diz: “O ponto de ebulição da água diminui, mas a água ainda ferve a 100 °C”.

[D5]: O ponto de ebulição da água é de 100°C (apenas).

⇒ Por acreditar que o ponto de fusão e fervura são coisas distintas, porém não perde a convicção de que a sempre irá ferver a 100 °C.

d) Sandra diz: “Eu concordo com Pedro. A água nunca atingirá o seu ponto de ebulição”.

⇒ Acredita-se que a água em elevadas altitudes seja mais difícil de ferver, ou seja, o ponto de ebulição se torna demasiadamente acentuado e pouco provável de que se alcance esse ponto. Apesar disso, é o contrário do que acontece, pois, o ponto de ebulição tende a diminuir. (*): *Não houve correspondência direta com nenhuma concepção do inventário).*

COMPETÊNCIAS CIENTÍFICAS ASSOCIADAS A:

- *Conhecer* a Temperatura de um Corpo ou de um Ambiente em Situações Típicas – [Descritor 3];
- *Aplicação* de Processos de Fervura de acordo com a Variação da Pressão – [Descritor 6]

PROFICIÊNCIA** – *Aplicar* um processo de aquecimento considerando que o ponto de ebulição da água varia conforme a variação da pressão, com isso, poder *inferir* que a «**temperatura da água fervente**» «varia para mais ou menos do valor de 100 °C» «**dependendo da altitude em relação ao nível do mar** pelo fato de variar a pressão atmosférica local. No caso de estar numa montanha alta, o ponto de ebulição fica mais baixo, com isso, torna-se mais fácil alcançar o processo de fervura, pois se atinge mais rapidamente o ponto de ebulição».

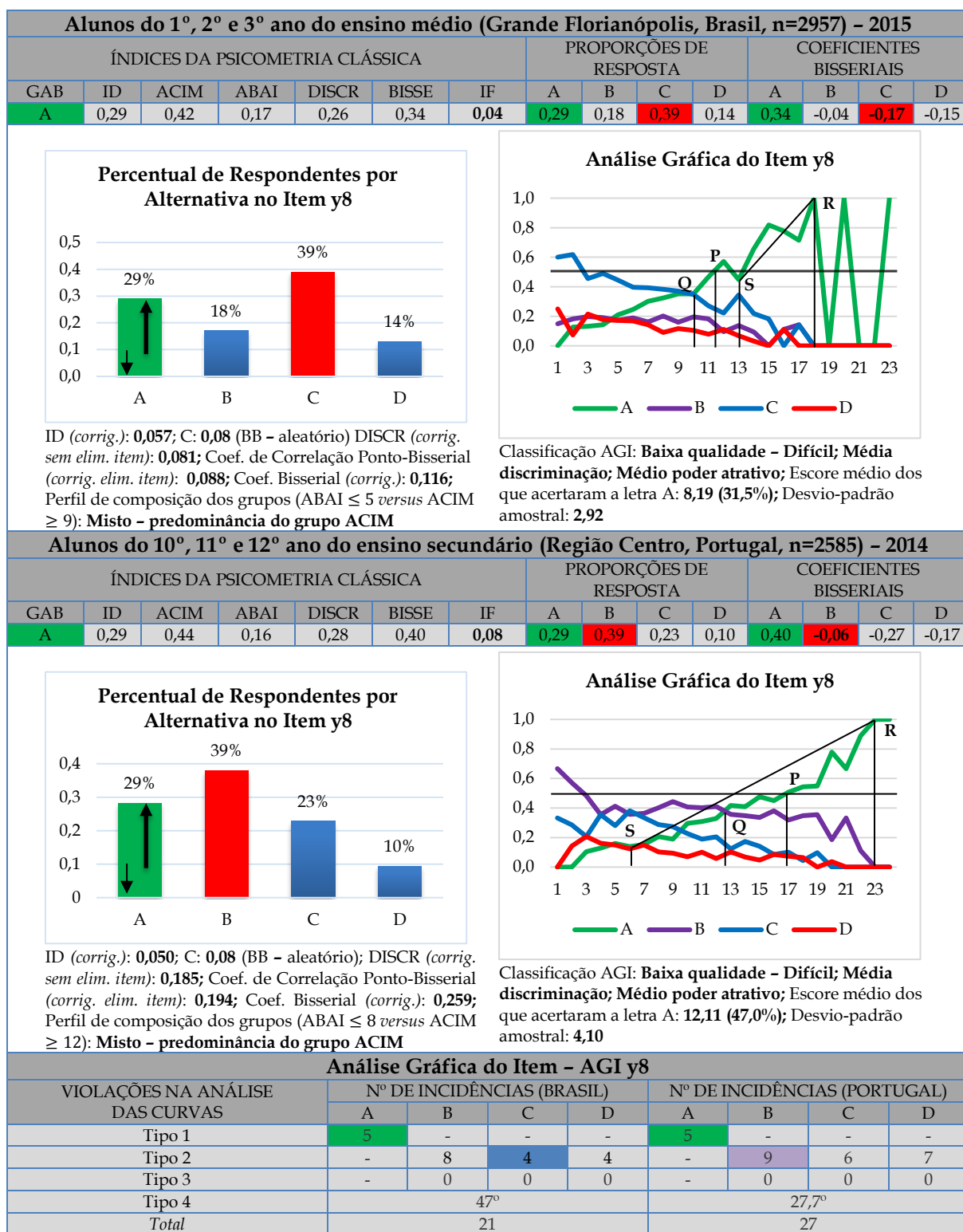
INDÍCIOS DE DEPENDÊNCIA LOCAL: O item y4 possui indícios de correspondência, apesar não seguir um mesmo contexto para a situação-problema. O enunciado busca aferir uma compreensão sobre a concepção da temperatura de ebulição, que por influência de um entendimento intuitivo sobre a variância do ponto de ebulição com relação altitude (região do contexto da situação) na qual ocorre o fenômeno, este poderá refletir na tomada de decisão.

*Esta proficiência não garante uma apreensão sobre a relação da temperatura de ebulição com a variação da pressão atmosférica local.

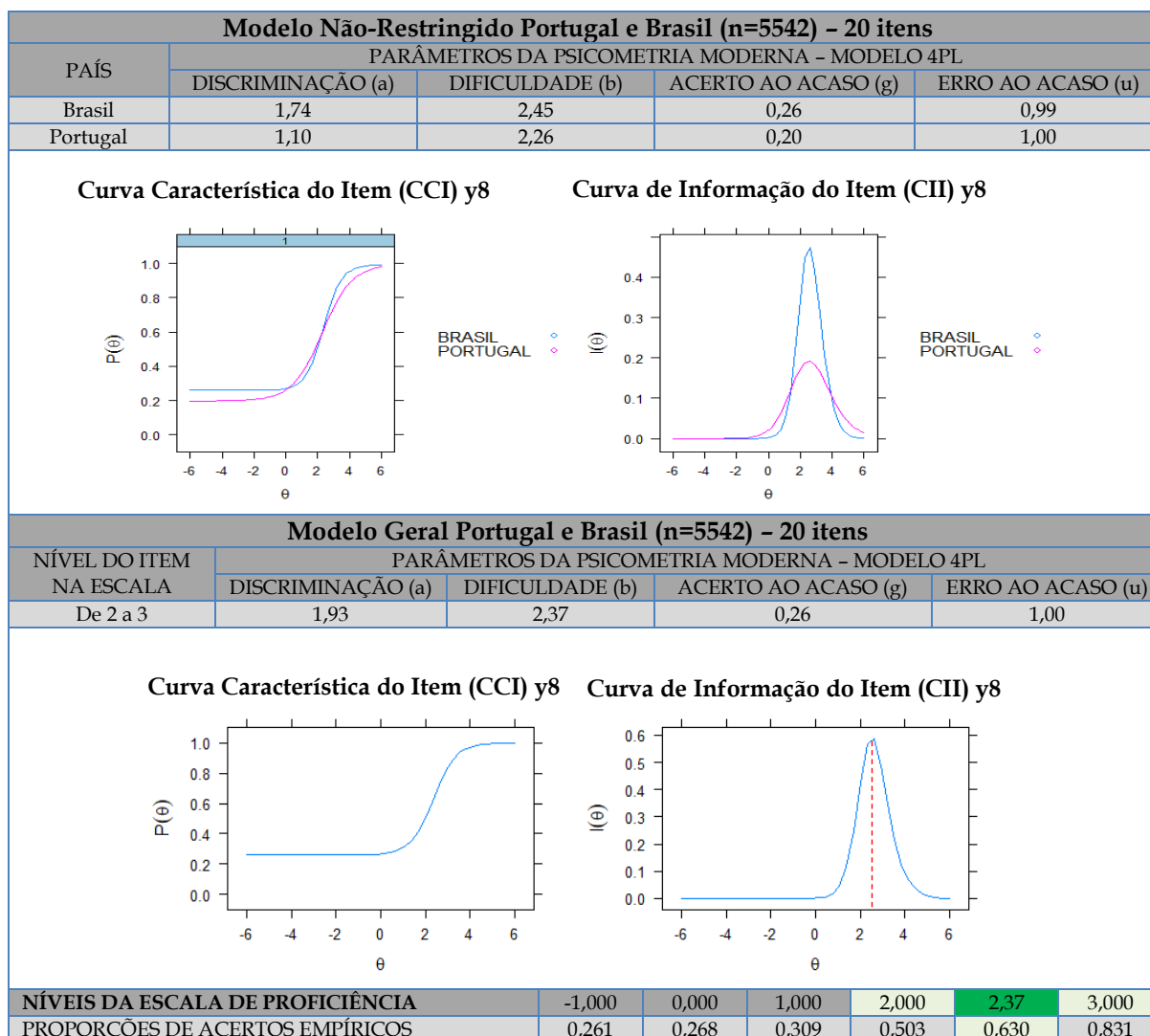
**Considera-se em negrito o subsunçor (ou parte dele) a ser evidenciado pelos respondentes de modo a se associar as informações do contexto do problema entre os outras que serão ou foram abordadas (conhecimentos potencialmente significativos), para assim representar uma assimilação conceitual.

PAINÉL DE ESTATÍSTICAS DO ITEM y8

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y8

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem uma mesma tipologia para o item em ambos os contextos sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **Marginal**, sujeito a reelaboração ($0,20 \leq DISCR < 0,30$) (Brasil e Portugal).

O percentual de acerto não foi predominante para nenhuma das populações, sendo a segunda opção para ambos. A letra A foi de 29% no Brasil “perdendo” para a letra C com 39%. Em Portugal com 29% “perdendo” para a letra B com 39%, apesar da letra C também revelar ser atrativa, com 23%, porém encontra-se na *zona de guessing*. A falta de predomínio percentual revela possíveis problemas no item.

O grupo de maior desempenho representa 42% e o de menor 17%, quanto aos acertos registrados. Em Portugal os valores percentuais que compõem esse grupo são próximos e iguais a 44% e 16%, isso significa que os índices de discriminação são equivalentes. Os bisseriais foram 0,34 e 0,40, e acompanham a mesma lógica discriminativa, como se espera. O perfil de composição sugerido em ambos os contextos é misto, mas com predominância relativa de alunos com bom desempenho, mas nem tantos assim.

O distrator C possui bisserial de -0,17 e -0,27, ou seja, no Brasil é misto e praticamente formado proporcionalmente pela mesma quantidade de alunos de baixo e alto desempenho (semelhante ao que ocorre com a letra B em Portugal tendo um bisserial igual a -0,06) devido a maior proximidade com o zero, enquanto em Portugal a composição é semelhante, porém, sutilmente composto por uma quantidade superior de alunos de mal desempenho. No Brasil alunos com 17 acertos assinalam a letra C, e entre os alunos portugueses com 21 e 22 acertos assinalam a letra B, isso mostra a força atrativa desses distratores, que precisam ser verificados.

A AGI mostra uma quantidade de violações para Portugal. Também é maior de oscilações no distrator B para ambas as populações entre os outros distratores, isso significa muitas incertezas entre os grupos de diversos níveis de desempenho. O poder discriminatório dado pelo ângulo de atividade de 47° para as proporções de respostas dos alunos brasileiros demonstra ser maior que para o contexto português.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **bom**, *sujeito a aprimoramento*, com *discriminação moderada* ($0,65 < a \leq 1,35$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade e discriminação relativamente próxima entre os contextos, com uma pequena superioridade nesses dois índices para os alunos brasileiros. O valor estimado para o acerto ao acaso

(*guessing*) foi de 0,26 e 0,20, muito próximo para os valores esperados, que seria de $\frac{1}{4}$ ou 0,25, uma probabilidade inferior a isso reforça uma maior coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência dos sujeitos. Não foram estimados erros ao acaso para os contextos. Comparando as CCI de cada contexto, percebe-se uma aproximação entre elas estando-as quase sobrepostas, isso sugere que os alunos com mesma proficiência θ em cada contexto possuem praticamente uma mesma probabilidade de acerto.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros, porém os portugueses demonstram contribuir na informação do item dentro de um mesmo intervalo e nível de proficiência (entre 2 e 3, aproximadamente), conforme sugere os parâmetros de dificuldade. A informação se concentra entorno de 2,5 desvio-padrão ao redor do posicionamento do item.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 2 a 3, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 3 itens-âncoras y_8 , y_{19} e y_{15} (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (2,37) e discriminação (1,93) classificando o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado** para uma reaplicação, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 63% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso foi estimado em 0,26, dentro do valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 26% de probabilidade de acertar sem possuir uma proficiência adequada. Não foi estimado erro ao acaso por descuido, ou seja, cometer equívocos mesmo tendo boa proficiência. Por fim, a elevação da CII indica uma boa discriminação e uma boa informação, sendo considerado um item quase-âncora por não atender ao critério de que o nível de anterior de desvio-padrão tem de ser menor que 0,50, que no caso foi de 0,503. Rabelo (2013, p. 146) faz uma ressalva a esse critério, considerado muito rígido e sugere a

adoção de que o anterior seja menor que 0,65, com isso se justifica o y8 pertencer a escala de proficiência.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O conhecimento que envolve o item y8 é de certa forma correspondente ao item y4, pois remete a ideia de diferentes temperaturas de fervura. Nesse caso, se discute a fervura de um chá numa montanha, e se explora a capacidade dos respondentes associarem com a altitude, ou seja, quanto maior a altitude, menor será a pressão atmosférica local e o ponto de ebulição tende a diminuir, e com isso, se torna mais fácil a sua fervura. O distrator C afirma que: *“O ponto de fusão da água diminui, mas a água ainda ferve a 100 °C”*, e o B diz: *“Não é verdade. A água ferve sempre à mesma temperatura”*, e isso reforça (de uma maneira ou outra) a ideia de que os alunos tendem a conceber o ponto de fervura da água inalterado, e seja independente das condições locais de altitude e pressão.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: Não se considera a necessidade de reelaboração do item ou entre os distratores, pois as informações estão claras e minimamente necessárias para uma boa interpretação e tomada de decisão.

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: O subsunçor a ser evidenciado se refere a compreensão de que **a temperatura da água fervente varia dependendo da altitude em relação ao nível do mar**. Os distratores sugerem que seja reforçado uma abordagem das relações entre temperatura e pressão, em especial, durante os processos de mudança de fase. Também considera necessário associar as relações da altitude com a variação da pressão atmosférica, assim como de como esses fatores alteram o ponto de ebulição dos líquidos. As considerações e sugestões apresentadas ao item y4 são em grande parte adequadas para esta situação-problema.

A9: ITEM y9

9. Samuel pega numa lata de refrigerante cola e uma garrafa plástica de refrigerante cola da geladeira, onde tinham sido colocados na noite anterior. Ele rapidamente coloca um termômetro no refrigerante contido na lata. A temperatura é de 7 °C. Quais são as temperaturas mais prováveis da garrafa de plástico e do refrigerante cola de seu interior?

a) Estão ambas abaixo de 7 °C.

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido)

⇒ Por se acreditar que a garrafa de plástico seja um bom condutor térmico, a temperatura tende a ser mais baixa que a da lata.

b) Estão ambas a 7 °C.

c) Estão ambas acima de 7 °C.

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por se acreditar que a garrafa de plástico seja um bom isolante térmico, a temperatura tende a ser mais elevada que a da lata, não conseguiu esfriar suficientemente desde a noite passada

d) O refrigerante cola está a 7°C, mas a garrafa está a mais de 7 °C.

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por se acreditar que o refrigerante e a garrafa plástica não trocam calor entre si, e com isso, não tendem a alcançar um equilíbrio térmico entre si.

[D1]: A temperatura é uma propriedade de um material ou objeto particular. (uma característica específica para cada tipo de material).

⇒ Por se acreditar que o refrigerante e a garrafa de plástico serem constituídos de materiais diferentes, cada um por si só tende a pontos de equilíbrio térmico.

e) Depende da quantidade de refrigerante cola e/ou do tamanho da garrafa.

[B7]: A temperatura de um objeto depende de seu tamanho.

⇒ Por acreditar que quanto maior o corpo, mais baixa será a sua temperatura, pois é mais difícil resfriar e alcançar o equilíbrio, por isso, mais baixa será a temperatura.

Não se percebe que o:

[D2]: Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não uma energia transitória. O frio também uma substância).

⇒ Por não fazer associação com a lata metálica, pois o problema se situa na garrafa de plástico e o seu refrigerante.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão* do Comportamento Térmico de acordo com as Propriedades Térmicas - [Descritor 17]

PROFICIÊNCIA*: *Inferir* uma «**igualdade de temperatura entre corpos com diferentes propriedades térmicas com base na Lei Zero da Termodinâmica**», «no caso, entrando em contato físico (sensação térmica) com o alumínio (lata), medindo a temperatura (termômetro) do líquido em seu interior (refrigerante) sendo 7 °C, e com isso concluindo sobre a temperatura de um recipiente plástico (garrafa) e do respectivo líquido em seu interior (refrigerante), na qual estavam sob as mesmas condições climáticas ambientais da vizinhança (interior da geladeira), portanto, também igual a 7 °C».

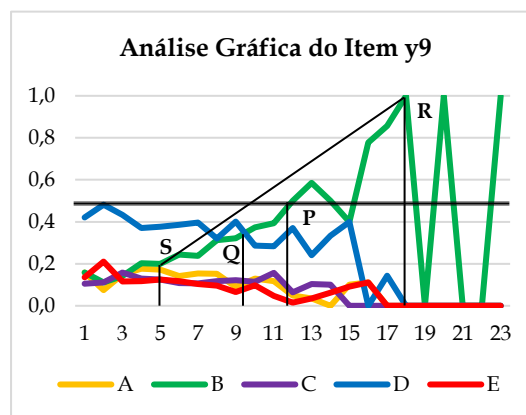
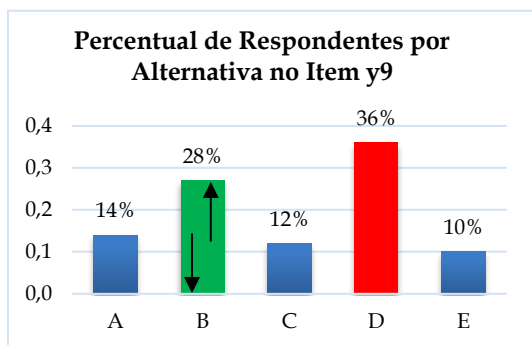
INDÍCIOS DE DEPENDÊNCIA LOCAL: Não foi identificado entre os itens anteriores.

**Considera-se em negrito o subunçor (ou parte dele) a ser evidenciado pelos respondentes de modo a se associar as informações do contexto do problema entre os outras que serão ou foram abordadas (conhecimentos potencialmente significativos), para assim representar uma assimilação conceitual.*

PAINÉL DE ESTATÍSTICAS DO ITEM y9

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

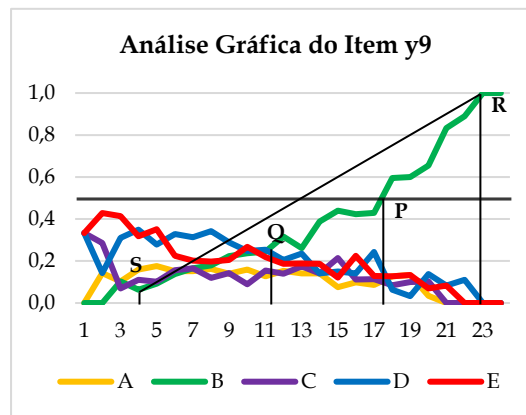
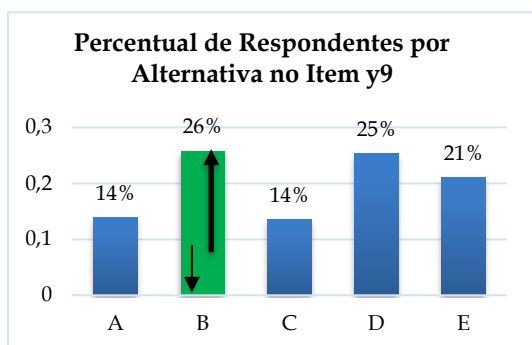
Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015																
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA					COEFICIENTES BISSERIAIS				
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	E	A	B	C	D	E
B	0,28	0,39	0,18	0,21	0,29	0,02	0,14	0,28	0,12	0,36	0,10	-0,10	0,29	-0,04	-0,09	-0,14



ID (corr.): 0,097; C: 0,10 (BB - aleatório); DISCR (corr. sem elim. item): 0,045; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,049; Coef. Bisserial (corr.): 0,065; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): Misto

Classificação AGI: **Baixa qualidade - Difícil; Baixa discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra B: 8,06 (31,0%); Desvio-padrão amostral: 2,98

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014																
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA					COEFICIENTES BISSERIAIS				
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	E	A	B	C	D	E
B	0,26	0,40	0,14	0,26	0,40	0,08	0,14	0,26	0,14	0,25	0,21	-0,09	0,40	-0,01	-0,20	-0,15

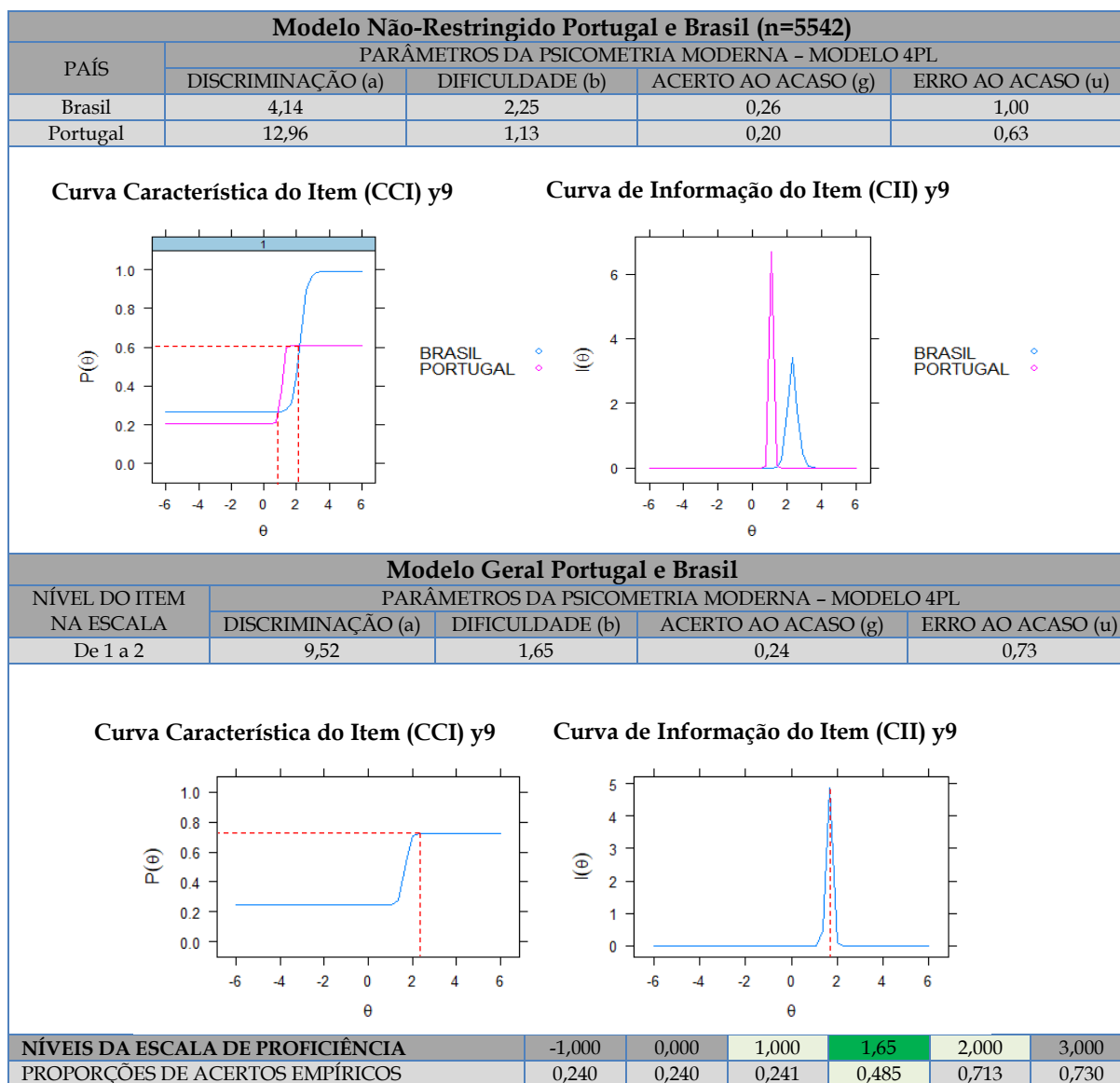


ID (corr.): 0,074; C: 0,03 (BB - aleatório); DISCR (corr. sem elim. item): 0,185; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,185; Coef. Bisserial (corr.): 0,251; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): Misto - predominância do grupo ACIM

Classificação AGI: **Média qualidade - Difícil; Média discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra B: 12,21 (47,0%); Desvio-padrão amostral: 4,18

Análise Gráfica do Item - AGI y9										
VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)					Nº DE INCIDÊNCIAS (PORTUGAL)				
	A	B	C	D	E	A	B	C	D	E
Tipo 1	-	7	-	-	-	-	3	-	-	-
Tipo 2	6	-	6	9	8	11	-	10	10	8
Tipo 3	0	-	1	0	0	0	-	0	0	0
Tipo 4	35º					30,1º				
Total	36					42				

Coeficientes Estatísticos Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y9

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem uma mesma tipologia para o item nos dois contextos sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **marginal** ($0,20 \leq DISCR < 0,30$), com necessidade de *reelaboração* (Brasil e Portugal).

A porcentagem absoluta do total de respostas na alternativa correta B ficou em 28% e 26% para os alunos brasileiros e portugueses. No Brasil a alternativa com maior porcentagem foi a letra D (36%), e em Portugal a letra D a que mais se aproximou com

25%, sendo a letra E também um forte atrator com 21%. A falta de um predomínio em marcações na alternativa correta sinaliza problemas no item.

Os grupos de maior e menor desempenho diante das respostas corretas no Brasil e em Portugal ficaram em 39% e 18%, e 40% e 14%, com um índice de discriminação um pouco maior entre alunos portugueses, porém são valores intermediários (entre 0,20 e 0,30). O bisserial é um outro índice de discriminação e os valores ficaram em 0,29 e 0,40, destacando mais entre alunos portugueses. Esse valor sugere um perfil de composição misto, com uma predominância maior relativamente maior de alunos com maior desempenho, sendo um pouco maior entre os alunos portugueses.

O distrator D teve um bisserial de -0,09 para o Brasil e -0,20 para Portugal, ou seja, uma composição de alunos praticamente equilibrado entre alunos bom e mal desempenho, sendo entre alunos portugueses, composto sutilmente com mais alunos de desempenho baixo. O ideal seria um bisserial mais próximo de -1, com predominância efetiva de alunos com desempenho inferior, mas não acontece. Esse distrator é atrativo para alunos brasileiros com 17 acertos, e para alunos portugueses com até 22 valores no escore bruto. Além disso, o distrator E possui um bisserial de -0,15, na qual também atrai alunos com bom desempenho, e isso não deve ser tolerável.

A AGI mostra uma quantidade de violações superior (comparado a Portugal) para a alternativa correta no contexto brasileiro, ou seja, parte dos alunos de bom desempenho oscilaram e foram atraídos por distratores entre os diversos níveis de escore. O poder de discriminação ou a intensidade com que as proporções de acerto fossem elevadas são próximas entre as populações, com uma leve superioridade para os brasileiros que sobem a partir do escore 5 (ponto S) e atingem o ápice em 18 acertos (ponto R) - ângulo de atividade igual a 35° , enquanto os portugueses são 4 (ponto S) e 23 (ponto R), porém com um intervalo de discriminação maior (distância de Q à R no eixo das abscissas).

O distrator D foi o que teve maior incerteza ou variações na proporção de marcações dessa alternativa no Brasil, porém muito próximo a quantidade de

violações do distrator E, sendo, entretanto, em menos proporções para essa última. Em Portugal, as violações com os distratores C e D foram equiparáveis, porém, o gráfico mostra que as proporções maiores foram em relação a D.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros conforme dizem os valores desse parâmetro. Os dois contextos apresentam uma boa discriminação conforme mostra a elevada inclinação das curvas. O acerto o acaso (*guessing*) foi de 0,26 e 0,20, muito próximo para os valores esperados, que seria de $\frac{1}{4}$ (0,25). Não foi estimado erros ao acaso para os brasileiros, mas para os portugueses o índice foi de 0,63, ou seja, 37% de cometerem equívocos por descuidos ainda que suas proficiências se eleve acima aproximadamente $\theta = 2,2$ (aproximadamente). De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os brasileiros são os que possuem uma maior probabilidade de acerto somente dentro de intervalo aproximadamente igual a $1,00 < \theta < 2,20$ (analisando comparativamente as CCIs).

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os portugueses, porém dentro de um intervalo pouco menor que 1 desvio-padrão (aproximadamente), enquanto para os brasileiros, apesar de estarem mais à direita, sob uma região de maior dificuldade, a informação se distribui dentro de um intervalo de 2 desvios-padrões aproximadamente entre os níveis de proficiência 1 e 2.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 1 a 2, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento técnico em uma proficiência representada por 5 itens-âncoras y16, y14, y26, y24 e y9 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (1,65) e discriminação (9,52) classificando o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado** para uma reaplicação, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 48,5% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso foi estimado em 0,24, dentro do valor do esperado que seria de $\frac{1}{4}$ ou 0,25, representando 24% de probabilidade de acertar sem possuir uma proficiência adequada. Foi estimado um índice de erro ao acaso de 0,73, ou seja, 27% de cometerem equívocos por descuido para além de um nível de proficiências igual a $\theta = 2,2$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação e uma boa informação para a escala de proficiência em um intervalo próximo de 1 desvio-padrão entre os níveis 1 e 2, sendo considerado um *item-âncora*.

DANDO SENTIDO AOS ERROS/DIFICULDADES: A questão tem a finalidade de inferir sobre o conhecimento de equilíbrio térmico de entre corpos diferentes em contato e sobre as propriedades térmicas entre a lata (metálica) e uma garrafa plástica. No caso, tanto a lata como a garrafa de plástico e seus conteúdos no interior estão há certo tempo dentro da geladeira, e é dito que o indivíduo segura a lata e a garrafa plástica na geladeira, e ao comparar as temperaturas por meio da sensação térmica, opta pela lata, que em seguida mede sua temperatura e obtém 7 °C. Espera-se que o aluno compreenda que a garrafa e o refrigerante que ficaram na geladeira estão em equilíbrio térmico, assim como a lata e o seu refrigerante, estando todos a uma temperatura de 7 °C. A alternativa D diz que: “O refrigerante cola está a 7°C, mas a garrafa está a mais de 7 °C”, isso leva a acreditar sobre o forte impulso sobre temperatura baseada na sensação térmica, além do desconhecimento sobre as propriedades de condutividade térmica diferentes entre o metal e o plástico. O distrator C diz: “Estão ambas acima de 7 °C”, que reforça a mesma convicção entre os alunos portugueses, com o agravante de que o plástico isola o refrigerante em seu interior, e faz com que se acredite que ambos ficam com uma temperatura superior a 7 °C.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: O fato de “pegar” na lata de metal e na garrafa de

plástico pode sugerir aos alunos que hajam trocas de calor, e isso influenciar na estimativa da temperatura para a garrafa de plástico e do refrigerante em seu interior, pois o metal tem maior facilidade de conduzir o calor. Dessa forma, sugere-se uma reelaboração no enunciado em que a medida realizada aconteça sem que haja o contato físico com os corpos.

CONSIDERAÇÕES DA PSICOLOGIA EDUCACIONAL DE AUSUBEL COMO CONDIÇÕES DE CONTORNO PARA A APRENDIZAGEM SIGNIFICATIVA: O *subsunçor* na situação-problema remete a um conhecimento preexistente de que haja uma **igualdade de temperatura entre corpos com diferentes propriedades térmicas com base na Lei Zero da Termodinâmica, considerando se encontram em um estado de equilíbrio térmico.** Dessa forma, a *Lei Zero da Termodinâmica* é um conhecimento mais estruturado que deve ser levado em consideração numa organização hierárquica dentro de uma sequenciação na abordagem dos conteúdos, dessa forma, indissociavelmente uma compreensão sobre o estado de equilíbrio térmico se configura como *conhecimento potencialmente significativo* para uma tomada de decisão de igualdade de temperatura entre os diferentes corpos.

Diante dos distratores evidenciados, sugere-se minimizar essas dificuldades explorando as limitações empíricas em considerar a sensação térmica como mecanismo de aferição da temperatura, apesar do enunciado afirmar que tenha sido utilizado um termômetro. A ideia seria conduzir a realização de uma sequência de atividades para investigar sobre os diferentes coeficientes de condutibilidade térmica dos materiais, explorar uma noção de temperatura baseada no tato, em seguida colocá-los sob mesmas condições ambientais para que alcancem o equilíbrio térmico e constatem que mesmo possuindo diferentes condutibilidades térmicas eles possuem a mesma temperatura, ainda que a sensação térmica possa induzir ao erro. Outra sugestão seria inicialmente introduzir em uma geladeira dois corpos de materiais diferentes, sendo um mais quente que o outro, juntamente com um termômetro no interior de cada objeto indicando as diferentes temperaturas iniciais, podendo se possível, deixar um dos termômetros recluso no interior da garrafa de plástico. Após

ter decorrido um razoável intervalo de tempo, se verificaria uma indicação de igualdade na graduação dos termômetros, isso constaria que o equilíbrio térmico sempre ocorre mesmo para materiais diferentes, e da mesma forma para o seu interior, reforçando a ideia que um recipiente isolante não impede de forma absoluta as trocas de calor. Questionando os alunos sobre isso, poderia ser explorado como causa do equilíbrio térmico as trocas de calor existentes entre o meio externo e o recipiente, que por sua possibilita o processo de condução térmica através do invólucro do recipiente, retirando ou fornecendo energia térmica, dependendo da temperatura interna.

Uma atividade experimental clássica e apropriada para trabalhar a sensação térmica seria a dos três recipientes do filósofo John Locke (1623-1704). Os três baldes de Locke pode ser reproduzida dentro de uma sala de aula ou em um laboratório. Consiste em um balde conter água fria ou água com gelo (balde 1), uma segunda com água um pouco aquecida (balde 2) e uma em equilíbrio com o ambiente ou em situação próxima a isso, ou seja, morna (balde 3). A finalidade da atividade é fazer os alunos aferir se água do balde 3 está fria ou relativamente quente. A sequência da atividade pode ser realizada por diversas combinações de modo que as mãos quando vem de baldes com temperaturas distintas, quando mergulhadas na água morna irão sugerir sensações térmicas distintas.

Uma sugestão para isso seria mergulhar as duas inicialmente no balde 3 durante alguns segundos, espera-se que elas digam que as duas mãos percebem da mesma forma a sensação térmica (morna), depois colocam-se as mãos dentro dos baldes 1 e 2, e volta-se para o balde 3. Quando isso ocorre os alunos deverão perceber que cada uma das mãos agora emite sensações térmicas distintas, espera-se que uma mão sinta a sensação de “quente” e a outra de “frio”. Esse é um ponto interessante para argumentar a necessidade e relevância do termômetro na medição da temperatura dos corpos.

Também é oportuno para explicar o funcionamento de diferentes termômetros, seus mecanismos e fenômenos térmicos envolvidos, destacando o conceito de grandeza termométrica. O termoscópio de Galileu também pode ser incluído nesse processo. Sugere-se que essas abordagens de ensino inicialmente sejam na modalidade

investigativa para os alunos, e posteriormente como verificativas e ilustrativas diante dos princípios abordados e discutidos. O conceito de temperatura deve ser explorado, podendo ainda ser utilizado o *Modelo Cinético Molecular* para auxiliar e possibilitar uma compreensão dualista em níveis microscópico (suposição do comportamento e agitação das partículas) e mesoscópico (baseado na sensação térmica e ainda na observação do registro da temperatura nos termômetros). Questões instigadoras também são relevantes para iniciar esse processo investigativo, como por exemplo: *Há um limite para o valor de temperatura mais elevada possível? E mais baixa? Qual a temperatura dos oceanos? De uma piscina? De um lago? Da sala de aula?* Sendo as situações mais promissoras aquelas mais próximas da realidade do aluno, portanto, mais significativas. Além de atividades práticas e experimentais enfatizadas, uma sugestão interessante seria fazer com que os alunos formulassem problemas e investigassem na modalidade de projetos de pesquisa, na qual articulam temas transversais e mobilizam diversos saberes, sendo, portanto, favoráveis para o desenvolvimento de competências, mas para isso seria fundamental identificar previamente um conjunto de situações-problemas promissoras, ou seja, situações que exigem diferentes habilidades e que envolvam o mesmo fenômeno ou contexto do problema.

A10: ITEM y10

10. Pedro coloca a lata de refrigerante cola sobre a mesa. Poucos minutos mais tarde, Carlos levanta a lata de refrigerante cola e diz que a superfície da mesa que estava debaixo da lata está mais fria do que o resto da mesa.

a) Gabriel diz: “O frio foi transferido do refrigerante cola para a mesa”.

[A3]: O quente e frio são diferentes, em vez de extremidades opostas de um **continuum** (as sensações térmicas são quantificáveis ou com caráter substancialista, ao invés de uma avaliação qualitativa e subjetiva).

⇒ Por acreditar que o frio seja algo quantificável (algo como energia térmica) que pode ser transferível e não como uma sensação térmica a ser avaliada qualitativamente em um continuum.

[A1]: O calor é uma substância (concepção substancialista).

⇒ Por acreditar que o frio seja uma forma de calor. ref. [C4].

[C4]: Calor e frio fluem como líquidos (concepção substancialista).

⇒ Por acreditar que o frio se “transfere”.

b) André diz: “Não há energia deixada na mesa pela lata.”

*Alternativa confusa: Não se percebe uma concepção alternativa, pois de fato, “a lata não deixou energia na mesa”, mas sim retirou energia térmica da mesa (!?).

c) Sabrina diz: “Um pouco de calor foi transferido* da mesa para o refrigerante cola”.

*A ideia de transferência de calor é controversa, pois somente se transfere aquilo que se possui, sendo uma energia térmica presente e se transferindo no meio da transição.

d) Elisa diz: “A lata faz com que o calor debaixo dela se desloque através dela para a mesa”.

*Alternativa confusa: Situação muito improvável de acontecer.

[A1]: O calor é uma substância (concepção substancialista).

⇒ Por acreditar que o calor se “desloca”.

[B3]: Percepções de quente e frio não estão relacionados com a transferência de energia.

⇒ Por não se perceber que a sensação de “frio” é devido a transferência de energia térmica da mesa para a lata.

[C4] menos provável: Calor e frio fluem como líquidos (concepção substancialista).

⇒ Por acreditar, em parte, pois o calor se “desloca” e faz sentir “frio” na mesa? Se fosse o “frio a se deslocar” ao invés do calor seria mais pertinente.

Não se percebeu entre as alternativas que:

[B6]: Um corpo frio não possui a capacidade de aquecer um outro corpo (somente corpos quentes “possuem calor” e podem aquecer, ou seja, se admite uma concepção ontológica substancialista).

⇒ Por acreditar que a mesa não foi aquecida pela lata, conforme diz na letra “b” (ao contrário como acredita Gonçalves (2012) e Louzada (2015)).

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir

uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por acreditar que eles tendem a mesma temperatura (também não foi identificado por Gonçalves (2012) e Louzada (2015))

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão sobre a Sensação Térmica explicada pela Propagação do Calor - [Descritor 13]*

PROFICIÊNCIA*: *Inferir sobre o sentido de «transferência de energia térmica que ocorre durante o processo de trocas de calor entre diferentes corpos sólidos com base na sensação térmica», no caso, «a lata de alumínio sobre a mesa de apoio retira um pouco de energia térmica por meio de calor na região da mesa em que houve contato físico, diferenciando a sensação térmica de frio entre essa região da mesa e o restante. De modo geral, se refere a uma compreensão de que o calor transita no sentido do corpo mais quente para o mais frio, ou seja, parte da energia térmica foi transferida da mesa para a lata de refrigerante».*

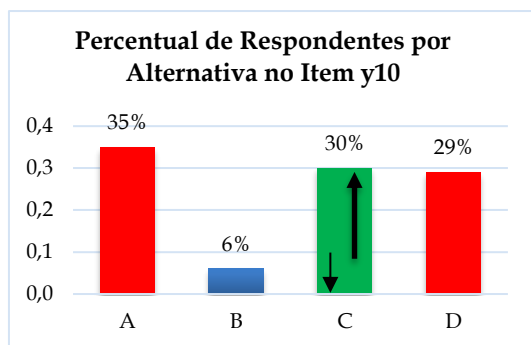
INDÍCIOS DE DEPENDÊNCIA LOCAL: Apesar do item y9 não se referir especificamente em um processo de transferência de energia, mas sim na caracterização de um estado termodinâmico para determinado sistema, ele também envolve uma compreensão construída a partir da sensação térmica quando diz que “...Samuel pega numa lata...”, com isso, esse é um fator que induz o respondente a imaginar a situação e a influenciá-lo. Como a tomada de decisão no item y10 remete a uma noção sobre sensação térmica, pode-se afirmar indício de dependência local entre os itens.

*Esta proficiência pode em parte aferir uma capacidade de compreensão que a mesa seja feita de um material de baixa condutividade térmica (geralmente madeira ou vidro), por esse motivo se diferencia a sensação térmica da região de contato com o restante.

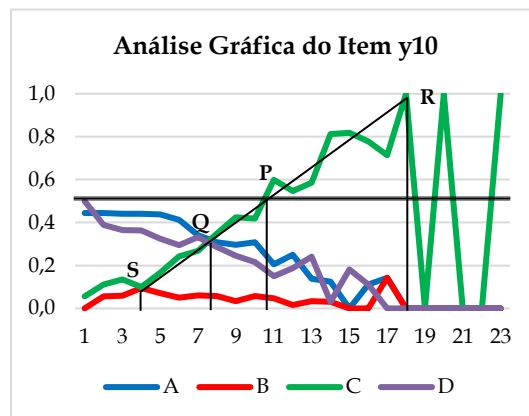
PAINÉL DE ESTATÍSTICAS DO ITEM y10

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
C	0,30	0,50	0,13	0,36	0,43	0,07	0,35	0,06	0,30	0,29	-0,20	-0,08	0,43	-0,17

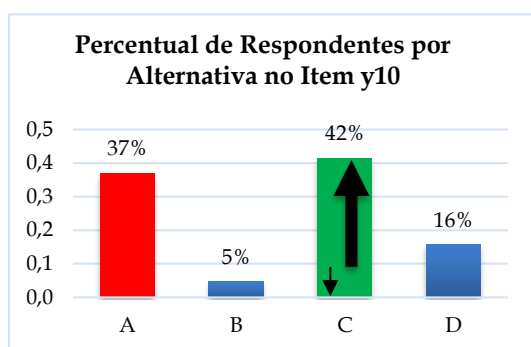


ID (corr.): 0,070; C: 0,10 (BB - aleatório); DISCR (corr. sem elim. item): 0,148; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,165; Coef. Bisserial (corr.): 0,217; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): **Misto - predominância do grupo ACIM**

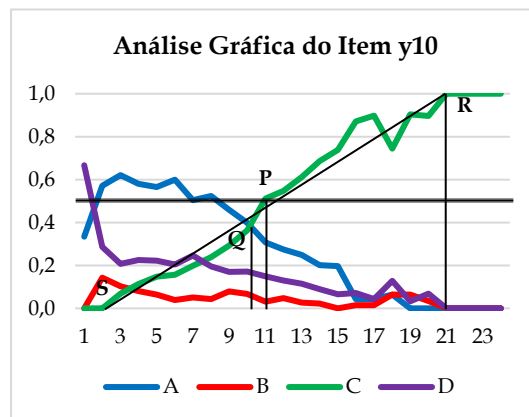


Classificação AGI: **Boa qualidade - Difícil; Média discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra C: 8,48 (32,6%); Desvio-padrão amostral: 2,88

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
C	0,42	0,70	0,18	0,52	0,60	0,18	0,37	0,05	0,42	0,16	-0,44	-0,13	0,60	-0,22



ID (corr.): 0,226; C: 0,17 (MB - aleatório); DISCR (corr. sem elim. item): 0,365; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,363; Coef. Bisserial (corr.): 0,458; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): **Predominância do grupo ACIM**

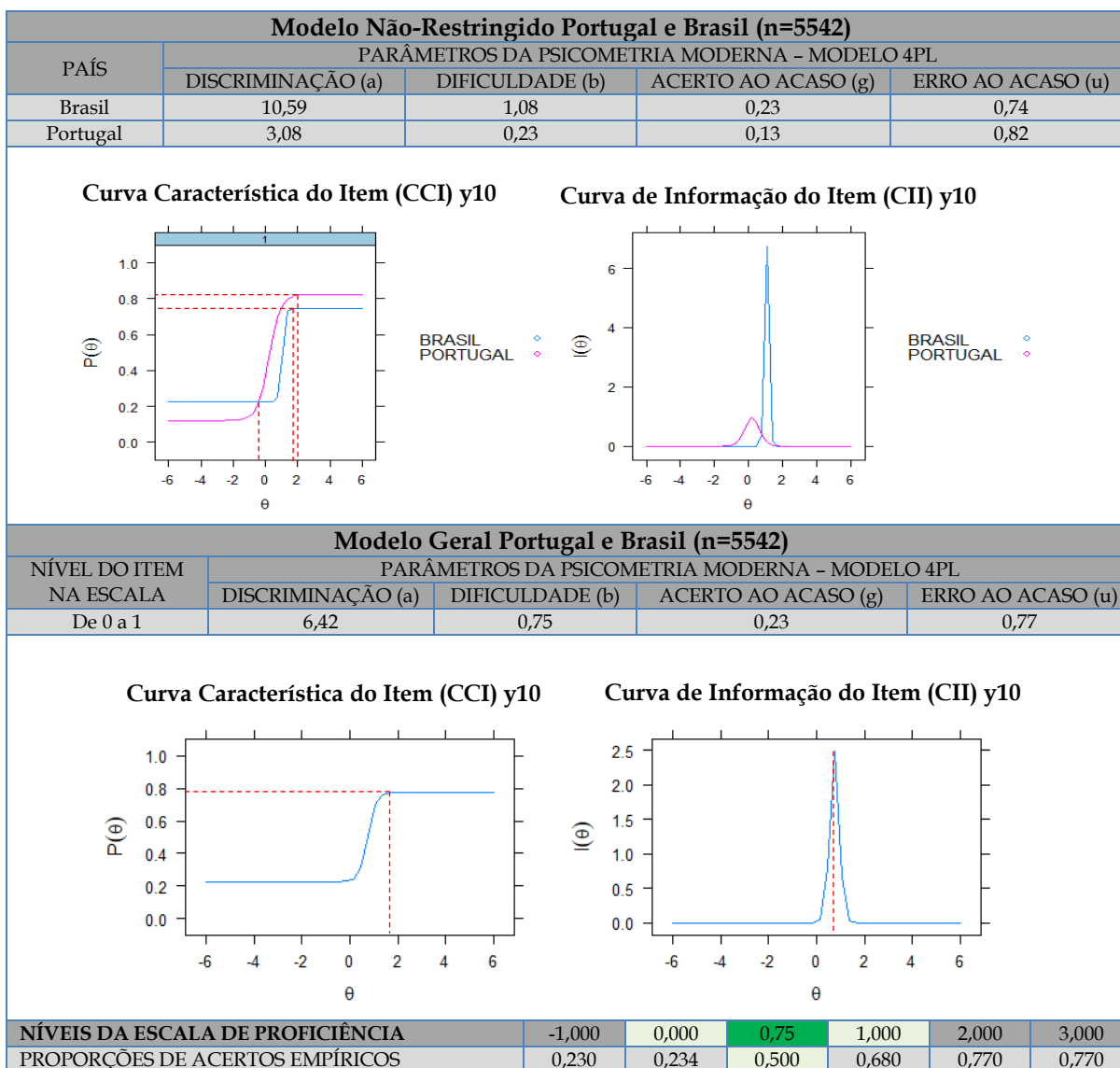


Classificação AGI: **Boa qualidade - Difícil; Alta discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra C: 12,45 (47,9%); Desvio-padrão amostral: 3,86

Análise Gráfica do Item - AGI y10

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	-	-	7	-	-	-	2	-
Tipo 2	5	7	-	4	6	8	-	6
Tipo 3	0	0	-	0	0	0	-	0
Tipo 4	31º				32,2º			
Total	23				22			

Coeficientes Estatísticos Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y10

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **bom, sujeito a aprimoramento** ($0,30 \leq DISCR < 0,40$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **adequado, disponível a reaplicação** ($DISCR \geq 0,4$).

As porcentagens totais das respostas mostram que não houve predominância absoluta das respostas corretas em ambos os contextos (30% e 42%), porém em Portugal se tem uma predominância relativa, na qual se refere a uma porcentagem maior da alternativa correta sobre os distratores, apesar de que não seja a maioria

(>50%). No Brasil o distrator A alcançou a maior porcentagem com 35%, sendo a letra C a segunda opção mais assinalada, e a letra D teve quase a mesma porcentagem com 29%, isso mostra a força atrativa dos distratores A e D para os alunos brasileiros. Em Portugal o distrator A também se apresentou com forte atração com 37%, isso ressalta investigar com atenção essa alternativa.

Os grupos de maior e menor desempenho ficaram com 50% e 13%, e 70% e 18%, ou seja, o grupo de maior desempenho prevaleceu nos dois contextos acertando proporcionalmente mais que o grupo inferior, dessa forma seguiu a ordem esperada em que $ABAIX < INTERM < ACIMA$ ou $A_{27\%}^+ < A_{46\%}^\pm < A_{27\%}^-$, em outras palavras, que o percentual de acerto do grupo inferior seja menor que o intermediário, e este seja menor que o grupo de maior desempenho, e isso é bom sinal, apesar de estar evidenciado uma grande dificuldade para o item.

Os bisseriais obtidos da alternativa C foram de 0,43 e 0,60, e isso sugere um perfil de composição com predominância de alunos com bom desempenho, como mostra os percentuais totais de acerto dentro de cada grupo. O distrator D teve certa influência para os respondentes brasileiros, e seu bisserial foi de -0,17, ou seja, um perfil misto de composição, com leve predominância de alunos com baixo desempenho.

O distrator A merece atenção e obteve valores para o coeficiente bisserial de -0,20 e -0,44, isso significa dizer que os alunos de menor desempenho foram os que assinalaram em sua maioria essa alternativa, sem mais expressivo essa discriminação para os alunos portugueses, apesar de que o ideal seria esse valor estar mais aproximado de -1. Observa-se ainda no gráfico que tiveram alunos brasileiros com 17 acertos que assinalaram essa alternativa, assim como alunos portugueses com 18 acertos.

A AGI revela uma quantidade maior de violações para os alunos brasileiros quanto ao acerto do item, comparado com os alunos portugueses, ou seja, houveram alguns decréscimos relevantes na proporção de acertos, principalmente entre os alunos com escores iguais a 16 e 17 valores. O distrator B foi o que mais teve oscilações

nas proporções em ambos os contextos, porém não convêm levar em grande consideração devido ao seu percentual total dentro das amostras, até mesmo pelos baixos picos de oscilação. O distrator A merece maior atenção. O poder de discriminação ficou equivalente nos dois contextos com um ângulo próximo de 30°.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Normais/Medianos** (proporção esperada de respondentes 40%; $-0,51 < b \leq 0,51$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros conforme dizem os valores desse parâmetro. As CCIs nos dois contextos sugerem uma boa discriminação conforme mostra a elevada inclinação. O acerto ao acaso (*guessing*) foi de 0,23 e 0,13 (valor esperado de $\frac{1}{4}$ ou 0,25), sugerindo que os portugueses foram menos propensos a “chutes”, seja por fadiga, pressa em terminar ou por desconhecimento, entre outros, isso indica coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência estimados para os sujeitos. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 0,74 e 0,82 (Brasil e Portugal), com isso se estima 26% e 18% a partir de um nível de proficiência em torno de $\theta = 2,0$, sendo os alunos portugueses com uma probabilidade percentual inferior de cometer esses erros a partir de um valor θ aproximado. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto quando $\theta > 0,4$ (aproximadamente).

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros, porém dentro de um intervalo de 1 desvio-padrão (aproximadamente), enquanto para os portugueses, o gráfico sugere a contribuição de informações para a interpretação da escala, porém com uma discriminação menor,

estando mais à esquerda, reforçando a perspectiva de uma maior facilidade para responder o item. Essas elevações reforçam o poder discriminativo do item e de contribuições para as informações interpretativas na escala, no caso, entre os níveis de proficiência 0 e 1.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 0 a 1, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y_3 , y_{13} , y_{23} , y_{10} e y_{22} (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (0,75) e discriminação (6,42) classificando o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 50% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso foi estimado em 0,23, dentro do esperado que seria de $\frac{1}{4}$ ou 0,25, representando 23% de probabilidade de acertar sem possuir uma proficiência adequada. Foi estimado um índice de erro ao acaso de 0,77, ou seja, 23% de cometerem equívocos por descuido para além de um nível de proficiências igual a $\theta = 1,8$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação e uma boa informação para a escala de proficiência em um intervalo próximo de 1,5 desvio-padrão entre os níveis 0 e 1, sendo considerado um *item-âncora*.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item afirma que a temperatura de uma região da mesa está mais fria após ter sido colocada uma lata de um refrigerante, supostamente com temperatura mais fria. No distrator A, a opinião de Gabriel diz que: “*O frio foi transferido do refrigerante cola para a mesa*”, primeiramente isso sugere uma falta de compreensão do que seja o “frio”, pois deve ser entendido como uma sensação térmica e não como energia transferida a ser cedida ou recebida, e em segundo lugar, o frio é uma percepção de perda de energia térmica por meio do tato ou contato físico com um corpo de temperatura mais baixa. Conceber a ideia de frio como algo a ser transferido pode remeter uma concepção substancialista do calor,

ou de que seja uma forma de “calor negativo” que produz frio, sendo uma aversão a noção de que o calor aqueça. No distrator D, Elisa afirma: “A lata faz com que o calor debaixo dela se desloque através dela para a mesa”, é um pouco mais controverso. Um equívoco seria associar o “frio” nesta região da mesa foi causada pela perda de calor na parte inferior da lata, quando na verdade ela recebe calor da mesa. Quando se diz “...que o calor debaixo dela se desloque...” e como consequência deixa a região da mesa mais fria, o equívoco pode estar na associação desse “calor” com a ideia da “existência de frio” na lata (reforçando dizer que frio não existe, é apenas uma percepção sobre a ausência/perda de energia térmica), que por sinal se “desloca”. Com isso, se tem os enganos de confundir calor com frio, e de que frio se desloca, não percebendo que a lata foi esquentada pela mesa a supostamente permanecer relativamente fria a lata.

SUGESTÕES: Buscar diferenciar os conceitos de calor e temperatura, sobretudo com a ideia de sensação térmica de frio, quente ou morno, visando uma compreensão sobre o processo de “trocas de calor”, ou melhor dizendo, da perda/ganho de energia térmica entre os corpos em contato. O uso de um termômetro dessa situação na vida real poderia evidenciar se de fato o refrigerante está aumentando, diminuindo ou permanecendo com a mesma temperatura, assim como verificar que entre os corpos do entorno/vizinhança que estão na temperatura ambiente tendem a um equilíbrio térmico (considerando um ambiente relativamente controlado, sem uma influência de variáveis climáticas, como vento, ar-condicionado, aquecedores, etc.). Inserir dentre as atividades com os alunos a experiência dos três baldes de Locke juntamente com uma discussão sobre a sensação térmica, temperatura média da pele humana e o campo de temperatura podem contribuir no esclarecimento, semelhante como Mattos & Drummond (2004) apresentam. Uma discussão sobre a temperatura média da pele humana, por exemplo, pode ser expressa com a fórmula definida por Hardy-Dubois modificada (Ibid, 2004):

$$T_{pele} = 0,07 \cdot (T_{testa} + T_{pé}) + 0,05 \cdot (T_{mão}) + 0,14 \cdot T_{antebraço} + 0,35 \cdot (T_{peito} + T_{costas}) + 0,19 \cdot T_{coxa} + 0,13 \cdot T_{canela}$$

E isso pode ajudar a entender que as diversas partes do corpo podem emitir uma sensação térmica diferente. Nisso, evidências de que diferentes pessoas possuem temperaturas sutilmente distintas na sua pele, por influência inclusive de características próprias (como: espessura da pele, gordura, pelos, etc.) também podem ser aspectos enriquecedores nessa discussão. E por fim, sugere-se esclarecer que o fluxo de calor ou potência calorífica de determinados tipos de materiais pode trazer elementos que diferenciam os efeitos observados em situações em que ocorrem o processo de perda/ganho de energia térmica.

A afirmação de que sensação térmica de frio se concentra numa região específica da mesa, no caso, na região do contato físico entre a lata e a mesa, também pode remeter uma ideia sobre condutividade térmica da mesa, apesar de que seja algo secundário, pois não interfere ou influencia na escolha das alternativas, porém pode influenciar em itens posteriores. Nesse caso, convém destacar que uma mesa na maioria das vezes é feita de madeira ou vidro, com isso possuem um valor médio de condutividade térmica a 27 °C (temperatura ambiente) de $K_{madeira} = 0,13 J \cdot s^{-1} \cdot m^{-1} \cdot K^{-1}$ (pinho) e $K_{vidro} = 0,79 J \cdot s^{-1} \cdot m^{-1} \cdot K^{-1}$, bem abaixo dos metais e valores razoavelmente próximos a da água. Em situações de maior influência da condutividade térmica na escolha das alternativas, é importante destacar a necessidade de compreensão sobre a Lei de Fourier.

A11: ITEM y11

11. Patrícia pergunta aos amigos: “Se eu colocar 100 gramas de gelo a 0 °C e 100 gramas de água a 0 °C dentro da geladeira, qual dos dois perderá a maior quantidade de calor?”

a) Cristina diz: “Os 100 gramas de gelo”.

⇒ É mais difícil esfriar o gelo que a água estando ambos a 0°C (gasta-se mais energia para esfriar/aquecer o gelo do que a água, assim como gasta-se mais energia para esfriar/aquecer a água que o vapor d’água). (*): Não houve correspondência direta com nenhuma concepção do inventário).

b) Bernardo diz: “Os 100 gramas* de água”.

c) Marcelo diz: “Nenhum dos dois, porque ambos contêm a mesma quantidade de calor”.

[A5]: O calor é proporcional à temperatura.

⇒ Por ser considerado apenas a temperatura e a mesma substância.

[D4]: Diferentes materiais, independentemente de suas massas, sempre irão receber a mesma quantidade de energia térmica (acredita-se que diferentes materiais possuam a mesma capacidade térmica, e o conceito de calor assume uma concepção substancialista).

⇒ Por acreditar que são a mesma substância e estão em uma mesma quantidade.

d) Mateus diz: “Não há resposta, porque o gelo não contém nenhum calor”.

[B6]: Um corpo frio não possui a capacidade de aquecer um outro corpo (somente corpos quentes “possuem calor” e podem aquecer, ou seja, se admite uma concepção ontológica substancialista)

⇒ Por acreditar que o gelo seja um corpo é um corpo na qual lhe foi retirado todo o calor, portanto, não possui mais calor e com isso não pode aquecer outro corpo.

e) Júlia diz: “Não há resposta, porque não podemos ter água a 0 °C”.

[D7]: A água no estado líquido não pode estar a 0°C.

⇒ Por acreditar que inevitavelmente a 0 °C a água se torna gelo.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão sobre a Propagação do Calor - [Descritor 9]*

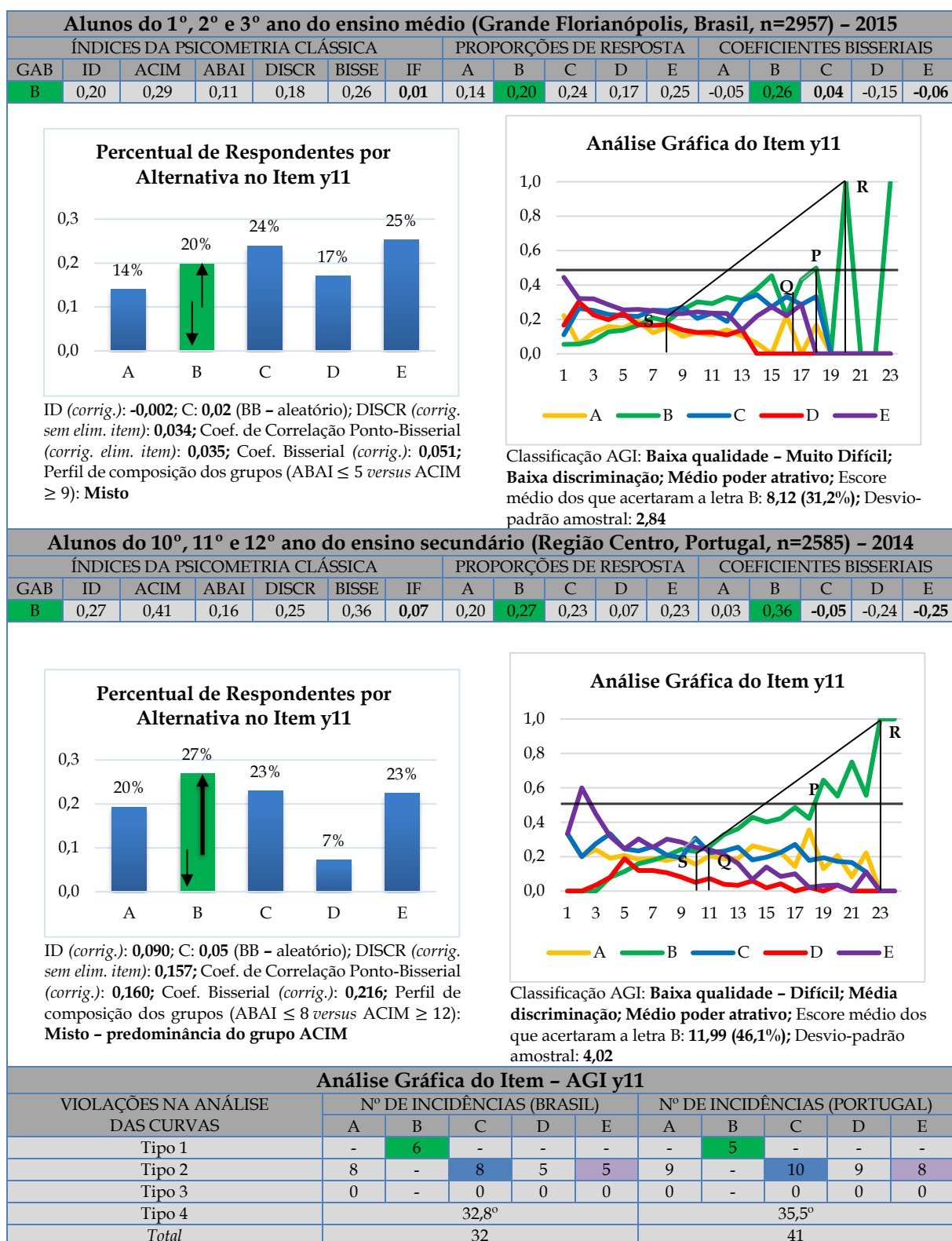
PROFICIÊNCIA: *Inferir* que corpos correspondentes a uma mesma substância, com mesma massa*, e quando submetidos a uma mesma fonte térmica, «o corpo em um estado com maior força de repulsão e menor força de coesão em sua estrutura molecular necessita perder uma quantidade maior de calor para alcançar o mesmo estado termodinâmico do outro corpo». No caso, «a água tende a perder uma maior quantidade de calor que o gelo» «até que se alcance o equilíbrio térmico dentro de uma geladeira, pois encontra-se em um estado termodinâmico mais “distante” que o gelo, portanto, quem perde mais calor são os 100 g de água».

INDÍCIOS DE DEPENDÊNCIA LOCAL: Não foi identificado entre os itens anteriores.

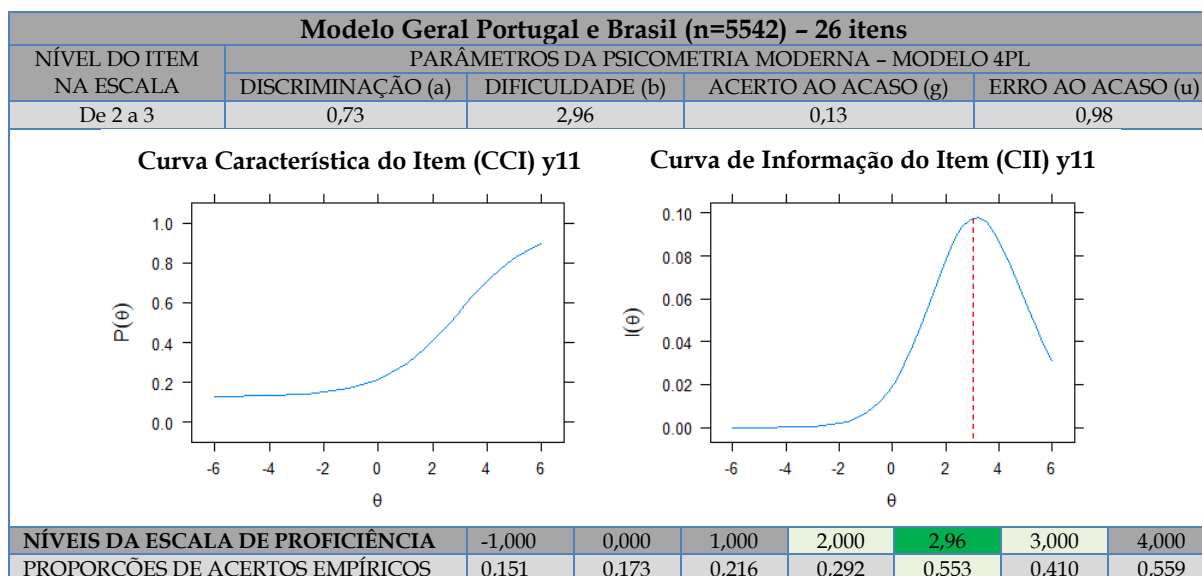
*A massa é uma informação parasita, pois a compreensão desta questão depende exclusivamente do estado físico, e assim, esta informação pode estar gerando equívocos.

PAINÉL DE ESTATÍSTICAS DO ITEM y11

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y11

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **inadequado, sujeito a rejeição** ($DISCR < 0,20$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **marginal, sujeito a reelaboração** ($0,20 \leq DISCR < 0,30$) (Portugal).

Os percentuais de acerto ficaram em 20% no Brasil (sendo a 3ª opção no contexto geral) e em 27% para Portugal. Nos dois contextos os distratores demonstram grande força atrativa para os respondentes, em especial, as alternativas C e E, isso demonstra problemas no item e a necessidade de verificação desses distratores.

Os grupos de maior e menor desempenho ficaram em 29% e 11% (Brasil), e 41% e 16% (Portugal), sendo o índice de discriminação percentual mais baixo entre os alunos brasileiros, ou seja, com mais dificuldade em diferenciar alunos bons e ruins. O bisserial da letra B foi de 0,26 e 0,36, sendo menor para o Brasil, como deveria ser. Os valores dos bisseriais sugerem um perfil de composição misto entre alunos de alto e baixo desempenho, com leve predominância de alunos maior desempenho, porém ainda longe do ideal que seria um valor mais próximo de 1.

O distrator C apresenta bisserial igual a 0,04 e -0,05, ou seja, quase igual a zero em ambos os contextos. Isso significa dizer que de forma igualitária é formado por alunos com alto e baixo desempenho, e isso é algo que não deve ser tolerado, pois alunos brasileiros com 18 acertos e portugueses com 22 acertos assinalaram essa alternativa. O distrator E foi de -0,06 e -0,25, ou seja, entre alunos brasileiros é praticamente igual a composição de alunos com desempenho bom e ruim, enquanto em Portugal prevalece um pouco mais de alunos de baixo desempenho, apesar de que não seja muito representativo, tendo alunos com escore igual a 22 sendo atraídos por essa alternativa.

Na AGI, chama atenção as oscilações dos distratores A e C, com mais intensidade para o C, e entre os alunos portugueses os distratores A, C e D, com mais intensidade proporcional nas amplitudes para a letra A. O poder de discriminação ficou baixo e valor próximo entre as amostras, de acordo com os ângulos de inclinação, sendo o um valor menor para o intervalo de discriminação entre os brasileiros.

Com base na estimação dos parâmetros da TRI através do modelo geral com 26 itens, o item y4 pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **bom**, *sujeito a aprimoramento*, com uma *discriminação moderada* ($0,65 \leq a \leq 1,35$).

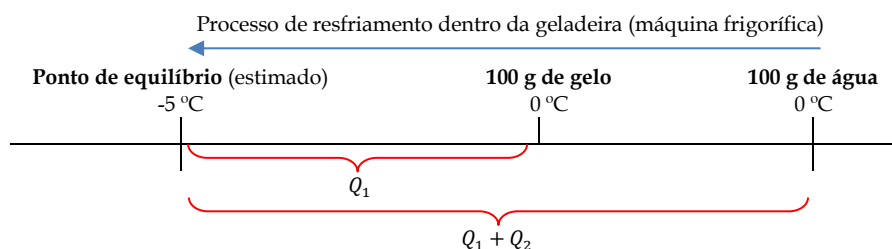
A CCI apresenta uma sigmoide com uma inclinação próxima de 45° , revelando discriminação moderada, ou seja, o item não distingue muito bem respondentes alta e baixa proficiência. Estando mais à direita, indica uma posição do item em uma alta habilidade θ , possuindo um percentual de acertos empíricos estimados em 55,3% entre aqueles que alcançam esse nível de habilidade θ . O índice de acerto ao acaso foi de 0,13, um pouco abaixo do valor esperado que seria de $1/5$ ou 0,20, isso representa um probabilidade de 13% dos respondentes acertarem o item sem possuírem habilidades suficientes. O índice de *guessing* um pouco abaixo do esperado indica uma boa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os respectivos níveis de proficiência alcançados. O índice de erro ao acaso foi de 0,98, isso representa apenas 2% de probabilidade de os alunos cometerem equívocos, ainda que possuam uma boa habilidade e esse nível se eleve. Por fim, a CII indica uma baixa

discriminação, com isso, o item y11 não apresenta informações relevantes para uma interpretação pedagógica, estando sujeito a rejeição.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item indaga sobre quem “perde mais calor”, gelo ou a água, porém elas possuem a mesma massa e a mesma temperatura, no caso 0 °C. Apesar de ser compreensível a situação, falar quem a “perdeu” ou “ganhou” calor pode induzir de forma implícita de que um corpo “possui” calor, e isso é um erro. Superando essa perspectiva, 0 °C é uma temperatura de transição de fases para o gelo (fusão) e a água (solidificação), que no caso correspondem a uma mesma substância. A resposta a essa questão é conhecer quem perde mais calor quando submetido a uma mesma taxa ou fluxo calorífico, considerando que para um frigorífico essa taxa é negativa, diferente para as máquinas térmicas.

Intuitivamente, no contexto do problema, é importante ter uma noção de que seja mais fácil esfriar o gelo do que a água, mas é preciso ter atenção entre resfriar mudando a temperatura e resfriar mudando de fase, pois os processos de mudança de fase geralmente exigem uma maior quantidade de energia para realizar uma conversão do que uma energia para variar sua temperatura estando o corpo na mesma fase. Diante disso, sabe-se que a água necessita de uma quantidade maior de perda de energia térmica para se converter plenamente a gelo ou para se solidificar plenamente, do que resfriar sem mudar de fase. Na situação apresentada os dois são submetidos a uma mesma fonte térmica (mesmo fluxo calorífico, no caso, supondo que dentro da máquina frigorífica eles estejam em condições idênticas quanto a tendência de perda de energia térmica) tendem a perder/ganhar a mesma quantidade de energia térmica, porém os processos de como ocorre isso e os efeitos causados são diferentes, pois dependem de suas propriedades térmicas específicas. Apesar disso, subentende-se que dentro de uma máquina térmica ambos tende a um equilíbrio térmico com sua vizinhança, dessa forma a quantidade de energia térmica cedida até esse ponto é maior para a água, que segue um “trajeto” maior até alcançar esse estado. Em outras palavras, espera-se que o gelo perca uma quantidade de calor sensível Q_1 , enquanto a água uma quantidade de calor latente para solidificar totalmente Q_2 mais Q_1 para

alcançar o equilíbrio com o meio, juntamente com o gelo. Vale ainda destacar que o valor da massa é uma variável parasita, pois para responder essa questão é necessário apenas saber o estado físico dos corpos.



Na situação, o gelo apenas tende a diminuir sua temperatura a partir de 0 °C em proporções que respeite o valor de seu calor específico “médio” ($c_{gelo} = 0,5 \frac{cal}{g \cdot ^\circ C}$), enquanto a água deva mudar de fase mantendo sua temperatura de 0 °C respeitando o valor de seu calor latente ou calor específico latente “médio” de solidificação ($L_{solidificação} = -80 \frac{cal}{g}$). Comparando os valores percebe-se que o gelo para variar 1 °C em cada grama precisa perder apenas 0,5 cal, em quanto a água para solidificar cada grama precisar perder 80 cal, ou seja, precisa perder mais energia para influenciar o estado térmico de cada grama, porém o enunciado não especifica dessa forma, e também não tem a intenção de que os alunos saibam e utilizem esses valores como base para a escolha, com exceção dos alunos que já tiveram aulas de calorimetria dentro da termodinâmica.

Quando se indaga “...qual dos dois irá perder a maior quantidade de calor? ”, e não fornece informação quanto ao tempo, com isso deve se subentendido os dois são supostamente colocados ao mesmo tempo dentro da geladeira. O tempo de perda de quantidade de calor para os corpos também deve subentendido até que atinjam o equilíbrio com o meio. Como a fonte é a mesma, é compreensível que perdem a mesma quantidade de calor “durante esse processo”, diante disso, a ausência de uma delimitação sobre esses aspectos pode estar gerando um problema na interpretação do enunciado.

Apesar dessas limitações identificadas, pode-se ainda inferir sobre alguns equívocos dos alunos que resolveram assinalar uma alternativa (talvez por se sentir

obrigado a isso devido responder de acordo com as imposições das alternativas), por exemplo, ao observar o distrator C, em que o Marcelo diz: *“Nenhum dos dois, porque ambos contêm a mesma quantidade de calor.”*, isso sugere que os alunos concebiam que os corpos “possuem calor”, que o calor seja uma propriedade do corpo, e isso é um erro. No distrator E, Júlia diz: *“Não há resposta, porque não podemos ter água a 0 °C”*, os alunos com essa convicção possuem a crença de que a água chega ao máximo próximo de 0 °C, sendo 0 °C apenas atribuída ao gelo, e isso sugere uma necessidade de compreender o processo de mudança de fase.

SUGESTÕES: A reelaboração no enunciado do item é um ponto-chave preliminar. Uma sugestão para contornar vieses interpretativos seria especificar qual dos dois necessitam perder mais calorias para mudar o estado térmico de cada uma grama, seja variando a temperatura ou mudando de fase, porém isso requer uma “malícia” para transpor numa linguagem compreensível e que faça sentido aos alunos (que respeite critérios de legibilidade quanto a faixa etária), ou seja, numa transposição didática adequada, porém fazendo isso, muda-se o problema. Para manter o problema, pode-se informar qual seria a perda até que ambos alcancem o equilíbrio dentro da geladeira, independentemente da massa (variável parasita). Essa talvez seria uma alternativa mais apropriada para o enunciado.

Diante dos equívocos dos distratores assinalados se recomenda explorar os processos de aquecimento e resfriamento incluindo variações de temperatura e mudanças de fase, de preferência utilizando inicialmente a água pura, depois alterando o seu estado de pureza, acrescentando, por exemplo sal ou outras substâncias visando alterar suas propriedades térmicas. Além das atividades experimentais, o uso de vídeos e a proposição de trabalhos em forma projetos na perspectiva interdisciplinar e/ou investigativa que explore o conceito de calor específico são caminhos promissores, pois a dificuldade de compreensão para os alunos e até mesmo para se ensinar essa grandeza física é algo que exige uma maior atenção. É importante ressaltar que esses valores (calor específico e específico latente) não são absolutos, mas **valores médios**. Uma forma de aprofundar essa discussão seria apresentando equações matemáticas que expressam essas médias, por exemplo, o

calor latente de condensação da água, no intervalo de temperatura entre -40°C e 40°C , pode ser obtido através da seguinte função:

$$L_{\text{água}} = (2500,8 - 2,36 \cdot T + 0,0016 \cdot T^2 - 0,00006 \cdot T^3) \frac{J}{g}$$

Outro exemplo, dentro do mesmo intervalo de temperatura, o calor latente de sublimação do gelo pode dado pela função:

$$L_{\text{gelo}} = (2834,1 - 0,29 \cdot T + 0,004 \cdot T^2) \frac{J}{g}$$

Compreender a relação entre as unidades Joule (J) e calorias (cal) se torna necessário, com isso a experiência de James Prescott Joule (24 de dezembro de 1818 - 11 de outubro de 1889) denominado como “equivalente mecânico” seria enriquecedor numa abordagem para o contexto histórico e epistemológico, na qual poderia exercer uma função de organizador prévio para minimizar dificuldades nesse âmbito específico.

Mudando o foco dos objetos que estão sendo resfriados para a máquina frigorífica que buscar “retirar” o calor de seu interior, se tem a perspectiva da 1ª Lei da Termodinâmica, em que, durante o processo de resfriamento, a geladeira necessita realizar um trabalho W para “extrair” toda a energia térmica Q do sistema até que se atinja e se mantenha o equilíbrio termodinâmico, pois a geladeira não é um sistema perfeitamente isolado termicamente. Dessa forma, ao explorar os objetos em seu interior juntamente como a geladeira realiza esse processo e mantém esse estado de equilíbrio são aspectos que ampliam e complementam para uma melhor compreensão do fenômeno.

A12: ITEM y12

12. Marta está fervendo água numa panela sobre o fogão. O que você acha que está no interior das bolhas que se formam na água em ebulição? Na maior parte:

a) Ar.

[D12]: As bolhas na água fervente contêm “ar”, “oxigênio” ou “nada”.

⇒ Por acreditar que as bolhas, por serem transparentes, devam conter ar, devido à semelhança.

b) Gás oxigênio e hidrogênio.

[D12]: As bolhas na água fervente contêm “ar”, “oxigênio” ou “nada”.

⇒ Por acreditar que a água por se composta de átomos de oxigênio e hidrogênio, ao ferver desassocia e libera esses átomos como moléculas de gás em forma de bolhas

c) Vapor de água.

d) Não há nada nas bolhas

[D12]: As bolhas na água fervente contêm “ar”, “oxigênio” ou “nada”.

⇒ Por acreditar que a bolha por ser transparente não contém matéria.

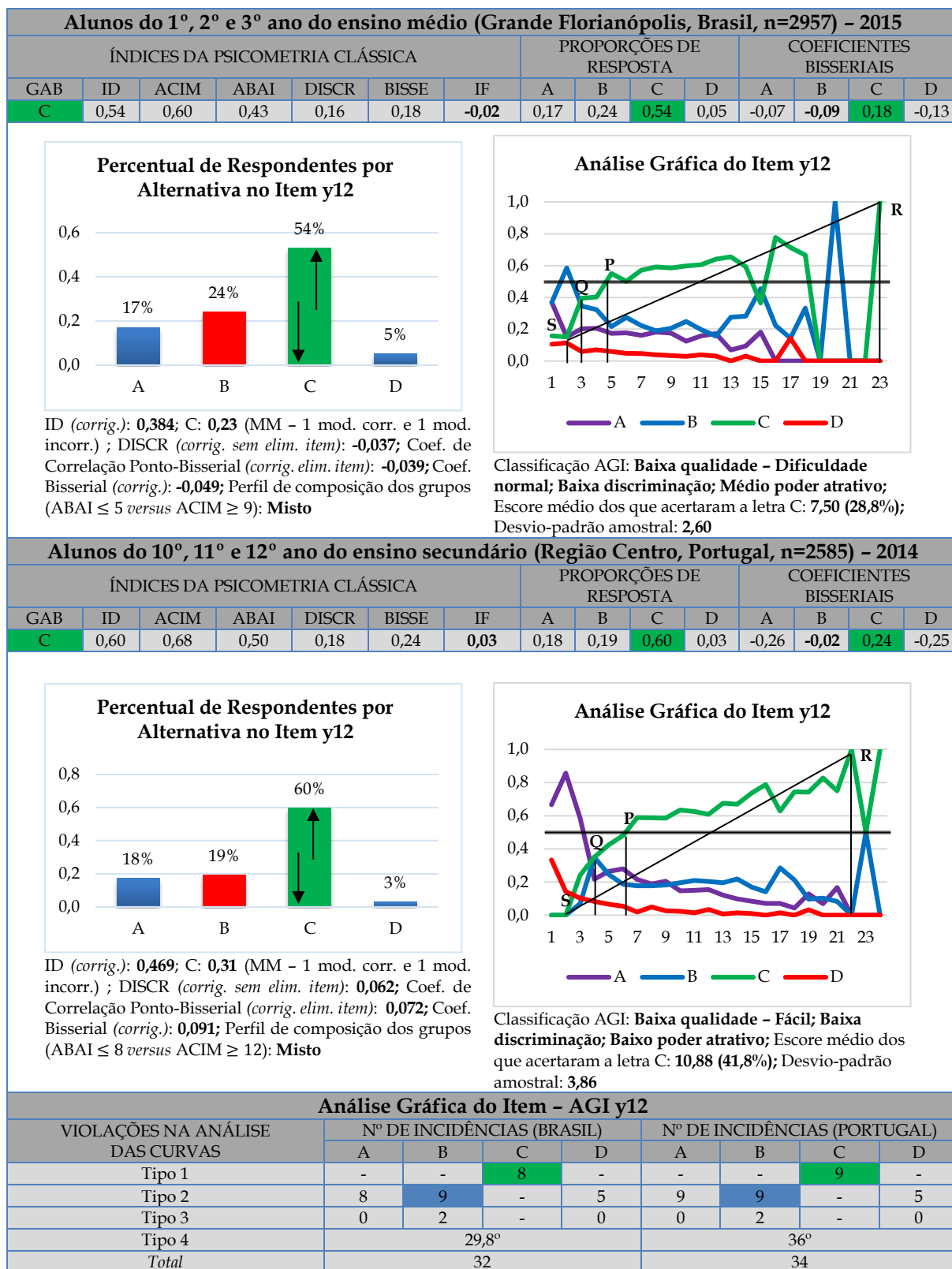
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão* sobre o Comportamento Térmico dos corpos durante a Mudança de Fase - [Descritor 8]

PROFICIÊNCIA: *Inferir* por meio da percepção visual que «as bolhas formadas em meio a água fervente representam na maior parte vapor d'água», sendo uma situação que ocorre habitualmente «em uma panela em aquecimento sobre o fogão que contêm água em processo de vaporização por ebulição».

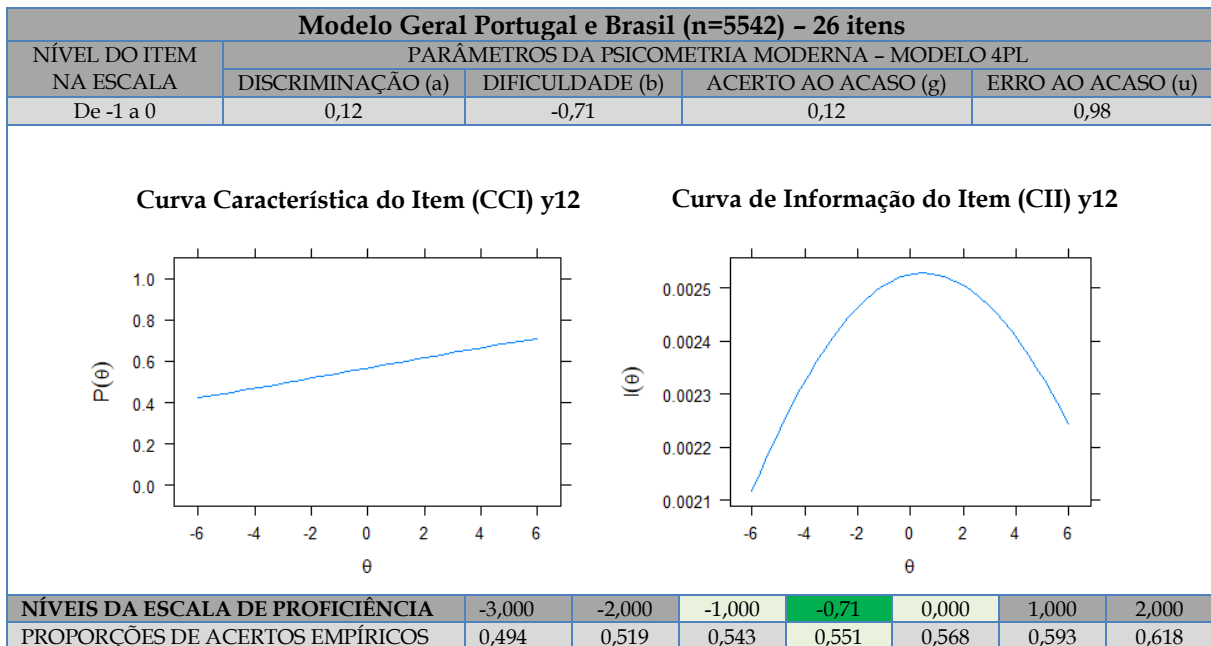
INDÍCIOS DE DEPENDÊNCIA LOCAL: Não foi identificado entre os itens anteriores.

PAINÉL DE ESTATÍSTICAS DO ITEM y12

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticas e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y12

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Normal/Mediano** (proporção esperada de respondentes 40%; $0,45 \leq ID < 0,55$) e **inadequado, sujeito a rejeição** ($DISCR < 0,20$) (Brasil); **Fácil** (proporção esperada de respondentes 20%; $0,55 \leq ID < 0,75$) e **inadequado, sujeito a rejeição** ($DISCR < 0,20$) (Portugal).

As porcentagens de acerto total foram predominantes nos dois contextos com 54% para Brasil e 60% para Portugal. Os distratores mais atrativos em ambos os contextos foram o A e o B, com mais ênfase no B com 24% (Brasil) e 19% (Portugal), porém, está na zona de *guessing*.

Os grupos de maior e menor desempenho tiveram uma porcentagem de acerto de 60% e 43% entre alunos brasileiros, e 68% e 50% entre os portugueses, ou seja, apresentam índices de discriminação equivalentes e uma elevada porcentagem de acertos entre alunos com baixo desempenho, isso é preocupante, pois os alunos do grupo inferior deveria ficar com baixa porcentagem de acerto, porém, também pode indicar que seja um item fácil até para alunos de baixa proficiência. O coeficiente bisserial dos alunos brasileiros foi de 0,18 e dos alunos portugueses 0,24, ou seja, são

compostos por alunos de desempenho superior e inferior de forma quase equivalente, com uma leve predominância de alunos com bom desempenho, e isso é algo aceitável.

O distrator B possui um bisserial de -0,09 e -0,02. Isso representa que em ambos os contextos o perfil de composição dos grupos foi formado por alunos de baixo e alto desempenho, entretanto representam relativamente um percentual baixo do total de respondentes. Se observa que alunos brasileiros com 20 acertos, e portugueses com 23 acertos foram atraídos, isso reforça uma necessidade de verificar esse distrator.

A AGI nos dois contextos mostrou violações equivalente entre os distratores A e B, assim como para a alternativa correta C. O poder de discriminação é sutilmente maior entre os respondentes portugueses, porém é considerado baixo.

Com base na estimação dos parâmetros da TRI através do modelo geral com 26 itens, o item y12 pode ser classificado como: **Fácil** (proporção esperada de respondentes 20%; $-1,27 < b \leq -0,52$) e **muito ruim**, *sujeito a reelaboração* ($0,0 < a \leq 0,35$).

A CCI não apresenta uma sigmoide, mas sim uma reta praticamente linear, revelando pouca discriminação, ou seja, o item não consegue distinguir bem respondentes alta e baixa proficiência. Estando mais à esquerda, indica uma posição do item em uma baixa habilidade θ , possuindo um percentual de acertos empíricos estimados em 55,1% entre aqueles que alcançam esse nível de habilidade θ . O índice de acerto ao acaso foi de 0,12, um pouco abaixo do valor esperado que seria de $\frac{1}{4}$ ou 0,25, isso representa um probabilidade de 12% dos respondentes acertarem o item sem possuírem habilidades suficientes. O índice de *guessing* um pouco abaixo do esperado indica uma boa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os respectivos níveis de proficiência alcançados. O índice de erro ao acaso foi de 0,98, isso representa apenas 2% de probabilidade de os alunos cometerem equívocos, ainda que possuam uma boa habilidade e esse nível se eleve. Por fim, a CII indica uma baixa discriminação, com isso, o item y4 não apresenta informações relevantes para uma interpretação pedagógica, estando sujeito a rejeição.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item indaga sobre a composição de bolhas formadas em água fervente, e se espera que os alunos percebam

que são constituídas de vapor d'água, pois as moléculas que ficam na parte inferior da panela são aquecidas mais rapidamente que o restante, por isso vaporizam rapidamente. O distrator B diz que é formado por gás oxigênio e hidrogênio, ou seja, moléculas conhecidas que fazem parte do ar que respiramos. O gás oxigênio representa em cerca de 20% enquanto o nitrogênio 78%, o primeiro é bem conhecido por representar o gás essencial para os seres humanos e o outro por ser um dos mais abundantes em nossa atmosfera, e essas características podem induzir o respondente achar que toda as bolhas representam esses tipos de moléculas.

SUGESTÕES: Atividades experimentais e vídeos explicativos são mecanismos que podem ajudar a esclarecer a formação e a composição desses tipos de bolhas. Vale ressaltar que essas bolhas mudam de tamanho à medida que sobem, e isso está associado a diferença de pressão do líquido sobre elas, com isso, uma revisão sobre o Teorema de Stevin se torna apropriado. Também existem outros fenômenos associados, como a convecção, a das diferentes densidades (bolha e água líquida) que justifica o fato delas subirem. Uma explicação melhor sobre a questão da densidade remete uma revisão sobre o Teorema de Arquimedes, discussões sobre densidade e massa específica, assim como densidade relativa e das relações com a temperatura, com isso se tem uma noção de para essa grandeza se adota habitualmente um valor médio. Outro aspecto complementar seria o “comportamento anômalo” da água entre 0 °C e 4 °C que explicaria cristalinis ou espaços vazios formados no processo de solidificação.

Tabela 49: Variação da densidade da água em função da temperatura.

Temperatura (°C)	Densidade (kg/m ³)
100	958,4
80	971,8
60	983,2
40	992,2
30	995,6502
25	997,0479
22	997,7735

20	998,2071
15	999,1026
10	999,7026
4	999,9720
0	999,8395

Fonte: "Densidade" (n.d.).

Considera-se que o *Modelo Cinético-Molecular* como mecanismo explicativo focado no processo de aquecimento das partículas e de mudança de fase, pode ser um caminho rico e desmistificador para uma melhor compreensão da formação de bolhas, de sua agitação e deslocamento em meio a água fervente.

A13: ITEM y13

13. Depois de cozinhar alguns ovos em água fervente, Mariana resfria-os ao colocá-los numa panela com água fria. Qual das seguintes afirmações explica o processo de resfriamento?

a) A temperatura é transferida do ovo para a água.

[C5]: **A temperatura pode ser transferida** (concepção substancialista, a temperatura é uma substância passível de se deslocar de um corpo para outro).

⇒ Por considerar que a temperatura seja uma substância passível de ser transferida.

b) O frio desloca-se da água para os ovos.

[A3]: **O quente e frio são diferentes, em vez de extremidades opostas de um continuum** (as sensações térmicas são quantificáveis ou com caráter substancialista, ao invés de uma avaliação qualitativa e subjetiva).

⇒ Por acreditar que o frio seja algo quantificável (algo como energia térmica) que pode ser transferível e não como uma sensação térmica a ser avaliada qualitativamente em um continuum.

[C4]: **Calor e frio fluem como líquidos** (concepção substancialista).

⇒ Por considerar que a temperatura seja uma substância passível de se deslocar e ser transferida.

c) Objetos quentes resfriam naturalmente.

[C7]: **Objetos quentes naturalmente se resfriam, objetos frios naturalmente se aquecem.** (tange a concepção animista, como se os objetos tivessem vontade própria de se aquecerem ou se resfriarem espontaneamente, independentemente de qualquer fator externo)

⇒ Por acreditar que os corpos possuem uma temperatura típica e própria, e ao ser aquecido, tende a retornar a sua temperatura normal. Ou considerar que um corpo quente, está em desequilíbrio em uma temperatura acima de sua temperatura típica.

d) A energia é transferida dos ovos para a água.

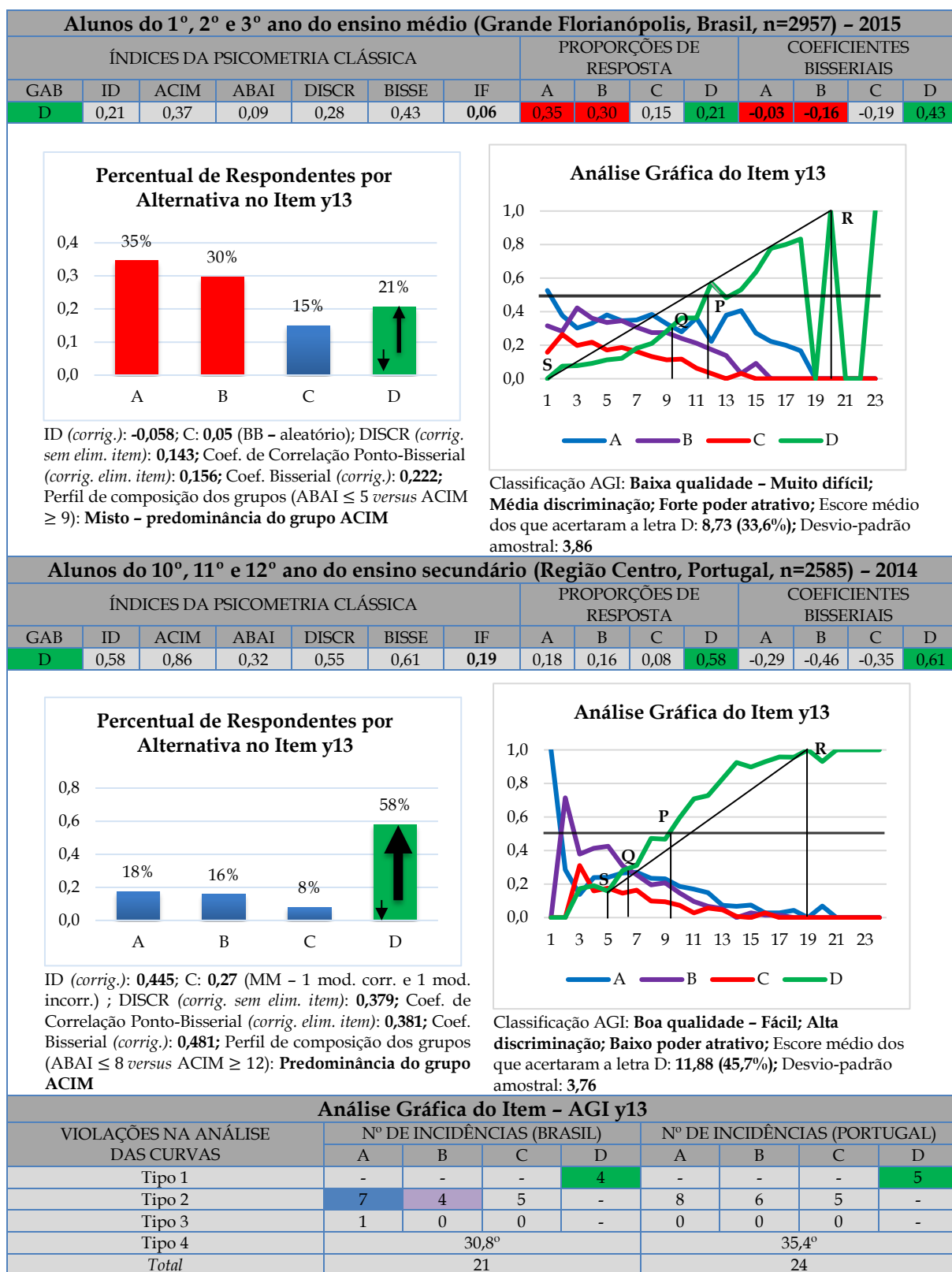
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão* sobre a Propagação do Calor - [Descritor 10]

PROFICIÊNCIA: *Inferir* sobre o sentido de «transferência de energia térmica que ocorre durante o processo de resfriamento de ovos cozidos quando submetidos em água fria», nessa situação que ocorre uma «transferência de energia térmica dos ovos para a água, ou seja, do corpo mais quente (maior temperatura) para o mais frio (menor temperatura)».

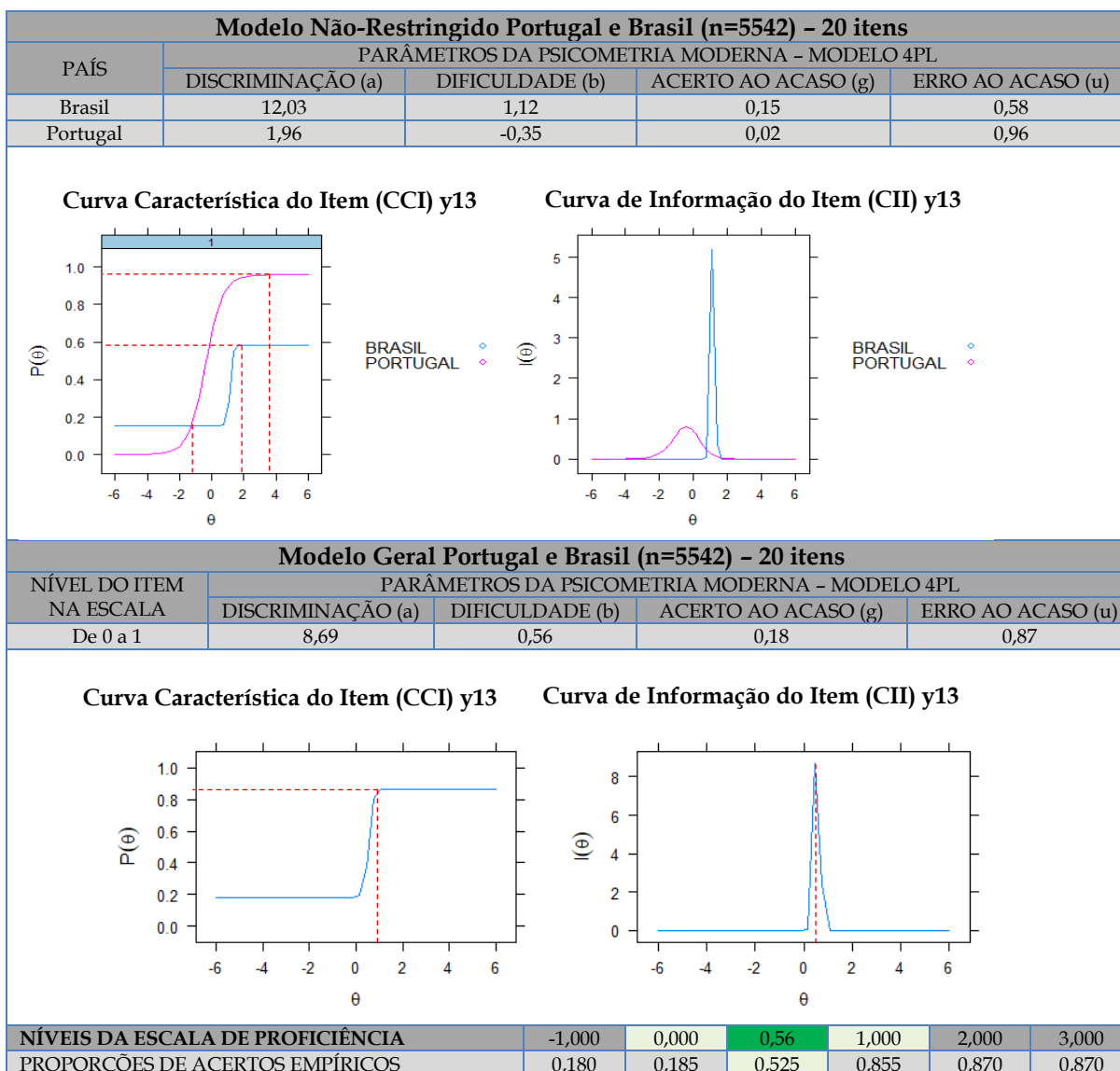
INDÍCIOS DE DEPENDÊNCIA LOCAL: O item y10 também invoca um conhecimento sobre o sentido no processo de transferência de energia que deve ir de uma fonte com temperatura mais elevada para uma fonte mais fria.

PAINÉL DE ESTATÍSTICAS DO ITEM y13

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y13

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **marginal, sujeito a reelaboração** ($0,20 \leq DISCR < 0,30$) (Brasil); **Fácil** (proporção esperada de respondentes 20%; $0,55 \leq ID < 0,75$) e **adequado, disponível a reaplicação** ($DISCR \geq 0,4$). Constata-se um comportamento discrepante entre os dois contextos para um diagnóstico sobre a aferição do conhecimento a ser avaliado, ou seja, há um indicativo de função diferenciada do item.

A porcentagem total para a alternativa D ficou em 21% para o Brasil, sendo a 3ª opção no percentual total, e com 58% para os alunos portugueses, ou seja, com predominância absoluta sobre os distratores. Os distratores A e B merecem atenção quanto as concepções dos alunos brasileiros, em que representam 35% e 30%, enquanto para os portugueses se tem 18% e 16%. O distrator C ficou com o menor percentual para ambos os grupos, com isso, considera-se relevante uma interpretação sobre os distratores A e B, com maior ênfase para o A.

Os grupos de maior e menor desempenho ficou com 37% e 9% (Brasil), e 86% e 32% (Portugal). Isso indica que alunos de bom desempenho possuem maior acerto, como era de se esperar, entretanto, em Portugal esse percentual é mais elevado, assim como a diferença, com isso demonstra ser uma amostra com maior discriminação, e como o grupo de baixo desempenho possui um percentual maior que o dos brasileiros, sugere que seja um item fácil até mesmo para quem tem maior dificuldade, pouco domínio de conhecimento, ou seja, baixa proficiência.

Os bisseriais da letra D ficaram com 0,43 e 0,61. Para uma interpretação sobre o perfil de composição dos grupos, sugere-se que sejam formados por uma maioria de alunos bom desempenho, sendo no contexto português uma maior predominância devido estar mais próximo de 1. Os bisseriais para A ficaram com -0,03 e -0,29, ou seja, os alunos que assinalaram essa alternativa foram praticamente iguais entre alunos brasileiros de baixo e bom desempenho, e entre os portugueses foi levemente mais acentuado entre os alunos de baixo desempenho. Pode-se observar que houve entre os de bom desempenho no Brasil, um percentual de alunos com 18 acertos que optaram pela alternativa A, e também incidências de pouca proporção entre os alunos portugueses com até 20 acertos. Para o distrator B o bisserial mostra o mesmo padrão entre os brasileiros com -0,16, e bem mais discriminante entre os portugueses com -0,46.

A AGI mostra poucas violações no crescimento da alternativa correta para os grupos, porém uma oscilação elevada no distrator A entre os alunos brasileiros comparativamente com os portugueses, e isso sinaliza maiores incertezas para essa

opção entre os brasileiros. O poder de discriminação para o grupo de Portugal ficou levemente superior ao do Brasil, mas pode ser considerado baixo de modo geral.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Normal/Mediano** (proporção esperada de respondentes 40%; $-0,51 < b \leq 0,51$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros conforme dizem os valores desse parâmetro. As CCIs nos dois contextos sugerem uma boa discriminação conforme mostra a elevada inclinação. O acerto ao acaso (*guessing*) foi de 0,15 e 0,02, valores abaixo do esperado que seria $\frac{1}{4}$ ou 0,25, isso significa que os alunos foram pouco propensos a sortear a alternativa ou escolher aleatoriamente ("chute"), seja por fadiga, pressa em terminar ou por desconhecimento, em especial, os alunos portugueses. O valor baixo de *guessing* também indica coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência alcançados, em especial, as habilidades (θ) estimadas para os sujeitos. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 0,58 e 0,96 (Brasil e Portugal). Com isso, estima-se que os alunos brasileiros a partir de um nível de proficiência em torno de $\theta = 1,9$ tendem a uma probabilidade de 42% de cometerem esses erros, enquanto os portugueses ficaram com uma probabilidade baixíssima de 4%. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros, porém dentro de um intervalo de 1 desvio-padrão (aproximadamente), enquanto para os portugueses, o gráfico sugere a contribuição de informações para a interpretação da escala, porém com uma discriminação menor,

estando mais à esquerda, reforçando a perspectiva de uma maior facilidade para responder o item. Essas elevações reforçam o poder discriminativo do item e de contribuições para as informações interpretativas na escala, no caso, entre os níveis de proficiência 0 e 1.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 0 a 1, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y_3 , y_{13} , y_{23} , y_{10} e y_{22} (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (0,56) e discriminação (8,69) classificando o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 52,5% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso foi estimado em 0,18, abaixo de um valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 18% de probabilidade de acertar sem possuir uma proficiência adequada. Foi estimado um índice de erro ao acaso de 0,87, ou seja, 13% de cometerem equívocos por descuido para além de um nível de proficiências igual a $\theta = 1,1$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação e uma boa informação para a escala de proficiência em um intervalo próximo de 1 desvio-padrão entre os níveis 0 e 1, sendo considerado um *item-âncora*.

DANDO SENTIDO AOS ERROS/DIFICULDADES: A questão busca identificar uma compreensão sobre o processo de trocas de calor, em especial, quanto ao sentido em que a energia térmica se transfere. É esperado que os respondentes compreendam que o ovo cozido ao ser submerso na água fria aqueça a água.

A descrição do distrator A diz que “a temperatura é transferida do ovo para a água”, isso evidencia um forte atrativo da concepção alternativa entre alunos brasileiros de que a temperatura seja equivalente a noção de calor, e isso não deve ser aceito. Essa concepção alternativa demonstra ser menos intenso entre alunos portugueses. Também houve um percentual significativo para o distrator B entre os alunos

brasileiros quando se considera que “o frio desloca-se da água para os ovos”. Admitir essa concepção demonstra ser ainda mais preocupante, tendo em vista que frio é uma sensação térmica, além do que há uma violação da 1ª Lei da Termodinâmica quanto a ideia de que a energia térmica deva ser transferida espontaneamente no sentido do corpo mais quente para o mais frio.

SUGESTÕES: Compartilha-se da noção de que as concepções alternativas não precisam, em especial, sobre calor e temperatura, não precisam necessariamente serem superadas ou abandonadas, mas sim de serem toleradas dependendo do contexto que se possam ser expressas, estando os alunos conscientes dessas diversas zonas e relações entre elas (Amaral & Mortimer, 2001; Mortimer, 1995). No sentido de compreensão dos conceitos se admite como um equívoco a ser esclarecido com o risco de distorção sobre o entendimento científico, diante disso, torna-se fundamental a atribuição de metodologias de ensino que visem diferenciar esses conceitos buscando novos perfis conceituais desassociando concepções equivocadas sobre calor e temperatura, em que muitas vezes são caracterizadas como obstáculos epistemológicos, como apresentado por Farias (2013).

Sugere-se que o conceito de temperatura possa ser explorado em uma modalidade investigativa tendo como vetor epistemológico uma concepção racionalista baseada no *Modelo Cinético-Molecular* (MCM) visando uma apreensão do sujeito por meio da observação indireta do fenômeno (Braga, 2010), ou seja, no sentido do racional para o real (Bachelard, 1978). É importante associar a concepção de temperatura com a noção de energia de vibração molecular, nisso envolve uma apreensão sobre o estado termodinâmico e a energia interna do sistema, e não o processo de perda ou ganho de energia térmica. O calor é um conceito abstrato, pois não é propriedade de um corpo, com isso é preciso que seja explorado a concepção de energia térmica em trânsito que se perde ou ganha, mas não no sentido de apropriação, e sim como a gente causador da mudança de estado termodinâmico ou de realizador de trabalho dentro de um sistema.

Calor pode ser associado com a ideia de fontes térmicas ou fluxo de energia térmica, pois é uma entidade física dinâmica e de transição entre um corpo emissor e

receptor. Em corpos físicos pode ser entendido como uma energia que se propaga molécula-por-molécula, mas também pode ser manifestada por ondas eletromagnéticas independente da matéria, nesse caso conhecido como “onda de calor”. Essa complexidade indica que seja um conceito que necessita ser trabalhado em distintos campos epistemológicos, sendo concebido como um comportamento corpuscular ou ondulatório, e isso demarca limitações para o uso do MCM como viés explicativo.

A14: ITEM y14

14. Janete diz que não gosta de se sentar em cadeiras de metal porque “elas são mais frias do que as de plástico”.

a) Rafael concorda e diz: “Elas são mais frias porque o metal é naturalmente mais frio do que o plástico”.

[D1]: A temperatura é uma propriedade de um material ou objeto particular.

(uma característica específica para cada tipo de material)

⇒ Por se acreditar que as cadeiras de metal possuem temperaturas típicas, diferentemente das de plástico.

b) Mário diz: “Elas não são mais frias, elas estão à mesma temperatura”.

c) Luís diz: “Elas não são mais frias, as de metal só parecem mais frias porque são mais pesadas”.

[B7]: A temperatura de um objeto depende de seu tamanho.

⇒ Por acreditar que o metal ao ser mais “pesado”, ou seja mais denso, ou que possui maior tamanho, que esses sejam aspectos que influenciam diretamente a temperatura de um corpo.

d) Sandra diz: “Elas são mais frias porque o metal tem menos calor para perder do que o plástico”.

[D2]: Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não um energia transitória. O frio também uma substância)

⇒ Por considerar que a possibilidade da cadeira de metal conter ou manter calor seja menor que a capacidade da cadeira de plástico.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão* do Comportamento Térmico de acordo com as Propriedades Térmicas – [Descritor 17]

PROFICIÊNCIA*: *Inferir* uma «igualdade de temperatura entre corpos com diferentes propriedades com base na Lei Zero da Termodinâmica», «no caso, entrando em contato físico com uma cadeira de metal e uma de plástico, considerando uma independência da sensação térmica, portanto, as cadeiras estão à mesma temperatura e não devem ser avaliadas com base na sensação térmica».

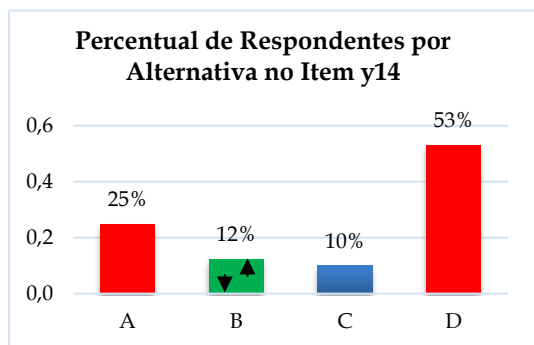
INDÍCIOS DE DEPENDÊNCIA LOCAL: O item y9 também busca aferir uma compreensão de equilíbrio térmico comparando semelhantemente corpos com diferentes propriedades térmicas (condutividade), entre o metal e o plástico, considerando, nesse caso, a necessidade de um juízo que independa da influência da sensação térmica. O item y10 também envolve influência da sensação térmica.

**Janete apresenta uma noção da sensação de frio está associada a noção de temperatura, e essa é uma perspectiva equivocada tolerada no enunciado, porém está reforçada na alternativa considerada correta evidenciando em parte um erro interpretativo associado, com isso se torna necessário fazer algumas críticas e considerações a esse respeito.*

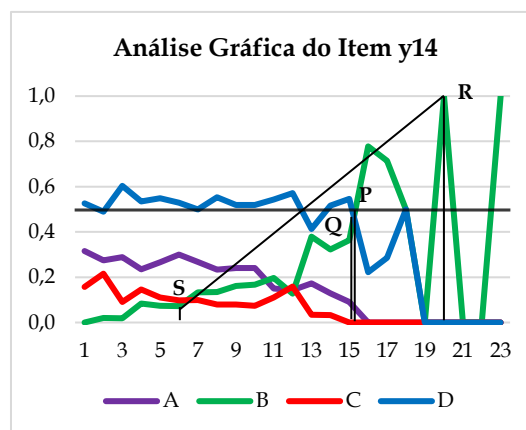
PAINÉL DE ESTATÍSTICAS DO ITEM y14

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
B	0,12	0,20	0,06	0,14	0,34	0,03	0,25	0,12	0,10	0,53	-0,09	0,34	-0,09	0,00

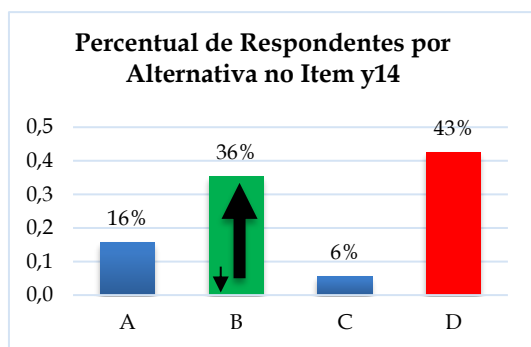


ID (corr.): -0,168; C: 0,21 (BM - 2 mod. incorr.); DISCR (corr. sem elim. item): 0,085; Coef. De Correlação Ponto-Bisserial (corr. elim. item): 0,089; Coef. Bisserial (corr.): 0,143; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): **Misto**

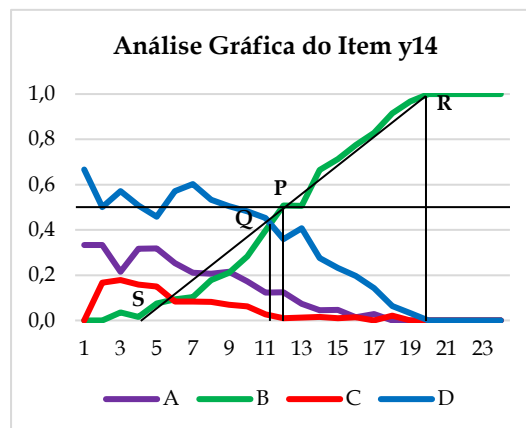


Classificação AGI: **Baixa qualidade - Muito difícil; Média discriminação; Forte poder atrativo; Escore médio dos que acertaram a letra B: 8,64 (33,2%); Desvio-padrão amostral: 3,17**

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
B	0,36	0,67	0,11	0,56	0,69	0,21	0,16	0,36	0,06	0,43	-0,34	0,69	-0,33	-0,34



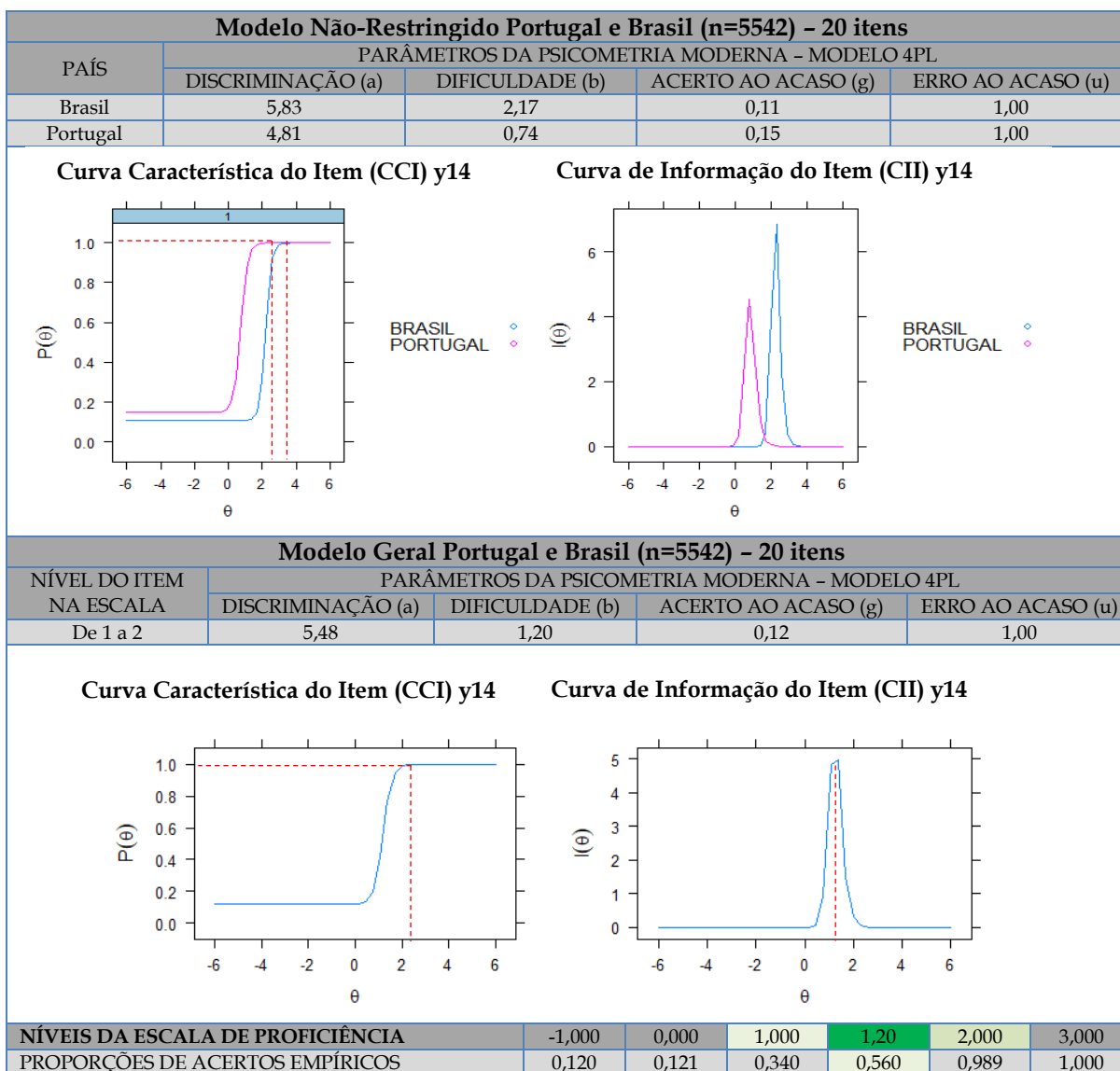
ID (corr.): 0,142; C: 0,17 (BB - aleatório); DISCR (corr. sem elim. item): 0,442; Coef. De Correlação Ponto-Bisserial (corr.): 0,436; Coef. Bisserial (corr.): 0,562; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): **Predominância do grupo ACIM**



Classificação AGI: **Boa qualidade - Difícil; Alta discriminação; Médio poder atrativo; Escore médio dos que acertaram a letra B: 13,11 (50,4%); Desvio-padrão amostral: 3,78**

Análise Gráfica do Item - AGI y14								
VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	-	9	-	-	-	2	-	-
Tipo 2	6	-	5	9	6	-	7	4
Tipo 3	0	-	0	1	0	-	0	0
Tipo 4	33,5°				38°			
Total	30				19			

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y14

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **bom, sujeito a aprimoramento** ($0,30 \leq DISCR < 0,40$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **adequado, disponível a reaplicação** ($DISCR \geq 0,4$).

O percentual total de acertos entre alunos brasileiros ficou em 12%, sendo 3ª opção, enquanto para alunos portugueses foi de 36%, ou seja, em nenhum cenário houve predominância da resposta correta, isso sinaliza de imediato problemas

relevantes nesse item. Os distratores D e B ficaram com 53% e 25% para os brasileiros, sendo para alunos portugueses o distrator D a opção igualmente mais atrativa entre todos com 43%, ou seja, torna-se relevante avaliar as possíveis causas da força atrativa desse distrator para ambos os contextos.

Os grupos de maior e menor desempenho ficou com 20% e 6% (Brasil), e 67% e 11% (Portugal). Os alunos de bom desempenho possuem maior percentual de acerto nos dois contextos, como era de se esperar, entretanto, em Portugal esse percentual é mais elevado e acima de 50%, com predomínio absoluto do total de respondentes, logo com uma maior diferença discriminativa, e como o grupo de baixo desempenho comparativamente possui um percentual maior que a dos brasileiros, isso sugere que seja um item fácil até mesmo para quem tem maior dificuldade, pouco domínio de conhecimento, ou seja, baixa proficiência.

Os valores dos coeficientes bisseriais ficaram com 0,34 e 0,69, sendo, portanto, bem mais discriminativo entre respondentes portugueses. Esses valores sugerem um perfil de composição de grupo misto entre os brasileiros, com um sutil predomínio de alunos com bom desempenho, e para os portugueses uma concentração relevante de alunos com bom desempenho.

O distrator D possui bisseriais de 0,00 e -0,34, ou seja, no contexto brasileiro essa alternativa foi assinalada igualmente entre alunos de alto e baixo desempenho, assim como para a opção A que teve -0,09. Para portugueses, a alternativa D sugere um perfil misto com leve predominância de alunos com baixo desempenho. No gráfico é possível identificar alunos brasileiros com 18 acertos sendo atraídos pelo distrator D, e entre alunos portugueses houve um decréscimo de percentual inicial representativo entre os grupos de desempenho, mais que culmina com atração até alunos com 18 valores no escore bruto.

A AGI destaca uma quantidade maior de violações para a letra A em ambos os contextos, porém não representam oscilações de grandes proporções. A curva de inclinação mostra um poder discriminativo para portugueses em cerca de 35,4° e para

brasileiros com 30,8°, sendo um baixo, porém indicando uma melhor discriminação para o contexto português.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros conforme dizem os valores desse parâmetro. As CCIs nos dois contextos sugerem uma boa discriminação conforme mostra a elevada inclinação. O acerto ao acaso (*guessing*) foi de 0,11 e 0,15, valores abaixo do esperado que seria $\frac{1}{4}$ ou 0,25, isso significa que os alunos foram pouco propensos a sortear a alternativa ou escolher aleatoriamente ("chute"), seja por fadiga, pressa em terminar ou por desconhecimento, em especial, os alunos brasileiros. Além disso, o baixo valor de *guessing* é um fator que revela uma melhor coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência (habilidades) alcançados ou escores padronizados obtidos (Z ou θ). Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos, talvez por "pegadinha". Quando isso acontece, deve-se verificar se o item possui uma tendência de indução ao erro, seja pela inclusão de variáveis desnecessárias, ou ainda, inversão de argumentação lógica no enunciado, supostamente não por falta de compreensão do respondente. Os índices desse parâmetro sugerem que não houveram de forma significativa a ocorrência desses equívocos, e isso sinaliza positivamente a qualidade do item. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros, porém dentro de um intervalo de 2 desvios-padrões (aproximadamente), enquanto para os portugueses, o gráfico sugere a contribuição de

informações para a interpretação da escala, porém com uma discriminação menor, estando mais à esquerda, reforçando a perspectiva de uma maior facilidade para responder o item. Essas elevações reforçam o poder discriminativo do item e precisão da escala, oferecendo contribuições para as informações interpretativas na escala, no caso, entre os níveis de proficiência 1 e 2.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 1 a 2, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y16, y14, y26, y24 e y9 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (1,20) e discriminação (5,48) classificando o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 56% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso foi estimado em 0,12, abaixo de um valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 12% de probabilidade de acertar sem possuir uma proficiência adequada. Não foi estimado um índice de erro ao acaso, isso representa indício de qualidade do item quanto a supostos equívocos as serem cometidos por alunos de elevada proficiência. Por fim, a elevação da CII indica uma boa discriminação, contribuindo assim para a precisão da escala fornecendo informações significativas a interpretação pedagógica da escala de proficiência em um intervalo próximo de 2 desvios-padrões entre os níveis 1 e 2, sendo considerado um *item-âncora*.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item busca aferir sobre a compreensão de equilíbrio térmico entre uma cadeira de plástico e outra de metal, independente da sensação térmica, com isso, espera-se que o aluno compreenda que ambas as cadeiras, por estarem supostamente em equilíbrio térmico, possuem as mesmas temperaturas, de acordo com a Lei zero da Termodinâmica. O distrator D se refere a opinião da personagem chamada Sandra, que diz "*Elas são mais frias porque o metal tem menos calor para perder do que o plástico*". Supostamente os alunos estão em

grande parte convencidos que o metal “possui calor” em quantidade menor a ceder que o plástico, ou seja, eles admitem que calor seja uma propriedade do corpo. O frio é uma sensação térmica associada a uma percepção de perda de calor da nossa pele, ou seja, quando encontramos em algo que sentimos frio provavelmente é porque esse corpo esteja numa temperatura inferior à nossa, essa percepção de perda também está associada ao fluxo de energia térmica, com isso, pode-se dizer que quanto maior a condutividade térmica entre os corpos, maior será também a sensação de frio. Diante disso, quando Sandra afirma que a cadeira de metal é mais fria e diz que isso ocorre porque “...o metal tem menos calor para perder do que o plástico” aumenta essa contradição. A própria contradição deve fazer sentido para os respondentes, portanto, procurando compreender isso se pode admitir ainda que este o equívoco remete a ideia de que “os corpos armazenam calor”, e quanto “menos calor armazenado”, mais frio esse corpo terá, ou seja, nesse sentido os alunos não estaria compreendendo o frio como sensação térmica, mas sim admitindo a noção de frio com temperatura, e a temperatura associada com uma quantidade de calor armazenada no corpo, ou seja, quanto mais calor o corpo tiver, mais quente ele será, e quanto menos calor o corpo tiver, mais frio ele será. Acredita-se que essa seja a concepção errônea mais provável e que justifica a forte atração do distrator para a grande parte dos respondentes, em especial, entre os alunos brasileiros.

SUGESTÕES: torna-se importante novamente reforçar uma melhor compreensão do que seja sensação térmica, e diferencia-la do conceito de temperatura ou calor, como sugerido nos itens 9 e 10. Exemplos que evidenciem um estado de equilíbrio térmico e outros que possam mostrar elementos ligados aos processos de trocas de calor podem contribuir na argumentação científica sobre o conceito de calor e as suas diferenças com temperatura.

Uma sugestão para isso seria observar o gelo em condições diferenciadas dentro de um mesmo intervalo de tempo, como por exemplo, em um congelador e depois em condições ambientais típicas ($\approx 27\text{ }^{\circ}\text{C}$), porém é necessário discutir e escolher entre diversas possibilidades.

Uma verificação da temperatura de cadeiras metálicas e de plástico (ou objetos com essas características) em condições análogas pode ser um mecanismo rico de discussão, dessa forma, contrapondo intuições baseadas na sensação térmica.

Uma crítica ao gabarito está no fato de que a alternativa correta Mário diz: “*Elas não são mais frias, elas estão à mesma temperatura*”, ou seja, o argumento de Mário faz uma associação entre sensação térmica e o valor da temperatura, sendo entretanto identificado um equívoco intrínseco, pois a sensação térmica nos metais é evidentemente de “mais frio”, porém isso não deve ser interpretado como se as cadeiras estivessem com temperatura mais baixas, portanto, é a resposta mais admissível diante das outras, porém carrega uma concepção errônea e necessita ser reformulada. Portanto, nas cadeiras de metais, de modo geral, se percebe que são mais frias devido ao coeficiente de condutividade térmica dos metais serem maiores que as do plástico, como foi comentado nos itens 9 e 10, dessa forma se torna relevante desenvolver atividades referentes a Lei de Fourier para a condução do calor.

Tabela 50: Condutividade térmica de materiais a 27 °C (300 K).

Material	Condutividade térmica $\langle K = W \cdot m^{-1} \rangle$
Alumínio	237
Cobre	401
Ferro	80,2
Ouro	317
Prata	429
Tungstênio	174
Grafite pirolítico	195 (<i>planar</i>)
Grafite pirolítico*	5,70 (<i>perpendicular</i>)*
Vidro	0,79 (<i>valor médio</i>)
Tijolo	0,6 (<i>valor médio</i>)
Madeira (pinho)	0,13 (<i>valor médio</i>)
Fibra de vidro	0,05
Epóxi	0,30 (<i>cargueada com sílica</i>)
Epóxi*	0,15 (<i>não cargueada</i>)
Espuma de poliestireno	0,03
Polipropileno	0,25
Espuma de poliuretano	0,02
Água	0,61
Ar	0,03

Fonte: “Condutividade Térmica” (n.d.).

A15: ITEM y15

15. Um grupo está a ouvir a previsão do tempo no rádio: "... à noite fará 5 °C, mais frio do que os 10 °C da noite passada".

a) Marta diz: "Isso significa que esta noite será duas vezes mais fria que a noite passada".

[A4]: Calor e temperatura são a mesma coisa

⇒ Por relacionar a temperatura com sensação térmica, no caso a sensação térmica é baseada em uma quantidade de calor recebida ou cedida.

[A5]: O calor é proporcional à temperatura

⇒ Por relacionar na mesma proporção os valores das temperaturas com a da sensação térmica (supostamente seria entender que o dobro de "frio" significa perder duas vezes mais calor que a noite anterior)

[B1]: A temperatura é a "intensidade" de calor.

⇒ Por relacionar quantitativamente a temperatura com sensação térmica, no caso a sensação térmica é baseada em uma quantidade de calor recebida ou cedida.

b) Alice diz: "Não está certo. 5 °C não é duas vezes mais frio que 10 °C".

c) Rui diz: "Está parcialmente correto, mas ela deveria dizer que 10 °C é duas vezes mais quente do que 5 °C".

[A3]: O quente e frio são diferentes, em vez de extremidades opostas de um continuum.

⇒ Por ser diferenciadas quantitativamente as sensações de frio e quente, ao invés de serem consideradas avaliações qualitativas e subjetivas dentro de uma mesma dimensão.

[A4]: Calor e temperatura são a mesma coisa.

⇒ Por relacionar a temperatura com sensação térmica, no caso a sensação térmica é baseada em uma quantidade de calor recebida ou cedida.

[A5]: O calor é proporcional à temperatura

⇒ Por relacionar na mesma proporção os valores das temperaturas com a da sensação térmica (supostamente seria entender que o duas vezes mais "quente" significa receber duas vezes mais calor)

[B1]: A temperatura é a "intensidade" de calor.

⇒ Por relacionar quantitativamente a temperatura com sensação térmica, no caso a sensação térmica é baseada em uma quantidade de calor recebida ou cedida.

d) Gabriel diz: "Está parcialmente correto, mas ela deveria ter dito que 5 °C é metade do frio de 10 °C".

[A4]: Calor e temperatura são a mesma coisa.

⇒ Por relacionar a temperatura com sensação térmica, remete, no caso, a sensação térmica como uma quantidade de calor recebida ou cedida.

[A5]: O calor é proporcional à temperatura.

⇒ Por relacionar na mesma proporção os valores das temperaturas com a da sensação térmica (supostamente seria entender que a metade do valor da temperatura é a "metade do frio" é, ou seja, é duas vezes mais quente)

[B1]: A temperatura é a "intensidade" de calor.

⇒ Por relacionar quantitativamente a temperatura com sensação térmica, no caso a sensação térmica é baseada em uma quantidade de calor recebida ou cedida.

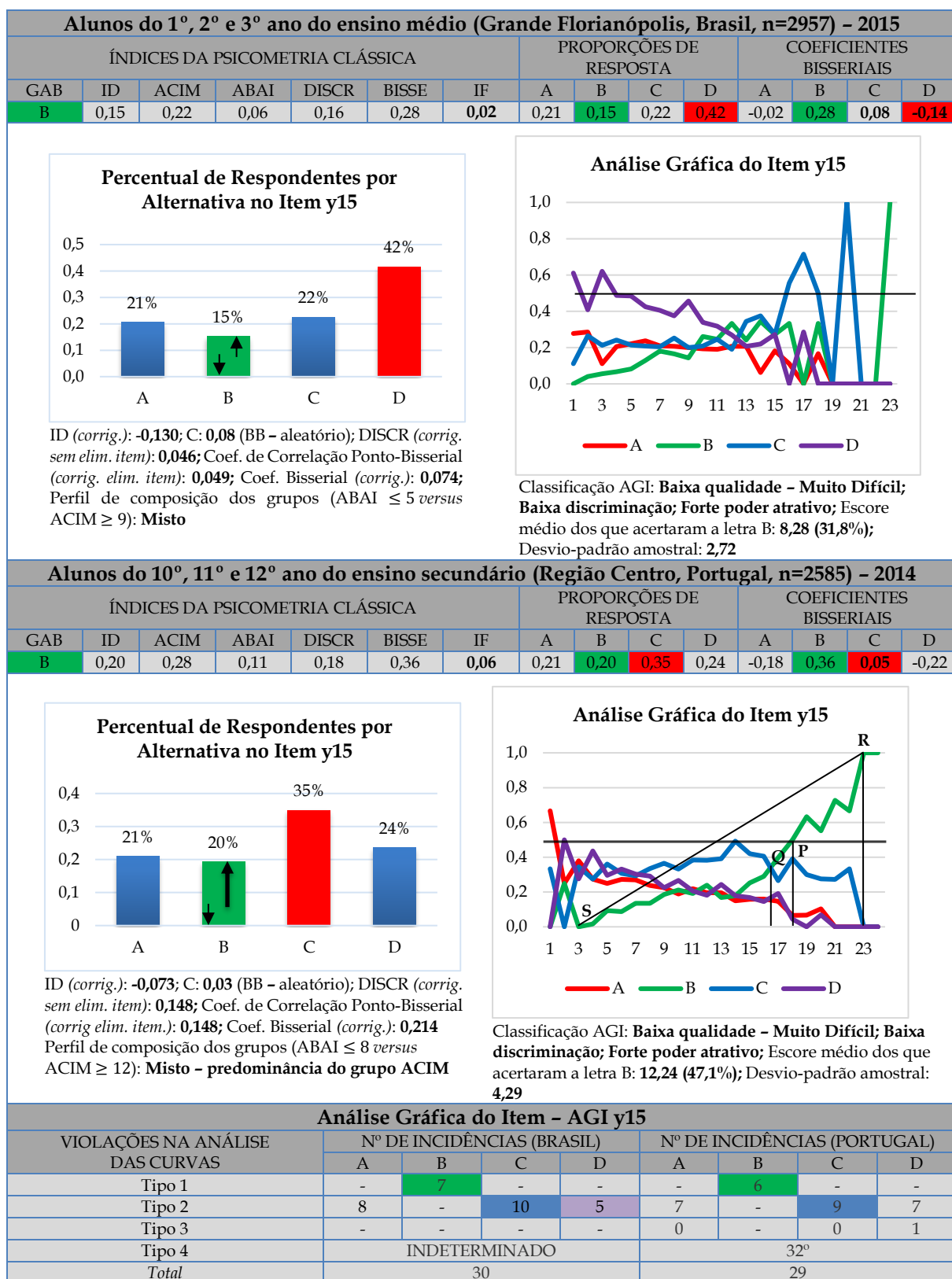
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Conhecer* a Temperatura de um Corpo ou de um Ambiente em Situações Típicas – [Descritor 4]

PROFICIÊNCIA: *Inferir* que um valor que corresponda «a metade de uma determinada temperatura não significa a metade de uma sensação térmica, pois a sensação térmica não depende da temperatura do ambiente e não se mede quantitativamente, mas sim das propriedades dos corpos em contato e avaliam essa sensação». No caso, «não faz sentido interpretar a metade ou dobro de quente ou frio para uma previsão de tempo, portanto, 5 °C não é duas vezes mais frio que 10 °C simplesmente por ter um valor de temperatura que corresponde à metade da outra quando comparada».

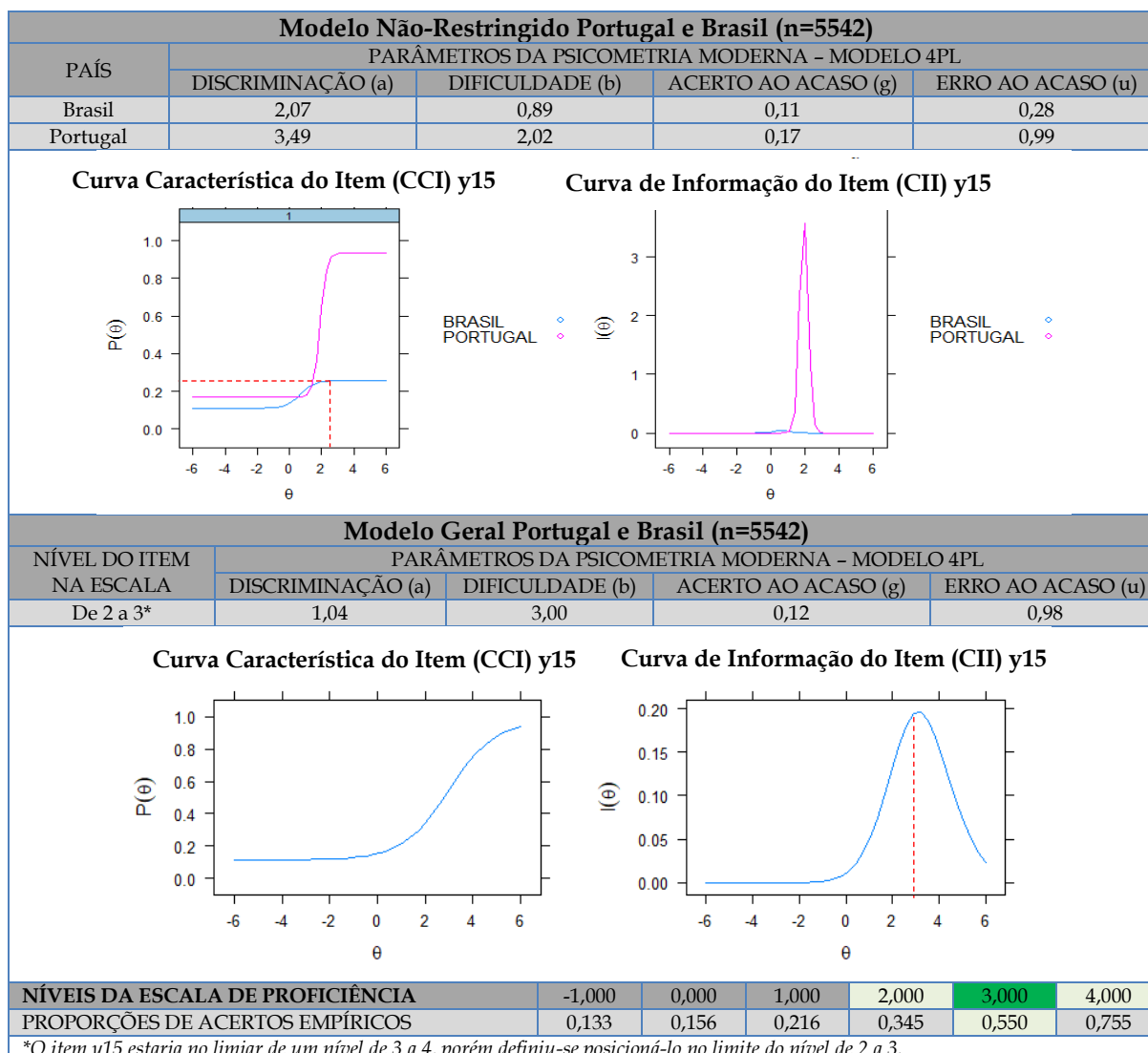
INDÍCIOS DE DEPENDÊNCIA LOCAL: Os itens y9 e y14 também buscam aferir uma compreensão de estado termodinâmico associando o conceito de temperatura de forma que independente da influência da sensação térmica. O item y10 envolve uma interpretação com base na sensação térmica.

PAINÉL DE ESTATÍSTICAS DO ITEM y15

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y15

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem uma mesma tipologia para o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **inadequado**, *sujeito a rejeição* ($DISCR < 0,20$). Esse é um forte indicativo inicial de eventuais problemas com o item.

O percentual total de acertos para o Brasil foi de apenas 15%, sendo a última opção entre as alternativas. Em Portugal também foi a última opção com apenas 20%. Os distratores mais fortes foram as alternativas C e D, sendo entre alunos brasileiros com percentuais de 22% e 42%, e para os portugueses 35% e 24%, ou seja, os grupos

foram atraídos para distratores distintos, mas no geral são os mesmos. É preciso verificar esses distratores.

Os grupos de maior e menor desempenho ficou com 22% e 6% (Brasil), e 28% e 11% (Portugal). Esses baixos percentuais entre alunos de alto e baixo desempenho sugere que seja um item difícil para todos, e a pouca diferença sendo um baixo índice de discriminação, pois o grupo intermediário é predominante nas duas amostras, e isso não é um bom sinal para a qualidade do item.

Os coeficientes bisseriais para alternativa correta foram de 0,28 e 0,36. No Brasil o desempenho dos alunos que assinalaram essa alternativa segue de modo geral com baixo percentual, e isso não pode ser admitido, apesar o bisserial sugerir um perfil de composição misto com leve predomínio de alunos mais acertados. Observa-se que para brasileiros e portugueses o distrator C também é positivo e bem próximo de zero, ou seja, sendo uma composição mista para os percentuais de 22% e 35% entre os respondentes. Todas as alternativas ficaram próximos de zero, isso demonstra uma questão com discriminação ruim entre os diferentes grupos de desempenho. De modo semelhante acontece com Portugal, também com uma discriminação baixa e bisseriais todos próximos de zero, bem distantes de 1, como deveria ser. Pode-se observar incidências de respondentes brasileiros com 20 acertos, e portugueses com 22, isso reforça a necessidade de verificar o distrator C.

Na AGI se observa que para os alunos portugueses a discriminação começa a ocorrer de forma mais intensa em torno dos grupos com 17 certos. Para ambos os contextos o maior número de violações ficou na alternativa C, sendo as oscilações proporcionais mais intensas para os brasileiros que findou por indeterminar o poder de discriminação devido a existência de múltiplos intervalos de discriminação. Alunos com bom desempenho foram atraídos nos dois contextos pelo distrator C, e isso merece ser investigado.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma

discriminação muito alta ($a > 1,70$) (Brasil); **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os portugueses conforme dizem os valores desse parâmetro. As CCIs nos dois contextos sugerem uma boa discriminação conforme mostra a elevada inclinação. O acerto ao acaso (*guessing*) foi de 0,11 e 0,17, valores abaixo do esperado que seria $\frac{1}{4}$ ou 0,25, isso significa que os alunos foram pouco propensos a sortear a alternativa ou escolher aleatoriamente ("chute"), seja por fadiga, pressa em terminar ou por desconhecimento, em especial, os alunos brasileiros. O índice de acerto "aleatório" também indica uma boa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os respectivos níveis de proficiência dos sujeitos. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 0,28 e 0,99 (Brasil e Portugal). Com isso, estima-se que os alunos brasileiros a partir de um nível de proficiência em torno de $\theta = 2,8$ tendem a uma probabilidade de 72% de cometerem esses erros, enquanto os portugueses ficaram com uma probabilidade baixíssima de 1%. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto, exceção em pequeno intervalo próximo de $\theta = 1$.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os portugueses, porém dentro de um intervalo quase 2 desvios-padrões (aproximadamente), enquanto para os brasileiros, o gráfico sugere praticamente nenhuma informação relevante. Essas elevações reforçam o poder discriminativo do item e de contribuições dos respondentes portugueses para as informações interpretativas na escala, no caso, precisamente no nível 3.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 2 a 3 (precisamente no nível 3), em que representa um patamar de competência

representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por apenas 2 itens-âncoras y8 e y19 (em ordem crescente de dificuldade) (ver escala de proficiências), sendo pouco informativo.

A CCI apresenta os parâmetros de dificuldade (3,00) e discriminação (1,04) classificando o item sendo: **Muito Difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **bom**, *sujeito a aprimoramento*, com uma *discriminação moderada* ($0,65 < a \leq 1,35$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 55% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso foi estimado em 0,12, abaixo de um valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 12% de probabilidade de acertar sem possuir uma proficiência adequada. Foi estimado um índice de erro ao acaso de 0,98, ou seja, apenas 2% de probabilidade de cometerem equívocos por descuido, isso representa um índice de boa qualidade para o item. Por fim, a elevação da CII indica uma boa discriminação, contribuindo para a precisão na interpretação do item em um intervalo próximo de 5 desvios-padrões entre os níveis 2 e 4. Ressalta-se que o item y15 não possui subsídios suficientes para ser um *item-âncora*.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item comenta sobre uma previsão de tempo e compara com o valor da temperatura da noite anterior, no caso, prevê 5 °C e diz que será mais frio que a noite passada que teve 10 °C. Espera-se que a ideia de magnitude do valor da temperatura não esteja associada com uma corresponde aferição estimada de sensação térmica. Apesar disso, a maioria concebeu que a metade do valor da temperatura prevista significa a metade do frio (letra C), ou que 10 °C corresponde ao dobro de “quentura” ou duas vezes mais quente, dessa forma isso sinaliza uma dificuldade de compreensão sobre o conceito de sensação térmica, além de equívocos com o conceito de temperatura.

SUGESTÕES: A necessidade de uma melhor compreensão sobre a sensação térmica se torna recorrente entre os itens do teste, com isso as sugestões dos itens 9, 10, 13 e 14 são em grande parte cabíveis. Apesar disso parece que seja promissor evidenciar dados empíricos relacionados a valores da temperatura, montando tabelas e elaborando gráficos, e para cada valor de temperatura serem propostas e desenvolvidas atividades na modalidade de ensino experimental envolvendo

situações que submeta o aluno a sentir essas variações e descreverem categoricamente suas percepções.

A16: ITEM y16

16. Tomás pega numa régua de metal e numa de madeira do seu estojo. Ele diz que sente a régua de metal mais fria do que a de madeira. Qual é explicação com que você concorda mais?

a) O metal retira energia da mão mais rapidamente do que a madeira.

b) A madeira é naturalmente uma substância mais quente do que o metal.

[D1]: A temperatura é uma propriedade de um material ou objeto particular. (uma característica específica para cada tipo de material).

⇒ Por considerar que a sensação de “quente” na madeira está associada a uma temperatura mais elevada comparativamente com o metal, e esta é uma propriedade do material.

[B2]: Pele ou toque pode determinar a temperatura.

⇒ Por fazer referência especulativa na determinação da temperatura com base na sensação térmica, no caso, que a madeira possui uma temperatura maior do que o metal.

c) A régua de madeira contém mais calor do que a régua de metal.

[A1] O calor é uma substância (*concepção substancialista*).

⇒ Por conter calor, considerar uma propriedade do corpo.

[D1]: A temperatura é uma propriedade de um material ou objeto particular. (uma característica específica para cada tipo de material).

⇒ Por considerar que a ideia da madeira conter mais calor que o metal, está associada a uma temperatura mais elevada comparativamente com o metal.

[D2]: Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não um energia transitória. O frio também uma substância).

⇒ Por considerar que a possibilidade da régua de metal conter ou manter calor assim como a de madeira.

d) Os metais são melhores irradiadores de calor do que a madeira.

[D2]: Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não um energia transitória. O frio também uma substância).

⇒ Por considerar que o metal possua a capacidade de intensificar o calor, conseqüentemente, irradiar.

e) O frio flui mais facilmente de um metal.

[A1] O calor é uma substância (*concepção substancialista*) –

⇒ Por considerar o frio fluindo como uma substância, remete também a uma noção substancialista de calor.

[C4]: Calor e frio fluem como líquidos (*concepção substancialista*).

⇒ Por considerar o frio fluindo.

[D2]: Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não um energia transitória. O frio também uma substância).

⇒ Por considerar que o metal possui a capacidade de atrair e intensificar o frio, e com isso, acreditar que o frio possa fluir facilmente a partir dele.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão* sobre a Propagação do Calor - [Descritor 11]

PROFICIÊNCIA*: *Inferir* sobre a «sensação térmica entre corpos com diferentes propriedades térmicas estando na mesma temperatura», no caso, «entrando em contato físico com uma régua de metal e uma de madeira, considera-se que a temperatura independe da sensação térmica»

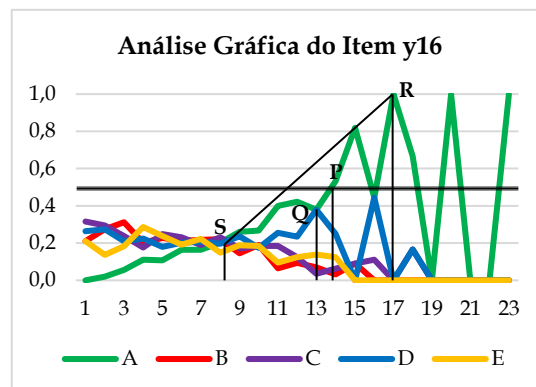
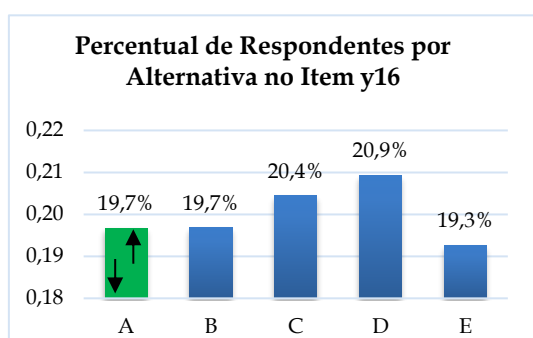
INDÍCIOS DE DEPENDÊNCIA LOCAL: Os itens y9, y10, e y14 também envolvem em parte a necessidade de uma compreensão da influência da sensação térmica com base nas diferentes propriedades térmicas dos materiais. O item y15 também remete a necessidade de diferenciar a sensação térmica com o conceito de temperatura.

**Compreende-se que devido ao elevado coeficiente de condutividade térmica dos metais, a régua de metal retira energia térmica da mão mais rapidamente do que a régua de madeira, e com isso se tenha uma percepção térmica diferenciada comparativamente.*

PAINÉL DE ESTATÍSTICAS DO ITEM y16

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

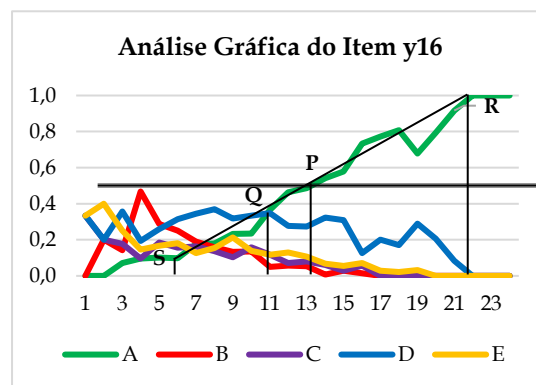
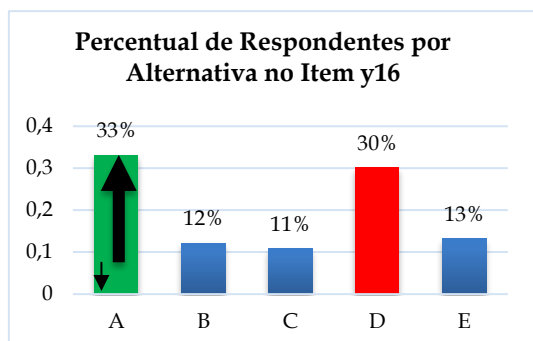
Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015																
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA					COEFICIENTES BISSERIAIS				
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	E	A	B	C	D	E
A	0,20	0,33	0,09	0,24	0,39	0,05	0,20	0,20	0,20	0,21	0,19	0,39	-0,14	-0,09	0,03	-0,11



ID (corrige.): -0,004; C: 0,0004 (BB - aleatório); DISCR (corrige. sem elim. item): 0,121; Coef. de Correlação Ponto-Bisserial (corrige. elim. item): 0,131; Coef. Bisserial (corrige.): 0,187; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): **Misto**

Classificação AGI: **Baixa qualidade - Muito difícil; Média discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra A: 8,63 (33,2%); Desvio-padrão amostral: 3,03

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014																
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA					COEFICIENTES BISSERIAIS				
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	E	A	B	C	D	E
A	0,33	0,59	0,14	0,45	0,57	0,16	0,33	0,12	0,11	0,30	0,13	0,57	-0,41	-0,24	-0,08	-0,21



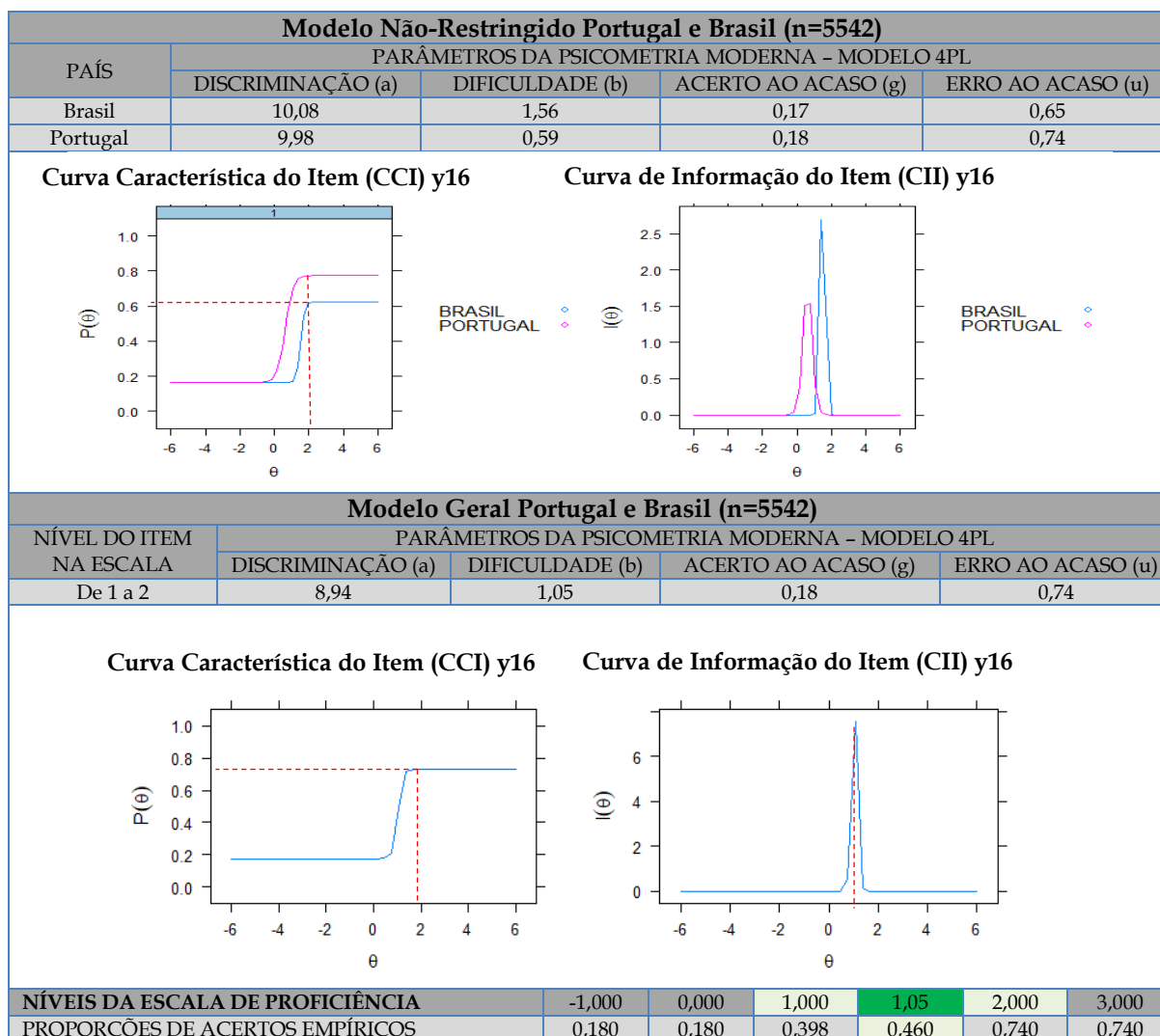
ID (corrige.): 0,167; C: 0,09 (BB - aleatório); DISCR (corrige. sem elim. item): 0,331; Coef. de Correlação Ponto-Bisserial (corrige. elim. item): 0,329; Coef. Bisserial (corrige.): 0,427; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): **Predominância do grupo ACIM**

Classificação AGI: **Boa qualidade - Difícil; Alta discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra A: 12,68 (48,8%); Desvio-padrão amostral: 3,94

Análise Gráfica do Item - AGI y16

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)					Nº DE INCIDÊNCIAS (PORTUGAL)				
	A	B	C	D	E	A	B	C	D	E
Tipo 1	6	-	-	-	-	2	-	-	-	-
Tipo 2	-	8	7	9	6	-	5	5	10	8
Tipo 3	-	0	0	1	0	-	0	0	0	0
Tipo 4	42º					29,4º				
Total	37					30				

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y16

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **marginal, sujeito a reelaboração** ($0,20 \leq DISCR < 0,30$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **adequado, disponível a reaplicação** ($DISCR \geq 0,4$) (Portugal).

A porcentagem de acerto ficou com 19,7% e 33%, sendo entre os alunos brasileiros o percentual mais baixo e praticamente igual entre todas as respostas, e isso sinaliza um item confuso. Para os portugueses houve um predomínio relativo da

alternativa correta, porém com uma forte atração da alternativa D com 30%, isso demonstra uma necessidade de verificar esse distrator.

Os grupos de maior e menor desempenho ficou com 33% e 9% (Brasil), e 59% e 14% (Portugal). Os alunos de bom desempenho possuem maior percentual de acerto nos dois contextos, como era de se esperar, entretanto, em Portugal esse percentual é mais elevado e acima de 50%, com predomínio absoluto do total de respondentes, logo com uma maior diferença discriminativa. Os grupos de baixo desempenho comparativamente possuem um percentual de acertos bem próximos e em torno de 10%, isso sugere que seja um item difícil.

Os bisseriais foram 0,39 e 0,57, ou seja, para aqueles que acertaram o item se tem um perfil de composição geral formado predominantemente pelo grupo de alto desempenho dentro de cada contexto, sendo um pouco mais acentuado entre os portugueses e para os brasileiros um pouco mais misto. Os bisseriais dos distratores C e D para os brasileiros ficou com -0,09 e 0,03, ou seja, muito próximo de zero resultando uma composição mista e bem equilibrado, tanto alunos com baixo e alto desempenho assinalando essa opção dentro do percentual de 20% para cada uma das alternativas. Para os portugueses o distrator D possui bisserial igual a -0,24, sendo também mista, com leve predomínio de alunos com baixo desempenho, mas não tão representativo, pois se encontra longe do -1 que seria o ideal. Observa-se alunos brasileiros com 18 acertos sendo atraídos, e portugueses com 21 acertos.

A AGI revela um maior número de violações de modo geral foi para o Brasil, em especial, no aumento percentual da alternativa correta para os brasileiros de bom desempenho, com destaque para os alunos de 16 acertos, na qual houve um decréscimo percentual elevado devido a atração do distrator D. O distrator D teve o maior número de oscilações para as duas amostras ao invés de um decréscimo mais acentuado, como deveria ser, isso demonstra incertezas e dúvidas para essa alternativa até entre alunos de bom desempenho.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de

respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma maior de dificuldade (mais à direita) para os alunos brasileiros. As CCIs nos dois contextos sugerem uma boa discriminação conforme mostra a elevada inclinação. O ajuste do modelo estimou os valores de 0,17 e 0,18 para o índice de acerto ao acaso (“*guessing*”) em cada contexto (Brasil e Portugal), sendo valores próximos ao esperado que seria de 1/5 ou 0,20 (por ser 5 alternativas), isso indica uma boa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência alcançados dos sujeitos, em especial, suas habilidades θ . Os índices de erro ao acaso (“*upper*”) foram de 0,65 e 0,74, na qual remetem a 35% e 26% de probabilidade de assinalar um distrator por descuido entre os alunos de elevada proficiência, sendo o menor valor para os alunos portugueses. Estima-se que essas probabilidades de erro ao acaso são a partir de uma proficiência igual a $\theta = 2,3$ (aproximadamente) em ambos os contextos, ou seja, alunos com proficiência $\theta > 2,3$ são os que possuem maiores tendências de cometer esses equívocos. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto.

A elevação na *Curva de Informação do Item* (CII) para cada contexto indica uma boa informação do item para ambos os respondentes, contribuindo para a precisão da escala, na qual está mais bem representada pelos alunos brasileiros em um intervalo de 1 desvio-padrão aproximadamente.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 1 a 2, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y16, y14, y26, y24 e y9 (em ordem crescente de dificuldade).

A CCI representa uma sigmoide bem inclinada e os parâmetros de dificuldade (1,05) e discriminação (8,94) classificam o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 46% entre aqueles que alcançam esse nível de habilidade θ . Foi estimado acertos ao acaso com probabilidade de 0,18, bem próximos ao valor esperado que seria de 1/5 ou 0,20. O índice de erros ao acaso foi de 0,74, isso remete a uma probabilidade de 26% de alunos com proficiência $\theta > 1,7$ cometerem equívocos por descuido, ainda que suas proficiências se elevem. Por fim, a CII indica uma boa discriminação na região em torno de $\theta = 1$ dentro de um intervalo de aproximadamente 1 desvio-padrão, com isso, o y16 apresenta informações relevantes para a interpretação e precisão da escala entre os níveis 1 e 2, sendo considerado um item-âncora.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item busca aferir sobre o porquê de um corpo de metal ser aparentemente mais frio que um de madeira, enquanto na verdade estão em uma mesma temperatura, ou seja, visa uma compreensão sobre a diferença entre sensação térmica e conceito de equilíbrio térmico, assim como de suas propriedades que possa estar influenciando esse tipo de percepção. O distrator D diz que a explicação para isso é que “os metais são melhores irradiadores de calor do que a madeira”. Acredita-se que a ideia de “irradiar” o calor esteja sendo compreendido como maior facilitador na condução do calor, ou seja, sendo um melhor “condutor”, que de fato não estaria errado. Apesar disso, irradiar e conduzir são processos distintos e não seria o argumento para justificar a ilusão da sensação térmica de mais frio. A explicação estaria relacionada na propriedade térmica do metal em perder e ganhar calor, em outras palavras, na elevada condutividade térmica dos metais, conforme é mostrada nos comentários pedagógicos do item 14.

SUGESTÕES: O equívoco em equiparar uma noção entre os processos de irradiar e conduzir aparentemente ser um fator a influenciar negativamente os alunos, ou atraí-los ao distrator D. Torna-se relevante destacar a diferença entre esses processos incluindo os de convecção. Os itens com indícios de dependência local que

sugerem envolver uma aferição sobre a compreensão diante do mesmo conhecimento apresentam algumas sugestões cabíveis. O contexto histórico e epistemológico é um “ingrediente” imprescindível numa sequenciação das propostas de atividades ou até mesmo concentrada dependendo do objetivo do ensino, pois favorecem uma melhor apropriação dos conceitos, de um panorama de como foram desenvolvidos e também para o conhecimento factual de datas e cientistas. A declaração dos objetivos é uma etapa fundamental que deve ser bem analisada, planejada e desenvolvida nas atividades. Visitas a museus, fábricas e centros de divulgação da ciência também são modalidades em potencial que podem ser desenvolvidas.

O distrator C se refere ao calor como propriedade de um corpo e demonstrou se atrativo aos alunos brasileiros, e isso se torna um erro relevante a ser considerado para os professores em seu planejamento de aula.

Como o fenômeno envolve processo de transmissão do calor, tem-se a: radiação, a convecção e a condução. A radiação térmica pode ser explorada por associando o conhecimento de ondas eletromagnéticas, possibilitando assim uma intradisciplinaridade. A Lei de Stefan-Boltzmann é uma outra vertente que pode ser explorada no campo intradisciplinar para explorar a radiação e confrontar com os outros processos de transmissão do calor. A convecção tem-se atividades que envolvem o aquecimento da água, que pode ser explorada utilizando objetos leves para melhor visualizar as correntes de convecção, e um contexto de vida real seria explorar o funcionamento de fornos elétricos, frigoríficos, interior de geladeiras, salas climatizadas, fluxos do movimento do ar nas proximidades de lagos e praias, etc. A dilatação térmica é também um fenômeno que explora a condução do calor e pode ser associada facilmente diante da compreensão sobre o conceito de condutividade térmica, que talvez seja de imediata a mais apropriada para ressaltar a explicação do fenômeno que envolve o item 16, na qual a resposta correta é a letra A que diz: “o metal retira energia da mão mais rapidamente do que a madeira”, que também pode ser promissores para confrontar a noção equivocada de calor como propriedade térmica.

A17: ITEM y17

17. Alzira pegou em duas garrafas de vidro contendo água a 20 °C e embrulhou-as em toalhas. Uma das toalhas estava molhada e a outra estava seca. Vinte minutos mais tarde, ela mediu a temperatura da água em cada garrafa. A água na garrafa embrulhada na toalha molhada estava a 18 °C e a água na garrafa com a toalha seca estava a 22 °C. A temperatura ambiente mais provável da sala durante esse tempo era:

a) 26 °C

b) 21 °C

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por acreditar que a lâ seca não entra em equilíbrio com o ambiente a 21 °C. A outra água está a 18 °C devido ao pano molhado lhe ter esfriado.

[D9]: Materiais como a lâ têm a capacidade de aquecer as coisas (concepção animista, pois entende-se que a lâ seja semelhante a um mamífero, sendo uma fonte térmica natural).

⇒ Por acreditar que a temperatura da sala seja 21 °C e água somente está a 22 °C devido a lâ lhe ter aquecido.

c) 20 °C

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por acreditar que a lâ seca não permanece em equilíbrio com o ambiente a 20 °C. A outra água está a 18 °C devido ao pano molhado lhe ter esfriado.

[D9]: Materiais como a lâ têm a capacidade de aquecer as coisas (concepção animista, pois entende-se que a lâ seja semelhante a um mamífero, sendo uma fonte térmica natural).

⇒ Por acreditar que a temperatura da sala seja 20 °C e água somente está a 22 °C devido a lâ lhe ter aquecido.

d) 18 °C

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por acreditar que a lâ seca não tende ao equilíbrio com o ambiente a 18 °C. A outra água está a 18 °C devido ao pano molhado lhe ter esfriado e buscado o equilíbrio partindo de 20 °C

[D9]: Materiais como a lâ têm a capacidade de aquecer as coisas (concepção animista, pois entende-se que a lâ seja semelhante a um mamífero, sendo uma fonte térmica natural).

⇒ Por acreditar que a temperatura da sala seja 18 °C e água somente está a 22 °C devido a lâ lhe ter aquecido.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão do Comportamento Térmico de acordo com as Propriedades Térmicas* – [Descritor 18]

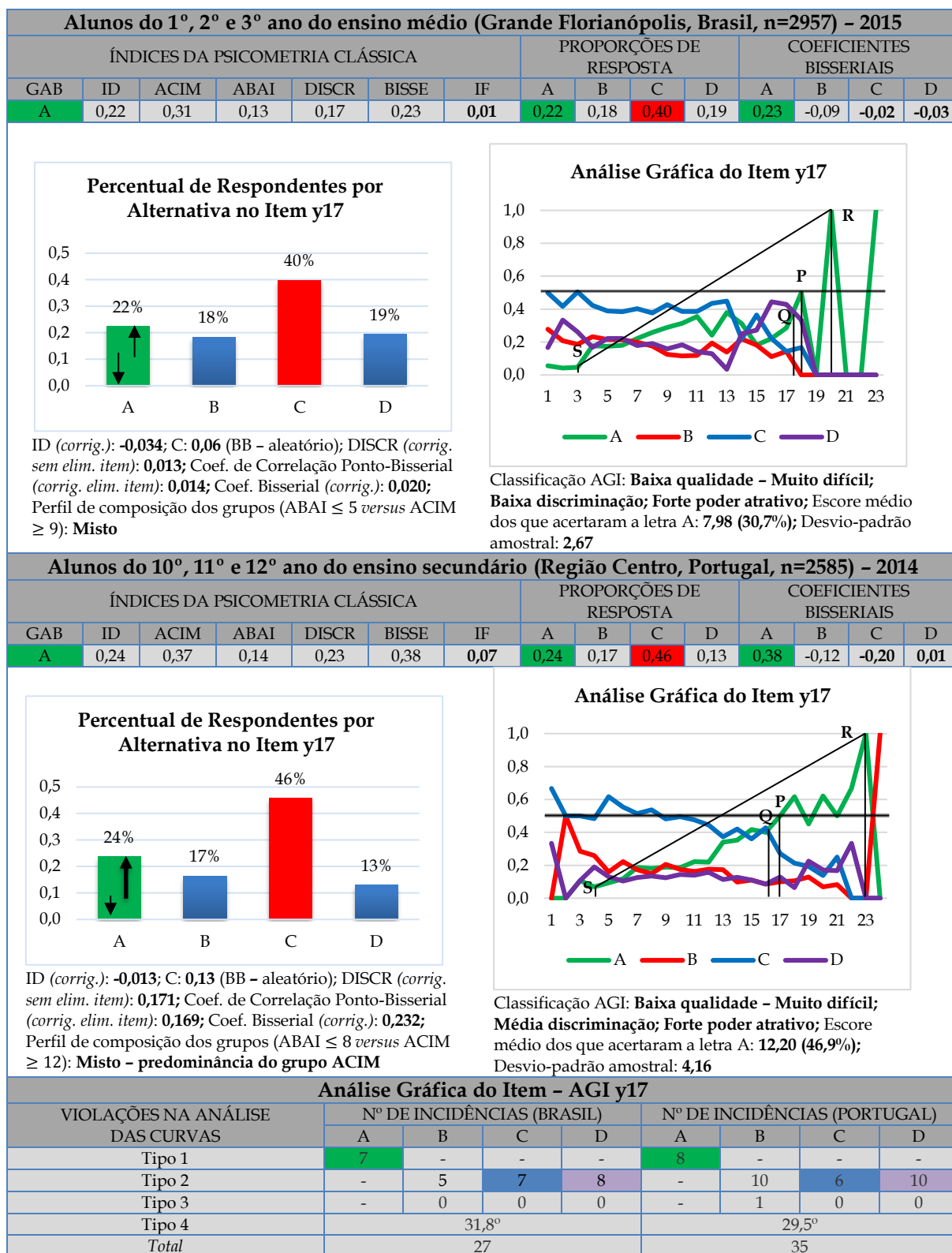
PROFICIÊNCIA*: *Inferir sobre a «temperatura ambiente com base na comparação de equilíbrio térmico entre corpos com diferentes propriedades térmicas», no caso, «duas garrafas contendo água a 20 °C embrulhadas em toalhas diferentes, uma seca* e outra molhada, após certo tempo adquirem temperaturas iguais a 18 °C e 22 °C, logo a temperatura mais provável do ambiente é de 26 °C». A toalha molhada estava com baixa temperatura e diminuiu a temperatura da água dentro da garrafa para 18 °C, porém como a garrafa com toalha seca alcançou 22 °C, a temperatura ambiente só pode ser maior ou igual a 22 °C, ou seja, por indução se pode afirmar que seja provavelmente 26 °C a temperatura da sala, pois é a única alternativa dentro dessas condições.*

INDÍCIOS DE DEPENDÊNCIA LOCAL: Não foi identificado entre os itens anteriores.

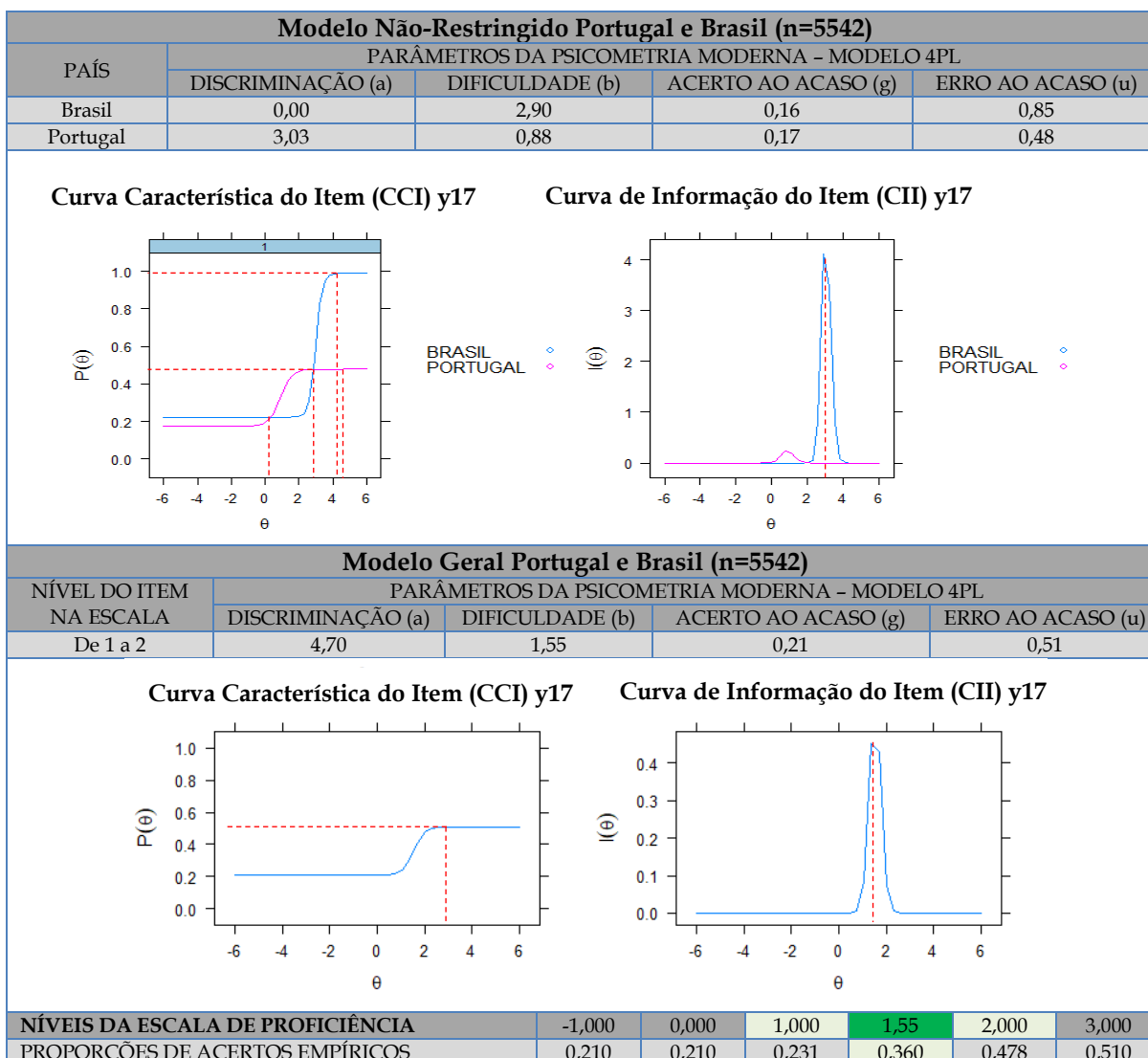
**O desconhecimento da temperatura da toalha seca poderia induzir a possibilidade de uma temperatura de equilíbrio em 21 °C, no caso, compreendendo que ela resfrie até 21 °C e o sistema garrafa-água se eleve até os 21 °C, e no caso a temperatura ambiente seria 21 °C. Mas como os objetos são idênticos, eles devem tanto ganhar como perder a mesma quantidade de energia térmica para um mesmo intervalo de tempo, dessa forma, se uma das garrafas diminuiu 2 °C e a outra elevou 2 °C, não há hipótese que o ambiente esteja a 21 °C nestas condições (o sistema garrada-água poderia se elevar até se equilibrar com a toalha de depois regredir até 21 °C), portanto, temperatura do ambiente só poderá ser igual ou superior a 22 °C, e por indução, a única resposta que obedeça esta condição seriam de 26 °C.*

PAINÉL DE ESTATÍSTICAS DO ITEM y17

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y17

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **inadequado, sujeito a rejeição** ($DISCR < 0,20$) (Brasil); **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **marginal, sujeito a reelaboração** ($0,20 \leq DISCR < 0,30$) (Portugal).

O percentual de acerto total observado foi de 22% e 24%, sendo nos dois contextos a segunda opção, perdendo para o distrator C que possui 40% e 46%, mais atrativo para os portugueses, com isso, deve ser investigado eventuais as causas desse evento.

Os grupos de desempenho ficaram com um percentual de 31% e 13% (Brasil), e 37% e 14% (Portugal), com um comportamento aparentemente equivalente entre os dois contextos. Os grupos de mais elevado desempenho correspondem a percentuais de acerto abaixo de 50%, isso sinaliza problemas com o item, ou talvez fortes dificuldades entre os respondentes.

Os coeficientes bisseriais foram 0,23 e 0,38, sugerem um perfil de composição misto entre os grupos de alto e baixo desempenho, com ele predominância de respondentes com bons desempenhos que assinalaram corretamente. O distrator C possui bisseriais -0,02 e -0,20, muito próximo a zero, ou seja, alunos com bom desempenho se sentem atraídos por esse distrator e esses valores sugerem composição mista entre os grupos de desempenho, e isso não pode ser tolerado. O gráfico mostra que o distrator D também atraiu alunos de bom desempenho que possuem 18 e 22 acertos dentro de cada contexto, e também precisa ser verificado.

A AGI mostra 7 e 8 violações para os grupos de acerto no conjunto todos os grupos de desempenho, porém os percentuais de acerto dentro de cada grupo são maiores para os portugueses. Apesar do percentual baixo, o distrator D apresenta maiores oscilações para ambas as amostras, sendo C também em destaque nessa condição para os portugueses. O intervalo de discriminação (de Q a R) se apresenta baixo e há uma violação de interseção com o distrator B no desempenho 24, porém é um nível com poucos indivíduos portugueses. O poder de discriminação revela uma inclinação baixa em cerca de 30°.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **inadequado**, *sem nenhuma discriminação, sujeito a rejeição* ($a = 0,00$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, *com evidências de ser reaplicado com uma discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros conforme dizem os valores desse parâmetro. As CCIs nos

dois contextos sugerem uma boa discriminação conforme mostra a elevada inclinação. O acerto ao acaso (*guessing*) foi de 0,16 e 0,17, próximo de um valor esperado de $\frac{1}{4}$ ou 0,25, sugerindo que os alunos foram menos propensos a “chutes”, seja por fadiga, pressa em terminar ou por desconhecimento, entre outros. Isso indica coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades e níveis de proficiência alcançadas. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 0,85 e 0,48 (Brasil e Portugal), com isso se estima 15% de probabilidade de alunos brasileiros cometerem equívocos com proficiência $\theta > 4,3$, e 52% para alunos portugueses a partir de um nível de proficiência $\theta > 4,7$. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os brasileiros são os que possuem uma maior probabilidade de acerto, com exceção no intervalo $0,40 < \theta < 3,00$ (aproximadamente).

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros e uma maior informação, porém dentro de um intervalo de 1 desvio-padrão (aproximadamente), enquanto para os portugueses o gráfico sugere uma baixa discriminação e pouca informação. O item y17 não atende aos critérios para ser considerado um item-âncora.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 1 a 2, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y16, y14, y26, y24 e y9 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (1,55) e discriminação (4,70) classificando o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 36% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso foi

estimado em 0,21, dentro do esperado que seria de $\frac{1}{4}$ ou 0,25, representando 21% de probabilidade de acertar sem possuir uma proficiência adequada. Foi estimado um índice de erro ao acaso de 0,51, ou seja, 49% de cometerem equívocos por descuido para além de um nível de proficiências igual a $\theta = 2,5$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação para o item y17 em um intervalo próximo de 1,5 desvio-padrão entre os níveis 1 e 2, entretanto, o item não apresenta subsídios necessários para ser considerado um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item exige uma análise comparativa sobre duas situações parecidas que envolve trocas de calor e compreensão sobre equilíbrio térmico entre uma toalha seca e molhada envolvida em garrafas com água inicialmente a mesma temperatura (20 °C), e espera-se com base nas respostas disponíveis concluir por indução que a temperatura ambiente seja acima de 21 °C, portanto, a letra A que afirma ser 26 °C. Como não é fornecida as informações sobre a temperatura inicial das toalhas, a análise se baseia em especulações da temperatura final após certo intervalo de tempo na qual uma esfria a água (18 °C) e outra esquenta (22 °C). O distrator C é o mais atrativo e afirma que a temperatura ambiente seria 20 °C. Um dos motivos para isso talvez seja que os respondentes imaginem que a temperatura inicial da água nas garrafas já corresponderia a temperatura ambiente. Outra possibilidade seria imaginar que a temperatura ambiente seria a média entre as temperaturas finais das duas garrafas no intervalo considerado.

SUGESTÕES: A compreensão deste problema remete a uma interpretação sobre as trocas de calor que ocorre em cada garrafa, supondo a diferença da influência de uma toalha molhada e uma seca, entretanto, o mais relevante é o fato das temperaturas finais sugerirem a processos opostos, um resfria e outro aquece. A indução é que as alternativas determinam a tomada de decisão, devido ao desconhecimento das temperaturas iniciais das toalhas e isso pode ser explorado e melhor compreendido por meio da elaboração de curvas de aquecimento e resfriamento para cada situação, e confrontado com as alternativas disponíveis.

A18: ITEM y18

18. Daniel pega simultaneamente em duas embalagens de leite achocolatado, uma fria, retirada da geladeira e outra morna, que tinha sido deixada sobre a mesa por algum tempo. Por que razão a embalagem retirada da geladeira parece mais fria do que a da mesa? Comparando com a embalagem morna, a embalagem fria:

a) Contém mais frio.

[A3]: O quente e frio são diferentes, em vez de extremidades opostas de um continuum (as sensações térmicas são quantificáveis ou com caráter substancialista, ao invés de uma avaliação qualitativa e subjetiva).

⇒ Por acreditar que o frio seja considerado como uma substância ou propriedade de um corpo que pode ser armazenada, ao invés de uma sensação térmica a ser avaliada qualitativamente em um continuum.

[B3]: Percepções de quente e frio não estão relacionados com a transferência de energia.

⇒ Por considerar que ele “contenha” mais frio (ideia equivocada de calor) ao invés de temperatura mais baixa, acredita-se que isso indique uma incapacidade de transferência térmica.

b) Contém menos calor.

[A3]: O quente e frio são diferentes, em vez de extremidades opostas de um continuum (as sensações térmicas são quantificáveis ou com caráter substancialista, ao invés de uma avaliação qualitativa e subjetiva).

⇒ Por acreditar que o frio seja associado a uma baixa quantidade de calor armazenado, ou seja, sendo uma substância ou propriedade de um corpo ao invés de uma sensação térmica a ser avaliada qualitativamente em um continuum.

[B3]: Percepções de quente e frio não estão relacionados com a transferência de energia.

⇒ Por considerar que ele “contenha” mais frio (ideia equivocada de calor) ao invés de temperatura mais baixa, acredita-se que isso indique uma incapacidade de transferência térmica.

c) É pior condutora de calor.

[C9]: A teoria cinética na verdade não explica a transferência de energia térmica. (As explicações são relatadas, mas não acreditadas pelo sujeito, ou seja, não se percebe que as diversas formas de energia que se propaga entre as partículas representam a energia térmica em trânsito).

⇒ Por ter uma noção de que o corpo possui uma capacidade de transferir (receber ou ceder) maior quantidade de energia térmica, porém, não demonstra uma compreensão sobre o fenômeno microscópico que explica a forma como ocorre a transferência. No caso, considera-se que os corpos frios (temperatura baixa) não conduzem muito bem o calor, e corpos mais quentes (temperatura elevada) conduzem melhor o calor (equívoco na relação entre sensação térmica e condutividade térmica).

d) Retira calor mais rapidamente da mão de Daniel.

e) Cede calor mais rapidamente para a mão de Daniel.

[C9]: A teoria cinética na verdade não explica a transferência de energia térmica. (As explicações são relatadas, mas não acreditadas pelo sujeito, ou seja, não se percebe que as

diversas formas de energia que se propaga entre as partículas representam a energia térmica em trânsito).

⇒ Por ter uma noção de que o corpo possui uma capacidade de transferir (receber ou ceder) maior quantidade de energia térmica, porém, não demonstra uma compreensão sobre o fenômeno microscópico que explica a forma como ocorre a transferência. No caso, considera-se que o corpo está mais frio (temperatura baixa) devido a receber calor mais rapidamente para as mãos do que se estivesse morno, ou seja, quanto mais calor você recebe por unidade de tempo, maior será a sensação de frio ao invés de quente (equivoco na relação entre sensação térmica e condutividade térmica)

Não se percebeu entre as alternativas que:

[A4]: Calor e temperatura são a mesma coisa.

⇒ Por não associar a sensação térmica ou temperatura com o calor, mas com relação a condutividade térmica.

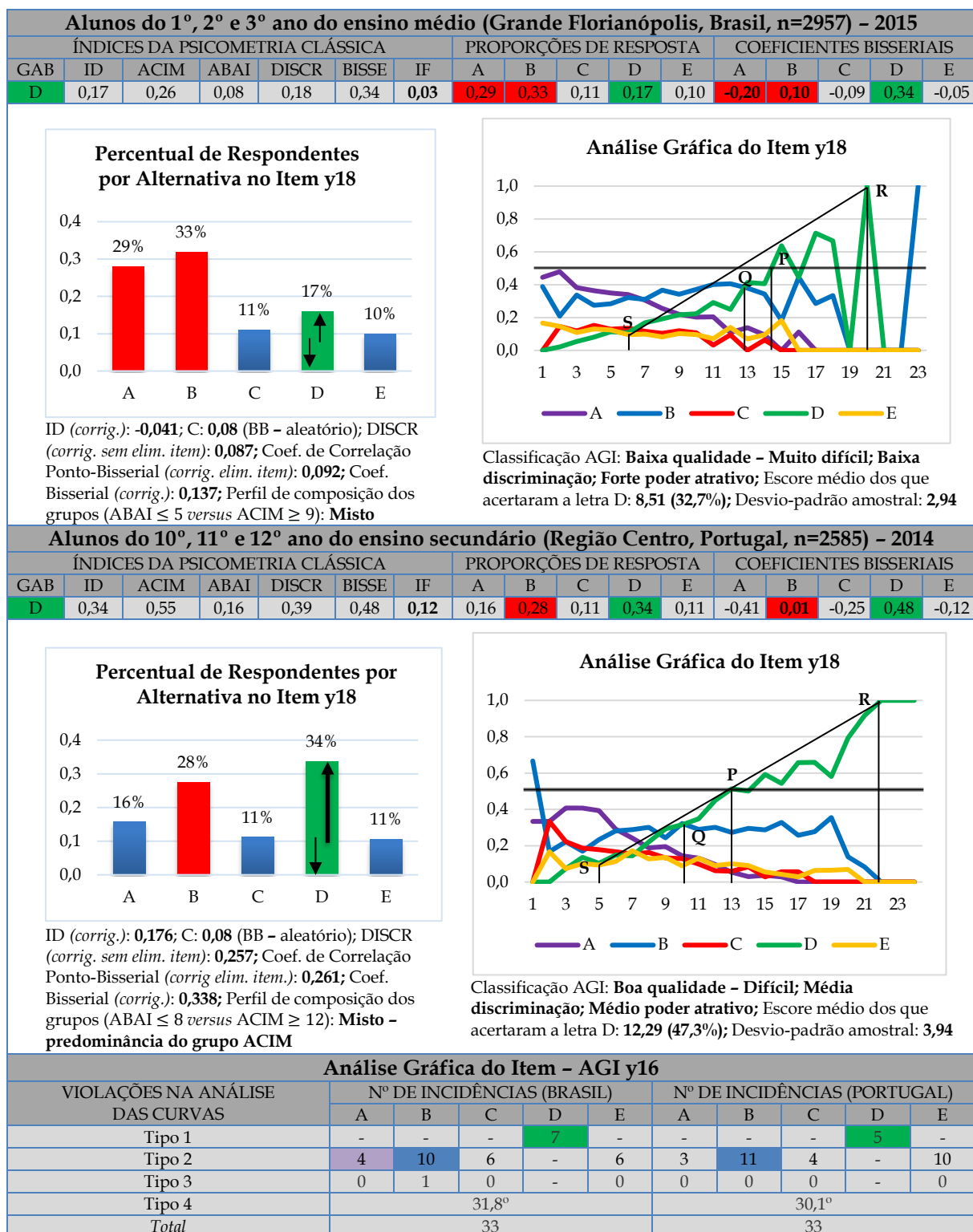
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão sobre a Sensação Térmica explicada pela Propagação do Calor - [Descritor 14]*

PROFICIÊNCIA: *Inferir por meio da «sensação térmica o porquê de um corpo parecer mais frio que outro idêntico e de mesmo material, porém com temperatura mais elevada», no caso, «duas embalagens de leite, sendo um retirado da geladeira e outra com um certo tempo deixado sobre a mesa, ao toca-los, se percebe mais frio a caixa de menor temperatura devido a uma maior diferença térmica com mão de Daniel, com isso se consegue retirar mais rapidamente energia térmica de sua mão».*

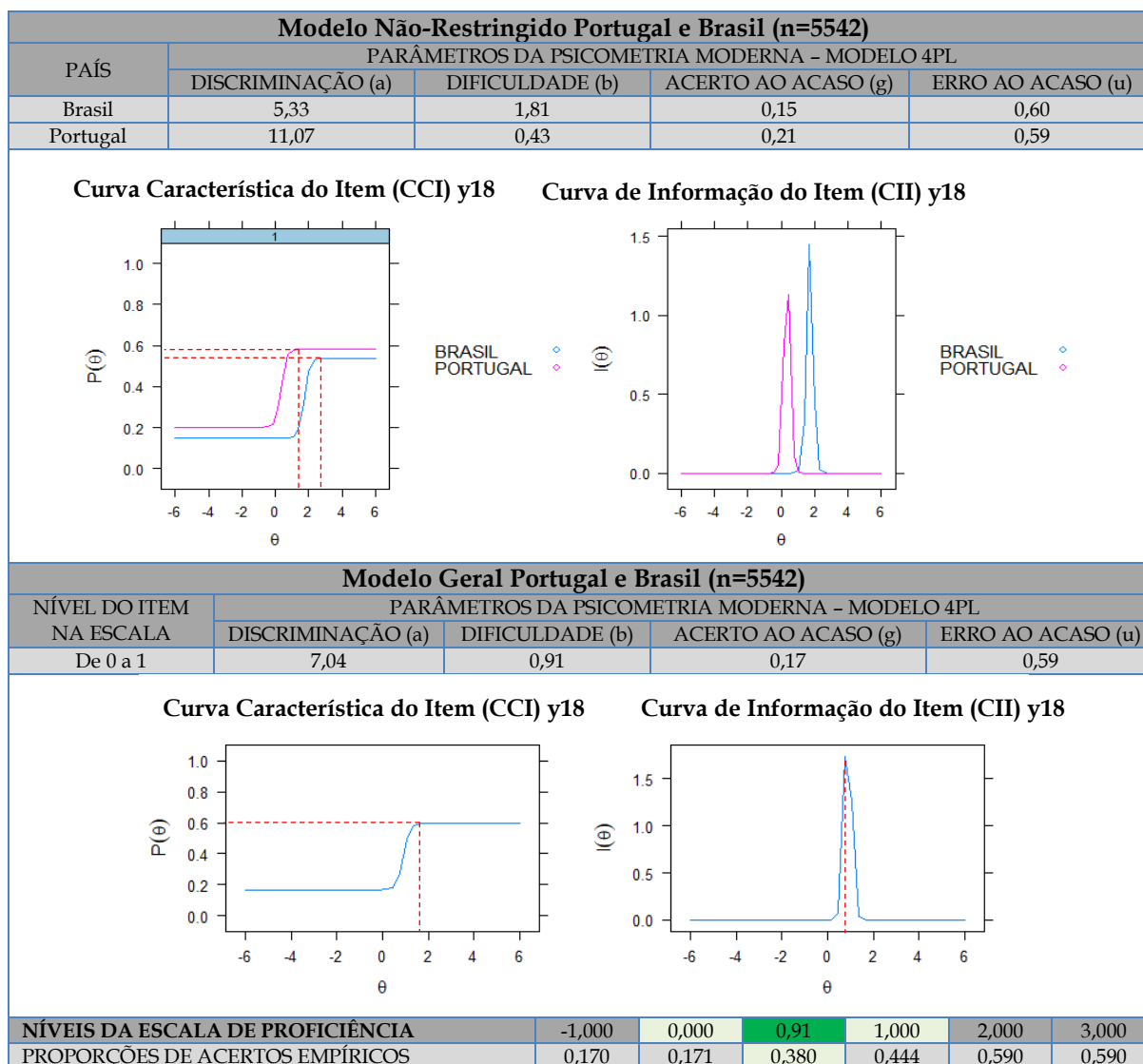
INDÍCIOS DE DEPENDÊNCIA LOCAL: Não foi identificado entre os itens anteriores.

PAINÉL DE ESTATÍSTICAS DO ITEM y18

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y18

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **inadequado, sujeito a rejeição** ($DISCR < 0,20$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **bom, sujeito a aprimoramento** ($0,30 \leq DISCR < 0,40$) (Portugal).

Os percentuais de acerto total foram de 16% e 34%, sendo a menor porcentagem para os brasileiros considerando a resposta correta a 3ª opção. De modo geral não houve predomínio absoluto (>50%) em ambos os contextos. Chama atenção a forte

atração do distrator B com 32% e 28%, isso remete a verificar eventuais causas para esse tipo de efeito. O distrator A também foi relevante para os brasileiros com 28%.

Os grupos de maior e menor desempenho que assinalaram corretamente ficaram com os valores percentuais de 26% e 8% (Brasil), e 55% e 16% (Portugal). Para os alunos brasileiros demonstra ser uma baixa média percentual entre os de bom desempenho, e para os portugueses há um predomínio de acertos com um percentual acima dos 50%, além de indicar uma maior discriminação pela diferença percentual entre esses grupos em seu contexto, em outras palavras, se pode afirmar que indica uma maior dificuldade para os alunos brasileiros comparativamente.

Os valores dos bisseriais foram de 0,34 e 0,48, ou seja, mais discriminativo para os portugueses, como indica as diferenças entre os grupos de maior e menor desempenho. Também indica um perfil de composição misto, ou seja, formado por alunos de alto e baixo desempenho, porém com uma predominância maior de alunos de bom desempenho na composição de grupo dos portugueses. O distrator B possui um bisserial de 0,10 e 0,01, ou seja, sugere um perfil misto composto de forma praticamente igual entre alunos de alto e baixo desempenho, e isso não é bom. Observa-se alunos brasileiros e portugueses com 23 e 21 acertos sendo atraídos por esse distrator.

A AGI revela maiores oscilações entre os percentuais de acerto para os brasileiros, com destaque na “queda” no grupo de 16 acertos, enquanto para alunos portugueses a oscilação foi mais suave percentualmente. O distrator B teve as maiores oscilações e não demonstrou um decréscimo relativamente linear, como deveria. O intervalo de discriminação (de Q a R) foi maior para o gráfico dos alunos portugueses, porém o poder de discriminação revela um ângulo baixo para ambos em cerca de 30°.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Normal/Mediano** (proporção esperada de

respondentes 40%; $-0,51 < b \leq 0,51$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros conforme dizem os valores desse parâmetro. As CCIs nos dois contextos sugerem uma boa discriminação conforme mostra a elevada inclinação. O acerto ao acaso (*guessing*) foi de 0,15 e 0,21, próximo de um valor esperado de 1/5 ou 0,20, sugerindo que os alunos brasileiros foram menos propensos a “chutes”, seja por fadiga, pressa em terminar ou por desconhecimento, entre outras, indicando assim uma coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ . Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 0,60 e 0,59 (Brasil e Portugal), com isso se estima 40% de probabilidade de alunos brasileiros cometerem equívocos com proficiência $\theta > 2,7$, e 41% para alunos portugueses a partir de um nível de proficiência $\theta > 1,7$, aproximadamente. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto.

A elevação na *Curva de Informação do Item* (CII) para cada contexto indica uma melhor discriminação para os brasileiros e de o item ser mais difícil de responder (por estar mais à direita), apesar de ser bem discriminativo também para os portugueses. As informações se distribuem em um intervalo de 1 desvio-padrão (aproximadamente) em torno do posicionamento do item em cada contexto entre os níveis 0 e 2. O item y18 não atende aos critérios para ser considerado um item-âncora.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 0 a 1, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y3, y13, y23, y10, y22 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (0,91) e discriminação (7,04) classificando o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 38%, ou seja, menos da metade dos respondentes que alcançam esse nível θ possuem habilidade suficiente para responder o item corretamente. O acerto ao acaso foi estimado em 0,17, abaixo do esperado que seria de $\frac{1}{4}$ ou 0,25, representando 17% de probabilidade de acertar sem possuir uma proficiência adequada, isso indica coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência dos sujeitos. Foi estimado um índice de erro ao acaso de 0,59, ou seja, 41% de cometerem equívocos por descuido para além de um nível de proficiências igual a $\theta = 2,5$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação para o item y18 em um intervalo próximo de 1,5 desvio-padrão entre os níveis 0 e 1, entretanto, o item não apresenta subsídios necessários para ser considerado um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item procura aferir uma compreensão sobre os processos de trocas de calor diante de um mesmo material, porém com temperaturas diferentes, e espera-se que o aluno compreenda que uma maior diferença de temperatura sugere que um maior fluxo de calor, e, portanto, uma sensação maior de frio para a caixa retirada da geladeira. O distrator B afirma que a sensação de mais frio é porque a caixa “contém menos calor”. Essa atração sugere um equívoco de que o calor seja uma propriedade do corpo, e ainda que a sensação de frio ou da noção de temperatura esteja relacionada ao conceito de ter mais ou menos calor. Os alunos brasileiros também foram fortemente atraídos pelo distrator A em que diz que a caixa retirada da geladeira “contém mais frio”. Esse equívoco remete uma compreensão errônea de que uma sensação térmica seja propriedade de um corpo, e que um corpo de temperatura mais baixa é devido a “possuir” mais frio.

SUGESTÕES: Apresentar a Lei de Fourier para condução térmica e demonstrar que mesmo para materiais com mesmas condutividades térmicas, o fluxo de calor \emptyset pode ser maior devido a ser diretamente proporcional a diferença de temperatura

entre o corpo com maior temperatura T_1 e o de menor temperatura T_2 , ainda que possuam dimensões iguais, ou seja, mesma área de contato A e espessura e .

$$\phi = \frac{A}{e} \cdot k \cdot (T_1 - T_2)$$

Ou seja:

$$\phi \sim (T_1 - T_2)$$

Numa atividade experimental seria interessante verificar as diferenças de temperatura com objetos de mesmo material e dimensões iguais variando somente as diferenças de temperatura entre eles. Obviamente seria rico por meio de uma modalidade investigativa, favorecendo entre os alunos os aspectos de problematização, previsão, descrição e discussão. O Modelo Cinético-Molecular também poderia auxiliar uma melhor compreensão e apreensão desse processo associando a observação indireta mesoscópica com o movimento das partículas. Isso também pode ser feito por atividade experimental demonstrativa ou laboratorial, assim como através do uso de vídeos e de simulações virtuais de objetos educacionais digitais. A escolha dessas metodologias deve ser analisada pelo tempo despendido para a realização da atividade e também se atende aos objetivos estabelecidos.

A19: ITEM y19

19. Rui sabe que a sua mãe faz sopa na panela de pressão porque esta cozinha mais rápido do que uma panela normal, mas não sabe o porquê disso (panelas de pressão têm a tampa selada, o que faz com que a pressão no interior atinja um valor maior que o da pressão atmosférica).

a) Armanda diz: “Isso acontece porque a pressão faz a água ferver acima de 100 °C”.

b) Carlos diz: “Isso ocorre porque a pressão alta gera calor extra”.

[B5]: Ponto de ebulição é a temperatura máxima que uma substância pode alcançar.

⇒ Por considerar outras causas para o rápido aquecimento sem mudar a temperatura de ebulição da sopa.

[D5]: O ponto de ebulição da água é de 100°C (apenas).

⇒ Por considerar o comportamento da sopa igual à da água, sendo o ponto de ebulição invariável, mesmo quando submetido a uma maior pressão que a atmosférica. Ref. [B5].

c) Fernanda diz: “É porque o vapor está numa temperatura maior do que a sopa fervente”.

[B5]: Ponto de ebulição é a temperatura máxima que uma substância pode alcançar.

⇒ Por considerar que a sopa não eleva seu ponto de ebulição ao comparar com o vapor mesmo com um aumento de pressão (*ou eleva, mas sendo uma temperatura inferior à do vapor)

[D5]: O ponto de ebulição da água é de 100°C (apenas).

⇒ Por considerar o comportamento da sopa igual à da água, sendo o ponto de ebulição invariável, mesmo quando submetido a uma maior pressão que a atmosférica. Ref. [B5].

[D8]: O vapor d’água está sempre a uma temperatura superior a 100°C.

⇒ Por acreditar que o vapor esta supostamente acima de 110 °C, pois quando muda-se de fase, a temperatura necessariamente deve ser superior a do ponto de ebulição.

d) Tomás diz: “É porque panelas de pressão propagam o calor mais uniformemente através dos alimentos”.

[B5]: Ponto de ebulição é a temperatura máxima que uma substância pode alcançar.

⇒ Por considerar outras causas para o rápido aquecimento sem mudar a temperatura de ebulição da sopa.

[D5]: O ponto de ebulição da água é de 100°C (apenas).

⇒ Por considerar o comportamento da sopa igual à da água, sendo o ponto de ebulição invariável, mesmo quando submetido a uma maior pressão que a atmosférica. Ref. [B5].

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Aplicação* de Processos de Fervura de acordo com a Variação da Pressão – [Descritor 7]

PROFICIÊNCIA: *Inferir* que a «temperatura da água fervente» «varia para mais ou menos do valor de 100 °C» «dependendo da variação da pressão em que fica submetida. No caso, estando em um recipiente fechado (panela de pressão) o

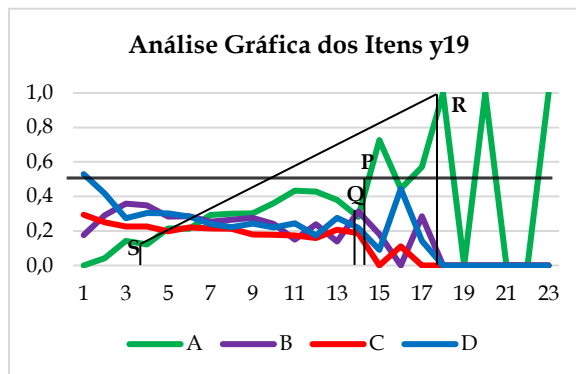
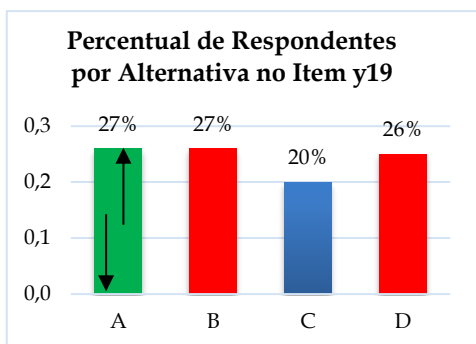
aquecimento aumenta o movimento e a força de repulsão entre as partículas, e isso resulta em uma maior incidência do choque entre elas, ou seja, a pressão interna fica maior que a pressão atmosférica externa e o ponto de ebulição da água contida na sopa fica mais baixo tornando mais fácil alcançar o processo de fervura e de cozinhamento».

INDÍCIOS DE DEPENDÊNCIA LOCAL: Os itens y4 e y8 contextualizam uma situação de aquecimento em que se busca compreender o ponto de ebulição da água, e dentro disso se torna necessário saber a relação do ponto de ebulição com a pressão, assim como o item y19.

PAINÉL DE ESTATÍSTICAS DO ITEM y19

Coeficientes Estatísticos e Gráficos de Teoria de Clássica dos Testes (TCT)

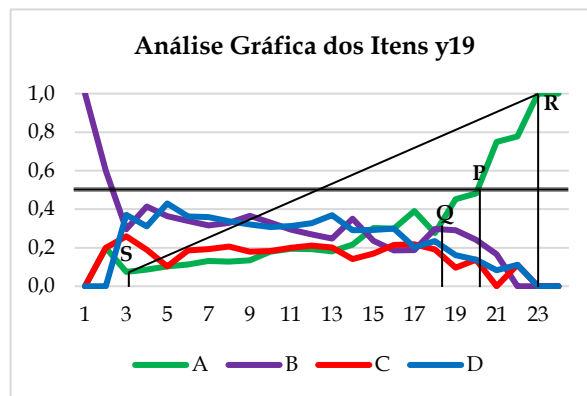
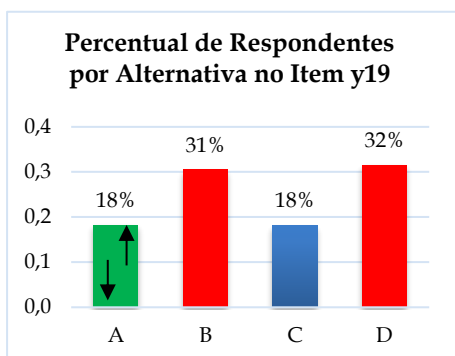
Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
A	0,27	0,36	0,15	0,22	0,29	0,02	0,27	0,27	0,20	0,26	0,29	-0,07	-0,05	-0,08



ID (corr.): 0,024; C: 0,01 (BB - aleatório) DISCR (corr. sem elim. item): 0,052; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,055; Coef. Bisserial (corr.): 0,074; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): Misto

Classificação AGI: **Baixa qualidade - Difícil; Baixa discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra A: 8,11 (31,2%); Desvio-padrão amostral: 2,79

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
A	0,18	0,27	0,12	0,15	0,32	0,05	0,18	0,31	0,18	0,32	0,32	-0,11	-0,01	-0,10



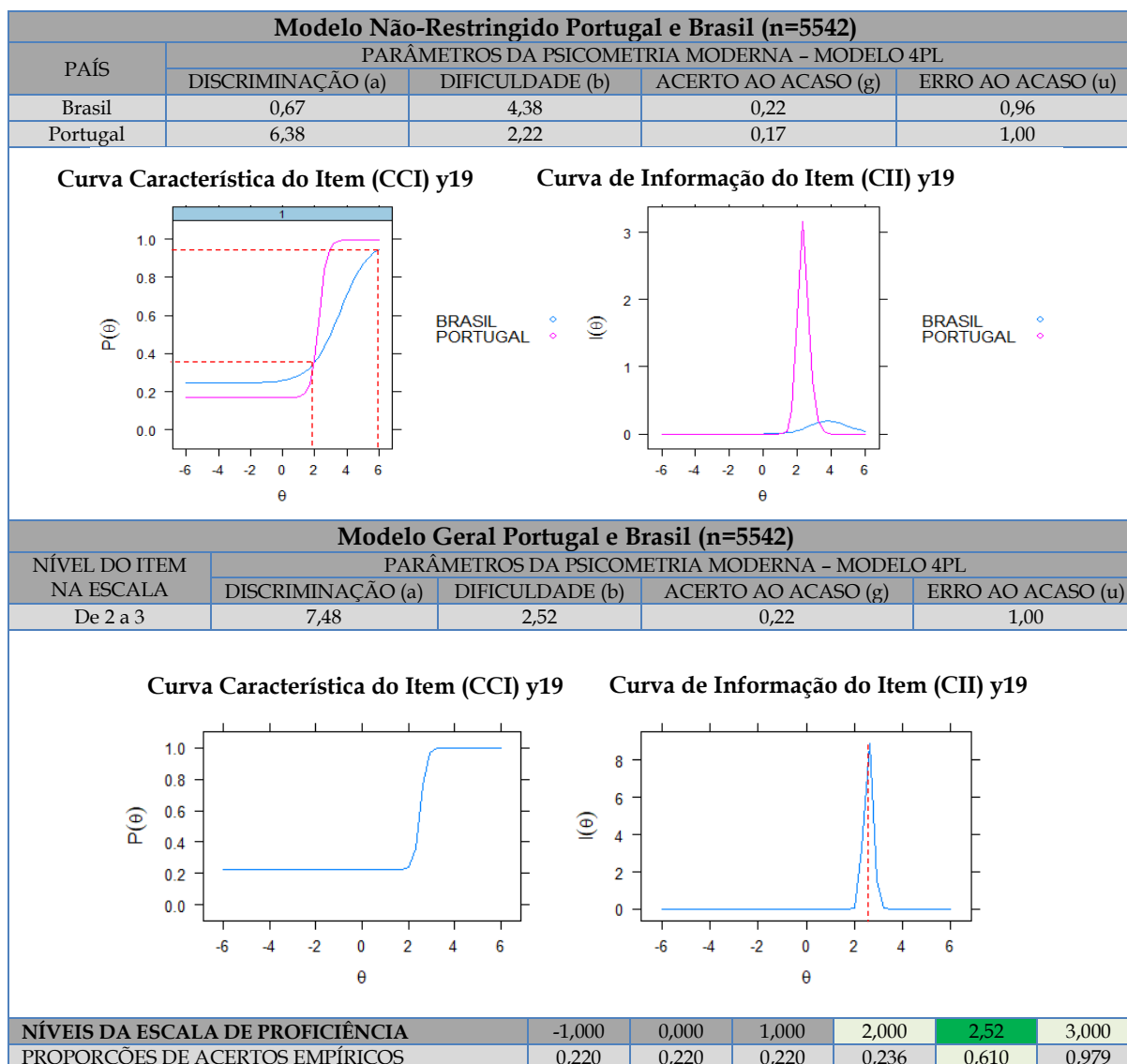
ID (corr.): -0,086; C: 0,03 (BB - aleatório); DISCR (corr. sem elim. item): 0,124; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,130; Coef. Bisserial (corr.): 0,190; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): Misto

Classificação AGI: **Baixa qualidade - Muito difícil; Baixa discriminação; Forte poder atrativo**; Escore médio dos que acertaram a letra A: 12,10 (46,5%); Desvio-padrão amostral: 4,40

Análise Gráfica do Item - AGI y19

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	8	-	-	-	6	-	-	-
Tipo 2	-	8	3	5	-	3	13	9
Tipo 3	-	0	0	1	-	0	0	0
Tipo 4	26,8º				23,5º			
Total	25				31			

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y19

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **marginal**, *sujeito a reelaboração* ($0,20 \leq DISCR < 0,30$) (Brasil); **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **inadequado**, *sujeito a rejeição* ($DISCR < 0,20$).

Os percentuais de acerto total para a alternativa correta foram de 27% e 18% (Brasil e Portugal). Para Portugal o percentual mais baixo é igual ao distrator C, em que são as duas opções mais inferiores percentualmente, estando prevalectido o distrator D com 32%, que corresponde quase ao mesmo valor da letra B, com 31%. No Brasil o

percentual de acerto se iguala ao distrator B, e também demonstra um percentual semelhante entre B e D, com 27% e 26%. Pode-se observar um certo equilíbrio entre as opções, e isso demonstra indícios de um item confuso, com distratores fortemente atrativos.

Os grupos de maior e menor desempenho correspondem a 36% e 15% (Brasil), e 27% e 12% (Portugal), ou seja, os acertos percentualmente em média dentro desses grupos são inferiores a 50% para as duas amostras, isso indica força dos distratores e problemas no item.

Os bisseriais possuem valores iguais a 0,29 e 0,32, sinalizando comportamentos correspondentes entre os contextos, tendo um perfil de composição de sujeitos de forma mista entre alunos de alto e baixo desempenho, com leve superioridade daqueles com bom desempenho. Os distratores B e D possuem bisseriais -0,07, -0,05 e -0,08 para os brasileiros, e -0,11, -0,01 e -0,10 entre os portugueses, ou seja, para todos eles o perfil de composição de grupos de alto e abaixo desempenho são formados de forma praticamente igual, isso sinaliza e reforça consideravelmente problemas no item, sendo confuso e de muitas incertezas entre os respondentes. Observa-se uma atração dos distratores B e D aos alunos brasileiros com 17 acertos e dos distratores C e D aos alunos portugueses com 22 acertos, com destaque para a alternativa D.

O AGI revela uma quantidade maior de oscilações e mais intensas percentualmente entre os alunos brasileiros na sequência geral dos grupos de desempenho, com destaque para o grupo de 16 acertos que foi fortemente atraído pelo distrator D. Os distratores B variou mais para os brasileiros, e o C entre os portugueses. O poder de discriminação ficou com valor baixo no ângulo de inclinação, com cerca de 25° aproximadamente para ambos.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **bom**, *sujeito a aprimoramento*, com uma *discriminação moderada* ($0,65 \leq a \leq 1,35$) (Brasil); **Muito difícil** (proporção esperada de respondentes

10%; $b \geq 1,28$) e **adequado**, com evidências de ser reaplicado com uma discriminação muito alta ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros, e com menor discriminação (com inclinação da curva mais elevada) conforme dizem os valores desse parâmetro. O acerto ao acaso (*guessing*) foi de 0,22 e 0,17, um pouco abaixo do valor esperado de $\frac{1}{4}$ ou 0,25, sugerindo que os alunos portugueses foram menos propensos a “chutes”, seja por fadiga, pressa em terminar ou por desconhecimento, entre outras, indicando assim uma maior coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ . Os índices para o parâmetro de erro ao acaso (“*upper*”) para cada grupo foram de 0,96 e 1,00 (Brasil e Portugal), com isso se estima apenas 4% de probabilidade de alunos brasileiros cometerem equívocos com proficiência $\theta > 6,00$ (aproximadamente), sendo praticamente não constatado, assim como para o contexto português. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto a partir de $\theta = 2,00$ (aproximadamente).

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os portugueses, e sendo um item ser mais difícil de responder (por estar mais à esquerda). Para os brasileiros demonstra ser pouco discriminativo e mais difícil. As informações se distribuem próximo de um intervalo de 3 desvios-padrões estando mais próximo do posicionamento do item no contexto geral (considerando conjuntamente Brasil e Portugal) entre os níveis 2 e 3.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 2 a 3, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 3 itens-âncoras y_8 , y_{19} e y_{15} (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (2,52) e discriminação (7,48) classificando o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; b

$\geq 1,28$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 61%, ou seja, mais da metade dos respondentes que alcançam esse nível possuem habilidade suficiente para responder o item corretamente. O acerto ao acaso foi estimado em 0,22, próximo ao esperado que seria de $\frac{1}{4}$ ou 0,25, representando 22% de probabilidade de acertar sem possuir uma proficiência adequada, isso indica ainda uma relativa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os níveis de proficiência alcançados e habilidades θ estimadas para os sujeitos. Não foi estimado para o grupo geral um índice de erro ao acaso ($u = 1,00$), com isso, ficam ausentes perspectivas diagnósticas e probabilidade da ocorrência de equívocos por descuido entre alunos de elevada proficiência. Por fim, a elevação da CII indica uma boa discriminação para o item y19 em um intervalo próximo de 1,5 desvio-padrão entre os níveis 2 e 3, sendo considerado, portanto, um item com subsídios necessários para ser considerado um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item busca aferir uma compreensão sobre o funcionamento de uma panela de pressão, especificamente sobre o fato de cozinhar mais rápido, espera-se que os alunos compreendam que o aumento da pressão eleva o ponto de ebulição, fazendo com que a fervura em temperatura mais elevada possibilite cozinhar mais rapidamente, conforme afirma Armanda na letra A: *“isso acontece porque a pressão faz a água ferver acima de 100 °C”*. O distrator B justifica com a opinião de Carlos que diz: *“isso ocorre porque a pressão alta gera calor extra”*, e o D com Tomás dizendo: *“é porque panelas de pressão propagam o calor mais uniformemente através dos alimentos”*. O primeiro equívoco associa o aumento de pressão com o aumento de calor, talvez devido a ideia de que a sensação de aumento de temperatura seja correspondente ao calor, ou seja, confunde-se o conceito de temperatura com calor, além de desconhecer a relação da variação da pressão com o ponto de ebulição. O outro equívoco parece estar associado na crença de que a pressão dentro da panela faz com que o calor (e não a temperatura) se concentre e se distribua de forma igual a todos os pontos no interior da panela, talvez porque isso de fato acontece fazendo com que a pressão interna se seja igual a todos os pontos, com isso, confunde-se associando a característica da pressão interna com a ideia de calor distribuído uniformemente.

SUGESTÕES: É relevante nesse caso explorar a relação do aumento da pressão com a elevação do ponto de ebulição, de forma análoga quando ocorre uma diminuição da pressão em região elevadas. O item 4 apresenta sugestões metodológicas para que se explore e facilite a compreensão da relação entre pressão e ponto de ebulição.

A20: ITEM y20

20. Patrícia acredita que o seu pai coloca um bolo na prateleira superior do forno elétrico porque a parte de cima é mais quente do que a de baixo.

a) Paula diz que é mais quente porque o calor sobe.

[C2]: **O calor propaga-se em meios materiais em sentido unicamente para cima.** (concepção animista, pois o calor possui “desejo” de se propagar em sólidos e fluidos no sentido para cima).

⇒ Por considerar que “suba” com relação ao sentido de condução térmica no suporte da bandeja.

[C3]: **O calor sobe** (concepção aristotélica animista, pois o calor possui “desejo” de estar em cima, é o seu lugar natural).

⇒ Por considerar que “suba” em relação ao sentido as massas de ar envolvidas no interior do forno elétrico.

b) Samuel diz que é mais quente porque bandejas de metal concentram mais calor.

[C9]: **A teoria cinética na verdade não explica a transferência de energia térmica.** (As explicações são relatadas, mas não acreditadas pelo sujeito, ou seja, não se percebe que as diversas formas de energia que se propaga entre as partículas representam a energia térmica em trânsito).

⇒ Por ter uma noção de que o corpo possui uma capacidade de transferir (receber ou ceder) maior quantidade de energia térmica, porém, não demonstra uma compreensão sobre o fenômeno microscópico que explica a forma como ocorre a transferência.

[D2]: **Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio.** (concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não um energia transitória. O frio também como uma substância).

⇒ Por considerar que o metal possua a capacidade de absorver o calor no interior do forno elétrico, ou de absorver todo o calor que “sobe” nas mais diversas maneiras.

c) Rubens diz é mais quente na parte superior do forno porque quanto mais quente o ar, menos denso ele é.

d) Tiago não concorda com eles e diz que não é possível que a parte superior do forno seja mais quente.

⇒ Acredita-se que a temperatura no interior do forno elétrico seja homogênea, ou seja, a temperatura é a mesma para todos os pontos. (*): *Não houve correspondência direta com nenhuma concepção do inventário).*

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão sobre a Propagação do Calor - [Descritor 12]*

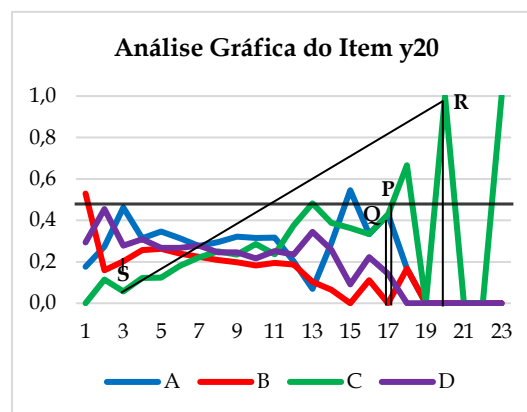
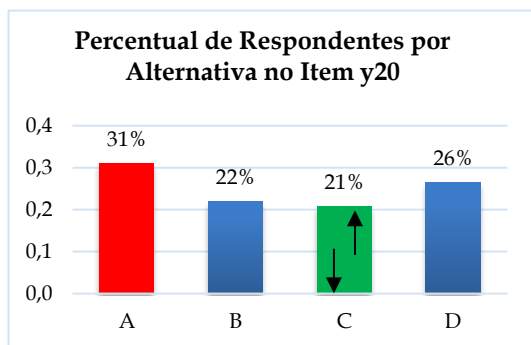
PROFICIÊNCIA: *Inferir que a «convecção do ar no interior do forno elétrico ocorre pelo deslocamento de massas de ar quente devido a diminuição de sua densidade», com isso, «a parte superior no interior do forno se torna mais eficiente para o aquecimento do bolo».*

INDÍCIOS DE DEPENDÊNCIA LOCAL: Não foi identificado entre os itens anteriores.

PAINÉL DE ESTATÍSTICAS DO ITEM y20

Coeficientes Estatísticos e Gráficos da Teoria de Clássica dos Testes (TCT)

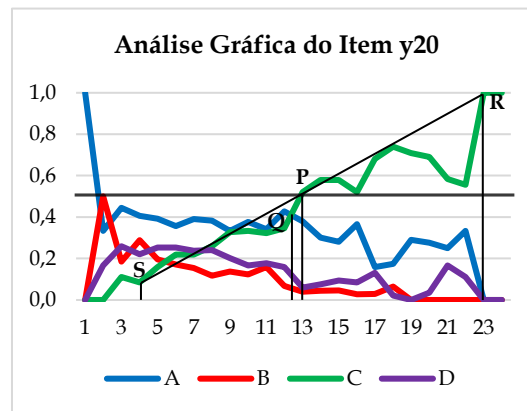
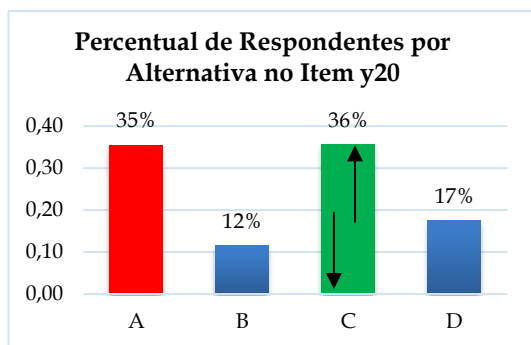
Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
C	0,21	0,10	0,28	0,18	0,28	0,02	0,31	0,22	0,21	0,26	-0,02	-0,08	0,28	-0,05



ID (corr.): -0,057; C: 0,01 (BB - aleatório); DISCR (corr. sem elim. item): 0,046; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,049; Coef. Bisserial (corr.): 0,069; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): **Misto**

Classificação AGI: **Baixa qualidade - Muito difícil; Baixa discriminação; Forte poder atrativo**; Escore médio dos que acertaram a letra C: **8,19 (31,5%)**; Desvio-padrão amostral: **2,85**

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014														
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS			
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D
C	0,36	0,53	0,21	0,32	0,41	0,09	0,35	0,12	0,36	0,17	-0,09	-0,27	0,41	-0,23



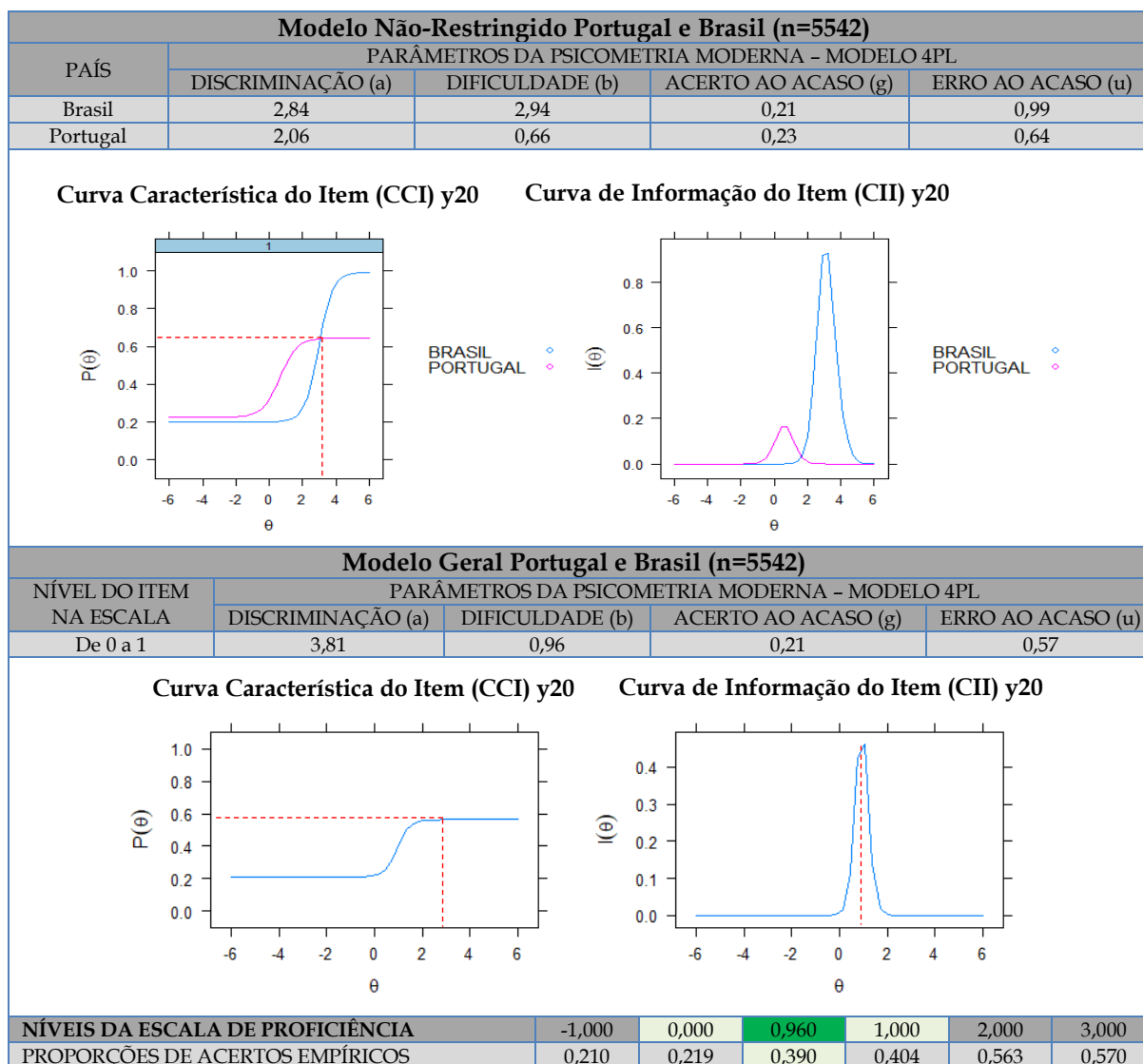
ID (corr.): 0,142; C: 0,09 (BB - aleatório); DISCR (corr. sem elim. item): 0,198; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,196; Coef. Bisserial (corr.): 0,252; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): **Misto - predominância do grupo ACIM**

Classificação AGI: **Baixa qualidade - Difícil; Baixa discriminação; Forte poder atrativo**; Escore médio dos que acertaram a letra C: **11,93 (45,9%)**; Desvio-padrão amostral: **3,98**

Análise Gráfica do Item - AGI y20

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	-	-	8	-	-	-	8	-
Tipo 2	9	6	-	7	8	8	-	11
Tipo 3	0	0	-	0	0	0	-	0
Tipo 4	30,8°				29°			
Total	30				35			

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y20

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **inadequado, sujeito a rejeição** ($DISCR < 0,20$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **bom, sujeito a aprimoramento** ($0,30 \leq DISCR < 0,40$) (Portugal).

O percentual de acertos total ficou em 21%, sendo a última opção entre os brasileiros, e 36% entre os portugueses estando no mesmo patamar percentual do distrator A, que coincidentemente é o mais atrativo para os alunos brasileiros.

Os grupos de maior e menor desempenho ficaram com 10% e 28% (Brasil), e 53% e 21% (Portugal), sinalizando uma diferença da função exercida pelo item em cada contexto. Os valores desses grupos para os brasileiros revelam uma porcentagem menor e pouco expressiva de indivíduos de alto desempenho que responderam corretamente, sendo os de menor desempenho acertando mais, e isso não deveria acontecer. Em Portugal há um predomínio absoluto (>50%) de alunos com bom desempenho acertando e uma diferença entre as porcentagens revelando uma melhor discriminação entre os grupos.

Os bisseriais foram de 0,28 e 0,41, sendo mais discriminativo para os portugueses, de acordo como sinaliza as diferenças percentuais, e isso sugere uma composição de indivíduos com predominância de alunos com bom desempenho respondendo corretamente em Portugal, apesar de que o ideal seja um valor mais próximo de 1. No Brasil o perfil é de um grupo mais misto, com leve superioridade na quantidade de indivíduos com bom desempenho, sendo pouco representativo na discriminação. Entre os brasileiros, o distrator A, assim como os outros, apresenta um bisserial muito próximo de zero, isso mostra que os alunos de bom desempenho têm sentido atração por esses equívocos da mesma forma que os de baixo desempenho, e isso não deveria acontecer. Para os portugueses o distrator A também apresenta um bisserial próximo de zero com valor de -0,09, ou seja, é também um perfil misto para aqueles que optam por essa alternativa, entretanto, para os outros distratores já um sutil predomínio de alunos com baixo desempenho devido o valor ser um pouco abaixo de zero, porém ainda longe do ideal que seria próximo a -1. A atração do distrator D para os alunos portugueses com 22 acertos ganha destaque, e merece uma atenção maior.

A AGI indica uma mesma quantidade de violações do tipo 1 entre os dois contextos, em que se espera que os percentuais de acerto devam subir à medida em que avançam os percentuais dos grupos com melhores escores brutos. A quantidade de violações do tipo 2 que se refere ao decréscimo percentual dos distratores mostra uma maior oscilação entre os brasileiros para a opção A, com 8 violações. Para os portugueses, uma maior incerteza para o distrator D, com 11 violações. Não houveram

violações do tipo 3 e o poder de discriminação para os dois contextos foi de um ângulo correspondente em cerca de 30° , sendo um valor baixo.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros, e com melhor discriminação conforme mostra a inclinação e os valores desse parâmetro. O acerto ao acaso (*guessing*) foi de 0,21 e 0,23 (Brasil e Portugal), próximo de um valor esperado de $\frac{1}{4}$ ou 0,25, sugerindo que os alunos foram menos propensos a “chutes”, seja por fadiga, pressa em terminar ou por desconhecimento, entre outros. Isso indica coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ e níveis de proficiência alcançadas. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 0,99 e 0,64 (Brasil e Portugal), com isso se estima apenas 1% de probabilidade de alunos brasileiros cometerem equívocos, porém há 36% de chance de erro para alunos com proficiência $\theta > 3,00$. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto, com exceção no intervalo $\theta > 3,00$ (aproximadamente), porém é praticamente ausente indivíduos com essa proficiência no contexto brasileiro.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros (mais elevado), maior dificuldade (está mais à direita) e uma alta precisão nas interpretações dentro desse contexto para o item, na qual se distribui em um intervalo de 2,5 desvios-padrões (aproximadamente).

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 0 a 1, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y3, y13, y23, y10, y22 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (0,96) e discriminação (4,70) classificando o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 39% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso (“*guessing*”) foi estimado em 0,21, dentro do esperado que seria de $\frac{1}{4}$ ou 0,25, representando 21% de probabilidade de acertar sem possuir uma proficiência adequada. Foi estimado um índice de erro ao acaso de 0,57, ou seja, 43% de cometerem equívocos por descuido para além de um nível de proficiências igual a $\theta = 2,9$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação para o item y20 em um intervalo próximo de 2,0 desvios-padrões entre os níveis 0 e 2, entretanto, o item não apresenta subsídios necessários para ser considerado um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item exige uma análise comparativa sobre duas situações parecidas que envolve trocas de calor e compreensão sobre equilíbrio térmico entre uma toalha seca e molhada envolvida em garrafas com água inicialmente a mesma temperatura (20 °C), e espera-se com base nas respostas disponíveis concluir por indução que a temperatura ambiente seja acima de 21 °C, portanto, a letra A que afirma ser 26 °C. Como não é fornecida as informações sobre a temperatura inicial das toalhas, a análise se baseia em especulações da temperatura final após certo intervalo de tempo na qual uma esfria a água (18 °C) e outra esquentada (22 °C). O distrator C é o mais atrativo e afirma que a temperatura ambiente seria 20 °C. Um dos motivos para isso talvez seja que os respondentes imaginem que a temperatura inicial da água nas garrafas já corresponderia a temperatura ambiente. Outra possibilidade seria imaginar que a temperatura

ambiente seria a média entre as temperaturas finais das duas garrafas no intervalo considerado.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item possui a finalidade de aferir uma compreensão sobre o movimento de convecção do ar no interior de um forno elétrico, sendo essa a causa para que haja um melhor aquecimento do bolo na parte superior, pois o ar quente tende a se concentrar nessa região. Na alternativa correta Rubens afirma que *“é mais quente na parte superior do forno porque quanto mais quente o ar, menos denso ele é”*, justificando adequadamente que ao aquecer o ar, ele se expande se tornando menos denso e com isso a massa de ar tende a se deslocar para cima. O distrator A possui uma forte atração na qual segue a afirmação de Paula que *“é mais quente porque o calor sobe”*. Esse equívoco pode estar relacionado a noção de que o calor possui uma condição natural de “subir”, semelhante como se observa a chama do fogo, direcionado para cima. O distrator D também chama a atenção pela atração entre os brasileiros que concordam com a negação de Tiago que afirma não concordar com eles dizendo que *“não é possível que a parte superior do forno seja mais quente”*, ou seja, supostamente podem estar tendo a crença que o aquecimento de se dá de forma igual e homogênea no interior do forno, não admitindo ou compreendendo que haja correntes de convecção sendo formadas.

SUGESTÕES METODOLÓGICAS E RECOMENDAÇÕES BASEADAS NA PSICOLOGIA DE AUSUBEL COMO CONDIÇÕES DE CONTORNO: Os subsunçores para uma compreensão e resolução dessa situação-problema remete ao domínio sobre os processos de transmissão do calor, nomeadamente sobre as condições que envolvem e explicam o processo de formação das correntes de convecção. Esse é um aspecto necessita de uma noção adequada sobre as formas de como o calor se propaga em fluidos. Esse tipo de abordagem deve confrontar com a suposta influência do senso comum de que o calor “sobe” da mesma forma como se percebe visualmente em uma chama de fogo, ou por exemplo, no caso de balão ser preenchido por ar quente e tender a se elevar.

Metodologicamente os mecanismos de ensino por investigação são defendidos como sendo potencialmente significativos por favorecerem um confronto cognitivo com o senso comum dos alunos. A proposição da formulação de questionamentos, elaboração de previsões, descrições e discussão, são consideradas elementos essenciais dentre as tarefas logicamente significativas. A escolha de métodos de ensino que se adequem a isso devem ser propostas levando em conta uma sequenciação das atividades lógica e estratégica para o andamento gradual na aquisição e apreensão dos conceitos, promovendo a autonomia dos alunos, contextos de vida real, e com isso é considerado indispensável uma organização prévia dos conteúdos pelo professor e de sua abordagem, estando alinhados com os objetivos preestabelecidos no plano de ensino. Sugerem-se atividades experimentais investigativas que podem ser reais/concretas ou virtuais/simuladas, mas para a necessidade de aprendizagem mecânica, a modalidade de verificação de leis e princípios, podendo ser útil também para uma retenção subsunsa e reconciliação integradora. Na arguição docente, as modalidades verificativas e ilustrativas podem servir para reforçar a compreensão dos alunos, mas devem ser utilizadas no momento adequado dentro da sequenciação de aprendizagem.

Quanto a organização do conteúdo considera-se relevante investigar, observar e discutir sobre os diversos processos de transmissão do calor, podendo ser favorável ainda na distinção do conceito de temperatura. O Modelo Cinético-Molecular (MCM) pode auxiliar esclarecendo e diferenciando o comportamento das partículas entre os processos de convecção e condução do calor. O MCM pode ser explorado por meio de recursos físicos, como aparatos demonstrativos, ou mesmo por meio de vídeos, dependendo da escolha feita pelo professor. No funcionamento do forno elétrico também se torna relevante compreender sobre as micro-ondas emitidas, e da forma como elas interagem com a matéria contida no interior do forno. Nesse caso existe a intradisciplinaridade do Ramo de Ondas, Eletromagnetismo e Termodinâmica, caracterizando assim um elo entre a radiação e a convecção como processos interligados para uma melhor apreensão da eficiência no aquecimento do bolo ele estando na parte interna e superior do forno.

A21: ITEM y21

21. Beatriz lê uma questão de múltipla escolha num manual: “Suar resfria-te porque o suor sobre a tua pele:

a) Umedece a superfície, e superfícies úmidas retiram mais calor do que superfícies secas”.

[A1]: O calor é uma substância (concepção substancialista).

⇒ Por considerar a possibilidade de “retirar calor”, como se fosse um corpo físico ou uma substância.

[B3]: Percepções de quente e frio não estão relacionados com a transferência de energia.

⇒ Por considerar que a sensação de resfriamento está relacionada a umidade da superfície da pele e não apenas ao fluxo de transferência térmica, sem levar em consideração a umidade da superfície.

b) Drena o calor dos poros e espalha-o pela superfície da pele”.

[A1]: O calor é uma substância (concepção substancialista).

⇒ Ref. [C4].

[A2]: O calor não é energia.

⇒ Ref. [A1] por ser uma substância física, tipo: o suor líquido é a materialização do calor.

[B3]: Percepções de quente e frio não estão relacionados com a transferência de energia.

⇒ Ref. [C4].

[C4]: Calor e frio fluem como líquidos (concepção substancialista).

⇒ Por considerar que o calor “escorre” e se “espalha”, ao invés de ser uma transferência de energia térmica.

[C9]: A teoria cinética na verdade não explica a transferência de energia térmica. (As explicações são relatadas, mas não acreditadas pelo sujeito, ou seja, não se percebe que as diversas formas de energia que se propaga entre as partículas representam a energia térmica em trânsito).

⇒ Por argumentar através de conhecimentos microscópicos, mas não utilizar a teoria cinética para explicar o processo de transferência de energia térmica.

c) Está à mesma temperatura do que a tua pele, mas está evaporando e assim afasta o calor dela”.

[A1]: O calor é uma substância (concepção substancialista).

⇒ Por considerar a possibilidade de “afastar calor”, como se fosse um corpo físico ou uma substância.

[C9]: A teoria cinética na verdade não explica a transferência de energia térmica. (As explicações são relatadas, mas não acreditadas pelo sujeito, ou seja, não se percebe que as diversas formas de energia que se propaga entre as partículas representam a energia térmica em trânsito).

⇒ Por considerar que o calor se “desloca” ou se “afasta” do corpo, ou invés de ser uma transferência de energia térmica.

d) É ligeiramente mais frio do que a tua pele devido à evaporação, e dessa forma o calor é transferido da tua pele para o suor”.

Indique a resposta que você diria para ela escolher.

Nota: A letra “c” e “d” possuem ideias próximas, sendo a principal diferença é que um diz “afastar o calor” e o outro “transferir o calor”. Essa sutileza pode gerar problemas de ambiguidade, e com isso causar confusões. Na realidade a sensação de frio é causada em grande parte pelo fato das gotas de suor absorverem (receberem) a energia térmica da pele, utilizando parte dessa energia, fora a colisões internas entre as outras partículas da gota de suor, para se desprenderem ao ambiente, ou seja, evaporarem. Esse desprendimento é a principal causa da sensação de resfriamento do corpo devido a energia térmica da pele ser transferido as gotas que evaporam e se dispersam no ambiente.*

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão sobre a Sensação Térmica explicada pela Propagação do Calor - [Descritor 15].*

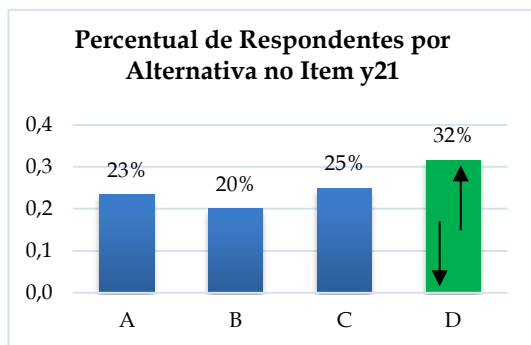
PROFICIÊNCIA: *Inferir sobre o sentido de «transferência de energia térmica que ocorre durante o processo de trocas de calor», no caso, entre a «pele e o suor, tendo uma compreensão sobre influência da evaporação na sensação térmica».*

INDÍCIOS DE DEPENDÊNCIA LOCAL: *Os itens y10 e y13 invocam a uma compreensão sobre o sentido de transferência de energia térmica, porém não associa aspectos peculiares envolvidos, como o fato do suor evaporar e influenciar na sensação térmica.*

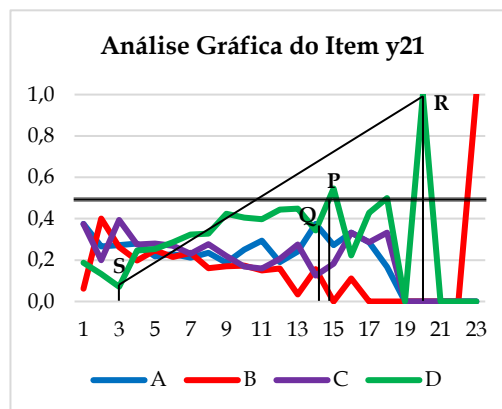
PAINÉL DE ESTATÍSTICAS DO ITEM y21

Coeficientes Estatísticos e Gráficos da Teoria de Clássica dos Testes (TCT)

Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015															
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS				
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D	
D	0,32	0,41	0,19	0,22	0,23	0,002	0,23	0,20	0,25	0,32	0,03	-0,09	-0,06	0,23	

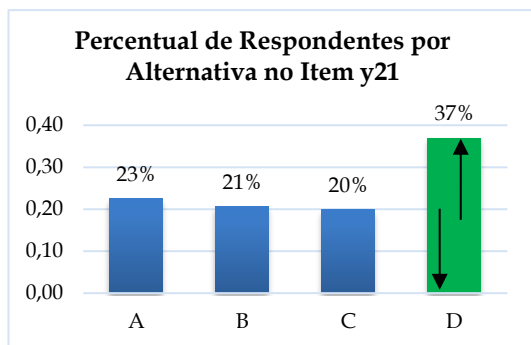


ID (corr.): 0,089; C: 0,01 (BB - aleatório); DISCR (corr. sem elim. item): 0,005; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,005; Coef. Bisserial (corr.): 0,007; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): **Misto**

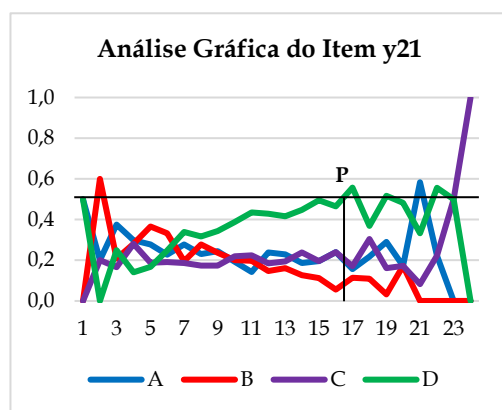


Classificação AGI: **Baixa qualidade - Difícil; Baixa discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra D: 7,85 (30,2%); Desvio-padrão amostral: 2,61

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014															
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA				COEFICIENTES BISSERIAIS				
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	A	B	C	D	
D	0,37	0,45	0,26	0,19	0,22	0,02	0,23	0,21	0,20	0,37	-0,04	-0,22	0,04	0,22	



ID (corr.): 0,160; C: 0,04 (BB - aleatório); DISCR (corr. sem elim. item): 0,050; Coef. de Correlação Ponto-Bisserial (corr. elim. item): 0,054; Coef. Bisserial (corr.): 0,069; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): **Misto**

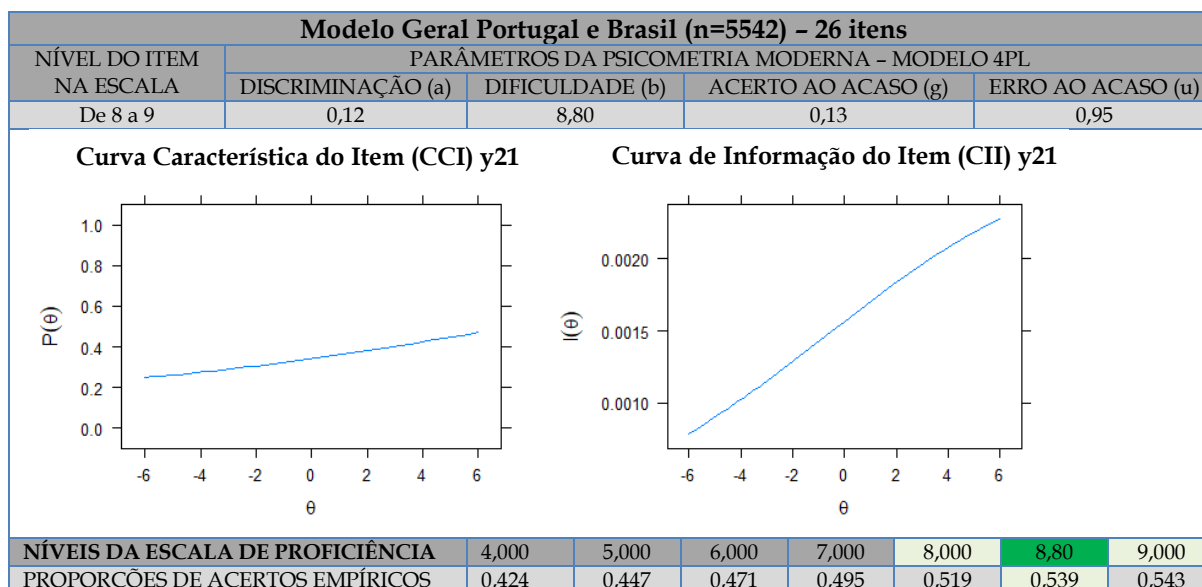


Classificação AGI: **Baixa qualidade - Difícil; Baixa discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra D: 11,18 (43,0%); Desvio-padrão amostral: 3,77

Análise Gráfica do Item - AGI y21

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)				Nº DE INCIDÊNCIAS (PORTUGAL)			
	A	B	C	D	A	B	C	D
Tipo 1	-	-	-	8	-	-	-	11
Tipo 2	9	9	8	-	9	7	14	-
Tipo 3	0	1	2	-	-	-	-	-
Tipo 4	31,7º				INDETERMINADO			
Total	37				41			

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y21

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **marginal, sujeito a reelaboração** ($0,20 \leq DISCR < 0,30$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **inadequado, sujeito a rejeição** ($DISCR < 0,20$) (Portugal).

Os percentuais de acerto (letra D) no âmbito geral foram de 32% e 37%, sem haver predomínio absoluto na escolha, mas com predomínio relativo diante dos distratores em ambos os contextos, em outras palavras, não foi maior que 50%, mas prevaleceu com uma maior porcentagem comparando com as outras alternativas. Observa-se que os distratores A, B e C ficaram próximos a 20% para as duas amostras, demonstrando uma funcionalidade equivalente do item.

Os grupos de maior e menor desempenho ficaram com 41% e 19% (Brasil), e 45% e 26% (Portugal), valores próximos, reforçando um comportamento correspondente do item para os dois contextos. Os valores de 41% e 45% revelam que quase a metade dos grupos de maior desempenho assinalaram a resposta correta. O ideal seria um predomínio acima de 50%, mas não foi evidenciado devido aos atratores.

Os bisseriais da alternativa correta foram de 0,23 e 0,22, sendo uma discriminação baixa e equivalente, com um perfil misto de composição de indivíduos formados por alunos de alto e baixo desempenho, com leve predominância daqueles com bom desempenho, porém, sendo pouco discriminativo. Os bisseriais dos distratores foram de 0,03, -0,09, e -0,06 entre os brasileiros, e -0,04, -0,22 e 0,04 para os portugueses, isso sugere um mesmo padrão na composição de grupo, sendo misto. O distrator B para Portugal foi sutilmente formado por um grupo com mais alunos de baixo desempenho, porém muito próximo a zero, sendo pouco discriminativo. A forte atração do distrator C para os alunos brasileiros e portugueses com bom desempenho merece uma atenção maior.

A AGI indica uma quantidade maior de violações da alternativa correta para os portugueses, sinalizando incertezas na escolha dessa alternativa, fato que não deveria ocorrer. As oscilações do distrator C também foram elevadas comparativamente com os brasileiros, porém, analisando os “saltos” nos gráficos, proporcionalmente sugere que isso foi mais intenso entre os brasileiros.

Com base na estimação dos parâmetros da TRI através do modelo geral com 26 itens, o item y21 pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **muito ruim**, *sujeito a reelaboração* ($0,0 < a \leq 0,35$).

A CCI apresenta uma sigmoide com uma baixa inclinação, revelando pouca discriminação, ou seja, o item não consegue distinguir respondentes alta e baixa proficiência. Estando mais à direita, indica uma posição do item em uma alta habilidade θ , possuindo um percentual de acertos empíricos estimados em 53,9% entre aqueles que alcançam esse nível de habilidade θ . O índice de acerto ao acaso foi de 0,13, um pouco abaixo do valor esperado que seria de $\frac{1}{4}$ ou 0,25, isso representa um probabilidade de 13% dos respondentes acertarem o item sem possuírem habilidades suficientes. O índice de *guessing* um pouco abaixo do esperado indica uma boa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com os respectivos níveis de proficiência alcançados. O índice de erro ao acaso foi de 0,95, isso representa 5% de probabilidade dos alunos cometerem equívocos, ainda que possuam

uma boa habilidade e esse nível se eleve. Por fim, a CII indica uma baixa discriminação, com isso, o item y21 não apresenta informações relevantes para uma interpretação pedagógica, estando sujeito a rejeição.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item tem a finalidade de verificar uma compreensão sobre o processo de resfriamento por meio do suor, nisso está envolvido uma noção sobre o princípio das trocas de calor entre a pele e o suor quando associada ao processo de evaporação. Intuitivamente é levado se tem como parâmetro a sensação térmica nesse evento, sendo necessário realizar uma interpretação adequada, ou seja, espera-se que os alunos percebam que a evaporação das gotas de suor influencia na sensação térmica por se resfriar nesse processo e assim retirar sutilmente o calor da pele. A alternativa correta afirma que a gota seja *“ligeiramente mais frio do que a tua pele [Beatriz] devido à evaporação, e dessa forma o calor é transferido da tua pele para o suor”*. A alternativa C afirma que isso ocorre devido a gota estar *“à mesma temperatura do que a tua pele [Beatriz], mas está evaporando e assim afasta o calor dela”*. Esse equívoco pode estar relacionado na crença que o suor e a pele estejam em equilíbrio térmico, não percebendo que a gota de suor quando transpira, entra em contato com o ar e resfria sutilmente ficando com uma temperatura inferior à da pele, e com isso acentuar a sensação de frio ao evaporar e receber o calor da pele ao mesmo tempo. O termo “afastar” o calor também transmite um significado equivocado, como se o fosse um corpo que pudesse ser deslocado.

SUGESTÕES METODOLÓGICAS BASEADAS NA PSICOLOGIA DE AUSUBEL COMO CONDIÇÕES DE CONTORNO: Os subsunçores relacionam-se sobre o processo de trocas de calor que explicam uma sensação térmica de fria para a pele. Para isso é necessário compreender que o frio é uma percepção de quando existem trocas de calor entre um corpo de maior e de menor temperatura, naturalmente sendo o fluxo indo no sentido do de menor temperatura. O processo de evaporação envolvido deve ser interpretado como um processo que atenua a sensação de frio pelo fato das partículas de suor se desprenderem da pele.

Metodologicamente pode ser proposta uma atividade significativa submetendo o aluno a uma experiência análoga a situação de um corpo suando e se resfriando. Para

isso pode ser utilizado um pouco água sobre a pele e explorar questionamentos instigadores que promovam outros problemas. Alguns recursos físicos como aparatos experimentais de laboratórios ou de demonstrações podem ser utilizados previamente para uma aquisição sobre os processos de condução térmica e de evaporação. Vídeos e/ou simuladores também são alternativas como objeto de ensino ou como recursos a serem utilizados em atividades propostas aos alunos, sendo importante que se priorize e promova conflitos de ideias e por supostas as concepções prévias errôneas. Os elementos que legitimam um ensino investigativo podem ser utilizados para uma melhor apropriação desses conceitos e processos, entretanto, essas atividades também pode exercer uma função de organizador prévio e/ou avançado, favorecendo uma retenção dos conceitos ou também para uma reconciliação integrativa. Como organizador prévio esses conhecimentos podem ser utilizados em uma aprendizagem mecânica, se for considerado necessário ou definido pelo professor

A organização do conteúdo pode explorar o conceito de calor teoricamente e experimentalmente dentro da modalidade investigativa, buscando identificar as condições para que esse processo ocorra. O MCM pode ser utilizando representando o comportamento dessas partículas, e de como ocorre o processo de evaporação durante as trocas de calor. Na sequenciação de aprendizagem se sugere uma verificação empírica da sensação térmica de frio pelos alunos para reforçar e enriquecer as discussões sobre esse processo. Uma noção sobre sensação térmica como conceito é um elemento que deve ser trabalhando previamente ou durante a atividade experimental com os alunos, vai depender da estratégia de ensino escolhida pelo docente.

A22: ITEM y22

22. Quando o André usa uma bomba para encher o pneu da bicicleta, observa que a bomba fica bem quente. Qual das explicações abaixo parece ser a melhor?

a) Energia foi transferida* para a bomba.

b) Temperatura foi transferida para a bomba.

[A4]: Calor e temperatura são a mesma coisa.

⇒ Por considerar que a temperatura foi transferida e não o calor, ou seja, temperatura e calor são a mesma coisa.

[B3]: Percepções de quente e frio não estão relacionados com a transferência de energia.

⇒ Por não considerar que a sensação de “quente” está relacionado com energia térmica.

[C5]: A temperatura pode ser transferida (*concepção substancialista, a temperatura é uma substância passível de se deslocar de um corpo para outro*).

⇒ Por considerar que a “temperatura” seja uma substância passível de ser transferida.

c) Calor propagou-se das suas mãos para a bomba.

[A1]: O calor é uma substância (*concepção animista*).

⇒ Por considerar que a mão “possui” (substância) calor e ela se “propaga” (possui vontade própria) para a bomba.

[A2]: O calor não é energia.

⇒ Por considerar que o calor é uma substância transferível.

[D2]: Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (*concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não um energia transitória. O frio também como uma substância*).

⇒ Por acreditar que o “calor das mãos” foi “atraído” pela bomba de metal.

d) O metal da bomba provoca o aumento de temperatura.

[B3]: Percepções de quente e frio não estão relacionados com a transferência de energia.

⇒ Por considerar que o aquecimento foi provocado diretamente pelo metal e não por um processo termodinâmico no interior da bomba durante o enchimento do pneu.

[B6]: Um corpo frio não possui a capacidade de aquecer um outro corpo (*somente corpos quentes “possuem calor” e podem aquecer, ou seja, se admite uma concepção ontológica substancialista*).

⇒ Por acreditar que os metais possuem calor e com isso aquecem a si próprios.

[D2]: Metal tem a capacidade de atrair, manter, intensificar ou absorver calor e frio. (*concepção substancialista do calor, sendo algo que pode ser atraído, absorvido, depositado e mantido dentro de um corpo, e não um energia transitória. O frio também como uma substância*).

⇒ Por considerar que ele atraiu e contém calor, e por isso aquece.

Nota: Isso ocorre sem a transferência de energia térmica entre o ar interno com o meio exterior (processo de compressão adiabática), porém, esse aquecimento se dá pela realização de trabalho externo comprimindo o ar e se convertendo na variação da energia interna, e com*

isso aquecendo a bomba, de modo que se possa afirmar que a energia em forma de trabalho de uma força externa na compressão interna do ar se transfere para a bomba e se converte em energia interna, que por fim, parte dela se converte novamente em energia térmica para a bomba.

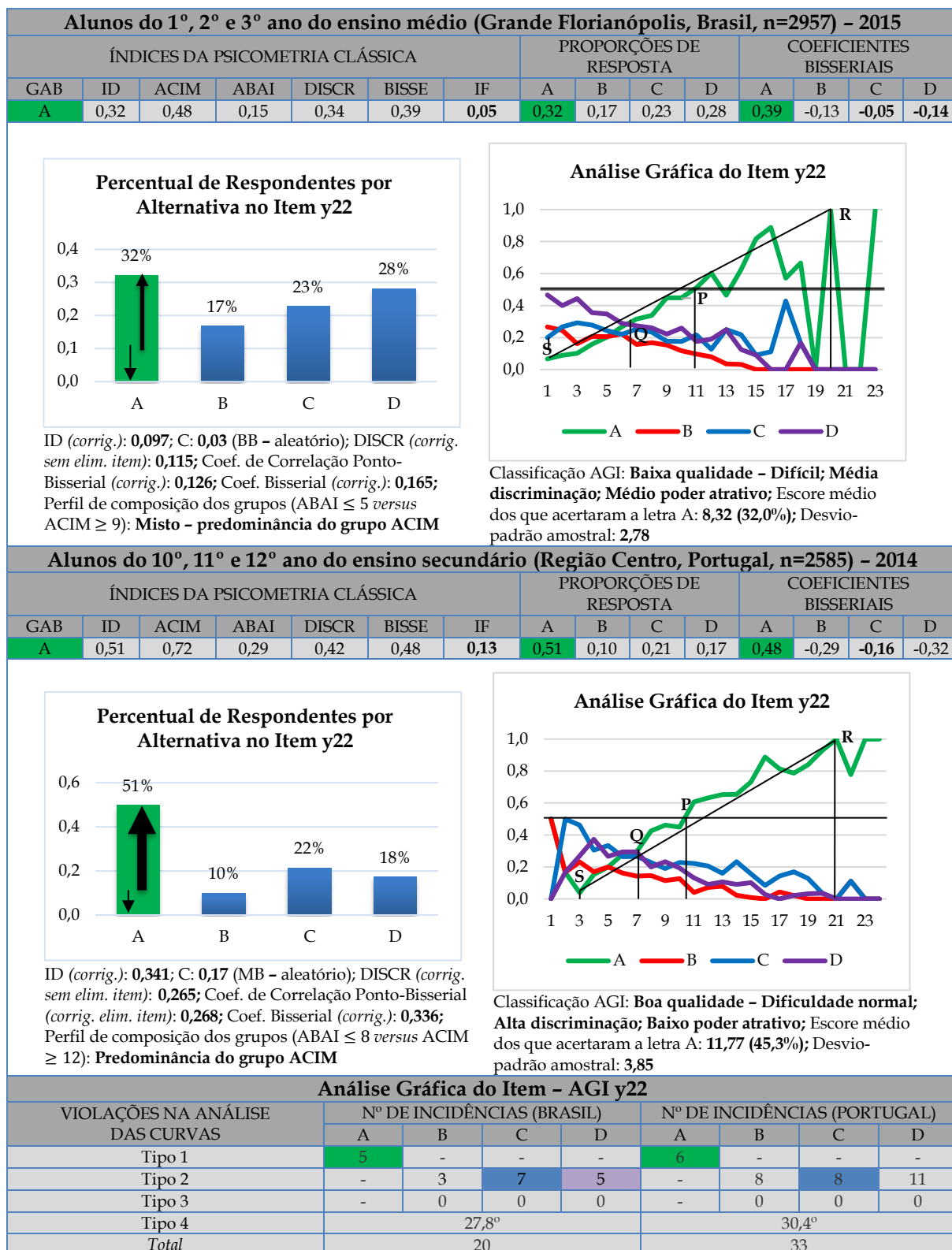
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão sobre a Sensação Térmica explicada pela Propagação do Calor - [Descritor 16]*

PROFICIÊNCIA: *Inferir que a «sensação térmica de “quente”» em uma «bomba durante enchimento de um pneu é causada pelo recebimento de energia térmica que advém do interior da bomba».*

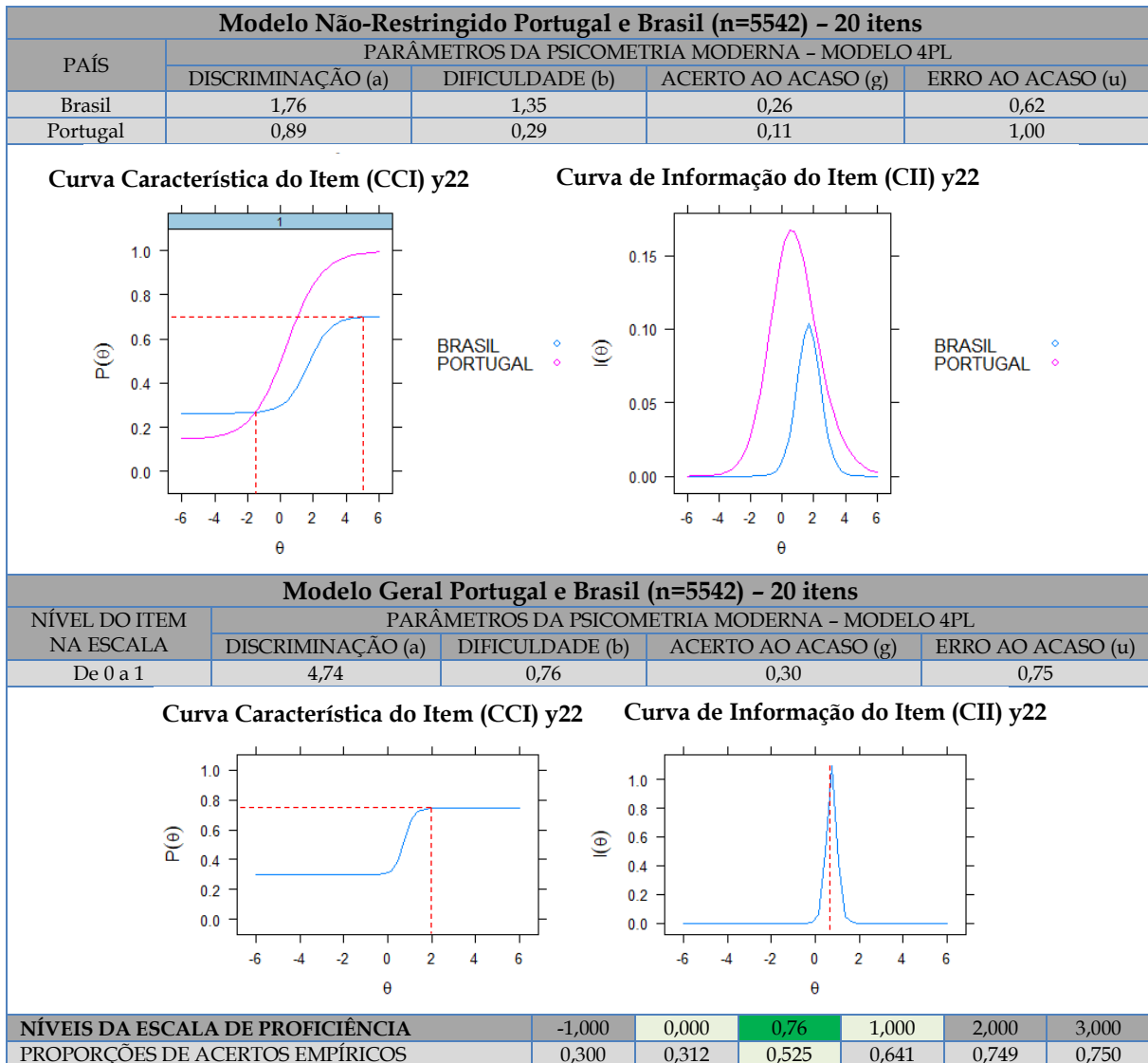
INDÍCIOS DE DEPENDÊNCIA LOCAL: Os itens y9, y10, y14, y16 incidem sobre a necessidade de compreensão sobre a noção de sensação térmica associando-a com as implicações das diferentes propriedades térmicas dos materiais. O item y15 incide sobre a necessidade de compreender e diferenciar a sensação térmica com o conceito de temperatura.

PAINÉL DE ESTATÍSTICAS DO ITEM y22

Coeficientes Estatísticos e Gráficos de Teoria de Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y22

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **bom, sujeito a aprimoramento** ($0,30 \leq DISCR < 0,40$) (Brasil); **Normal/Mediano** (proporção esperada de respondentes 40%; $0,45 \leq ID < 0,55$) e **adequado, disponível a reaplicação** ($DISCR \geq 0,4$) (Portugal).

Os percentuais de acerto total obtidos foram de 32% e 51%, sendo predominantes nos dois contextos diante dos distratores. Na amostra do Brasil se observa em destaque uma atração para as alternativas C e D com 23% e 28%, e em Portugal essas opções ficaram com 22% e 18%.

Os grupos de alto e baixo desempenho que assinalaram a alternativa correta ficaram com 48% e 15% (Brasil), e 72% e 29% (Portugal). O percentual dos alunos portugueses com bom desempenho demonstra um quantitativo elevado de acertos, como se espera, enquanto no Brasil se aproxima da metade dos indivíduos desse grupo, demonstrando uma maior dificuldade para esse item comparativamente.

Os coeficientes bisseriais para a alternativa correta foram de 0,39 e 0,48, sendo uma discriminação relativamente boa, sugerindo um perfil de composição com predomínio de alunos com maiores números de acertos, sendo mais acentuada essa diferença entre os alunos portugueses. Os distratores C e D possuem valores bisseriais iguais a -0,05 e -0,14 (Brasil), e -0,16 e -0,32 (Portugal), indicando assim um perfil de composição misto formado por alunos de bom e mal desempenho, com leve predominância de alunos baixo desempenho, sendo a alternativa D entre os portugueses, com um pouco mais discriminação, porém longe do ideal que seria mais próximo de -1. Observa-se alunos brasileiros com 18 acertos sendo atraídos por esses distratores, e entre os portugueses alunos com 22 acertos optam pelo distrator C. Isso reforça verificar eventuais causas para compreender a razão dessa atração.

Na AGI destacam-se as violações de decrescimento do distrator C em ambos os contextos. Apesar da maior incidência na alternativa D entre os alunos portugueses, o gráfico sugere que sejam oscilações menos intensas. O poder de discriminação foi baixo com um ângulo em cerca de 30° para as duas amostras.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Normal/Mediano** (proporção esperada de respondentes 40%; $-0,51 < b \leq 0,51$) e **bom**, sujeito a aprimoramento, com uma discriminação moderada ($0,65 \leq a \leq 1,35$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros, e uma melhor discriminação, conforme mostra a inclinação e os valores desse parâmetro, apesar de se estimar uma boa discriminação para ambos

os contextos. O acerto ao acaso (*guessing*) foi de 0,26 e 0,11, estando Portugal com valor bem abaixo do esperado que seria de $\frac{1}{4}$ ou 0,25, tendo apenas 11% de chances de acerto sem possuir habilidades suficientes para isso, sugerindo assim que os alunos portugueses foram menos propensos a “chutes”, seja por fadiga, pressa em terminar ou por desconhecimento, entre outros. Isso indica maior coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ e níveis de proficiência alcançadas. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 0,62 e 1,00 (Brasil e Portugal), com isso se estimam 38% de probabilidade de alunos brasileiros cometerem equívocos com $\theta > 5,00$ (aproximadamente), sendo inexistente erros ao caso para os portugueses, de acordo com o ajuste do modelo. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto, com exceção no intervalo $\theta < -1,50$ (aproximadamente), sendo entretanto basicamente uma zona de *guessing* para as habilidades θ no contexto brasileiro.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os portugueses (mais elevado), menor dificuldade (está mais à esquerda) e uma alta precisão das informações do item nas interpretações da escala de proficiência, na qual se distribui em um intervalo de 8,0 desvios-padrões (aproximadamente). A CII dos respondentes brasileiros sugere uma maior dificuldade, boa discriminação e contribuições na escala.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 0 a 1, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento técnico em uma proficiência representada por 5 itens-âncoras y_3 , y_{13} , y_{23} , y_{10} , y_{22} (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (0,76) e discriminação (4,74) classificando o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b$

$\leq 1,27$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 52,5% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso (“*guessing*”) foi estimado em 0,30, um pouco acima do valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 30% de probabilidade de acertar sem possuir uma proficiência adequada. Isso significa perda de coerência em responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ . Foi estimado um índice de erro ao acaso de 0,75, ou seja, 25% de cometerem equívocos por descuido para além de um nível de proficiências $\theta > 2,0$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação para o item y22 em um intervalo próximo de 2,0 desvios-padrões entre os níveis 0 e 2, sendo item com subsídios necessários a ser considerado um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item tem a finalidade aferir uma compreensão intuitiva sobre o processo de transformação adiabática em um contexto de enchimento de pneu na qual se utiliza uma bomba. No caso ele quer compreender o porquê de a bomba ter aquecido durante o enchimento. Espera-se que os alunos compreendam apenas que o fato de estar “quente” é devido a ele ter recebido energia térmica, ou seja, foca apenas no conhecimento sobre a sensação térmica e condições que envolve trocas de calor (variação de energia interna da bomba para as paredes do cilindro metálico). Sabe-se, porém, que é um processo não envolve trocas de calor com o meio externo, em que existe uma transformação de energia por meio de trabalho realizado por uma força externa em aumento de energia interna dentro da bomba, que por sua vez é a responsável pelo aquecimento ou fornecimento de energia para a bomba, mas o item não exige um domínio sobre isso. Para responder o problema basta analisar as alternativas e saber que a sensação de “quente” ocorre quando se recebe energia, sendo esse o traço latente que se mede. Esse fenômeno é regido pela 1ª Lei da termodinâmica que representa uma relação de conservação de energia. A alternativa correta afirma que a causa do aquecimento foi devida a “*energia ter sido transferida para a bomba*”. Essa afirmação pode confundir o respondente, achando que a transferência de energia foi das mãos para a bomba ao invés de seu interior para

bomba (paredes do cilindro), conforme diz a alternativa C. O distrator C afirma que o aquecimento da bomba ocorreu devido o “*calor ter se propagado das mãos para a bomba*”. Esse tipo de equívoco pode estar relacionado no desconhecimento da variação da energia interna na bomba, ou ainda em imaginar uma bomba metálica (geralmente é de metal, e a alternativa D reforça isso) sendo aquecida pela fricção da mão na bomba durante o enchimento.

SUGESTÕES DE ASPECTOS A SEREM APRIMORADOS/ REELABORADOS AO ITEM: A resposta pode fazer referência ao fato que o aquecimento é explicado por transformações adiabáticas ocorridas em seu interior, que no caso transfere a energia para as paredes de metal.

SUGESTÕES METODOLÓGICAS BASEADAS NA PSICOLOGIA DE AUSUBEL COMO CONDIÇÕES DE CONTORNO: O subsunçor identificado está relacionado na compreensão de sobre a sensação térmica e o processo de trocas de calor envolvido entre a bomba e as mãos, apesar de que pela observação das alternativas disponíveis reforça um entendimento minimamente necessário sobre a sensação térmica. É importante enfatizar e se certificar de que há uma compreensão sobre o sentido na qual o calor estar se propagando.

Compreender a situação-problema de sentir a sensação de “quente” no processo de encher um pneu utilizando uma bomba envolve uma série de habilidades cognitivas que vem a caracterizar uma família de situações-problemas, e assim, identificar esse conjunto de habilidades cognitivas e os conteúdos envolvidos dentro desse contexto de vida real atende a concepção sobre a competência científica que precisa ser desenvolvida.

Na concepção de Ausubel a organização do conteúdo do conhecimento térmico deve ser feita de forma hierárquica considerando como conhecimento mais inclusivo o MCM e as Leis da Termodinâmica. O conhecimento que o item exige se torna necessário compreender que a bomba está “quente” apenas porque recebeu calor ou foi transferida energia para ela. Esse tipo de compreensão não envolve uma explicação sobre o porquê ela aqueceu ou de que forma isso aconteceu. Uma compreensão sobre

as causas da ocorrência desse fenômeno remete a entender os processos térmicos envolvidos que fornecem energia a bomba, com isso é preciso seguir para a abordagem dos processos de transformações gasosas, em especial, a sobre o processo de transformação adiabática, na qual a bomba é submetida. A 1ª Lei da Termodinâmica se associa na explicação desse processo e articula uma compreensão mais significativa sobre a origem da energia que aquece a bomba e do sentido de propagação do calor, que é da bomba para as mãos, e assim resultando na sensação de “quente”. Esses conteúdos fazem parte da família de situações-problemas que precisam ser desenvolvidas.

A metodologia na concepção ausubeliana sugere que a necessidade de mecanismos potencialmente significativos, nisso encontram-se meios e recursos didáticos que possam facilitar a articulação do conhecimento, escolhas de técnicas adequadas e instrumentos que favoreçam a aquisição de conceitos. Dentro dessa concepção é considerado como finalidade promover o desenvolvimento de competências científicas baseadas em conceitos térmicos específicos, e a preocupação nesse âmbito envolve explorar confrontar equívocos intuitivos dos alunos, que no caso é não somente entender que é “quente” porque foi fornecido calor para a bomba.

A23: ITEM y23

23. Porque vestimos agasalhos quando está frio?

a) Para manter o frio fora da roupa.

[A3]: O quente e frio são diferentes, em vez de extremidades opostas de um *continuum* (as sensações térmicas são quantificáveis ou com caráter substancialista, ao invés de uma avaliação qualitativa e subjetiva).

⇒ Por considerar o frio uma substância inerte que está fora da roupa e impedida de entrar, ao invés de ser uma sensação térmica.

b) Para gerar calor.

[D9]: Materiais como a lã têm a capacidade de aquecer as coisas (concepção animista, pois entende-se que a lã seja semelhante a um mamífero, sendo uma fonte térmica natural).

⇒ Por acreditar que a temperatura se eleva devido a lã lhe ter aquecido, ou fornecido energia térmica.

c) Para reduzir a perda de calor.

d) As três respostas anteriores estão corretas.

⇒ *Confusão sobre o conceito de calor, incluindo as concepções [A3] e [D9].

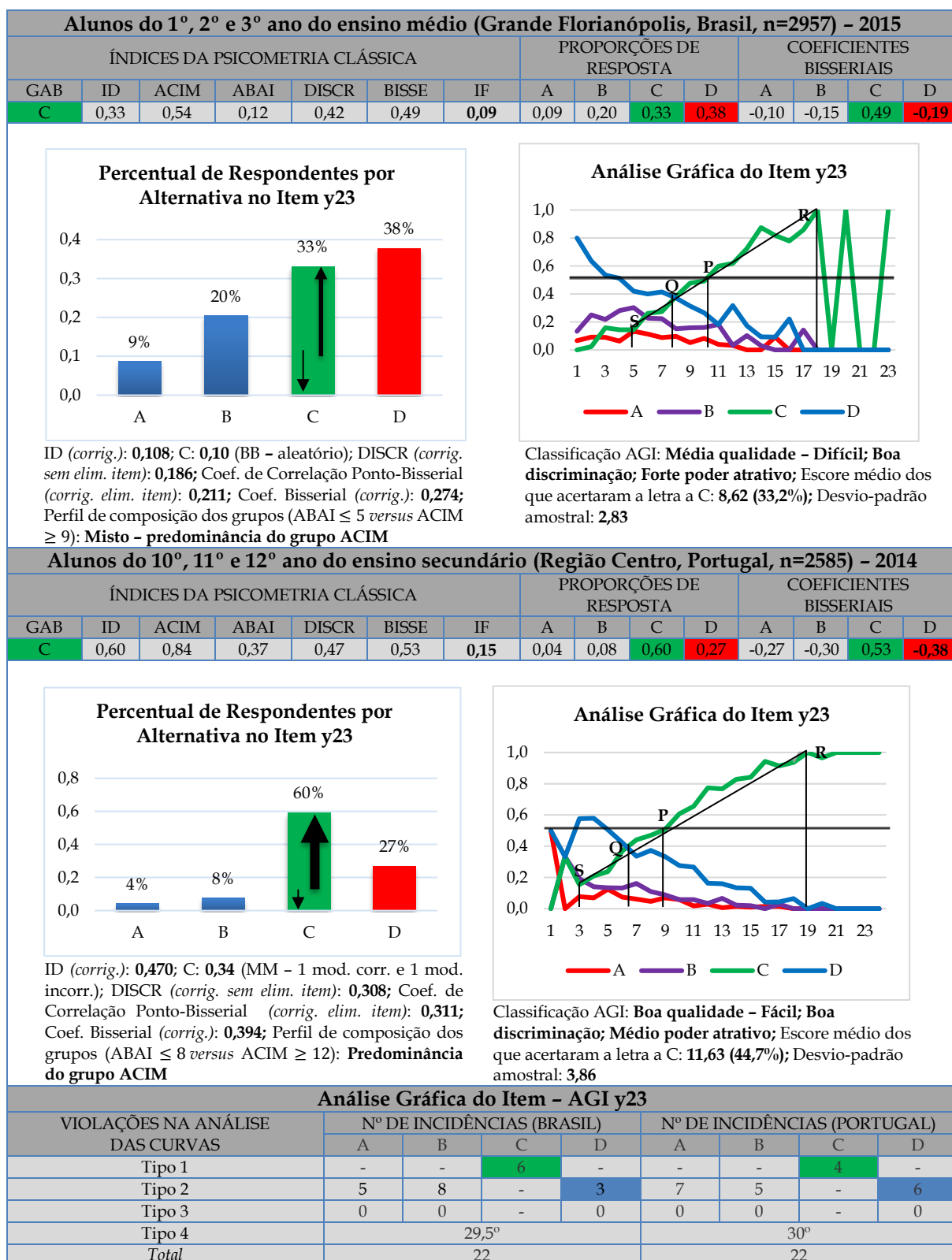
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão* do Comportamento Térmico de acordo com as Propriedades Térmicas - [Descritor 19]

PROFICIÊNCIA: *Inferir* que a «sensação térmica de frio da nossa pele pode ser evitada ao dificultar o processo de trocas de calor com o meio externo, e assim reter a energia térmica que se emite continuamente pelo corpo utilizando um agasalho como isolante térmico», dessa forma «o agasalho serve para reduzir a perda de calor com o ambiente».

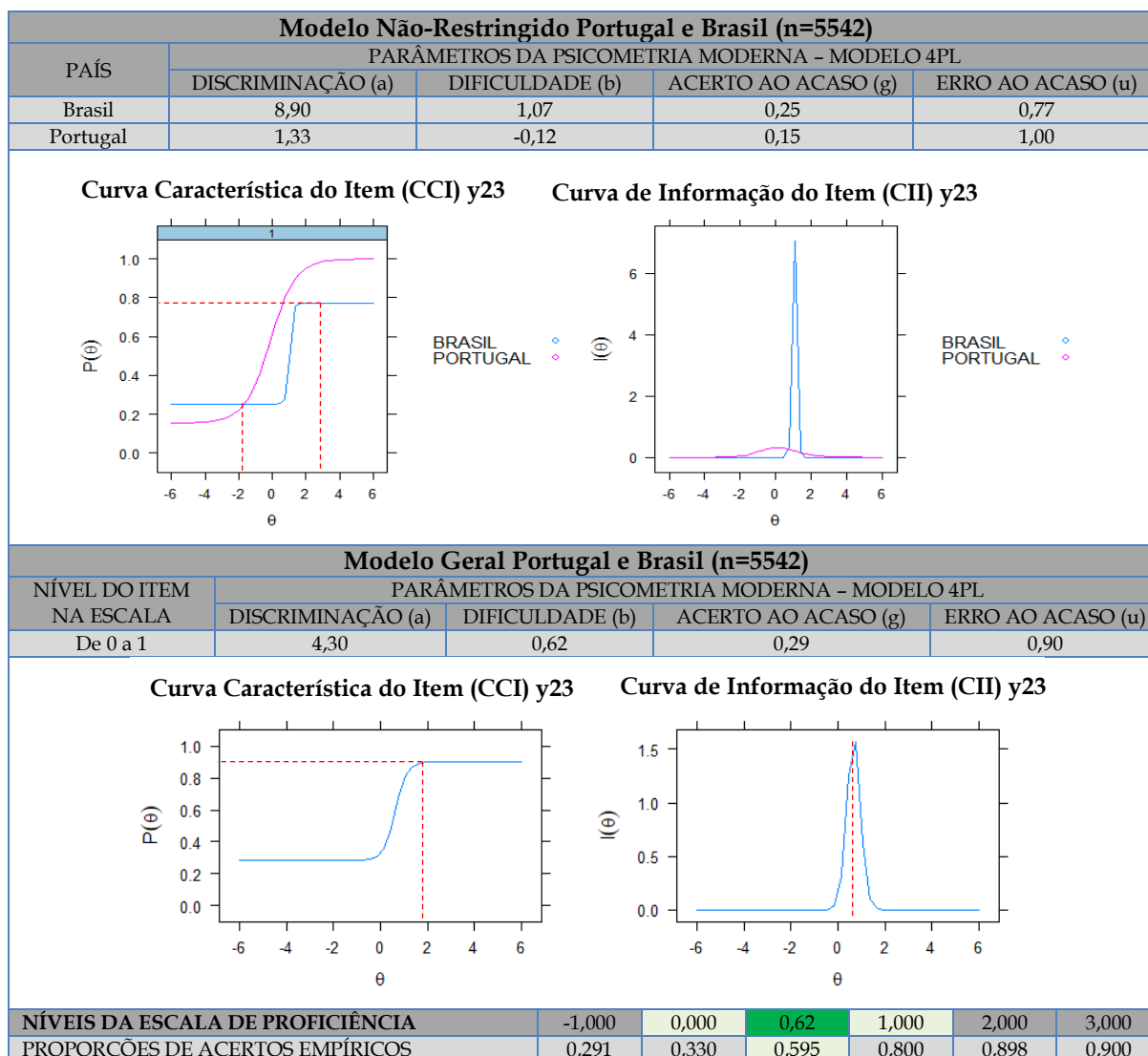
INDÍCIOS DE DEPENDÊNCIA LOCAL: Não foi identificado entre os itens anteriores.

PAINÉL DE ESTATÍSTICAS DO ITEM y23

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y23

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **adequado**, disponível a reaplicação ($DISCR \geq 0,4$) (Brasil); **Fácil** (proporção esperada de respondentes 20%; $0,55 \leq ID < 0,75$) e **adequado**, disponível a reaplicação ($DISCR \geq 0,4$) (Portugal).

Os percentuais de acerto totais em cada contexto foram 33% e 60%, sendo em Portugal a alternativa predominante, e no Brasil sendo o distrator D com 38%. Esse mesmo distrator foi a 2ª opção entre os alunos portugueses com 27%, sinalizando uma importância de verificar essa atração.

Os grupos de maior e menor desempenho ficaram com 54% e 12% (Brasil), e 84% e 37% (Portugal), demonstrando um predomínio de acertos entre os alunos de bom desempenho nos dois contextos.

Os valores dos coeficientes bisseriais foram de 0,35 e 0,53, sendo mais discriminativo entre os portugueses, e esse sugere um perfil de composição de grupos formado predominantemente por alunos com maior desempenho, sendo no Brasil um perfil misto, com leve superioridade de alunos com bom desempenho. O distrator D possui bisserial com valor de -0,19 e -0,38. O bisserial mais baixo entre os brasileiros indica uma composição mista formada por alunos de alto e baixo desempenho, enquanto para os portugueses sendo mais predominante entre alunos com baixo desempenho. O gráfico mostra pouca atração desse distrator para os alunos portugueses de alto desempenho. Para os brasileiros o gráfico mostra certa atração do distrator B entre alunos com 17 acertos.

A AGI aponta mais números de oscilações entre os alunos brasileiros que acertaram a alternativa do que os portugueses, isso revela mais incertezas para opção. Destaca-se as incertezas no distrator B entre os brasileiros com relação as violações do tipo 2 em que deveria haver um decréscimo mais linear. O poder de discriminação foi baixo formando um ângulo em torno de 30° para ambas as amostras, tendo um intervalo de discriminação mais alargado para o grupo de alunos portugueses.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Normal/Mediano** (proporção esperada de respondentes 40%; $-0,51 < b \leq 0,51$) e **bom**, sujeito a aprimoramento, com uma discriminação moderada ($0,65 \leq a \leq 1,35$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior para os brasileiros (mais à direita), e com melhor discriminação conforme a elevada inclinação e os valores desse parâmetro. O índice de acerto ao acaso (*guessing*) foi de 0,25 e 0,15, estando Portugal com valor bem abaixo do esperado de $\frac{1}{4}$ ou 0,25, tendo apenas 15% de chances de acertar sem possuir uma habilidade suficiente para isso, sugerindo

assim que os alunos portugueses foram menos propensos a “chutes”, seja por fadiga, pressão em terminar ou por desconhecimento, entre outros. Isso indica maior coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ e níveis de proficiência alcançadas. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 0,77 e 1,00 (Brasil e Portugal), com isso se estimam 23% de probabilidade de alunos brasileiros cometerem equívocos com $\theta > 2,00$ (aproximadamente), sendo inexistente erros ao caso para os portugueses, de acordo com o ajuste do modelo. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto, com exceção no intervalo $\theta < 2,50$ (aproximadamente), sendo entretanto uma zona de *guessing* para as habilidades θ no contexto brasileiro, ou seja, o acerto seria com base na probabilidade de “chute” e não por evidências de habilidades cognitivas disponíveis.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros (mais elevado), porém possuindo uma maior dificuldade (está mais à esquerda), isso representa uma boa precisão das informações do item nas interpretações da escala de proficiência, na qual se distribui em um intervalo curto de 1,0 desvio-padrão (aproximadamente).

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 0 a 1, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento técnico em uma proficiência representada por 5 itens-âncoras y3, y13, y23, y10, y22 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (0,62) e discriminação (4,30) classificando o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em

59,5% entre aqueles que alcançam esse nível de habilidade θ . O índice de acerto ao acaso (“*guessing*”) foi estimado em 0,29, um pouco acima do valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 29% de probabilidade de acertar sem possuir uma proficiência adequada. Isso significa perda de coerência em responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ . Foi estimado um índice de erro ao acaso de 0,90, ou seja, 10% de cometerem equívocos por descuido para além de um nível de proficiências $\theta > 2,00$ (aproximadamente). Por fim, a elevação da CII indica uma boa discriminação para o item y23 em um intervalo próximo de 1,5 desvios-padrões entre os níveis 0 e 2, sendo item com subsídios necessários a ser considerado um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item tem a finalidade de que se reconheça o gasalho como isolante térmico e o papel que esse tipo de propriedade exerce, em especial, quando se veste. O entendimento do uso de agasalhos envolve o conhecimento sobre trocas de calor entre os corpos e o ambiente, na qual espera-se que os respondentes compreendam que essa vestimenta evita as trocas de calor com o meio externo, com isso, mantém o corpo aquecido. A alternativa correta afirma que o seu uso serve “*para reduzir a perda de calor*”, no caso, do corpo e o ambiente, em que, supostamente o corpo esteja numa temperatura mais elevada. O distrator D afirma que “*as três respostas anteriores estão corretas*”, ou seja, não nega a alternativa C, porém consideram ainda verdadeiras as alternativas A e B, afirmando que o uso do agasalho serve “*para manter o frio fora da roupa*” e “*para gerar calor*”. Esses equívocos podem estar relacionados na crença de que “frio” e “temperatura” sejam sinônimos, demonstrando uma falta de atenção ou compreensão sobre a sensação térmica, apesar de que em parte, a palavra “manter” remeta a certo entendimento sobre isolamento térmico. Quando se acredita que o agasalho “gere calor”, isso sugere uma falta de compreensão sobre o conceito de calor, ou ainda que isolantes sejam fontes de calor.

SUGESTÕES METODOLÓGICAS BASEADAS NA PSICOLOGIA DE AUSUBEL COMO CONDIÇÕES DE CONTORNO: O subsunçor identificado no entendimento sobre o papel exercido pelos isolantes térmicos, tendo como situação o

uso de agasalhos. Os distratores sugere um entendimento elementar sobre o conceito de temperatura (letra A) e calor (letra B).

Compreender a situação-problema da necessidade no uso de agasalho envolve habilidades cognitivas que caracterizam uma família de situações-problema. Entre elas está a compreensão sobre o processo de trocas de calor existente entre o corpo e o ambiente, na qual o corpo nessas condições supostamente deva estar a uma temperatura interior ao meio externo. Os distratores reforçam a escolha na alternativa correta negando que “frio” e “temperatura” são conceitos diferentes, assim como um isolante não deva ser entendido como fonte térmica. Dentro disso se exige habilidades cognitivas sobre esses conceitos dentro dessa situação.

Na concepção de Ausubel a organização do conteúdo remete a necessidade de uma noção prévia sobre os conceitos de temperatura, calor e sensação térmica, assim como no conhecimento factual de que o corpo emite calor para o ambiente espontaneamente e que isolantes térmicos sejam meios que possuem propriedades que dificultam esse processo. Considera-se que a partir desse tipo de entendimento se torna favorável uma compreensão sobre o papel exercido no uso de agasalhos. O conhecimento mais inclusivo está na compreensão sobre isolantes térmicos e suas propriedades, o princípio das trocas de calor e da sensação térmica, que por fim seguem para os conceitos básicos de calor e temperatura. Esse tipo de sistematização tende a favorecer a associação do subsunçor com a compreensão da situação-problema.

Entre os mecanismos potencialmente significativos, pode-se utilizar recursos didáticos que demonstrem o papel exercido pelos isolantes térmicos, como por exemplo caixas de isopor contendo gelo e outro um objeto sendo fonte térmica. O uso de termômetros comparando a temperatura interior e exterior como reforçar o papel exercido pelos isolantes térmicos, e em seguida explorar uma compreensão sobre as suas propriedades. O MCM pode auxiliar um entendimento do porquê essas partículas dificultam o processo de condução térmica, com isso a temperatura interna tende a se manter constante, análogo ao uso de agasalhos. Explorar o funcionamento

de uma garrafa térmica também uma alternativa. A elaboração de tarefas potencialmente significativas pode ser feita priorizando os elementos fundamentais de um ensino investigativo, no caso, fazendo os alunos questionarem, realizarem previsões, descreverem o funcionamento desses aparatos com base nas observações dos termômetro, e assim promover discussões entre os alunos.

A24: ITEM y24

24. Vítor pega alguns picolés do congelador, onde os tinha colocado no dia anterior, e diz que os palitos de madeira estão a uma temperatura mais alta do que a parte do “sumo gelado” que têm o sabor.

a) Débora diz: “Está certo, porque os palitos de madeira não ficam tão frios quanto o sumo gelado”.

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido).

⇒ Por considerar acreditar que a madeira não tende a mesma temperatura do sumo gelado.

[D1]: A temperatura é uma propriedade de um material ou objeto particular. (uma característica específica para cada tipo de material) – por considerar que a sensação de “quente” na madeira está associada a uma temperatura mais elevada comparativamente com o metal.

⇒ Por acreditar que a madeira possui temperatura própria e que não será igual ao do sumo gelado.

b) Ivan diz: “Está certo, porque o sumo gelado contém mais frio do que a parte de madeira”.

[A3]: O quente e frio são diferentes, em vez de extremidades opostas de um continuum (as sensações térmicas são quantificáveis ou com caráter substancialista, ao invés de uma avaliação qualitativa e subjetiva).

⇒ Por acreditar que o frio seja um propriedade do corpo (“sumo gelado”) ao invés de ser uma sensação térmica.

[C6]: Objetos com temperaturas diferentes, que estão em contato entre si, ou em contato com o ar à temperatura diferente, não necessariamente tendem a atingir uma mesma temperatura. (a tendência natural de que um processo termodinâmico alcance um estado de equilíbrio térmico não é compreendido)

⇒ Por considerar acreditar que o sumo gelado comparado com a madeira, mesmo estando em contato com um mesmo ambiente, está mais frio (no caso, “contém” mais frio)

[D1]: A temperatura é uma propriedade de um material ou objeto particular. (uma característica específica para cada tipo de material) – por considerar que a sensação de “quente” na madeira está associada a uma temperatura mais elevada comparativamente com o metal.

⇒ Por acreditar que o sumo gelado possui temperatura própria e inferior à da madeira.

c) Márcio diz: “Está errado, eles apenas parecem diferentes porque o palito contém mais calor”.

[A3]: O quente e frio são diferentes, em vez de extremidades opostas de um continuum (as sensações térmicas são quantificáveis ou com caráter substancialista, ao invés de uma avaliação qualitativa e subjetiva).

⇒ Por acreditar que o calor seja um propriedade do corpo (“palito de madeira”) ao invés de ser uma sensação térmica.

[D1]: A temperatura é uma propriedade de um material ou objeto particular. (uma característica específica para cada tipo de material) – por considerar que a sensação de “quente” na madeira está associada a uma temperatura mais elevada comparativamente com o metal.

⇒ Por acreditar que o sumo gelado possui temperatura própria e inferior à da madeira, no entanto, se justifica que a temperatura elevada do palito seja porque contém mais calor (possível concepção substancialista subjacente).

d) Ana diz: “Eu acho que eles estão à mesma temperatura porque estão juntos”. Indique com qual deles você concorda.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão* do Comportamento Térmico de acordo com as Propriedades Térmicas – [Descritor 17]

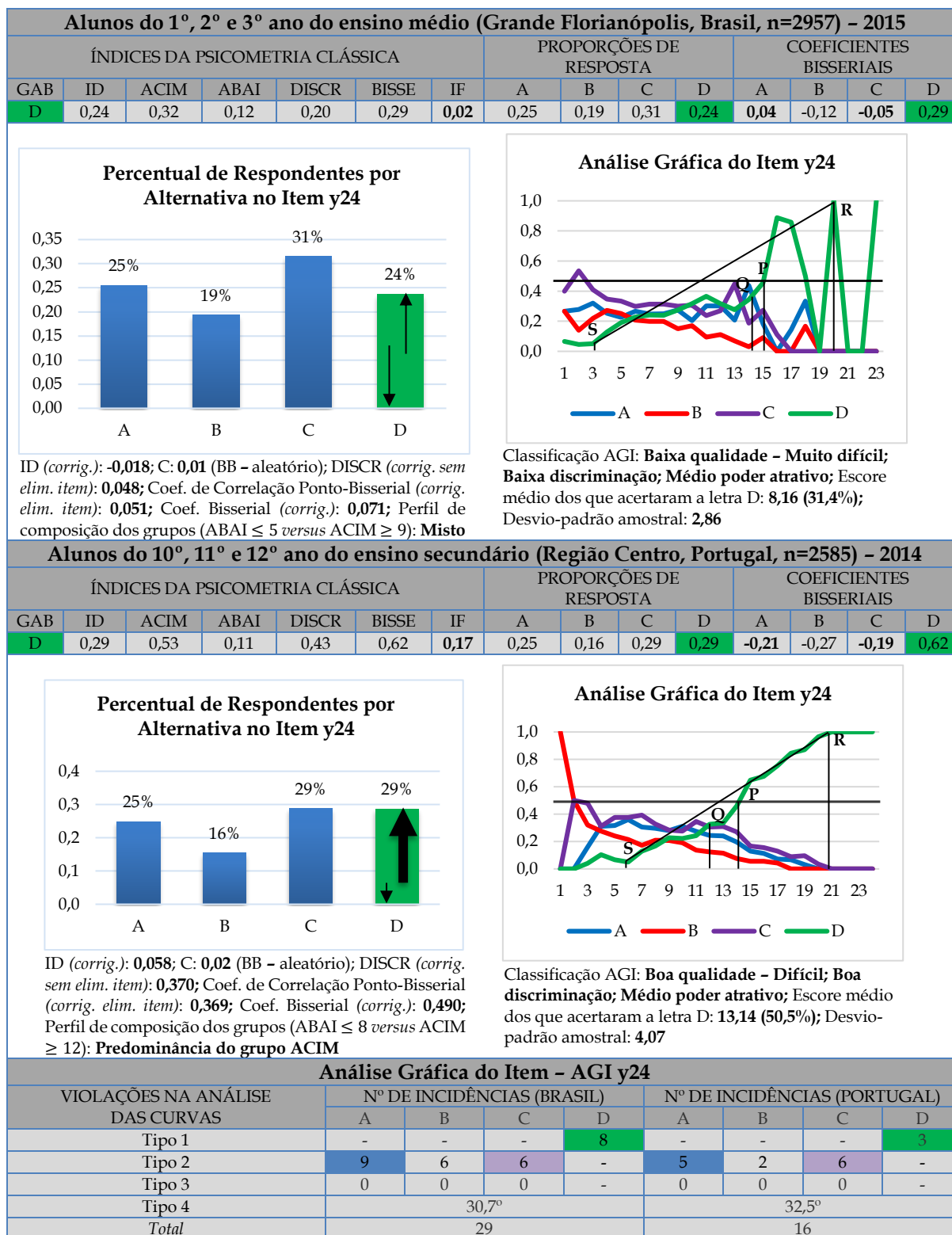
PROFICIÊNCIA: *Inferir* uma «igualdade de temperatura entre corpos com diferentes propriedades com base na Lei Zero da Termodinâmica», «no caso, entrando em contato físico [sensação térmica] com os palitos de madeira do picolé e o sumo gelado».

INDÍCIOS DE DEPENDÊNCIA LOCAL: Os itens y9, y10, y14 e y16 envolvem uma necessidade de determinar intuitivamente a temperatura entre corpos com propriedades térmicas de diferentes (plástico, metal, vidro e madeira) tendo como parâmetro de juízo a sensação térmica. Considera-se que o item y14 busca aferir relativamente com o mesmo traço latente do item y24. Como o item y15 também remete a uma compreensão sobre a sensação térmica, considera-se que esse entendimento supostamente possa ser transferido na tomada de decisão do y24.

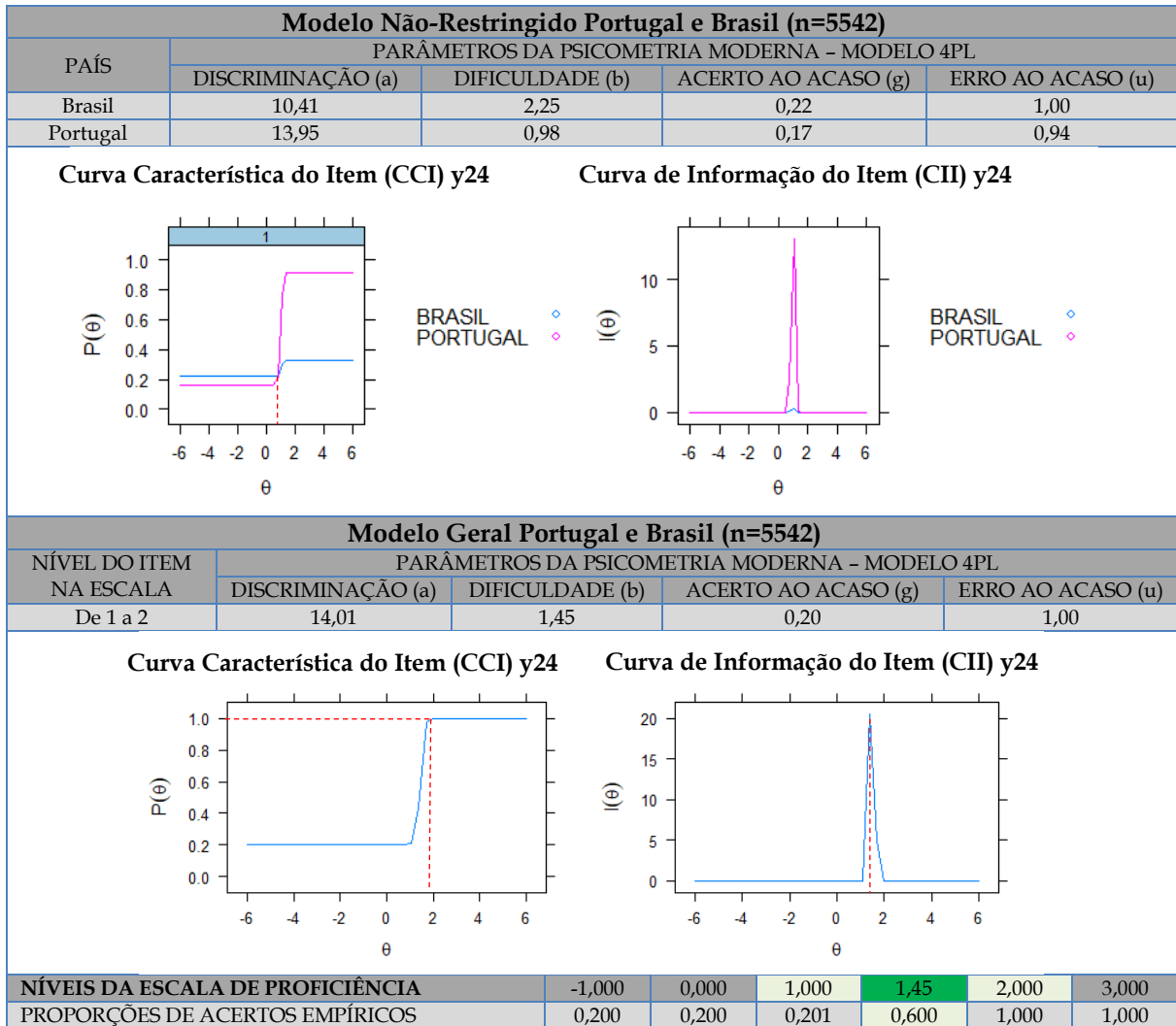
Considera-se a necessidade inferir compreendendo uma independência da sensação térmica, pois os palitos de madeira estão à mesma temperatura que o sumo gelado e suas temperaturas não devem ser avaliadas com base na sensação térmica.

PAINÉL DE ESTATÍSTICAS DO ITEM y24

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y24

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $0,05 \leq ID < 0,25$) e **marginal, sujeito a reelaboração** ($0,20 \leq DISCR < 0,30$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$); e **adequado, disponível a reaplicação** ($DISCR \geq 0,4$) (Portugal).

Os percentuais totais de acerto entre os contextos foram de 24% para os brasileiros sendo a 3ª opção, e entre os portugueses 29%, igualando com o distrator C. Observa-se um certo padrão comparativamente para as duas amostras, sem predomínio da alternativa correta e uma distribuição percentual na escolha entre todas as alternativas, isso evidencia problemas no item e/ou dúvidas entre os respondentes

diante das escolas disponíveis. O distrator C fica em destaque seguido da letra A para ambos os contextos, indicando um comportamento funcional do item de modo correspondente.

Os grupos de maior e menor desempenho alcançaram percentuais de 32% e 12% (Brasil), 53% e 11% (Portugal). A diferença percentual desses grupos entre os portugueses é maior, ou seja, apresenta uma melhor discriminação. Observa-se uma atração dos distratores A e B para alunos brasileiros com desempenho de 18 acertos.

O bisseriais de acerto foram de 0,29 e 0,62, sendo mais discriminativo para os portugueses, de acordo com o índice de discriminação referente a diferença percentual entre os grupos de alto e bom desempenho. Isso significa que quem mais acertou para portugueses fazem parte do grupo de alto desempenho, como se deseja, fato que não aconteceu entre os brasileiros. Os bisseriais alcançados para as letras A, B e C foram de 0,04, -0,12 e -0,05 entre os brasileiros, e -0,21, -0,27 e -0,19 entre os portugueses. Em geral, os perfis de composição entre os grupos de alto e bom desempenho que assinalaram essas alternativas se apresentam de modo misto, ou seja, sem um predomínio mais significativo para cada lado, apesar de que se composto um pouco mais de indivíduos com baixo desempenho entre os portugueses, como deveria ser, mas ainda longe do valor de -1 que seria o ideal.

A AGI revela uma maior incerteza para os alunos que optaram pela alternativa correta (violações do tipo 1) comparado com os portugueses. E entre os distratores se destaca a alternativa A, na qual apresenta 8 oscilações sendo o de maior proporção para os alunos brasileiros com nível de 18 acertos. O poder de discriminação ficou próximo de 31° para ambas as amostras, sendo baixa.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,52 < b \leq 1,27$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior para os brasileiros (mais à direita), porém o contexto português apresenta uma melhor discriminação conforme a elevada inclinação e os valores desse parâmetro. O índice de acerto ao acaso (*guessing*) foi de 0,22 e 0,17, estando Portugal com valor abaixo do esperado de $\frac{1}{4}$ ou 0,25, tendo apenas 17% de chances de acertar sem possuir uma habilidade suficiente para isso, sugerindo assim que os alunos portugueses foram menos propensos a “chutes”, seja por fadiga, pressa em terminar ou por desconhecimento, entre outros. Isso indica maior coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ e níveis de proficiência alcançadas. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos. Os índices desse parâmetro para cada grupo foram de 1,00 e 0,94 (Brasil e Portugal), com isso se estimam apenas 4% de probabilidade de alunos portugueses cometerem equívocos, sendo inexistente erros ao caso para os portugueses, de acordo com o ajuste do modelo. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto, com exceção no intervalo $\theta < 1,00$ (aproximadamente), sendo entretanto uma zona de *guessing* para as habilidades θ no contexto brasileiro, ou seja, o acerto seria com base na probabilidade de “chute” e não por evidências de habilidades cognitivas disponíveis.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os portugueses (mais elevado) em um intervalo curto de 1,0 desvio-padrão (aproximadamente), porém os brasileiros possuem uma maior dificuldade, conforme mostram os valores.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 1 a 2, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y_{16} , y_{14} , y_{26} , y_{24} e y_9 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (1,45) e discriminação (14,01) classificando o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 60% entre aqueles que alcançam esse nível de habilidade θ . O índice de acerto ao acaso (“*guessing*”) foi estimado em 0,20, um pouco abaixo do valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 20% de probabilidade de acertar sem possuir uma proficiência adequada. Isso significa uma boa coerência em responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ . Não foi estimado um índice de erro ao acaso, ou seja, que se cometa equívocos por distração ou descuido. Por fim, a elevação da CII indica uma boa discriminação para o item y24 em um intervalo próximo de 1,0 desvio-padrão entre os níveis 1 e 2, sendo item com subsídios necessários a ser considerado um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item possui a finalidade de aferir intuitivamente qual a temperatura mais elevada entre dois corpos diferentes (palito de madeira e o sumo gelado que têm o sabor) com base na sensação térmica diante em um estado de equilíbrio térmico. Espera-se que os respondentes compreendam que os corpos estejam em equilíbrio térmico e com isso possuam a mesma temperatura, apesar da sensação térmica ser distinta comparativamente entre eles devido as suas diferentes propriedades térmicas em conduzir/transfere o calor. Na alternativa correta, Ana diz: “*Eu acho que eles [palito de madeira e o sumo gelado] estão à mesma temperatura porque estão juntos*”, sugerindo uma compreensão adequada sobre o equilíbrio térmico, de acordo com a Lei Zero da Termodinâmica. No distrator A Débora diz: “*Está certo, porque os palitos de madeira não ficam tão frios quanto o sumo gelado*”. Esse tipo de equívoco pode ser associado a ideia de que determinados materiais alcançam um equilíbrio térmico diferentemente um do outro por terem propriedades diferentes, no caso, o palito de madeira se “*equilibra*” ou para de resfriar a uma temperatura superior à do sumo, mesmo estando sob mesmas condições climáticas. Na alternativa C, Márcio diz que: “*está errado, eles apenas parecem diferentes porque o palito contém mais calor*”. Nesse caso se admite uma percepção térmica

diferenciada, mas justifica-se erroneamente considerando que o calor possa ser uma propriedade do corpo, assim determinados materiais podem “conter” mais ou menos calor que outro.

A25: ITEM y25

25. Gabriela fala sobre um programa de televisão que viu na noite anterior: “Eu vi físicos fazerem supercondutores magnéticos a uma temperatura de $-260\text{ }^{\circ}\text{C}$ ”.

a) Jorge duvida disso: “Deves estar errada. Não se podem ter temperaturas tão baixas”.

⇒ Acredita-se que haja um limite inferior para as temperaturas bem acima de $-260\text{ }^{\circ}\text{C}$. (*): *Não houve correspondência direta com nenhuma concepção do inventário*.

b) Cátia discorda de Jorge e diz: “Sim, podem. Não existe um limite inferior para baixas temperaturas”.

[B8]: Não há limite para a temperatura mais baixa.

⇒ Por acreditar que a temperatura seja uma grandeza física capaz de assumir qualquer valor

c) Leonardo acredita que Gabriela está certa: “Eu acho que o íman estava perto da temperatura mais baixa possível”.

d) Gabriela não está segura: “Eu acho que os supercondutores são bons condutores de calor, logo não se pode esfriá-los a tão baixas temperaturas”.

[D3]: Objetos que facilmente se tornam quentes não são facilmente resfriados.

⇒ Por acreditar que ímas são provavelmente quentes ou se tornam quentes e dificilmente serão resfriados a temperaturas tão baixas. Acredita-se ainda que haja um limite inferior para as temperaturas bem acima de $-260\text{ }^{\circ}\text{C}$.

Indique quem você acha que está correto.

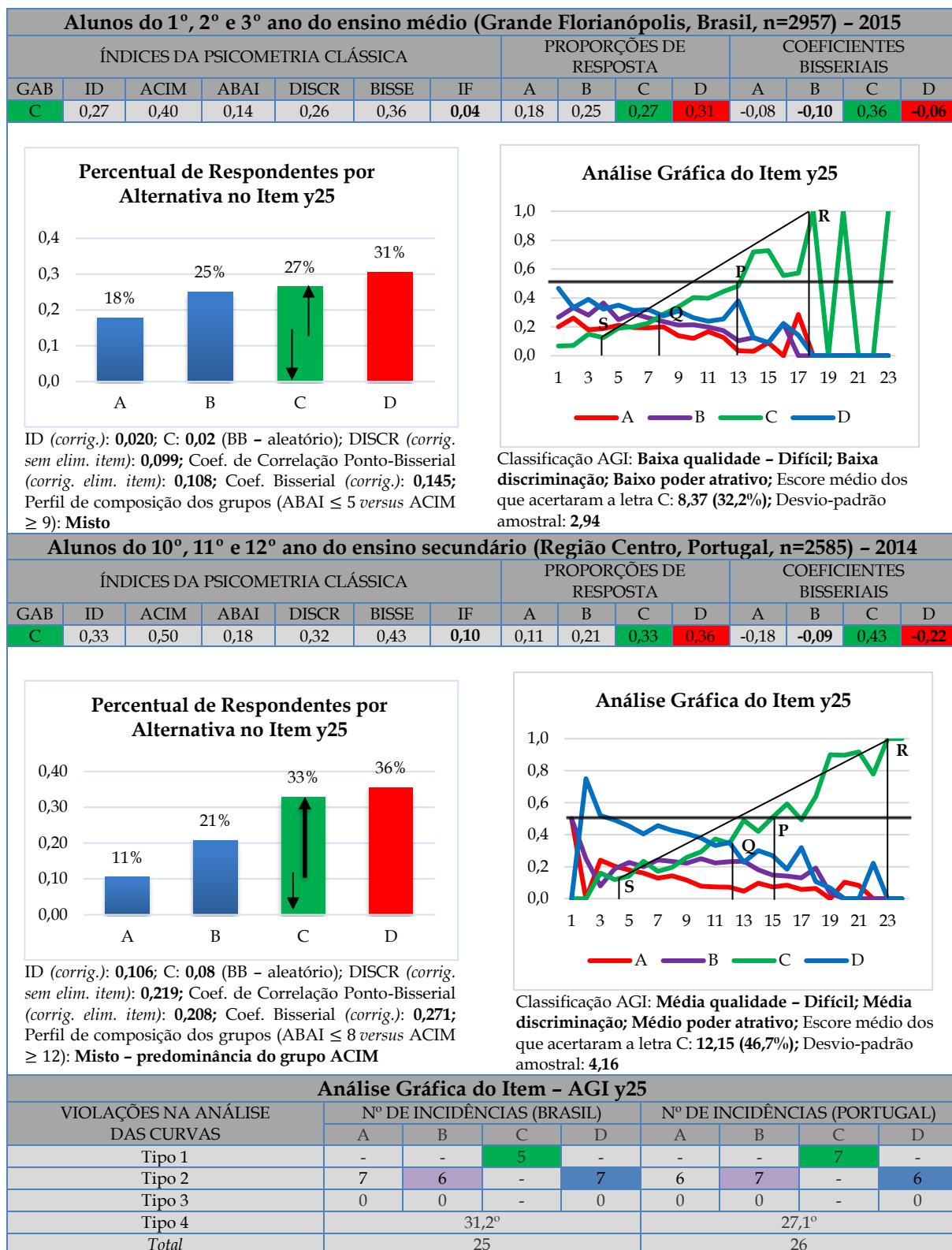
COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Conhecer* a Temperatura de um Corpo ou de um Ambiente em Situações Típicas – [Descritor 2]

PROFICIÊNCIA: *Recordar* que os «corpos podem alcançar até a temperatura hipotética termodinâmica mínima do Zero Absoluto, ou seja, 0 K ou $-273,5\text{ }^{\circ}\text{C}$ », no caso, «supercondutores são ímãs construídos por físicos que podem alcançar $-260\text{ }^{\circ}\text{C}$ ».

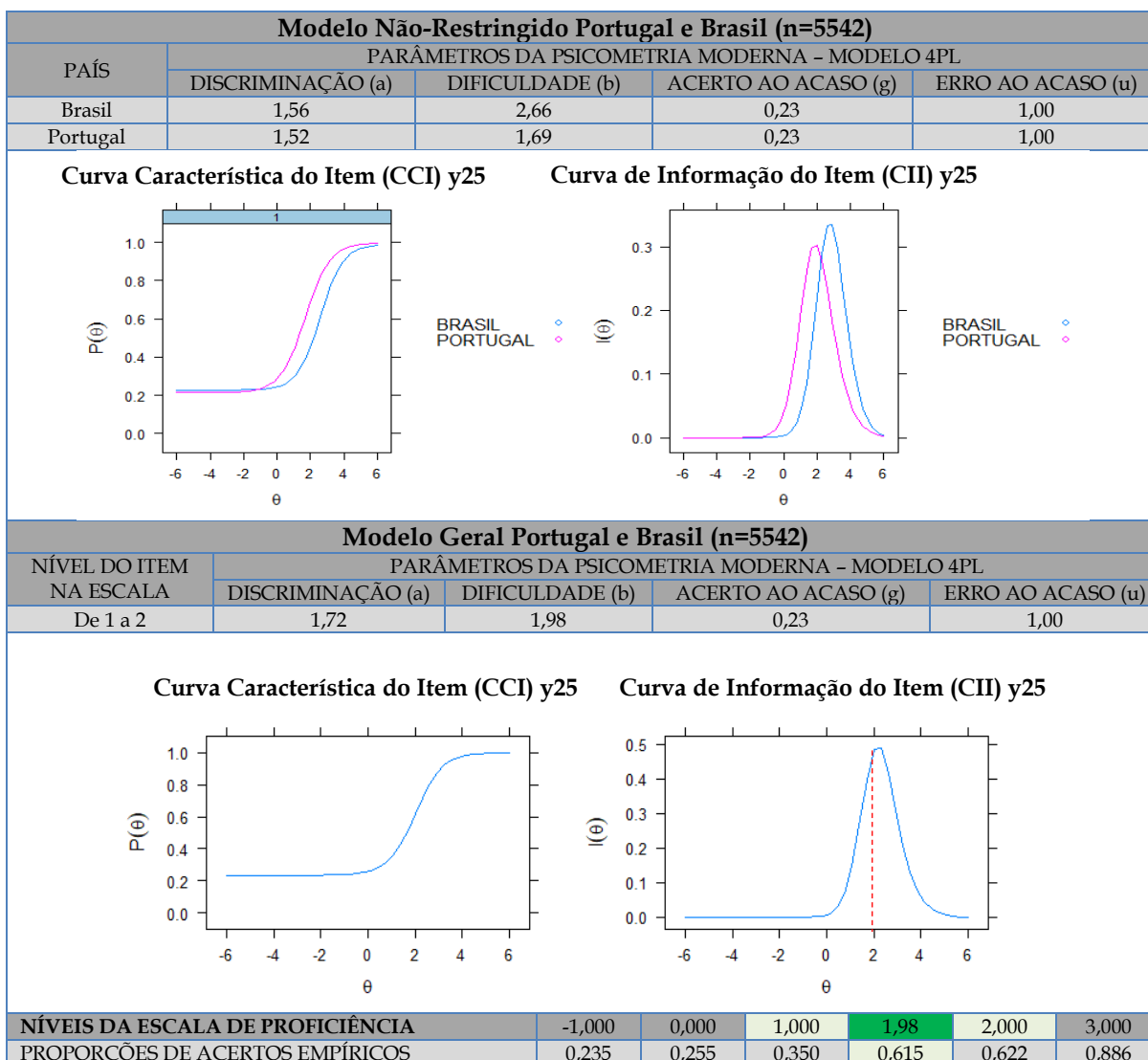
INDÍCIOS DE DEPENDÊNCIA LOCAL: O item y1 busca aferir um conhecimento de que os corpos podem adquirir temperaturas abaixo de $0\text{ }^{\circ}\text{C}$, apesar de que não explora um conhecimento sobre a temperatura mais baixa possível, ainda assim pode ser um indício de dependência local com o item y25.

PAINÉL DE ESTATÍSTICAS DO ITEM y25

Coeficientes Estatísticos e Gráficos de Teoria de Clássica dos Testes (TCT)



Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y25

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **marginal, sujeito a reelaboração** ($0,20 \leq DISCR < 0,30$) (Brasil); **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$); e **bom, sujeito a aprimoramento** ($0,30 \leq DISCR < 0,40$) (Portugal).

Os percentuais de acerto totais no Brasil e em Portugal foram de 27% e 33%, não sendo predominantes em nenhum dos contextos, isso sugere problemas no item, incertezas entre as alternativas disponíveis ou dificuldades de compreensão entre os respondentes. Os percentuais seguem um mesmo padrão, sinalizando um mesmo

comportamento funcional do item, dentre eles o distrator D segue com maior atração com 31% e 36%, e indica uma necessidade de verificar qualitativamente esse tipo de efeito.

Os grupos de maior e menor desempenho foram de 40% e 14% (Brasil), e 50% e 18% (Portugal). Fazendo a diferença percentual entre esses grupos se constata uma maior discriminação entre os alunos portugueses. Observa-se uma atração dos distratores A e D aos respondentes brasileiros com 18 acertos, e para o distrator D aos portugueses com 22 acertos, com isso a opção D apresenta-se em destaque para uma verificação desse tipo de efeito.

Os bisseriais de acerto nos dois contextos foram de 0,36 e 0,43 sugerindo um perfil de composição de grupos misto entre os alunos de alto e baixo desempenho, com uma relativa superioridade entre aqueles com maior acerto, sendo mais discriminativo para os alunos portugueses. Entre os distratores A, B e D foram obtidos bisseriais iguais a -0,08, -0,10 e -0,06 entre os respondentes brasileiros, e -0,18, -0,09 e -0,22 para os portugueses. Isso mostra que não houve um predomínio de alunos de baixo desempenho optando por essas alternativas, apesar de mostrar uma rejeição levemente superior na alternativa D entre os portugueses, porém sendo pouco discriminativo por ser ainda muito próximo de zero.

A AGI mostra um pouco mais de oscilações para a alternativa correta entre os portugueses, em especial para alunos com 17 e 22 acertos (violações do tipo 1). As violações do tipo 2, referentes a verificação de um decrescimento mais linear entre os distratores à medida em que se avançam os níveis de desempenho são relativamente equivalentes nos dois contextos. O poder de discriminação para os portugueses foi de obtido um ângulo de inclinação igual a $27,1^\circ$, sendo um pouco inferior ao dos brasileiros, porém ambos podem ser considerados baixos.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **muito bom, sujeito a aprimoramento**, com uma *discriminação alta* ($1,35 \leq a \leq 1,70$) (Brasil); **Muito difícil** (proporção esperada de

respondentes 10%; $b \geq 1,28$) e **muito bom**, *sujeito a aprimoramento*, com uma *discriminação alta* ($1,35 \leq a \leq 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior para os brasileiros (mais à direita), e com melhor discriminação conforme a elevada inclinação e os valores desse parâmetro, sendo os valores próximos com o contexto português. O índice de acerto ao acaso (*guessing*) coincidiu o mesmo valor para ambos os contextos sendo uma probabilidade igual a 0,23, estando próximo do valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 23% de chances de acertar sem possuir uma habilidade suficiente para isso, seja por fadiga, pressa em terminar ou por desconhecimento, entre outros. Esse valor indica uma relativa coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ e níveis de proficiência alcançadas. Algumas dessas causas também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance no teste com chances de cometerem equívocos, entretanto não foram estimados valores para esse parâmetro. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros (mais elevado), porém possuindo uma maior dificuldade (está mais à esquerda), porém ambos os contextos contribuem para uma boa precisão das informações do item nas interpretações da escala de proficiência, na qual se distribui em um intervalo de 6,0 desvios-padrões (aproximadamente).

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 1 a 2, em que representa um patamar de competência representado por habilidades cognitivas do conhecimento técnico em uma proficiência representada por 5 itens-âncoras y_{16} , y_{14} , y_{26} , y_{24} e y_9 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (1,98) e discriminação (1,72) classificando o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado** para uma *reaplicação*, com uma *discriminação muito alta* ($a > 1,70$). A

dificuldade posiciona o item em um percentual de acertos empíricos estimados em 61,5% entre aqueles que alcançam esse nível de habilidade θ . O índice de acerto ao acaso (*"guessing"*) foi estimado em 0,23, um pouco abaixo do valor esperado que seria de $\frac{1}{4}$ ou 0,25, representando 23% de probabilidade de acertar sem possuir uma proficiência adequada. Isso significa relativa coerência em responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ . Não foi estimado um índice de erro ao acaso (*"upper"*) com um valor de 1,00. Por fim, a elevação da CII indica uma boa discriminação para o item y25 em um intervalo próximo de 4,0 desvios-padrões entre os níveis 0 e 4, indicando boa informação na interpretação do item, porém não apresenta subsídios necessários a ser considerado um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item visa identificar um conhecimento factual sobre a temperatura mais baixa hipoteticamente considerada, no caso o Zero Absoluto. Espera-se que o aluno saiba que em qualquer situação ou estado termodinâmico possível a menor temperatura na escala de graduação Celsius seja de $-273,15\text{ }^{\circ}\text{C}$, no caso da informação divulgada em um programa de televisão sobre os supercondutores se deve admitir a possibilidade de possuírem a temperatura de $-260\text{ }^{\circ}\text{C}$ por esta não alcançar esse limite inferior. Na alternativa correta Leonardo acredita na afirmação de Gabriela ao dizer que acha *"que o íman estava perto da temperatura mais baixa possível"*. No distrator D se assegura que Gabriela esteja com dúvidas em relação a essa afirmação, pois se acha que *"os supercondutores são bons condutores de calor, logo não se pode esfriá-los a tão baixas temperaturas"*. Esse equívoco supostamente pode estar relacionado na crença de que o calor não deva existir quando as temperaturas estejam muito baixas, ou ainda que um corpo ao estar muito *"frio"* perca a eficiência de conduzir o calor, ainda que seja um bom condutor. O distrator A nega a possibilidade de existir algum corpo com essa temperatura, enquanto o B considera possível, porém demonstra não compreender que o menor grau de agitação seria supostamente um estado em que cessa a agitação térmica.

A26: ITEM y26

26. Alguns amigos lembram-se das brincadeiras de crianças. Patrícia diz: “Eu costumava envolver as minhas bonecas em cobertores, mas nunca entendi porque elas não aqueciam”.

a) Marcelo afirma: “É porque os cobertores eram provavelmente maus isoladores térmicos”.

⇒ Considera-se que elas não aqueciam devido ao cobertor de lã não isolar muito bem o calor ou por conduzir de forma eficiente o calor. Subentende-se que as bonecas “possuem” calor, mas o cobertor não as mantém aquecidas, permitindo que elas se esfriem com o ambiente. (*): *Não houve correspondência direta com nenhuma concepção do inventário*.

b) Laura replica: “É porque os cobertores eram provavelmente maus condutores”.

[C8]: Calor flui mais lentamente através de condutores fazendo com que os sinta quentes.

⇒ Por acreditar que os cobertores conduzem lentamente o calor, e por isso não foi percebido as bonecas aquecidas.

c) Fátima replica: “É porque as bonecas eram feitas de um material que não mantém bem o calor”.

[B6]: Um corpo frio não possui a capacidade de aquecer um outro corpo (somente corpos quentes “possuem calor” e podem aquecer, ou seja, se admite uma concepção ontológica substancialista).

⇒ Por acreditar que o material da boneca é feito de um material frio que não retém muito bem o calor.

d) Cecília diz: “É porque as bonecas eram feitas de um material que levava muito tempo a aquecer”.

[D9]: Materiais como a lã têm a capacidade de aquecer as coisas (concepção animista, pois entende-se que a lã seja semelhante a um mamífero, sendo uma fonte térmica natural).

⇒ Por considerar, ainda que de forma lenta, que o cobertor de lã estará aquecendo (fornecendo energia térmica) para a boneca.

[D10]: Alguns materiais são difíceis de serem aquecidos: eles são mais resistentes ao aquecimento.

⇒ Por considerar que o material da boneca demora para ser aquecido.

e) Joana afirma: “Vocês estão todos errados”.

COMPETÊNCIA CIENTÍFICA ASSOCIADA A: *Compreensão* do Comportamento Térmico de acordo com as Propriedades Térmicas – [Descritores 19, 20 e 21]

PROFICIÊNCIA: *Inferir* que «isolantes térmicos não aquecem corpos, apenas dificultam eventuais trocas de calor, sendo eficientes em reter a energia térmica caso o corpo coberto seja um emissor desse tipo de energia em potencial», dessa forma «o uso de cobertores em bonecas não irá aquece-las, pois geralmente elas já se encontram em equilíbrio térmico com o ambiente».

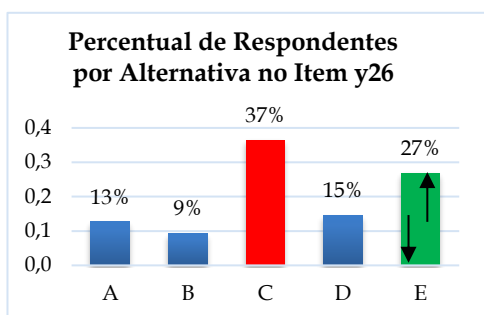
INDÍCIOS DE DEPENDÊNCIA LOCAL: O item y23 procura verificar um conhecimento sobre o papel exercido do agasalho em sua utilização como

vestimenta, para isso se torna necessário reconhecê-lo como um isolante térmico, assim como de suas propriedades térmicas como condições explicativas dos efeitos causados. Sabe-se que quando se é vestido por uma pessoa, o calor cedido pelo corpo é retido pelo agasalho e mantém ele aquecido ao dificultar que se transfira para o meio externo, enquanto para uma boneca nada acontece, pois ela não emite calor como um corpo humano, e geralmente se encontra em equilíbrio térmico com o ambiente em que se encontra. Uma compreensão sobre o item y23 pode ser um indício de dependência local com o item y26.

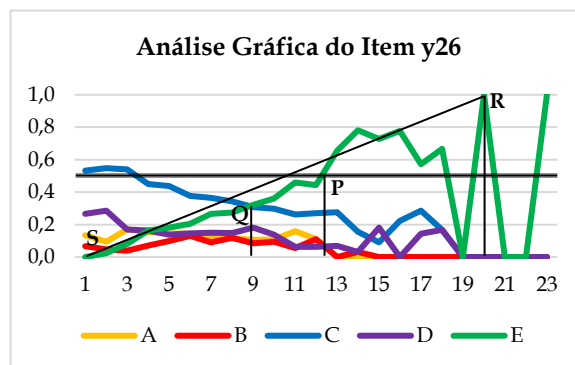
PAINÉL DE ESTATÍSTICAS DO ITEM y26

Coeficientes Estatísticos e Gráficos da Teoria Clássica dos Testes (TCT)

Alunos do 1º, 2º e 3º ano do ensino médio (Grande Florianópolis, Brasil, n=2957) - 2015																
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA					COEFICIENTES BISSERIAIS				
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	E	A	B	C	D	E
E	0,27	0,40	0,13	0,27	0,38	0,05	0,13	0,09	0,37	0,15	0,27	-0,07	-0,01	-0,14	-0,06	0,38

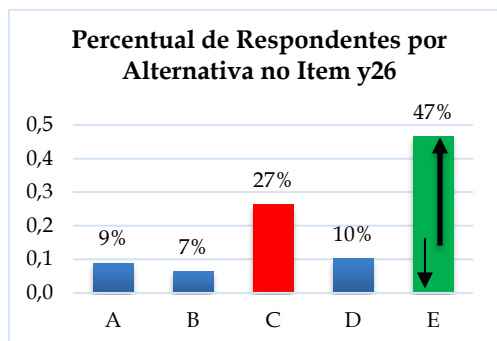


ID (corr.): **0,086**; C: **0,10** (BB - aleatório); DISCR (corr. sem elim. item): **0,115**; Coef. de Correlação Ponto-Bisserial (corr. elim. item): **0,126**; Coef. Bisserial (corr.): **0,169**; Perfil de composição dos grupos (ABAI ≤ 5 versus ACIM ≥ 9): **Misto**

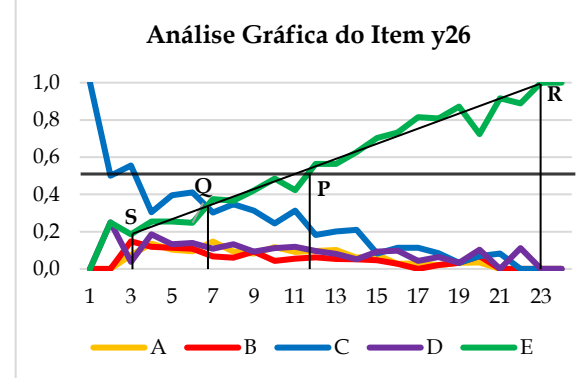


Classificação AGI: **Baixa qualidade - Difícil; Média discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra E: **8,44 (32,5%)**; Desvio-padrão amostral: **2,93**

Alunos do 10º, 11º e 12º ano do ensino secundário (Região Centro, Portugal, n=2585) - 2014																
ÍNDICES DA PSICOMETRIA CLÁSSICA							PROPORÇÕES DE RESPOSTA					COEFICIENTES BISSERIAIS				
GAB	ID	ACIM	ABAI	DISCR	BISSE	IF	A	B	C	D	E	A	B	C	D	E
E	0,47	0,66	0,30	0,36	0,41	0,10	0,09	0,07	0,27	0,10	0,47	-0,12	-0,16	-0,27	-0,11	0,41



ID (corr.): **0,342**; C: **0,21** (MM - 1 mod. corr. e 1 mod. incorr.); DISCR (corr. sem elim. item): **0,208**; Coef. de Correlação Ponto-Bisserial (corr. elim. item): **0,217**; Coef. Bisserial (corr.): **0,273**; Perfil de composição dos grupos (ABAI ≤ 8 versus ACIM ≥ 12): **Misto - predominância do grupo ACIM**

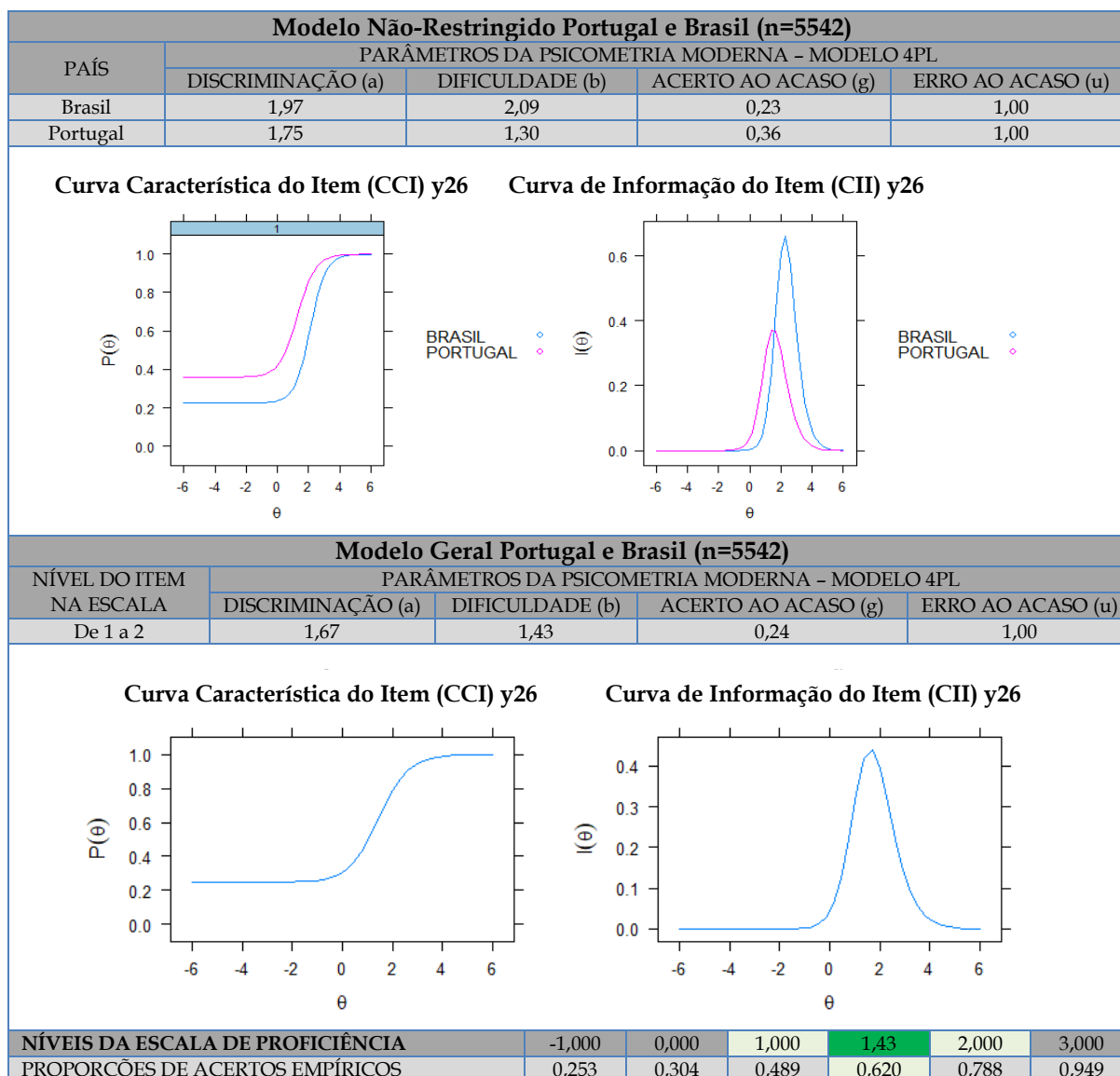


Classificação AGI: **Boa qualidade - Dificuldade normal; Média discriminação; Médio poder atrativo**; Escore médio dos que acertaram a letra E: **11,65 (44,8%)**; Desvio-padrão amostral: **4,00**

Análise Gráfica do Item - AGI y26

VIOLAÇÕES NA ANÁLISE DAS CURVAS	Nº DE INCIDÊNCIAS (BRASIL)					Nº DE INCIDÊNCIAS (PORTUGAL)				
	A	B	C	D	E	A	B	C	D	E
Tipo 1	-	-	-	-	5	-	-	-	-	8
Tipo 2	3	7	5	8	-	10	7	11	11	-
Tipo 3	0	0	0	0	-	0	0	1	0	-
Tipo 4	21,9º					20,2º				
Total	28					47				

Coeficientes Estatísticos e Gráficos da Teoria de Resposta ao Item (TRI)



COMENTÁRIO PEDAGÓGICO DO ITEM y26

Com base na estimação dos parâmetros da TCT os indicadores das dificuldades e discriminação sugerem o item sendo: **Difícil** (proporção esperada de respondentes 20%; $0,25 \leq ID < 0,45$) e **marginal, sujeito a reelaboração** ($0,20 \leq DISCR < 0,30$) (Brasil); **Normal/Mediano** (proporção esperada de respondentes 40%; $0,45 \leq ID < 0,55$) e **bom, sujeito a aprimoramento** ($0,30 \leq DISCR < 0,40$) (Portugal).

De acordo os índices de dificuldade e discriminação o item pode ser classificado como: difícil e regular, sujeito a reelaboração (Brasil); e normal e bom, sujeito a aprimorar (Portugal).

Os percentuais de acerto totais no Brasil e em Portugal foram de 27% e 47%, sendo predominante para os respondentes portugueses diante das outras opções, ao contrário do outro contexto em que foi o distrator C foi o mais atrativo, observando atração das alternativas C e D entre alunos brasileiros com 18 acertos, e entre portugueses uma atração de D para alunos portugueses com 22 acertos.

Os grupos de maior e menor desempenho são representados por um percentual de 40% e 13% (Brasil), e 66% e 30% (Portugal), sendo o contexto português um pouco mais discriminativo com uma maioria de alunos (>50%) de alto desempenho acertando.

Os valores dos coeficientes bisseriais foram de 0,38 e 0,41, mostrando uma discriminação relativamente igual entre os contextos e isso sugere um perfil de composição de grupos misto, com sutil predominância de alunos com bom desempenho, porém ainda distante do ideal que seria mais próximo de 1. Os distratores A, B, C e D possuem bisseriais iguais a -0,07, -0,01, -0,14 e -0,06 entre os brasileiros, e para os portugueses -0,12, -0,16, -0,27 e -0,11, como os valores estão próximos a zero isso mostra que tantos alunos do grupo de alto e baixo desempenho foram atraídos por esses equívocos, sendo um pouco mais acentuado a alunos de baixo desempenho na opção C para os portugueses, porém ainda longe do que se espera que seria próximo de -1, ainda que sendo o mais atrativo entre as outras opções.

A AGI mostra mais violações do tipo 1 entre os portugueses, com destaque para o decréscimo no nível de respondentes com 20 acertos, porém, proporcionalmente, os decréscimos percentuais entre os brasileiros demonstram serem mais intensos. As oscilações entre os distratores também foram mais numerosas entre os portugueses (violações do tipo 2). Uma noção sobre o poder de discriminação sinaliza que ambas as amostras possuem um valor baixo em torno de 20° no ângulo de inclinação.

Com base na estimação dos parâmetros da TRI através do modelo não-restringido o item pode ser classificado como: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Brasil); **Muito difícil** (proporção esperada de

respondentes 10%; $b \geq 1,28$) e **adequado**, com *evidências de ser reaplicado* com uma *discriminação muito alta* ($a > 1,70$) (Portugal).

A *Curva Característica do Item* (CCI) sugere uma dificuldade maior (mais à direita) para os brasileiros, e com melhor discriminação conforme mostra a inclinação e os valores desse parâmetro. O acerto ao acaso (*guessing*) foi de 0,23 e 0,36 (Brasil e Portugal), próximo de um valor esperado de $1/5$ ou 0,20, sugerindo que os alunos foram menos propensos a “chutes”, seja por fadiga, pressa em terminar ou por desconhecimento, entre outros, principalmente os alunos portugueses. Isso significa que foram estimados 36% de probabilidade de alunos portugueses acertarem a alternativa não possuindo habilidade suficiente, sugerindo uma falta de coerência entre responder uma alternativa (fácil, mediana ou difícil) de acordo com suas habilidades θ e níveis de proficiência alcançadas. Algumas dessas causas (ansiedade, cansaço, “pegadinha” ou indução ao erro, etc.) também explicam os erros ao acaso, nesse caso, entretanto, consideram-se alunos de boa performance (alta habilidade θ) no teste com chances de cometerem equívocos. Apesar disso, não foram estimados valores para esse parâmetro. De um modo geral, entre os alunos brasileiros e portugueses de mesma proficiência θ , os portugueses são os que possuem uma maior probabilidade de acerto.

A elevação na *Curva de Informação do Item* (CII) indica uma melhor discriminação para os brasileiros (mais elevado), maior dificuldade (está mais à direita) e uma alta precisão nas interpretações dentro desse contexto para o item, na qual se distribui em um intervalo de 2,0 desvios-padrões (aproximadamente). A CII do contexto português também apresenta boa discriminação e informações a contribui na interpretação da escala de proficiências, assim como na sua precisão, sendo o item y26 apresentando subsídios que o justificam como um item-âncora, de acordo com os critérios estabelecidos.

Na *escala de proficiência* é levado em consideração o contexto geral, ou seja, alunos portugueses e brasileiros. Nisso o item se encontra em um nível de 1 a 2, em que representa um patamar de competência representado por habilidades cognitivas

do conhecimento térmico em uma proficiência representada por 5 itens-âncoras y16, y14, y26, y24 e y9 (em ordem crescente de dificuldade) (ver escala de proficiências).

A CCI apresenta os parâmetros de dificuldade (1,43) e discriminação (1,67) classificando o item sendo: **Muito difícil** (proporção esperada de respondentes 10%; $b \geq 1,28$) e **muito bom**, *sujeito a aprimoramento*, com uma *discriminação alta* ($1,35 \leq a \leq 1,70$). A dificuldade posiciona o item em um percentual de acertos empíricos estimados em 62% entre aqueles que alcançam esse nível de habilidade θ . O acerto ao acaso (“*guessing*”) foi estimado em 0,24, um pouco acima do valor esperado que seria de 1/5 ou 0,20, representando 24% de probabilidade de acertar sem possuir uma proficiência adequada. Não foi estimado um índice de erro ao acaso, que seria de respondentes cometerem equívocos por descuido para além de um nível de proficiências com probabilidade máxima de acertos, ainda que sua proficiência θ se eleve. Por fim, a elevação da CII indica uma boa discriminação para o item y26 em um intervalo próximo de 4,0 desvios-padrões entre os níveis 0 e 4, apresentando informações relevantes na interpretação da escala e contribuindo para a sua precisão, sendo com isso um item com subsídios necessários que o justificam com um *item-âncora* na construção da escala.

DANDO SENTIDO AOS ERROS/DIFICULDADES: O item busca identificar uma compreensão sobre o agasalho como isolante térmico, de que o uso em bonecas de plástico não causa um efeito de aquecimento. Espera-se que o aluno compreenda que as bonecas não transferem calor para o meio externo como ocorrem numa situação análoga com as pessoas. Na alternativa correta Joana afirma que “*todos errados*”, obviamente que apenas faz sentido se forem verificadas as outras opções. Na alternativa A, Marcelo afirma: “*é porque os cobertores eram provavelmente maus isoladores térmicos*”. Esse equívoco sugere que o agasalho seja um bom condutor térmico, com isso pode haver a crença que ele transfira o calor da boneca para o ambiente e com isso não reter a energia térmica que justificaria ela está aquecida. Na letra B, Laura replica: “*é porque os cobertores eram provavelmente maus condutores*”. Nessa afirmação se admite adequadamente o agasalho como um isolante térmico, porém isso não justifica o fato da boneca não ser aquecida, ou seja, é uma afirmação vaga e incompleta. No distrator

C, Fátima replica: *“É porque as bonecas eram feitas de um material que não mantém bem o calor”*. Existem supostamente dois erros implícitos nessa afirmação. O primeiro erro busca explicar o aquecimento da boneca com base nas propriedades físicas do material de que é feita, nisso pode estar implícito a crença de que o agasalho “forneça” calor a boneca, mas ela não o retém por ser de um material que possui uma baixa capacidade de absorver o calor e/ou manter a temperatura. O segundo equívoco declarado está na admissão da possibilidade de retenção do calor (independentemente das propriedades térmicas), sugerindo um entendimento que o calor seja uma propriedade do corpo e não um processo de energia em trânsito. No distrator D, Cecília diz: *“É porque as bonecas eram feitas de um material que levava muito tempo a aquecer”*. Esse equívoco remete possivelmente aos mesmos erros da alternativa C, porém enfatiza que a boneca possua propriedades térmicas que dificultam a absorção do calor e com isso demore para ser aquecida (supostamente pelo agasalho), ou seja, tenha uma baixa condutividade térmica e um elevado calor específico.

B - População e Amostra

B1: População - Região Centro, Portugal.

Tabela 51: Dados da População (Escolas Secundárias) de acordo com o Distrito, Concelho, Nome da Escola e Livros Didáticos de Física e Química Adotados (vigência 2008-2014).

N	Distrito	Concelho	Designação	Distrito	Livro Didáticos de Física e Química			
					FA	FB	QA	QB
1	Aveiro	Águeda	Escola Secundária Marques de Castilho, Águeda	1	9	-	8	-
2	Aveiro	Águeda	Escola Secundária Adolfo Portela, Águeda (*não tem química)	1	8	-	12	-
3	Aveiro	Albergaria-a-Velha	Escola Secundária de Albergaria-a-Velha	1	11	14	12	16
4	Aveiro	Anadia	Escola Básica e Secundária de Anadia	1	11	14	13	16
5	Aveiro	Aveiro	Escola Secundária Homem Cristo, Aveiro	1	10	-	10	-
6	Aveiro	Aveiro	Escola Secundária Dr. Jaime Magalhães Lima, Esgueira, Aveiro	1	11	-	12	-
7	Aveiro	Aveiro	Escola Secundária Dr. Mário Sacramento, Aveiro	1	9	-	8	-
8	Aveiro	Aveiro	Escola Secundária José Estevão, Aveiro	1	11	14	6	16
9	Aveiro	Aveiro	Escola Artística do Conservatório de Música de Calouste Gulbenkian, Aveiro	-	-	-	-	-
10	Aveiro	Estarreja	Escola Secundária de Estarreja	1	9	14	2	16
11	Aveiro	Ílhavo	Escola Secundária de Gafanha da Nazaré, Ílhavo.	1	6	14	13	16
12	Aveiro	Ílhavo	Escola Secundária Dr. João Carlos Celestino Gomes, Ílhavo.	1	11	14	8	16
13	Aveiro	Mealhada	Escola Secundária da Mealhada	1	8	-	13	-
14	Aveiro	Murtosa	Escola Básica e Secundária Padre António Morais da Fonseca, Murtosa	1	8	-	8	-
15	Aveiro	Oliveira do Bairro	Escola Secundária de Oliveira do Bairro	1	11	-	12	-
16	Aveiro	Ovar	Escola Secundária de Esmoriz, Ovar.	1	9	-	13	-
17	Aveiro	Ovar	Escola Secundária Dr. José Macedo Fragateiro, Ovar.	1	8	-	6	-
18	Aveiro	Ovar	Escola Secundária Júlio Dinis, Ovar.	1	9	-	8	-
19	Aveiro	Sever do Vouga	Escola Básica e Secundária de Sever do Vouga	1	8	-	8	-
20	Aveiro	Vagos	Escola Secundária de Vagos	1	8	13	12	14
21	Castelo Branco	Belmonte	Escola Básica e Secundária Pedro Álvares Cabral, Belmonte	2	8	-	6	-
22	Castelo Branco	Castelo Branco	Escola Secundária Amato Lusitano, Castelo Branco	2	8	13	13	16
23	Castelo Branco	Castelo Branco	Escola Básica e Secundária de Alcains, Castelo Branco	2	8	-	8	-
24	Castelo Branco	Castelo Branco	Escola Secundária Nuno Álvares, Castelo Branco	2	11	-	13	-
25	Castelo Branco	Covilhã	Escola Secundária Frei Heitor Pinto, Covilhã	2	11	-	12	-
26	Castelo Branco	Covilhã	Escola Secundária Campos de Melo, Covilhã	2	8	-	13	-

27	Castelo Branco	Covilhã	Escola Secundária Quinta das Palmeiras, Covilhã	2	11	-	8	-
28	Castelo Branco	Fundão	Escola Secundária do Fundão	2	11	-	12	-
29	Castelo Branco	Idanha-a-Nova	Escola Básica e Secundária José Silvestre Ribeiro, Idanha-a-Nova	2	1	-	1	-
30	Castelo Branco	Oleiros	Escola Básica e Secundária Padre António de Andrade, Oleiros.	2	4	-	8	-
31	Castelo Branco	Penamacor	Escola Básica e Secundária Ribeiro Sanches, Penamacor	2	11	14	13	16
32	Castelo Branco	Proença-a-Nova	Escola Básica e Secundária Pedro da Fonseca, Proença-a-Nova	2	11	-	8	-
33	Castelo Branco	Sertã	Escola Secundária da Sertã	2	9	14	13	16
34	Castelo Branco	Vila de Rei	Escola Básica do Centro de Portugal, Vila de Rei	2	10	-	13	-
35	Coimbra	Arganil	Escola Secundária de Arganil	3	8	-	6	-
36	Coimbra	Cantanhede	Escola Secundária de Cantanhede	3	8	14	12	16
37	Coimbra	Coimbra	Escola Secundária Jaime Cortesão, Coimbra	3	8	-	8	-
38	Coimbra	Coimbra	Escola Secundária D. Duarte, Coimbra	3	11	-	8	-
39	Coimbra	Coimbra	Escola Básica e Secundária Quinta das Flores, Coimbra	3	11	13	1	16
40	Coimbra	Coimbra	Escola Secundária Avelar Brotero, Coimbra	3	11	-	12	-
41	Coimbra	Coimbra	Escola Secundária D. Dinis, Coimbra	3	11	14	13	16
42	Coimbra	Coimbra	Escola Secundária Infanta D. Maria, Coimbra	3	10	-	2	-
43	Coimbra	Coimbra	Escola Secundária José Falcão, Coimbra	3	8	-	12	-
44	Coimbra	Figueira da Foz	Escola Secundária Dr. Bernardino Machado, Figueira da Foz	4	8	14	8	16
45	Coimbra	Figueira da Foz	Escola Secundária Cristina Torres, Figueira da Foz	4	5	-	13	-
46	Coimbra	Figueira da Foz	Escola Secundária Dr. Joaquim de Carvalho, Figueira da Foz	4	11	-	8	-
47	Coimbra	Lousã	Escola Secundária da Lousã	4	11	-	12	-
48	Coimbra	Mira	Escola Secundária Dr.ª Maria Cândida, Mira	4	11	-	12	-
49	Coimbra	Miranda do Corvo	Escola Básica e Secundária José Falcão, Miranda do Corvo	4	9	-	8	-
50	Coimbra	Montemor-o-Velho	Escola Secundária de Montemor-o-Velho	4	11	-	13	-
51	Coimbra	Oliveira do Hospital	Escola Secundária de Oliveira do Hospital	4	11	-	8	-
52	Coimbra	Pampilhosa da Serra	Escola Básica de Pampilhosa da Serra	-	-	-	-	-
53	Coimbra	Penacova	Escola Básica e Secundária de Penacova	4	11	-	8	-
54	Coimbra	Soure	Escola Básica e Secundária Martinho Árias, Soure	4	11	14	12	16
55	Coimbra	Tábua	Escola Secundária de Tábua	4	5	-	13	-
56	Coimbra	Vila Nova de Poiares	Escola Básica e Secundária Dr. Daniel de Matos, Vila Nova de Poiares	4	8	-	12	-
57	Guarda	Aguiar da Beira	Escola Básica e Secundária Padre José Augusto da Fonseca, Aguiar da Beira	5	11	-	13	-
58	Guarda	Almeida	Escola Básica e Secundária de Vilar Formoso, Almeida	5	11	-	8	-

59	Guarda	Almeida	Escola Básica e Secundária Dr. José Casimiro Matias, Almeida	5	8	-	8	-
60	Guarda	Celorico da Beira	Escola Básica e Secundária Sacadura Cabral, Celorico da Beira	5	10	-	10	-
61	Guarda	Figueira de Castelo Rodrigo	Escola Secundária de Figueira de Castelo Rodrigo	5	2	-	3	-
62	Guarda	Fornos de Algodres	Escola Básica e Secundária de Fornos de Algodres	5	9	14	8	16
63	Guarda	Gouveia	Escola Secundária de Gouveia	5	8	-	12	-
64	Guarda	Guarda	Escola Secundária Afonso de Albuquerque, Guarda	5	11	-	13	-
65	Guarda	Guarda	Escola Básica e Secundária da Sé, Guarda	5	11	14	13	16
66	Guarda	Mêda	Escola Básica e Secundária de Mêda	5	8	-	8	-
67	Guarda	Pinhel	Escola Secundária de Pinhel	5	11	-	12	-
68	Guarda	Sabugal	Escola Secundária de Sabugal	5	11	-	12	-
69	Guarda	Seia	Escola Secundária de Seia	5	11	14	12	14
70	Guarda	Trancoso	Escola Secundária Gonçalo Anes Bandarra, Trancoso	5	11	-	8	-
71	Leiria	Alvaiázere	Escola Básica e Secundária Dr. Manuel Ribeiro Ferreira, Alvaiázere	6	9	-	12	-
72	Leiria	Ansião	Escola Básica e Secundária Dr. Pascoal José de Mello, Ansião	6	11	-	12	-
73	Leiria	Batalha	Escola Básica e Secundária da Batalha	6	5	14	13	16
74	Leiria	Figueiró dos Vinhos	Escola Secundária de Figueiró dos Vinhos	6	9	14	2	16
75	Leiria	Leiria	Escola Secundária Domingos Sequeira, Leiria	6	11	14	13	16
76	Leiria	Leiria	Escola Básica e Secundária Henrique Sommer, Maceira, Leiria	6	11	-	13	-
77	Leiria	Leiria	Escola Secundária Afonso Lopes Vieira, Leiria	6	11	-	13	-
78	Leiria	Leiria	Escola Secundária Francisco Rodrigues Lobo, Leiria	6	10	-	10	-
79	Leiria	Marinha Grande	Escola Secundária José Loureiro Botas, Vieira de Leiria, Marinha Grande	6	8	-	6	-
80	Leiria	Marinha Grande	Escola Secundária Pinhal do Rei, Marinha Grande	6	11	14	12	16
81	Leiria	Marinha Grande	Escola Secundária Eng. Acácio Calazans Duarte, Marinha Grande	6	11	-	12	-
82	Leiria	Pombal	Escola Básica e Secundária da Guia, Pombal	6	1	14	1	14
83	Leiria	Pombal	Escola Secundária de Pombal	6	8	-	12	-
84	Leiria	Porto de Mós	Escola Secundária de Mira de Aire, Porto de Mós	6	9	-	8	-
85	Leiria	Porto de Mós	Escola Secundária de Porto de Mós	6	10	-	13	-
86	Viseu	Carregal do Sal	Escola Secundária de Carregal do Sal	7	11	-	12	-
87	Viseu	Castro Daire	Escola Secundária de Castro Daire	7	11	14	12	14
88	Viseu	Mangualde	Escola Secundária Dr.ª Felismina Alcântara, Mangualde	7	8	-	8	-
89	Viseu	Mortágua	Escola Secundária Dr. João Lopes de Morais, Mortágua	7	11	14	13	16
90	Viseu	Nelas	Escola Básica e Secundária Eng. Dionísio Augusto Cunha, Canas de Senhorim, Nelas	7	8	-	13	-
91	Viseu	Nelas	Escola Secundária de Nelas	7	11	-	13	-

92	Viseu	Oliveira de Frades	Escola Básica e Secundária de Oliveira de Frades	7	11	14	12	16
93	Viseu	Penalva do Castelo	Escola Básica e Secundária de Penalva do Castelo	7	11	-	13	-
94	Viseu	Santa Comba Dão	Escola Secundária de Santa Comba Dão	7	11	-	13	-
95	Viseu	São Pedro do Sul	Escola Secundária de São Pedro do Sul	7	8	-	13	-
96	Viseu	Sátão	Escola Secundária Frei Rosa Viterbo, Sátão	7	11	-	12	-
97	Viseu	Tondela	Escola Secundária de Molelos, Tondela	-	-	-	-	-
98	Viseu	Tondela	Escola Secundária de Tondela	7	11	14	12	14
99	Viseu	Vila Nova de Paiva	Escola Secundária de Vila Nova de Paiva	7	9	-	8	-
100	Viseu	Viseu	Escola Secundária Alves Martins, Viseu	7	11	-	1	-
101	Viseu	Viseu	Escola Secundária Emídio Navarro, Viseu	7	11	14	10	16
102	Viseu	Viseu	Escola Secundária Viriato, Abraveses, Viseu	7	11	-	12	-
103	Viseu	Vouzela	Escola Secundária de Vouzela	7	8	-	13	-

*O primeiro LDF mais adotado em Portugal continental (vermelho); o segundo (amarelo); terceiro (azul); escolas vocacionais (lilás).

B2: População – Região da Grande Florianópolis, Brasil.

Tabela 52: Identificação de todas as escolas da Região da Grande Florianópolis, de acordo com o quantitativo de alunos (estimado) e modalidade de educação desenvolvida na unidade de ensino (referente a 2014).

Município	N	Unidade Escolar	Quant. De Alunos	Ensino
Àguas Mornas	1	EEB Coronel Antônio Lehmkuhl	379	EF
Àguas Mornas	2	EEB Cons. Manoel Philippi	320	EF
Angelina	3	EEB Nossa Senhora	369	EF
Angelina	4	EEF João Frederico Heck	62	EF
Angelina	5	EEB Norberto Teodoro de Melo	193	EF
Anitápolis	6	EEB Altino Flores	325	EF
Antônio Carlos	7	EEB Altamiro Guimarães	1190	EF
Biguaçu	8	EEB Cônego Rodolfo Machado	709	EF
Biguaçu	9	EEB Prefeito Avelino Muller	495	EF
Biguaçu	10	EEB Prof ^o Alexandre Sérgio Godinho	503	EF
Biguaçu	11	EEB Prof ^o José Brasília	688	EF
Biguaçu	12	EEF Teófilo Teodoro Regis	134	EF
Biguaçu	13	EEM Prof ^a Maria da Glória Viríssimo de Faria	854	EM
Biguaçu	14	EEB Prof ^a Eloísa Maria Prazeres de Faria		
Biguaçu	15	EEB Emérita Duarte Silva e Souza	791	EF
Biguaçu	16	EEB Prof ^a Tânia Mara Faria e Silva Locks	922	EF
Biguaçu	17	EEB Joaquim João Cardoso	118	EF
Biguaçu	18	EEB Prof ^a Maria de Lourdes Scherer	193	EF
Biguaçu	19	EEF Santo Antônio	184	EF
Biguaçu	20	EEF Areias de Cima	95	EF
Biguaçu	21	EEF Santa Cruz	16	EF
Biguaçu	22	EIEF Kakupe	13	EF
Biguaçu	23	EIEF Taguató	28	EF
Biguaçu	24	EIEF Wherá Tupã – Poty Djá	50	EF
Florianópolis	25	CEJA (Centro de Educação de Jovens e Adultos)	1830	EJA
Florianópolis	26	CEDUP- Dr. Jorge Lacerda	139	Técnico
Florianópolis	27	EEB Aderbal Ramos da Silva	1190	EF
Florianópolis	28	EEB América Dutra Machado	424	EF
Florianópolis	29	EEM Prof. Acácio Garibaldi São Thiago	139	EM
Florianópolis	30	EEB Prof ^o Anísio Teixeira (EM PROCESSO DE EXTINÇÃO FUNCIONANDO NA EEB JULIO DA COSTA NEVES)	0	EM
Florianópolis	31	EEM Antônio Paschoal Apóstolo	339	EM
Florianópolis	32	EEB Prof^o Aníbal Nunes Pires	378	EF
Florianópolis	33	EEF Baldicero Filomeno	136	EF

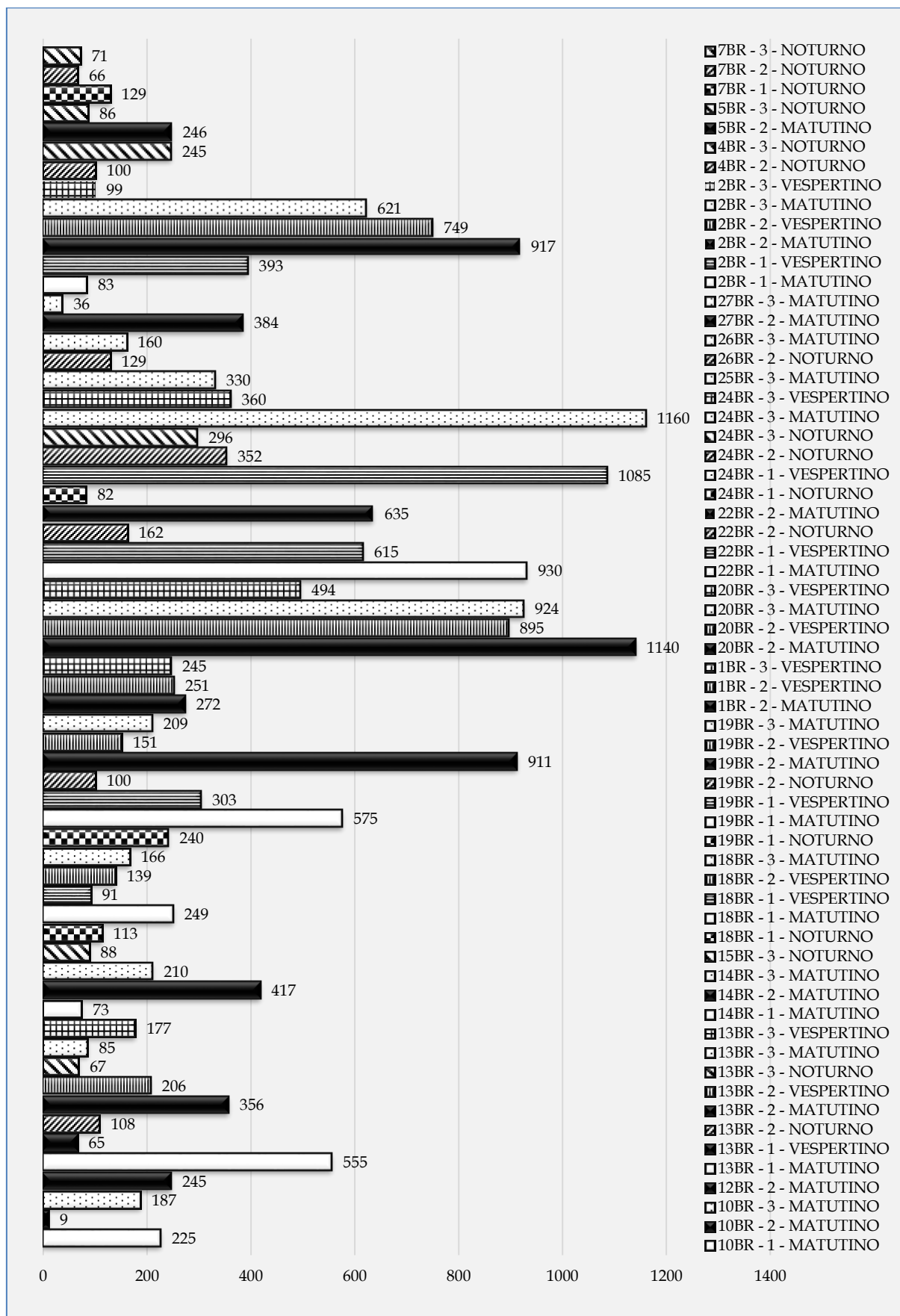
Florianópolis	34	EEM Henrique Veras	80	EM
Florianópolis	35	EEB Dayse Werner Salles	230	EF
Florianópolis	36	EEB Dom Jaime de Barros Câmara	549	EF
Florianópolis	37	EEB Dr. Paulo Fontes	145	EM
Florianópolis	38	EEB Edith Gama Ramos	552	EF
Florianópolis	39	EEF General José Vieira da Rosa	163	EF
Florianópolis	40	EEB Getúlio Vargas	920	EF
Florianópolis	41	EEB Profº Henrique Stodieck	661	EF
Florianópolis	42	EEB Hilda Theodoro Vieira	402	EF
Florianópolis	43	EEB Ildefonso Linhares		EF
Florianópolis	44	EEB Intendente José Fernandes	1672	EF
Florianópolis	45	EEB Irineu Bornhausen	429	EF
Florianópolis	46	EEB Jornalista Jairo Callado	478	EF
Florianópolis	47	EEB José Boiteux	465	EF
Florianópolis	48	EEF Júlio da Costa Neves	694	EF
Florianópolis	49	EEB Jurema Cavalazzi	306	EF
Florianópolis	50	EEB Profª Laura Lima	796	EF
Florianópolis	51	EEB Lauro Muller	536	EF
Florianópolis	52	EEB Leonor de Barros	730	EF
Florianópolis	53	EEB De Muquem	753	EF
Florianópolis	54	EEM Jacó Anderle	1677	EM
Florianópolis	55	EEB Padre Anchieta	871	EF
Florianópolis	56	EEB Pero Vaz de Caminha	540	EF
Florianópolis	57	EEB Porto do Rio Tavares	447	EF
Florianópolis	58	EEB Presidente Roosevelt	325	EF
Florianópolis	59	EEB Rosa Torres de Miranda	476	EF
Florianópolis	60	EEB Rosinha Campos	214	EF
Florianópolis	61	EEB Tenente Almachio	1067	EI
Florianópolis	62	EEF Severo Honorato da Costa	98	EF
Florianópolis	63	EEB Simão José Hess	1041	EF
Florianópolis	64	EEB João Gonçalves Pinheiro (<i>em processo de extinção funcionando na eem ver. Oscar 222emiss da conceição</i>)	878	EM
Florianópolis	65	EEB Presidente Castelo Branco (<i>em processo de extinção funcionando na eem ver. Oscar 222emiss da conceição</i>)	0	EM
Florianópolis	66	EEB Januária Teixeira da Rocha	130	EF
Florianópolis	67	EEM Vereador Oscar Manoel da Conceição	0	EM
Florianópolis	68	EEB Feliciano Nunes Pires (Colégio policial Militar)	-	EF
Florianópolis	69	Instituto Estadual de Educação	5.345 regulares aproximadamente e 2.500 em projetos	EI
Governador Celso Ramos	70	EEB Profº Maria Amália Cardoso	140	EM

Governador Celso Ramos	71	EEB Dr. Aderbal Ramos da Silva	674	EF
Palhoça	72	EEB Gov Ivo Silveira	1804	EF
Palhoça	73	EEB Irmã Maria Teresa	1459	EM
Palhoça	74	EEB João Silveira	1232	EF
Palhoça	75	EEB José Maria Cardoso da Veiga	735	EF
Palhoça	76	EEB Padre Vicente Ferreira Cordeiro	666	EF
Palhoça	77	EEF Dom Jaime de Barros Câmara	941	EF
Palhoça	78	EEB Gov. Pedro Ivo Figueiredo de Campos	464	EF
Palhoça	79	EEB Henrique Estefano Koerich	990	EF
Palhoça	80	EEB Prof. Claudete Maria Hoffmann Domingos	607	EF
Palhoça	81	EEB Profª Maria do Carmo Souza	806	EF
Palhoça	82	EEB Prof. Ursulina Senna de Castro	945	EF
Palhoça	83	EEB Nicolina Tancredo	697	EF
Palhoça	84	EIEF Itaty	38	EF
Palhoça	85	EEB Prof. Benonívio João Martins	1027	EF
Palhoça	86	EEB Senador Renato Ramos da Silva	1064	EF
Palhoça	87	EEF Venceslau Bueno	1456	EF
Palhoça	88	EEB Vicente Silveira	666	EF
Palhoça	89	EIEF Pira Rupá	17	EF e EJA Médio
Palhoça	90	EEF Maria Clementina de Souza Lopes	571	EF
Rancho Queimado	91	EEB Marilda Lênia Araújo	260	EF
Rancho Queimado	92	EEB Roberto Schutz	-	-
Santo Amaro da Imperatriz	93	EEB Prof. Silveira de Matos	372	EF
Santo Amaro da Imperatriz	94	EEB Nereu Ramos	845	EF
Santo Amaro da Imperatriz	95	EEB Anísio Vicente de Freitas	478	EF
Santo Amaro da Imperatriz	96	EEB Profª Zulma Becker	304	EF
São Bonifácio	97	EEB São Tarcísio	322	EF
São Pedro de Alcântara	98	EEB Gama Rosa	444	EF
São José	99	CEJA São José	1208	EJA
São José	100	EEB Bela Vista	225	EF
São José	101	EEB Cecília Rosa Lopes	577	EF
São José	102	EEB Francisco Tolentino	680	EF
São José	103	EEB Pres. Juscelino Kubitschek	1053	EF
São José	104	EEB Nossa Senhora da Conceição	889	EF
São José	105	EEB Prof Maria José Barbosa Vieira(ESCOLA JOVEM SÃO JOSÉ)	1283	EM
São José	106	EEB Profª Joaquim Santiago	489	EF
São José	107	EEB Prof Laércio Caldeira de Andrada	762	EF
São José	108	EEB Wanderley Jr.	1509	EF
São José	109	EEB Aldo Câmara da Silva	481	EF

São José	110	EEB Prof Valdete Luci Martins Porto	395	EF
São José	111	EEF José Matias Zimmermman	589	EF
São José	112	EEF Cristo Rei	379	EF
São José	113	EEF Dr. Homero de Miranda Gomes	399	EF
São José	114	EEB Profª Laurita Dutra de Souza	596	EF
São José	115	EEB Profª Maria do Carmo Lopes	883	EF
São José	116	EEF Prof Marcília de Olivera	708	EF
São José	117	EEB Prof Américo Vespúcio Prates	386	EF
São José	118	EEF Prof Oswaldo Rodrigues Cabral	537	EF
São José	119	EEF São Miguel	436	EF

Fonte: Concedido gentilmente pela Secretaria de Estado da Educação de Santa Catarina, Diretoria de Ensino Superior, Assessoria de Análise e Estatística – Referente ao ano de 2014.

B3: Quantitativo de alunos da amostra por escola e turno - Brasil



B4: Controle no Processo de Recolha de Dados - Portugal

Tabela 53: Controle logístico dos materiais distribuídos por escola, livros didáticos, sujeitos estimados e participantes.

Código das Escolas Secundárias	Total de Prof.	Supost. a colaborar	Retorno do TCLE	Retorno da Grelha	Kits	Total 10º	10º part.	Alunos 10º	Total 11º	11º part.	Alunos 11º	Total 12º	12º part.	Alunos 12º	FA	FB	QA	QB
E39	6	6	3	1	7	3	3	60	3	3	69	1	1	18	11	13	1	16
E40	10	2	1	0	13	6	4	104	6	3	84	1	1	9	11	-	12	-
E46	9	2	0	1	17	5	0	0	6	2	50	6	1	23	11	-	8	-
E54	6	6	0	0	5	2	2	41	2	2	45	1	2	8	11	14	12	16
E101	6	6	3	3	7	3	3	69	3	3	68	2	2	16	11	14	10	16
E1	5	1	0	0	5	2	2	54	2	1	26	2	2	11	9	-	8	-
E10	11	1	5	6	9	3	1	23	3	2	29	3	2	27	9	14	2	16
E36	9	9	1	3	10	4	3	64	5	3	62	1	1	16	8	14	12	16
E37	2	2	0	0	2	1	1	8	1	1	25	-	-	-	8	-	8	-
E42	10	1	3	4	8	6	6	143	-	0	0	2	2	34	10	-	2	-
E38	3	3	2	3	3	2	2	32	2	1	4	-	0	0	11	-	8	-
E41	2	2	2	1	3	1	1	23	2	2	39	0	0	0	11	14	13	16
E47	4	4	0	0	5	3	3	58	2	2	58	0	0	0	11	-	12	-
E50	4	4	2	5	8	3	3	58	3	4	57	2	2	2	11	-	13	-
E100	8	8	2	4	17	12	11	331	0	0	0	5	5	122	11	-	1	-
E7	9	-	0	0	10	4	3	54	4	2	54	3	4	21	9	-	8	-
E18	7	-	2	3	9	3	2	43	3	3	45	3	2	18	9	-	8	-
E43	8	8	0	3	11	4	4	86	5	4	56	2	2	16	8	-	12	-
E44	3	3	1	0	3	2	1	19	0	0	0	1	1	14	8	14	8	16
E45	10	1	0	3	8	3	1	28	3	4	79	2	2	40	5	-	13	-

C - Validade no *Design* da Pesquisa

C1: Análise de Riscos sobre a Validade Interna: Estudo Piloto

Quadro 80: Análise e considerações diante dos riscos à validade interna do *design* experimental do estudo piloto.

RISCOS A VALIDADE INTERNA			
N	Tipos de ameaças à validade interna	Justificativa	
		Quanto aos participantes	
1	História	✘	Independente do pesquisador os grupos de cada classe de aula irão passar supostamente pelos mesmos eventos externos ou forma de ensino (admitindo ser o mesmo professor e à mesma forma de ensino a todos da classe), apesar de haver peculiaridades de aprendizagem para cada sujeito, tendo em vista que o ensino se baseia pelo mesmo programa curricular.
2	Interações com a Seleção (Seleção/História)	✘	Quando os grupos de indivíduos que participam são provenientes de meios (sociais, habitacionais, culturais) distintos, fato que poderá traduzir em influências nos resultados, dado que cada um deles é susceptível de vivenciar acontecimentos particulares, diferentes daqueles que são experienciados pelos outros. Este é um risco real, naturalmente possível, porém se admite com pouca influência, sendo sujeitos no mesmo nível escolar (10º ano), mesmo turno, provavelmente adaptados ao contexto sociocultural e de níveis socioeconômicos equivalentes, devido à baixa desigualdade social existente.
3	Maturação	✘	Os indivíduos alunos já se encontram em uma mesma faixa etária em seus grupos (turmas), portanto não há ameaça de indivíduos amadurecerem mais que outros.
4	Interações com a Seleção (Seleção/Maturação)	✘	Quando os grupos se apresentam diferentes, do ponto de vista da maturação, por exemplo, são selecionados grupos de voluntários, em que o GE é mais velho ou mais novo que o GC. No caso do pré-teste e pós-teste isso não ocorre, pois, a seleção reúne alunos da mesma classe de aula e são considerados grupo único.
5	Regressão Estatística	✘	Não há seleção preferencialmente a indivíduos ou grupos com pontuações extremas. Entretanto, por serem inclusos na base de dados indivíduos de níveis de escolarização diferentes, supostamente tendem a uma pontuação superior (como se espera), ou seja, uma tendência de indivíduos com escores elevados (ou muito baixos) a regressarem para a média em medições sucessivas (repetidas) da mesma variável, fato que não tem nada a ver com o tratamento. Os que possuem elevados escores possuem maior índice de erro, e em medições posteriores os mais baixos tendem a melhorar e os mais elevados a um decréscimo, com isso, o tratamento pode ter sido ineficaz e se constatar efeitos significativos devidos a regressão. Esse, portanto termina sendo um risco a ser considerado.
6	Seleção Diferencial	✘	Quando a escolha de indivíduos não é aleatória na formação de grupos dentro do estudo, com isso, as diferenças entre eles podem emergir resultados explicando-os total ou parcialmente. No caso, os grupos de indivíduos foram escolhidos por conveniência de aproximação de um raio de 90 km com a Universidade de Coimbra, incidindo sobre outros distritos. Entretanto não houve <i>designação/sorteio/seleção/randomização</i> de indivíduos nem <i>destinação aleatória</i> ¹⁰⁵ de grupos, sendo, portanto, os indivíduos escolhidos dentro de seu conglomerado (escola e turma). A não <i>seleção aleatória</i> dos indivíduos se configura, portanto, em um risco de validade interna a ser considerado. Uma possibilidade seria referenciar e numerar categoricamente cada indivíduo nos grupos, com isso, ao selecioná-los, poderia ocorrer uma aleatorização, mas não foi adotado esse procedimento.

¹⁰⁵ *Destinados* aleatoriamente significa que são escolhidos grupos de forma aleatória e não indivíduos, que nesse caso seria *seleção aleatória*. Como os alunos fazem parte de grupos específicos (escolas e turmas), a organização foi por *destinação aleatória*. Apesar disso houve conveniência e não sorteio de escolas a serem analisadas, dando preferência as escolas mais próximas ao centro de Coimbra (Região Centro - Portugal) e de Florianópolis (SC - Brasil).

7	Mortalidade Experimental	✓	O abandono de indivíduos durante a pesquisa será supostamente mínimo (com exceção no caso de morte, mudança de residência, desinteresse, etc.), pois os indivíduos mantêm uma obrigação e compromisso escolar, ou seja, há um interesse grande e pessoal dos indivíduos permanecerem nos locais onde a investigação será realizada, sendo um risco mínimo existente.
N	Tipos de ameaças à validade interna	Justificativa	
		Quanto ao tratamento experimental	
8	Difusão ou Imitação do Tratamento	✓	Refere-se ao fato de os sujeitos que fazem parte de grupos diferentes poderem comunicar entre si, reduzindo, em virtude desta comunicação, as diferenças de tratamento pretendidas pelo investigador. Tem mais probabilidades de acontecer quando o espaço físico, que separa os grupos em estudo, é de tal maneira reduzido ou inexistente, que lhes permite estabelecer contatos. Desta forma, os sujeitos do GC poderão aprender as atividades que supostamente foram destinadas pelo investigador ao GE. Este fato poderá invalidar o estudo, pois as diferenças que se esperavam observar entre eles, serão, assim, minimizadas ou mesmo anuladas. No caso, foi verificado que os espaços físicos são adequados, as salas isolam os grupos, contudo, o tratamento considerado são as aulas realizadas sobre termodinâmica que podem ser repetidas por um professor de uma classe para outra, supostamente divergirem de um professor para outro. No grupo único não há grupos de controle, portanto esse risco é inexistente.
9	Desmoralização Compensatória/ Ressentida (por parte de indivíduos sujeitos a condições menos agradáveis)	✓	Não há esse risco pelo fato de que o projeto é de grupo único, ou seja, não há um Grupo Experimental (GE) sendo privilegiado por um tratamento e um Grupo de Controle (GC) ressentido pela ausência de tratamento, até porque todos recebem o "tratamento". O tratamento sendo as aulas de termodinâmica ministradas e assistidas pelos alunos, obviamente que o ensino varia para cada grupo, entretanto é exclusivo para os indivíduos de cada grupo. Nisso se inclui os LDFs adotados e utilizados (eventualmente), sendo desconhecida as suas diferenças e vantagens, levando em conta que todos os grupos de alguma forma possuam e usufruem de um LDF inviabilizando assim qualquer ressentimento de benefícios nesse sentido, com exceção se houverem sujeitos marginalizados com tratamento diferenciado, ou que não possuam LDF, ou que determinados LDFs são considerados mais eficazes e estão privilegiando determinados alunos (em escolas específicas), ou que grupos com alto desempenho possuam benefícios, ou mesmo determinados grupos possuam professores renomados e de grande prestígio, ou seja, se de alguma forma o GC se sinta inferiorizado/marginalizado quanto ao GE. Este é um fator a se considerar quando existe GE e GC.
10	Rivalidade Compensatória (por parte dos indivíduos sujeitos a condições menos agradáveis)	✓	Não há benefício para nenhum dos grupos, ou seja, nenhum dos grupos irá usufruir/desfrutar de importantes benefícios sociais, portanto não há motivos para que um GC tente minimizar/compensar as diferenças de desempenho por rivalidade. Sobretudo, o <i>design</i> não apresenta GC, portanto esse risco é inexistente.
11	Igualização Compensatória do Tratamento	✓	Não há necessidade de o pesquisador tomar medidas para criar igualdade entre os grupos de GC e GE, pois ambos estão dissociáveis, possuem localidades distintas, inviabilizando qualquer sentimento de desvalorização, até mesmo porque ambos recebem um tratamento semelhante, e ocorre apenas uma única observação, portanto, não há GC e não possibilidade de acontecer esse risco ou na necessidade de igualização compensatória pelo investigador.
N	Tipos de ameaças à validade interna	Justificativa	
		Quanto ao procedimento experimental	
12	Testing (Prática do teste ou testagem)	✗	Para que os participantes não se tornem familiarizados com teste, o pesquisador irá submetê-los com intervalo de tempo de no mínimo 4 meses e máximo de 6 meses. E para garantir maior eficácia, os itens serão trocados de ordem para o pós-teste. No caso, o estudo teve em média um intervalo superior a 6 meses e não foram trocados a ordem dos itens, sendo, portanto, um risco a ser considerado.
13	Instrumentação	✓	Relacionado a falta de calibração, perda de validade ou fidedignidade, modificações no agente humano de recolha de dados. No caso do método for a «observação», o cansaço do observador pode provocar mudanças na recolha dos dados. Este risco será reduzido quando houver um maior comprometimento de quem aplica os testes (observação). A comunicação entre os indivíduos durante a aplicação do teste (observação) não deve ser tolerada, e nem qualquer exposição que favoreça a

			ocorrência do fenômeno de “pensamento coletivo”. Nisso se inclui a proibição de levar o teste para fora do ambiente de aplicação e de fornecer a chave-de-correção. Para isso, foi fornecido uma cartilha para aplicação do teste TCE, descrevendo passos que evitem esses riscos, considerando ainda que as análises do teste e dos itens levam em consideração processos de calibração e análise de consistência e fidedignidade, e que não foi fornecido a chave-de-correção aos professores. A chave-de-correção também não obtém facilmente, não estando disponibilizada de forma completa em artigos, e uma das formas de se obter está no site https://www.physport.org/ , que por sua vez verificam se o solicitante é um investigador na área da educação, pois a difusão da chave-de-correção compromete a validade do instrumento. Esse, portanto, é um risco relativamente controlado.
14	Interações com a Seleção (Seleção/Instrumentação)	✘	A seleção resulta em grupos com características diferentes, aos quais são administradas medidas diferentes. Pode se tratar, por exemplo, de alunos de níveis escolares distintos a que se passam ou de forma intermediária, ou a forma avançada do teste. Tais formas podem ser muito fáceis ou muito difíceis para eles, conduzindo assim, aos efeitos de “teto” ou “assoalho”. Se administra o mesmo teste e os alunos se encontram em um mesmo nível escolar, portanto esse risco é descartado.
N	Tipos de ameaças à validade interna	Justificativa	
		Quanto a fatores externos	
15	Ambiguidade da Direção da Causa	✘	Existem situações em que é difícil ao investigador saber qual é a causa e qual é o efeito numa dada pesquisa, ou se haverá porventura uma terceira variável que é realmente a causa da relação verificada entre as variáveis consideradas inicialmente como independentes e dependentes. Esta é uma ameaça presente na verificação de relacionar as pontuações dos testes com a eficácia dos manuais escolares, tendo em vista fatores relacionados à forma como foram conduzidos os ensinos e a forma como se deu a aprendizagem.

Fonte: Elaborado e adaptado a partir de Creswell (2010) e Vieira (n.d.).

C2: Análise de Riscos sobre a Validade Externa: Estudo Piloto

Quadro 81: Análise e considerações diante dos riscos à validade externa do *design* experimental do estudo piloto.

Riscos à Validade Externa - Validade da População (características dos sujeitos que influenciam o estudo)			
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas às pessoas	
1	Interação entre a Seleção e o Tratamento	✓	O risco é mínimo, tendo em vista que as características dos participantes são equivalentes, não possuidores de dificuldades físicas ou psíquicas, e dentro de uma mesma faixa etária. Não há o risco no sentido da utilização de amostra em níveis infra-humanos (cobaias) para o nível humano. Dessa forma, o pesquisador deve apenas restringir as afirmações sobre grupos aos quais os resultados não poderão ser generalizados a grupos com características diferentes.
2	Generalização da População Acessível (Amostra) para a População-Alvo	✗	Para generalizar é preciso assegurar a representatividade dos sujeitos. Logo, a generalização só é verdadeiramente legítima quando a amostra for representativa da população-alvo. A amostra do projeto-piloto não corresponde significativamente a população-alvo (Região Centro de Portugal), sendo cerca de 4%, entretanto, não se busca a generalização de resultados nessa etapa inicial, mas sim em ajustes dos itens e diagnósticos preliminares. Esse é um risco existente, porém não pertinente com as finalidades estabelecidas.
3	Interação entre Variáveis Personológicas e o Tratamento (variável independente)	✗	As variáveis de personalidade podem agir com certos níveis da variável independente e não com outros níveis da mesma. Se isto acontecer, pode se falar de generalização apenas para esses níveis. As variáveis que envolvem competências socioafetivas e questões subjetivas da personalidade dos sujeitos são aspectos não identificados que estão influenciando, e, portanto, se configuram riscos associados.
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas aos locais	
4	Interação entre o Local e o Tratamento	✓	Devido às características do local dos participantes em um experimento, pode não ser possível generalizar para indivíduos outros locais. Com isso, torna-se necessário conduzir experimentos adicionais em novos locais para ver se ocorrem os mesmos resultados do que no local original. No estudo em questão, foi observado um padrão a nível intranacional (dentro de um contexto específico) em cada país, e sendo uma amostra representativa de cerca de 20% da população estabelecida, sendo fatores controlar esse risco. Além disso, as análises relacionadas a <i>Função Diferencial do Item</i> (DIF) tratam as diferenças entre os contextos, turnos e sexo, sendo assim um risco relativamente tratado e não deve ser considerado um fator de risco iminente.
Riscos à Validade Externa: Validade Ecológica (condições que influenciam a realização do estudo)			
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas à falta de clareza nas operacionalizações com as variáveis	
5	Descrição Explícita da Variável Independente	✓	O risco existe quando a variável independente não for convenientemente descrita (operacionalmente), de modo a permitir uma replicação fiel do estudo e a possibilitar uma generalização dos resultados. Operacionalizações pouco claras da variável independente poderão constituir um óbice à generalização dos resultados e à comparação entre estudos que se debruçaram teoricamente sobre o mesmo problema científico. Dessa forma, todas as operacionalizações são apresentadas em um esquema geral, explicitando os instrumentos a serem utilizados, e posteriormente, as formas de suas medidas. Sendo definidos e descritos as operacionalizações, e assumindo um compromisso com a clareza dos procedimentos das análises esse risco é desconsiderado.
6	Medição da Variável Dependente	✓	Se a variável dependente não for corretamente definida (operacionalmente), qualquer investigador que queira repetir a investigação poderá chegar a conclusões diferentes, pelo fato de não ter definido operacionalmente essa variável da mesma maneira. Desse modo, tal como acontece com a variável independente, uma operacionalização pouco clara ou mesmo frágil da variável dependente poderá conduzir a comparações erradas entre estudos ou mesmo à generalização de resultados que não são válidos. Apesar disso, declara-se comprometimento e clareza nas operacionalizações com as variáveis tornando esse risco mínimo.
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas aos tratamentos	
7	Interferência de Múltiplos Tratamentos	✗	A aplicação de muitos tratamentos ao mesmo grupo de sujeitos podem interferir uns com os outros, tornando difícil a percepção do efeito particular de cada um deles, acarretando confusão na generalização. Dessa forma, a interação entre diferentes tratamentos, quando aplicados num mesmo estudo, pode levantar questionamentos sobre qual o mais eficaz, havendo dúvidas quanto às conclusões

			verdadeiramente generalizáveis. Admitindo no estudo que a intervenção do professor no ensino da termodinâmica seja o tratamento, ou seja, suas metodologias, práticas de ensino e sua influência em múltiplos hábitos de aprendizagens envolvidos ao ministrar as aulas de termodinâmica, evidentemente existem distintas a formas de conduzir esse processo em cada sala de aula da amostra (por escola, professor e contexto escolar), sendo ainda um aspecto não observado diretamente, portanto é um risco a ser considerado.
8	Efeito <i>Hawthorne</i> (Efeito Experimental)	✓	Esse efeito foi descoberto por Hawthorne, numa fábrica de material telefônico, perto de Chicago. Consiste na modificação do comportamento dos sujeitos quando eles sabem que estão a participar numa investigação. As pessoas podem mudar o seu comportamento quando sabem que estão a participar numa investigação, o que poderá constituir um obstáculo à generalização dos resultados. Este risco será minimizado pelo fato de que os alunos não sabem que estão sob forma de tratamento, tendo em vista o tratamento é uma atividade normal de seu cotidiano. Suponha-se que esse efeito possa causar maior interferência no caso da aplicação do método de observação comportamental <i>in loco</i> ou de possíveis entrevistas, na qual os indivíduos estarão conscientes de participarem de uma investigação, ou percebem que a permanência do investigado em seu meio como uma interferência. O mesmo serve para a aplicação de questionários ou testes, entretanto a aplicação do instrumento de pré-teste e pós-teste ocorre sem a interferência de uma pessoa estranha de seu habitual convívio ou grupo social, pois será aplicado por seu professor, e não desconfiar que estão a participar de uma investigação, salvo se o professor lhe informar. O risco no GE é considerado mínimo.
9	Efeito da Novidade e da Ruptura	✓	Para determinadas pessoas, e não para outras, certos tratamentos poderão constituir algo como “inovador”, que essa sensação de novidade já é por si um fator de mudança, podendo ainda acontecer que esse “efeito da novidade” desapareça à medida que o tempo vai passando, e com isso deixe de ser verificado a sua eficácia. Há ainda a possibilidade de resistência dos sujeitos a um dado tratamento, demorando algum tempo até que os efeitos deste comecem a verificar-se, ou seja, ser de tal forma desconhecido que, inicialmente, se torna difícil avaliar a sua eficácia. Os seus verdadeiros efeitos não se manifestam num primeiro momento, mas surgem numa fase posterior. Quer num caso ou em outro, surgem dúvidas quanto significância do tratamento diante dos resultados para a população-alvo. Este tipo de risco é pouco provável pelo fato dos indivíduos não perceberem que estarão sob determinada forma de tratamento, pois o tratamento faz parte de suas atividades normais, pois não se configura como uma nova ou inovadora proposta de ensino a ser analisada.
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas as reações dos sujeitos nas observações	
10	Sensibilização ao Pré-teste	✓	A aplicação de um pré-teste pode predispor o sujeito a se comportar de determinada maneira, o que não aconteceria se não existisse o pré-teste. Por este fato, não é legítimo generalizar as conclusões de uma investigação em que houve pré-teste para outra em que tal não aconteceu. Trata-se, portanto, de um risco que deriva do fato de os sujeitos poderem mudar de opinião/perspectiva/visão sobre o assunto que se está a estudar (variável dependente) depois de serem confrontados com o instrumento de recolha de dados. De fato, pode haver determinadas questões que levem os sujeitos a refletir e a mudar a sua resposta habitual, sendo que essa mudança inicial nada tem a ver com o tratamento. Esta possibilidade dificulta, por exemplo, a comparação de resultados entre estudos com e sem pré-teste. Pode-se afirmar que o tratamento (aulas de termodinâmica) se relaciona com a temática do teste, além disso, o teste TCE propõe itens com situações comuns de seu cotidiano, sendo, portanto, um risco de sensibilidade e de choque minimizado.
11	Sensibilização ao Pós-teste	✓	Tal como o risco indicado anteriormente, o confronto com a recolha de dados durante o pós-teste pode modificar o comportamento do indivíduo, mudança essa que não se fica a dever ao tratamento. Por essa razão, depois se torna difícil decidir que tipo de conclusões se poderão generalizar. Esse risco é mínimo pelas mesmas razões apontadas na sensibilização ao pré-teste.
12	Efeito do Experimentador (Efeito de <i>Rosenthal</i>)	✓	Esse é um risco que se relaciona com os efeitos, deliberados ou inconscientes, que o investigador pode exercer sobre os sujeitos. Dada a necessária neutralidade do investigador num estudo de natureza quantitativa, estas influências constituem possíveis fontes de enviesamento dos resultados. Entre elas podem estar tanto as características físicas e pessoais do investigador (sexo, idade, simpatia, maneira de vestir), como as atitudes que toma para com os sujeitos (atenção diferencial, comportamentos de reforço desiguais, etc.). Com isso, certas características pessoais do investigador poderão dificultar a comparação entre estudos e a própria generalização dos resultados. Ser homem ou mulher pode fazer a diferença durante a recolha de dados (p. ex.: pesquisa feita com mulheres vítimas de violência doméstica), ou ainda, ser jovem ou de meia-idade pode interferir na

			vontade com que os sujeitos respondem a questões da vida privada (p. ex.: relação conjugal). Este é um risco é mínimo pelos seguintes fatores: (1) não há questões que colocam em constrangimento os indivíduos, e envolve somente questões sobre o conhecimento da Física e nada ligado a sua vida pessoal; (2) a aferição do investigador não põe em risco quaisquer prejuízos aos participantes, pois o teste será aplicado por seu professor (experimentador) e somente analisado posteriormente, portanto não há mediação ou contato do investigador com os alunos; (3) a perspectiva de análise será feita com base na matriz de resposta dos resultados, seguindo os critérios estatísticos psicométricos como objetos na tomada de decisão, permanecendo neutro nesse processo.
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas a situações passadas ou futuras	
13	Interação da História e do Tratamento	✘	Um tratamento pode ter resultados diferentes de eficácia dependendo do momento em que foi administrado. Logo, pode não ser legítima a generalização dos resultados do momento A para o momento B. Dessa forma, certos tratamentos poderão ser mais eficazes em determinados momentos (históricos, culturais, políticos, etc.) do que em outros, pelo que se torna pouco aconselhável 'generalizar às cegas', quando se desconfia que estes 'efeitos de época' podem ocorrer sobre os resultados. Nesse caso, se recomenda que o pesquisador replique o estudo em épocas posteriores para determinar se ocorrem os mesmos resultados de que no tempo anterior. Como esta investigação é caracterizada como um estudo transversal e não longitudinal este é um risco iminente a ser considerado para o desenho experimental.
14	Interação Tempo de Medição/Tratamento	✘	O efeito do tratamento pode não se conseguir avaliar logo no primeiro pós-teste, mas sim passadas algumas semanas. Esse risco existe quando se generaliza do primeiro pós-teste para outra medição realizada algumas semanas depois da aplicação do tratamento. Dessa forma, o tempo que medeia entre a aplicação do tratamento e a recolha de dados no pós-teste pode constituir uma variável que interfere nas conclusões obtidas. Diferentes estudos sobre o mesmo problema científico poderão, por isso, conduzir a conclusões distintas (porque os sujeitos mudam com o tempo, porque o efeito do tratamento vai-se desvanecendo, etc.). No atual estudo, o tempo é levado em consideração e tratado como uma variável de controle, entretanto, o momento da aplicação não segue rigorosamente de forma emparelhada dentro de cada contexto, pois o professor aplicou no momento mais oportuno, ultrapassando o período entre 4 a 6 meses recomendado após o tratamento, sendo, portanto, riscos previstos, com tentativa de controle, porém deve ser admitido no estudo.

Fonte: Elaborado e adaptado a partir de Creswell (2010) e Vieira (n.d.).

C3: Análise de Riscos sobre a Validade Interna – Portugal e Brasil

Quadro 82: Análise e considerações diante dos riscos à validade interna dos *designs* de Portugal e Brasil.

RISCOS A VALIDADE INTERNA			
N	Tipos de riscos à validade interna	Justificativa	
		Quanto aos participantes	
1	História	✘	Independente do pesquisador os grupos de cada classe de aula irão passar supostamente pelos mesmos eventos externos ou forma de ensino (admitindo ser o mesmo professor e a mesma forma de ensino a todos da classe), apesar de haver peculiaridades de aprendizagem para cada sujeito, tendo em vista que o ensino se baseia pelo mesmo programa curricular. Apesar disso, não é possível ignorar desconsiderar esse risco, tendo em vista que o <i>design</i> prevê uma análise comparativa a sujeitos de níveis diferenciados de escolaridade sem que garanta minimamente que tenham passado pelas mesmas experiências (mesmos professores e sistemas de ensino), sendo mais ameno no contexto português devido ao 10º e 11º ano corresponderem a ciclo de ensino, e com isso serem geralmente de competência de um mesmo professor.
2	Interações com a Seleção (Seleção/História)	✘	Quando os grupos de indivíduos que participam são provenientes de meios (sociais, habitacionais, culturais) distintos, fato que poderá traduzir em influências nos resultados, dado que cada um deles é susceptível de vivenciar acontecimentos particulares, diferentes daqueles que são experienciados pelos outros. Este é um risco real, naturalmente possível, as diferentes localidades das escolas é um fator agravante, ainda que pertencente a uma mesma região, pois os diferentes níveis de escolaridade acarretarem coletivos sociais distintos no pensamento e valores. Esse risco é de maior preocupação e influência no contexto brasileiro, pois os sujeitos de turnos diferentes se caracterizam com perfis distintos pelo estilo de vida e horário de estudo, muitas vezes associados com atividades de trabalho, além de evidências das desigualdades sociais existentes.
3	Maturação	✘	Os indivíduos alunos já se encontram em uma mesma faixa etária em seus grupos (turmas), portanto não há ameaça de indivíduos amadurecerem mais que outros. Esse é um risco a ser considerado de forma acentuada no contexto brasileiro, em especial para os grupos do turno noturno, na qual se torna mais difícil de atender rigorosamente a esse critério.
4	Interações com a Seleção (Seleção/Maturação)	✘	Quando os grupos se apresentam diferentes, do ponto de vista da maturação, por exemplo, são selecionados grupos de voluntários, em que o GE é mais velho ou mais novo que o GC. No caso o risco é inexistente na perspectiva de análise da TRL, pois os grupos não são submetidos a um pré-teste seguido de pós-teste, são grupos estáticos não equivalentes, sendo considerado uma única base de dados. Apesar disso, na TCI, foi estabelecido um <i>design</i> comparativo entre grupos distintos separados níveis de escolaridade, e pode remete a inclusão de indivíduos com maturação uns em relação aos outros, sem contar a falta de controle entre alunos que repetem a mesma série de escolaridade.
5	Regressão Estatística	✘	Não há seleção preferencialmente a indivíduos ou grupos com pontuações extremas. Entretanto, por serem inclusos na base de dados indivíduos de níveis de escolarização diferentes, supostamente tendem a uma pontuação superior (como se espera), ou seja, uma tendência de indivíduos com escores elevados (ou muito baixos) a regredirem para a média em medições sucessivas (repetidas) da mesma variável, fato que não tem nada a ver com o tratamento. Os que possuem elevados escores possuem maior índice de erro, e em medições posteriores os mais baixos tendem a melhorar e os mais elevados a um decréscimo, com isso, o tratamento pode ter sido ineficaz e se constatar efeitos significativos devidos a regressão. Esse, portanto termina sendo um risco a ser considerado.
6	Seleção Diferencial	✘	Quando a escolha de indivíduos não é aleatória na formação de grupos dentro do estudo, com isso, as diferenças entre eles podem emergir resultados explicando-os total ou parcialmente. No caso, os grupos de indivíduos foram escolhidos por conveniência de aproximação de um raio de 90 km com a Universidade de Coimbra, incidindo sobre outros distritos. Entretanto não houve

			<i>designação/sorteio/seleção/randomização</i> de indivíduos nem <i>destinação aleatória</i> ¹⁰⁶ de grupos, sendo, portanto, os indivíduos escolhidos dentro de seu conglomerado (escola e turma). A não <i>seleção aleatória</i> dos indivíduos se configura, portanto, em um risco de validade interna a ser considerado. Uma possibilidade seria referenciar e numerar categoricamente cada indivíduo nos grupos, com isso, ao selecioná-los, poderia ocorrer uma aleatorização, mas não foi adotado esse procedimento.
7	Mortalidade Experimental	✓	O abandono de indivíduos durante a pesquisa será supostamente mínimo, pois os indivíduos mantêm uma obrigação e compromisso escolar, ou seja, há um interesse grande e pessoal dos indivíduos permanecerem nos locais onde a investigação será realizada. Esse seria um risco em potencial para amostra menores, e quando efetivamente se realiza um pré e pós-teste aos mesmos grupos, fato que não ocorreu para as amostra nos dois contextos.
N	Tipos de riscos à validade interna	Justificativa	
		Quanto ao tratamento experimental	
8	Difusão ou Imitação do Tratamento	✓	Refere-se ao fato de os sujeitos que fazem parte de grupos diferentes poderem comunicar entre si, reduzindo, em virtude desta comunicação, as diferenças de tratamento pretendidas pelo investigador. Tem mais probabilidades de acontecer quando o espaço físico, que separa os grupos em estudo, é de tal maneira reduzido ou inexistente, que lhes permite estabelecer contatos. Desta forma, os sujeitos do GC poderão aprender as atividades que supostamente foram destinadas pelo investigador ao GE. Este fato poderá invalidar o estudo, pois as diferenças que se esperavam observar entre eles, serão, assim, minimizadas ou mesmo anuladas. No caso, foi verificado que os espaços físicos são adequados, as salas isolam os grupos, contudo, o tratamento considerado são as aulas realizadas sobre termodinâmica que podem ser repetidas por um professor de uma classe para outra, supostamente divergirem de um professor para outro. No <i>design</i> , os grupos experimentais já passaram pelo tratamento a certo tempo decorrido e se encontram em níveis de escolaridade diferentes, sendo esse risco praticamente inexistente.
9	Desmoralização Compensatória/ Ressentida (por parte de indivíduos sujeitos a condições menos agradáveis)	✓	Não há esse risco pelo fato de que o projeto é de grupo único, ou seja, não há um Grupo Experimental (GE) sendo privilegiado por um tratamento e um Grupo de Controle (GC) ressentido pela ausência de tratamento, até porque todos recebem o “tratamento”. O tratamento sendo as aulas de termodinâmica ministradas e assistidas pelos alunos, obviamente que o ensino varia para cada grupo, entretanto é exclusivo para os indivíduos de cada grupo. Nisso se inclui os LDFs adotados e utilizados (eventualmente), sendo desconhecida as suas diferenças e vantagens, levando em conta que todos os grupos de alguma forma possuem e usufruem de um LDF inviabilizando assim qualquer ressentimento de benefícios nesse sentido, com exceção se houverem sujeitos marginalizados com tratamento diferenciado, ou que não possuam LDF, ou que determinados LDFs são considerados mais eficazes e estão privilegiando determinados alunos (em escolas específicas), ou que grupos com alto desempenho possuam benefícios, ou mesmo determinados grupos possuam professores renomados e de grande prestígio, ou seja, se de alguma forma o GC se sinta inferiorizado/marginalizado quanto ao GE. Nenhum desses aspectos foi detectado, e não há possibilidade de haver influência caso eventualmente indivíduos do GC se sintam ressentidos por algum motivo.
10	Rivalidade Compensatória	✓	Não há benefício para nenhum dos grupos, ou seja, nenhum dos grupos irá usufruir/desfrutar de importantes benefícios sociais, portanto não há motivos para que um GC tente minimizar/compensar as diferenças de desempenho por rivalidade. Sobretudo, o <i>design</i> não possibilita a chance do GC tentar compensar as diferenças num eventual pós-teste, portanto esse risco é inexistente.
11	Igualização Compensatória do Tratamento	✓	Não há necessidade de o pesquisador tomar medidas para criar igualdade entre os grupos de GC e GE, pois ambos estão dissociáveis, possuem localidades distintas, inviabilizando qualquer sentimento de desvalorização, até mesmo porque ambos

¹⁰⁶ *Destinados* aleatoriamente significa que são escolhidos grupos de forma aleatória e não indivíduos, que nesse caso seria *seleção aleatória*. Como os alunos fazem parte de grupos específicos (escolas e turmas), a organização foi por *destinação aleatória*. Apesar disso houve conveniência e não sorteio de escolas a serem analisadas, dando preferência as escolas mais próximas ao centro de Coimbra (Região Centro - Portugal) e de Florianópolis (SC - Brasil).

			recebem um tratamento semelhante, e ocorre apenas uma única observação. O GC não tem possibilidade de tentar minimizar as diferenças de desempenho (buscar uma igualização ou aproximação com o desempenho do GE, supostamente beneficiado pelo tratamento) ou da necessidade em compensar o GC pelo investigador, buscando uma igualização com o GE, simplesmente porque não existe um pós-teste.
N	Tipos de riscos à validade interna	Justificativa	
		Quanto ao procedimento experimental	
12	Testing (Prática do teste ou testagem)	✓	Para que os participantes não se tornem familiarizados com teste, o pesquisador irá submetê-los com intervalo de tempo de no mínimo 4 meses e máximo de 6 meses. E para garantir maior eficácia, os itens serão trocados de ordem para o pós-teste. Na análise da TRI consideram-se um grupo único para Brasil e Portugal (na mesma métrica), com isso, pode-se admitir que não houve um pré-teste seguido de um pós-teste, sendo, portanto, descartado esse risco nessa perspectiva. Mesmo para o <i>design</i> da TCT em que são analisados alunos de níveis diferentes de escolaridade, não há esse risco pelo fato de não considerar alunos não que tiveram contato com o teste duas vezes.
13	Instrumentação	✓	Relacionado a falta de calibração, perda de validade ou fidedignidade, modificações no agente humano de recolha de dados. No caso do método for a «observação», o cansaço do observador pode provocar mudanças na recolha dos dados. Este risco será reduzido quando houver um maior comprometimento de quem aplica os testes (observação). A comunicação entre os indivíduos durante a aplicação do teste (observação) não deve ser tolerada, e nem qualquer exposição que favoreça a ocorrência do fenômeno de “pensamento coletivo”. Nisso se inclui a proibição de levar o teste para fora do ambiente de aplicação e de fornecer a chave-de-correção. Para isso, foi fornecido uma cartilha para aplicação do teste TCE, descrevendo passos que evitem esses riscos, considerando ainda que as análises do teste e dos itens levam em consideração processos de calibração e análise de consistência e fidedignidade, e que não foi fornecido a chave-de-correção aos professores. A chave-de-correção também não obtém facilmente, não estando disponibilizada de forma completa em artigos, e uma das formas de se obter está no site https://www.physport.org/ , que por sua vez verificam se o solicitante é um investigador na área da educação, pois a difusão da chave-de-correção compromete a validade do instrumento. Esse, portanto, é um risco relativamente controlado.
14	Interações com a Seleção (Seleção/Instrumentação)	✗	A seleção resulta em grupos com características diferentes, aos quais são administradas medidas diferentes. Pode se tratar, por exemplo, de alunos de níveis escolares distintos a que se passam ou de forma intermediária, ou a forma avançada do teste. Tais formas podem ser muito fáceis ou muito difíceis para eles, conduzindo assim, aos efeitos de “teto” ou “assoalho”. Apesar de ser administrado o mesmo teste aos sujeitos, se considerar uma análise na base de dados com alunos de diferentes níveis de escolarização no final do ciclo da educação básica, com isso, esse é um risco plausível a se considerar no estudo.
N	Tipos de riscos à validade interna	Justificativa	
		Quanto a fatores externos	
15	Ambiguidade da Direção da Causa	✗	Existem situações em que é difícil ao investigador saber qual é a causa e qual é o efeito numa dada pesquisa, ou se haverá porventura uma terceira variável que é realmente a causa da relação verificada entre as variáveis consideradas inicialmente como independentes e dependentes. Esta é uma ameaça presente na verificação de relacionar as pontuações dos testes com a eficácia dos manuais escolares, tendo em vista fatores relacionados à forma como foram conduzidos os ensinamentos e a forma como se deu a aprendizagem.

Fonte: Elaborado e adaptado a partir de Creswell (2010) e Vieira (n.d.).

C4: Análise de Riscos sobre a Validade Externa – Portugal e Brasil

Quadro 83: Análise e considerações diante dos riscos à validade externa dos *designs* de Portugal e Brasil.

RISCOS A VALIDADE EXTERNA			
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas às pessoas	
1	Interação entre a Seleção e o Tratamento	✓	O risco é mínimo, tendo em vista que as características dos participantes são equivalentes, não possuidores de dificuldades físicas ou psíquicas, e dentro de uma mesma faixa etária. Não há o risco no sentido da utilização de amostra em níveis infra-humanos (cobaias) para o nível humano. Dessa forma, o pesquisador deve apenas restringir as afirmações sobre grupos aos quais os resultados não poderão ser generalizados a grupos com características diferentes.
2	Generalização da População Acessível (Amostra) para a População-Alvo	✓	Para generalizar é preciso assegurar a representatividade dos sujeitos. Logo, a generalização só é verdadeiramente legítima quando a amostra for representativa da população-alvo. A amostra com base no total das escolas corresponde a cerca de 20% a população-alvo em cada contexto, dessa forma favorece a generalização de resultados. Pondera-se ao fato de não corresponder a critérios de diversidade regional da população (seleção de escolas/indivíduos com distribuição mais homogênea). Entretanto, considera-se um risco é minimizado pelo tamanho amostra.
3	Interação entre Variáveis Personológicas e o Tratamento (var. ind.).	✗	As variáveis de personalidade podem agir com certos níveis da variável independente e não com outros níveis da mesma. Se isto acontecer, pode se falar de generalização apenas para esses níveis. As variáveis que envolvem competências socioafetivas e questões subjetivas da personalidade dos sujeitos são aspectos não identificados que estão influenciando, e, portanto, se configuram riscos associados.
RISCOS A VALIDADE EXTERNA: Validade Ecológica			
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas aos locais	
4	Interação entre o Local e o Tratamento	✓	Devido às características do local dos participantes em um experimento, pode não ser possível generalizar para indivíduos outros locais. Com isso, torna-se necessário conduzir experimentos adicionais em novos locais para ver se ocorrem os mesmos resultados do que no local original. No estudo em questão, foi observado um padrão a nível intranacional (dentro de um contexto específico) em cada país, e sendo uma amostra representativa de cerca de 20% da população estabelecida, sendo fatores controlar esse risco. Além disso, as análises relacionadas a <i>Função Diferencial do Item</i> (DIF) tratam as diferenças entre os contextos, turnos e sexo, sendo assim um risco relativamente tratado e não deve ser considerado um fator de risco iminente.
RISCOS A VALIDADE EXTERNA: Validade Ecológica			
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas à falta de clareza nas operacionalizações com as variáveis	
5	Descrição Explícita da Variável Independente	✓	O risco existe quando não for convenientemente descrita (operacionalmente) esta variável, de modo a permitir uma replicação fiel do estudo e a possibilitar uma generalização dos resultados. Operacionalizações pouco claras da variável independente poderão constituir um óbice à generalização dos resultados e à comparação entre estudos que se debruçaram teoricamente sobre o mesmo problema científico. Dessa forma, todas as operacionalizações são apresentadas em um esquema geral, explicitando os instrumentos a serem utilizados, e posteriormente, as formas de suas medidas. Sendo definidos e descritos as operacionalizações, e assumindo um compromisso com a clareza dos procedimentos das análises esse risco é desconsiderado.
6	Medição da Variável Dependente	✓	Se a variável dependente não for corretamente definida (operacionalmente), qualquer investigador que queira repetir a investigação poderá chegar a conclusões diferentes, pelo fato de não ter definido operacionalmente essa variável da mesma maneira. Desse modo, tal como acontece com a variável independente, uma operacionalização pouco clara ou mesmo frágil da variável dependente poderá conduzir a comparações erradas entre estudos ou mesmo à generalização de resultados que não são válidos. Apesar disso, declara-se comprometimento e clareza nas operacionalizações com as variáveis tornando esse risco mínimo.
RISCOS A VALIDADE EXTERNA: Validade Ecológica			
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas aos tratamentos	
7	Interferência de Múltiplos Tratamentos	✗	A aplicação de muitos tratamentos ao mesmo grupo de sujeitos podem interferir uns com os outros, tornando difícil a percepção do efeito particular de cada um deles, acarretando confusão na generalização. Dessa forma, a interação entre diferentes tratamentos, quando aplicados num mesmo estudo, pode levantar

			questionamentos sobre qual o mais eficaz, havendo dúvidas quanto às conclusões verdadeiramente generalizáveis. Admitindo no estudo que a intervenção do professor no ensino da termodinâmica seja o tratamento, ou seja, suas metodologias, práticas de ensino e sua influência em múltiplos hábitos de aprendizagens envolvidos ao ministrar as aulas de termodinâmica, evidentemente existem distintas a formas de conduzir esse processo em cada sala de aula da amostra (por escola, professor e contexto escolar), sendo ainda um aspecto não observado diretamente, portanto é um risco a ser considerado.
8	Efeito <i>Hawthorne</i> (Efeito de Tratamento)	✓	Esse efeito foi descoberto por Hawthorne, numa fábrica de material telefônico, perto de Chicago. Consiste na modificação do comportamento dos sujeitos quando eles sabem que estão a participar numa investigação. As pessoas podem mudar o seu comportamento quando sabem que estão a participar numa investigação, o que poderá constituir um obstáculo à generalização dos resultados. Este risco será minimizado pelo fato de que os alunos não saberem que estão sob forma de tratamento, tendo em vista o tratamento é uma atividade normal de seu cotidiano. Suponha-se que esse efeito possa causar maior interferência no caso da aplicação do método de observação comportamental <i>in loco</i> ou de possíveis entrevistas, na qual os indivíduos estarão conscientes de participarem de uma investigação, ou percebem que a permanência do investigado em seu meio como uma interferência. O mesmo serve para a aplicação de questionários ou testes, entretanto a aplicação do instrumento de pré-teste e pós-teste ocorre sem a interferência de uma pessoa estranha de seu habitual convívio ou grupo social, pois será aplicado por seu professor, e não desconfiam que estão a participar de uma investigação, salvo se o professor lhe informar. O risco no GE é considerado mínimo.
9	Efeito da Novidade e da Ruptura	✓	Para determinadas pessoas, e não para outras, certos tratamentos poderão constituir algo como “inovador”, que essa sensação de novidade já é por si um fator de mudança, podendo ainda acontecer que esse “efeito da novidade” desapareça à medida que o tempo vai passando, e com isso deixe de ser verificado a sua eficácia. Há ainda a possibilidade de resistência dos sujeitos a um dado tratamento, demorando algum tempo até que os efeitos deste comecem a verificar-se, ou seja, ser de tal forma desconhecido que, inicialmente, se torna difícil avaliar a sua eficácia. Os seus verdadeiros efeitos não se manifestam num primeiro momento, mas surgem numa fase posterior. Quer num caso ou em outro, surgem dúvidas quanto significância do tratamento diante dos resultados para a população-alvo. Este tipo de risco é pouco provável pelo fato dos indivíduos não perceberem que estarão sob determinada forma de tratamento, pois o tratamento faz parte de suas atividades normais, pois não se configura como uma nova ou inovadora proposta de ensino a ser analisada.
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas as reações dos sujeitos nas observações	
10	Sensibilização ao Pré-teste	✓	A aplicação de um pré-teste pode predispor o sujeito a se comportar de determinada maneira, o que não aconteceria se não existisse o pré-teste. Por este fato, não é legítimo generalizar as conclusões de uma investigação em que houve pré-teste para outra em que tal não aconteceu. Trata-se, portanto, de um risco que deriva do fato de os sujeitos poderem mudar de opinião/perspectiva/visão sobre o assunto que se está a estudar (variável dependente) depois de serem confrontados com o instrumento de recolha de dados. De fato, pode haver determinadas questões que levem os sujeitos a refletir e a mudar a sua resposta habitual, sendo que essa mudança inicial nada tem a ver com o tratamento. Esta possibilidade dificulta, por exemplo, a comparação de resultados entre estudos com e sem pré-teste. Pode-se afirmar que o tratamento (aulas de termodinâmica) se relaciona com a temática do teste, além disso, o teste TCE propõe itens com situações comuns de seu cotidiano, sendo, portanto, um risco de sensibilidade e de choque minimizado.
11	Sensibilização ao Pós-teste	✓	Tal como na ameaça indicada anteriormente, o confronto com a recolha de dados durante o pós-teste pode modificar o comportamento do indivíduo, mudança essa que não se fica a dever ao tratamento. Por essa razão, depois torna-se difícil decidir que tipo de conclusões se poderão generalizar. Esse risco é inexistente por não haver pós-teste.
12	Efeito do Experimentador (Efeito de <i>Rosenthal</i>)	✓	Esse é um risco que se relaciona com os efeitos, deliberados ou inconscientes, que o investigador pode exercer sobre os sujeitos. Dada a necessária neutralidade do investigador num estudo de natureza quantitativa, estas influências constituem possíveis fontes de enviesamento dos resultados. Entre elas podem estar tanto as características físicas e personológicas do investigador (sexo, idade, simpatia, maneira de vestir), como as atitudes que toma para com os sujeitos (atenção diferencial, comportamentos de reforço desiguais, etc.). Com isso, certas características pessoais do investigador poderão dificultar a comparação entre estudos e a própria generalização dos resultados. Ser homem ou mulher pode fazer a diferença durante a recolha de dados (p. ex.: pesquisa feita com mulheres vítimas

			de violência doméstica), ou ainda, ser jovem ou de meia-idade pode interferir na vontade com que os sujeitos respondem a questões da vida privada (p. ex.: relação conjugal). Este é um risco é mínimo pelos seguintes fatores: (1) não há questões que colocam em constrangimento os indivíduos, e envolve somente questões sobre o conhecimento da Física e nada ligado a sua vida pessoal; (2) a aferição do investigador não põe em risco quaisquer prejuízos aos participantes, pois o teste será aplicado por seu professor (experimentador) e somente analisado posteriormente, portanto não há mediação ou contato do investigador com os alunos; (3) a perspectiva de análise será feita com base na matriz de resposta dos resultados, seguindo os critérios estatísticos psicométricos como objetos na tomada de decisão, permanecendo neutro nesse processo.
N	Tipos de riscos à validade externa	Justificativa	
		Inferências erradas relacionadas a situações passadas ou futuras	
13	Interação da História e do Tratamento	✘	Um tratamento pode ter resultados diferentes de eficácia dependendo do momento em que foi administrado. Logo, pode não ser legítima a generalização dos resultados do momento A para o momento B. Dessa forma, certos tratamentos poderão ser mais eficazes em determinados momentos (históricos, culturais, políticos, etc.) do que em outros, pelo que se torna pouco aconselhável 'generalizar às cegas', quando se desconfia que estes 'efeitos de época' podem ocorrer sobre os resultados. Nesse caso, se recomenda que o pesquisador replique o estudo em épocas posteriores para determinar se ocorrem os mesmos resultados de que no tempo anterior. Como esta investigação é caracterizada como um estudo transversal e não longitudinal este é um risco iminente a ser considerado para o desenho experimental.
14	Interação Tempo de Medição e Tratamento	✘	O efeito do tratamento pode não se conseguir avaliar logo no primeiro pós-teste, mas sim passadas algumas semanas. Esse risco existe quando se generaliza do primeiro pós-teste para outra medição realizada algumas semanas depois da aplicação do tratamento. Dessa forma, o tempo que medeia entre a aplicação do tratamento e a recolha de dados no pós-teste pode constituir uma variável que interfere nas conclusões obtidas. Diferentes estudos sobre o mesmo problema científico poderão, por isso, conduzir a conclusões distintas (porque os sujeitos mudam com o tempo, porque o efeito do tratamento vai-se desvanecendo, etc.). No atual estudo, o tempo é levado em consideração e tratado como uma variável de controle, entretanto, o momento da aplicação não segue rigorosamente de forma emparelhada dentro de cada contexto, pois o professor aplicou no momento mais oportuno, ultrapassando o período entre 4 a 6 meses recomendado após o tratamento, sendo, portanto, riscos previstos, com tentativa de controle, porém deve ser admitido no estudo.

Fonte: Elaborado e adaptado a partir de Creswell (2010) e Vieira (n.d.).

De modo geral, é importante destacar as seguintes observações:

- Não é possível que um estudo tenha validade externa se ele não tiver validade interna. Ou seja, não se podem generalizar os resultados de um dado estudo para a população a que pertence a amostra, se as conclusões obtidas não forem credíveis.
- Uma pesquisa bastante controlada, feita em condições de grande rigor experimental, tem habitualmente uma enorme validade interna, mas isso não significa que tenha uma grande validade externa.
- O rigor experimental poderá conduzir a uma 'artificialização' das condições de observação empírica da relação entre as variáveis, pouco tendo a ver com as condições normais da vida das pessoas (para as quais se pretende generalizar os resultados).

- A escolha do melhor plano de investigação, tendo como ponto de partida o problema científico e a natureza das variáveis, visa ajudar o investigador a seguir um rumo no seu trabalho, que lhe permita, por um lado, delinear conclusões em que possa confiar (validade interna) e, por outro, usar as suas 'descobertas' para melhorar a vida das pessoas em geral (e não apenas a daquelas que participaram na sua pesquisa - validade externa).
- Deverá haver, por isso, uma tentativa de equilíbrio (possível) entre a validade interna e a validade externa de um determinado estudo

D - A Física Térmica nos Programas Curriculares

D1: Portugal

D1.1: Programa de Físico-Química A, versão 2005.

Quadro 84: Capacidades a serem desenvolvidas na unidade temática de termodinâmica, de acordo com o programa de Físico-Química A, versão 2005.

N	Das fontes de energia ao utilizador
	AL 1 - Rendimento no aquecimento
1	Distinguir calor, temperatura e energia interna
2	Determinar a quantidade de energia necessária para aumentar a temperatura de uma certa massa de uma substância
3	Identificar os elementos constituintes de um circuito eléctrico e efectuar a sua montagem
4	Associar a cada elemento do circuito eléctrico a respectiva função
5	Determinar a potência fornecida por uma resistência eléctrica
6	Estabelecer balanços energéticos em sistemas termodinâmicos, identificando as parcelas que correspondem à energia útil e à energia dissipada no processo
7	Determinar o rendimento do processo
N	Do Sol ao aquecimento
	1. Energia - do Sol para a Terra
8	Explicar que a temperatura média da Terra é em grande parte determinada pela radiação que ela recebe do Sol, mas que esta também emite energia, pois, caso contrário, ficaria cada vez mais quente
9	Identificar um sistema termodinâmico como aquele em que são apreciáveis as variações de energia interna
10	Indicar que todos os corpos irradiam energia
11	Relacionar a potência total irradiada por uma superfície com a respectiva área e a quarta potência da sua temperatura absoluta (Lei de Stefan-Boltzmann)
12	Identificar a zona do espectro electromagnético em que é máxima a potência irradiada por um corpo, para diversos valores da sua temperatura (deslocamento de Wien)
13	Relacionar as zonas do espectro em que é máxima a potência irradiada pelo Sol e pela Terra com as respectivas temperaturas
14	Identificar situações de equilíbrio térmico
15	Explicitar o significado da Lei Zero da Termodinâmica
16	Explicar que, quando um sistema está em equilíbrio térmico com as suas vizinhanças, as respectivas taxas de absorção e de emissão de radiação são iguais
17	Determinar a temperatura média de equilíbrio radiativo da Terra com um todo a partir do balanço entre a energia solar absorvida e a energia da radiação emitida pela superfície da Terra e atmosfera
18	Interpretar o valor real da temperatura média da Terra, a partir da absorção e emissão de radiação por alguns gases presentes na atmosfera
N	2. A energia no aquecimento/ arrefecimento de sistemas
19	Distinguir os mecanismos de condução e convecção
20	Relacionar quantitativamente a condutividade térmica de um material com a taxa temporal de transmissão de energia como calor
21	Distinguir materiais bons e maus condutores do calor com base em valores tabelados de condutividade térmica
22	Interpretar a 1ª Lei da Termodinâmica a partir da Lei Geral da Conservação da Energia
23	Interpretar situações em que a variação de energia interna se faz à custa de trabalho, calor ou radiação
24	Estabelecer balanços energéticos em sistemas termodinâmicos
25	Calcular o rendimento de processos de aquecimento/arrefecimento
26	Explicitar que os processos que ocorrem espontaneamente na Natureza se dão sempre num determinado sentido - o da diminuição da energia útil do Universo (2ª Lei da Termodinâmica)
N	AL 1.1 - Absorção e emissão de radiação
27	Analisar transferências e transformações de energia em sistemas
28	Relacionar o poder de absorção de radiação com a natureza das superfícies
29	Reconhecer que a radiação incidente num corpo pode ser parcialmente absorvida, reflectida ou transmitida
30	Relacionar as taxas de emissão e de absorção da radiação de um corpo com a diferença entre a sua temperatura e a do ambiente que o rodeia
N	AL 1.2 - Energia eléctrica fornecida por um painel
31	Explicitar que a conversão fotovoltaica da energia solar consiste na transformação de energia radiante numa diferença de potencial entre os polos do painel fotovoltaico
32	Determinar a potência eléctrica fornecida por painel fotovoltaico
33	Identificar a existência de uma resistência exterior que optimiza o rendimento de um painel fotovoltaico
34	Explicar que, para maximizar o rendimento de um painel fotovoltaico, este deve estar orientado de forma a receber o máximo de radiação incidente (orientação a Sul e inclinação conveniente)

35	Explicar que, para dimensionar um sistema de conversão fotovoltaico, é necessário ter em consideração a potência média solar recebida por unidade de superfície terrestre, durante o dia (ou número médio de horas de luz solar por dia) e a potência a debitar
N	AL 1.3 – Capacidade térmica mássica
36	Analisar transferências e transformações de energia num sistema
37	Estabelecer balanços energéticos em sistemas termodinâmicos, identificando as parcelas que correspondem à energia útil e à energia dissipada no processo
38	Associar o valor (alto ou baixo) da capacidade térmica mássica ao comportamento térmico do material
39	Aplicar o conceito de capacidade térmica mássica à interpretação de fenómenos do dia a dia
N	AL 1.4 – Balanço energético num sistema termodinâmico
40	Identificar mudanças de estado físico: fusão, vaporização, condensação, solidificação e sublimação
41	Identificar a quantidade de energia necessária à mudança de estado físico de uma unidade de massa de uma substância como uma característica desta
42	Associar o valor, positivo ou negativo, da quantidade de energia envolvida na mudança de estado físico, às situações em que o sistema recebe energia ou transfere energia para as vizinhanças, respectivamente
43	Estabelecer um balanço energético, aplicando a Lei da Conservação da Energia

Fonte: Adaptado de Simões, Simões, Caldeira, Bello, & Pina (2005).

D1.2: Programa de Físico-Química A, versão 2014.

Quadro 85: Capacidades a serem desenvolvidas na unidade temática de termodinâmica, de acordo com o programa de Físico-Química A, versão 2014.

N	Energia, fenómenos térmicos e radiação
1	Compreender os processos e mecanismos de transferências de energia entre sistemas termodinâmicos, interpretando-os com base na Primeira e na Segunda Leis da Termodinâmica.
2	Distinguir sistema, fronteira e vizinhança e definir sistema isolado.
3	Identificar um sistema termodinâmico como aquele em que se tem em conta a sua energia interna.
4	Indicar que a temperatura é uma propriedade que determina se um sistema está ou não em equilíbrio térmico com outros e que o aumento de temperatura de um sistema implica, em geral, um aumento da energia cinética das suas partículas.
5	Indicar que as situações de equilíbrio térmico permitem estabelecer escalas de temperatura, aplicando à escala de temperatura Celsius.
6	Relacionar a escala de Celsius com a escala de Kelvin (escala de temperatura termodinâmica) e efetuar conversões de temperatura em graus Celsius e kelvin.
7	Identificar calor como a energia transferida espontaneamente entre sistemas a diferentes temperaturas.
8	Descrever as experiências de Thompson e de Joule identificando o seu contributo para o reconhecimento de que o calor é energia.
9	Distinguir, na transferência de energia por calor, a radiação – transferência de energia através da propagação de luz, sem haver contacto entre os sistemas – da condução e da convecção que exigem contacto entre sistemas.
10	Indicar que todos os corpos emitem radiação e que à temperatura ambiente emitem predominantemente no infravermelho, dando exemplos de aplicação desta característica (sensores de infravermelhos, visão noturna, termómetros de infravermelhos, etc.).
11	Indicar que todos os corpos absorvem radiação e que a radiação visível é absorvida totalmente pelas superfícies pretas.
12	Associar a irradiância de um corpo à energia da radiação emitida por unidade de tempo e por unidade de área.
13	Identificar uma célula fotovoltaica como um dispositivo que aproveita a energia da luz solar para criar diretamente uma diferença de potencial elétrico nos seus terminais, produzindo uma corrente elétrica contínua.
14	Dimensionar a área de um sistema fotovoltaico conhecida a irradiância solar média no local de instalação, o número médio de horas de luz solar por dia, o rendimento e a potência a debitar.
15	Distinguir os mecanismos de condução e de convecção.
16	Associar a condutividade térmica à taxa temporal de transferência de energia como calor por condução, distinguindo materiais bons e maus condutores do calor.
17	Interpretar o significado de capacidade térmica mássica, aplicando-o na explicação de fenómenos do quotidiano.
18	Interpretar o conceito de variação de entalpias de fusão e de vaporização.
19	Determinar a variação de energia interna de um sistema num aquecimento ou arrefecimento, aplicando os conceitos de capacidade térmica mássica e de variação de entalpia (de fusão ou de vaporização), interpretando o sinal dessa variação.
20	Interpretar o funcionamento de um coletor solar, a partir de informação selecionada, e identificar as suas aplicações.
21	Interpretar e aplicar a Primeira Lei da Termodinâmica.
22	Associar a Segunda Lei da Termodinâmica ao sentido em que os processos ocorrem espontaneamente, diminuindo a energia útil.
23	Efetuar balanços energéticos e calcular rendimentos
N	AL 3.1. Radiação e potência elétrica de um painel fotovoltaico
N	Objetivo geral: Investigar a influência da irradiância e da diferença de potencial elétrico no rendimento de um painel fotovoltaico.

24	Associar a conversão fotovoltaica à transferência de energia da luz solar para um painel fotovoltaico que se manifesta no aparecimento de uma diferença de potencial elétrico nos seus terminais.
25	Montar um circuito elétrico e efetuar medições de diferença de potencial elétrico e de corrente elétrica.
26	Determinar a potência elétrica fornecida por um painel fotovoltaico.
27	Investigar o efeito da variação da irradiância na potência do painel, concluindo qual é a melhor orientação de um painel fotovoltaico de modo a maximizar a sua potência.
28	Construir e interpretar o gráfico da potência elétrica em função da diferença de potencial elétrico nos terminais de um painel fotovoltaico, determinando a diferença de potencial elétrico que otimiza o seu rendimento.
N	AL 3.2. Capacidade térmica mássica
	Objetivo geral: Determinar a capacidade térmica mássica de um material.
29	Identificar transferências de energia.
30	Estabelecer balanços energéticos em sistemas termodinâmicos, identificando as parcelas que correspondem à energia útil e à energia dissipada.
31	Medir temperaturas e energias fornecidas, ao longo do tempo, num processo de aquecimento.
32	Construir e interpretar o gráfico da variação de temperatura de um material em função da energia fornecida, traçar a reta que melhor se ajusta aos dados experimentais e obter a sua equação.
33	Determinar a capacidade térmica mássica do material a partir da reta de ajuste e avaliar a exatidão do resultado a partir do erro percentual.
N	AL 3.3. Balanço energético num sistema termodinâmico
	Objetivo geral: Estabelecer balanços energéticos e determinar a entalpia de fusão do gelo
34	Prever a temperatura final da mistura de duas massas de água a temperaturas diferentes e comparar com o valor obtido experimentalmente.
35	Medir massas e temperaturas.
36	Estabelecer balanços energéticos em sistemas termodinâmicos aplicando a Lei da Conservação da Energia, interpretando o sinal positivo ou negativo da variação da energia interna do sistema.
37	Medir a entalpia de fusão do gelo e avaliar a exatidão do resultado a partir do erro percentual

Fonte: Fiolhais et al. (2014).

D1.3: Mapeamento da Física Térmica nos Exames Nacionais em Portugal (2006-2016).

O quadro elaborado informa com maiores detalhes a distribuição e incidência sobre essas temáticas. Convém lembrar que a partir de 2011 as questões foram agrupadas em blocos temáticos de conteúdos específicos, apesar de não ter sido formulado um título que pudesse indicar. Para facilitar o acesso ou uma possível verificação por um leitor interessado foi identificado o agrupamento de cada item no indicativo subscrito (inferior direito) da referência numérica dos itens no exame que estão no quadro. A letra “q” se refere ao fato de que a questão de física encontra-se dentro do agrupamento das questões de química, seja pela incorporação de caráter «multidisciplinar» (quando a questão de física está apenas “misturada” entre as questões de química) ou “Q” pela associação «interdisciplinar» (considera-se que a habilidade científica na questão envolve em parte a coordenação de conhecimentos físicos e químicos, ainda que não seja uma condição eminentemente necessária para a resolução), para uma melhor identificação visual foram destacados com um realce de cor no texto. Tendo como exemplo o item codificado como “2_{(6)Q}”, lê-se: “item 2, sexto agrupamento das questões de química, sendo esta uma questão com características interdisciplinares. É possível observar um certo destaque para questões com esse tipo de característica na temática “absorção e emissão da radiação”.

Como não foi disponibilizado o relatório nacional quanto ao EM-2012 pelo Gabinete de Avaliação Educacional (GAVE), entretanto, estando as questões disponíveis e divulgadas, foi possível verificar a temática experimental relacionado à determinação da *capacidade térmica mássica*. Ao lado do respectivo ano, o asterisco se refere a um exame realizado em época especial, e segue com os níveis de escolarização em que o exame foi direcionado e aplicado. O símbolo “/” se refere ao fato de estar escrito na folha do exame o termo “ou”, i. e., “10º ou 11º ano”, subentendendo que o exame seja para um nível *ou* outro, muito provavelmente tenha na realidade sido aplicado às duas séries, i.e., “10º; 11º ano”. Essas informações são baseadas na folha da prova dos exames, não podendo ter sido confirmado.

Tabela 54: Mapeamento na abordagem de temáticas relacionadas à termodinâmica no ensino experimental exigidas pelo programa curricular português, que no caso foram cobradas aos alunos nos exames nacionais de 2006 a 2016.

Temáticas	Série	Fase	Numeração dos Itens Identificados no contexto da Termodinâmica de acordo com Exames Nacionais (Física e Química A) de 2006 a 2016																		
			<i>(Decreto-Lei n.º 74/2004)</i>							<i>(Decreto-Lei n.º 139/2012)</i>											
			2006 11º	2007 11º/12º	2008 11º/12º	2009 11º/12º	2010 10º; 11º	2011 10º; 11º	2011* 10º; 11º	2012 10º; 11º	2012* 10º; 11º	2013 10º; 11º	2013* 10º; 11º	2014 11º	2014* 11º	2015 11º	2015* 11º	2016 11º	2016* 11º	Tot. (n= 205)	
Rendimento no aquecimento	10º	1ª					1.5	2 ⁽⁵⁾			2.1 ⁽⁴⁾	1.2 ⁽⁴⁾	1.2 ⁽²⁾						8		
		2ª					3.4	3 ⁽⁵⁾			3 ⁽¹⁾	2.2 ⁽⁴⁾	2 ⁽²⁾						5		
	11º	1ª	3.1			5.2.1	1.5	2 ⁽⁵⁾		2.1 ⁽⁴⁾	1.2 ⁽⁴⁾	1.2 ⁽²⁾		2 ⁽²⁾			3.2 ⁽²⁾	1.2 ⁽⁴⁾	2.1 ⁽²⁾	19	
		2ª			6	3.3	3.4	3 ⁽⁵⁾		3 ⁽¹⁾	2.2 ⁽⁴⁾	2 ⁽²⁾		3 ⁽⁴⁾				3 ⁽⁴⁾	2.2 ⁽²⁾	3 ⁽²⁾	9
	12º	1ª	3.1			5.2.1															2
		2ª			6	3.3				4 ⁽¹⁾		3.2 ⁽¹⁾				2.2 ⁽⁵⁾		1.1 ⁽³⁾			2
Absorção e emissão da radiação	10º	1ª							1 ⁽¹⁾											15	
		2ª						2 ^{(6)Q}	2 ⁽¹⁾		2.1 ⁽⁴⁾	1.1 ⁽²⁾	1 ⁽²⁾							12	
	11º	1ª			4.2	1.1			2 ^{(6)Q}	3 ⁽¹⁾	2.2 ⁽⁴⁾	1.2 ⁽²⁾	2 ⁽²⁾	2.2 ^{(5)Q}	3 ⁽⁴⁾	4 ⁽¹⁾	1 ⁽¹⁾	2 ⁽¹⁾	1.2 ⁽⁴⁾	26	
		2ª	1.4.1			3.4			1 ⁽¹⁾		3.1 ⁽⁶⁾	1.3 ⁽²⁾	2.1 ^{(3)q}								27
	12º	1ª	1.4.2	1.2	1.2	4.1 _Q			4 ⁽¹⁾							3 ⁽¹⁾					4
		2ª	1.4.3 _Q	2.2.3 _Q	3.2.2 _Q	4.3 _Q			5 ⁽¹⁾	1 ⁽¹⁾						3.1 ⁽²⁾					
		1.5						1.3 ⁽²⁾	5 ⁽¹⁾		4 ^{(5)Q}			2.1 ⁽⁵⁾							
				4.2	1.1			2 ^{(2)Q}													

É relativa a necessidade de que as maiores dificuldades em conceitos básicos que os alunos encontraram no início e/ou durante o ensino secundário estejam presentes no exame, mas pode sinalizar uma preocupação dos elaboradores em questões sensíveis e significativas para o aprendizado. Essa é uma questão que deve, sobretudo, ser vigiada e avaliada pelos professores durante o processo da aprendizagem. Também se torna importante verificar se esses conteúdos e habilidades cognitivas exigidos são de fato essenciais para a unidade temática da *Termodinâmica*, se cobre o construto a ser avaliado, e se o conhecimento exigido reflete aquilo que de fato eles devem saber. Obviamente essa decisão já foi realizada com base no programa quando foram definidos os conteúdos, porém não se percebe o critério de escolha para esses padrões dentro dos programas, apesar das orientações explicitarem de forma genérica as situações, os contextos e processos metodológicos sugeridos a serem desenvolvidos. Também não foi encontrado uma matriz de referência para a elaboração dos itens de modo a descrever de forma clara as habilidades e capacidades que se pretende avaliar, o critério para isso parece estar oculto, implicitamente entende-se que se baseia na matriz curricular ou programa curricular da disciplina.

D1.3.1: Algumas Considerações sobre os Exames Nacionais

Observa-se o uso dos mesmos termos semelhantes e por vezes idênticos ao título das temáticas obrigatórias dentro dos enunciados, dessa forma é possível uma identificação imediata da categoria e provavelmente remete a uma facilidade cognitiva no direcionamento dos processos de resolução dos problemas, considerando ainda que são fornecidas as fórmulas/equações/relações, admite-se o uso de calculadoras digitais (específicas), e um pouco mais de 5 min para resolver cada questão (algumas são abertas e possuem maior peso pela exigência descritiva das etapas de resolução). Antes de 2014 os exames eram destinados a dois níveis de escolaridade praticamente a mesma faixa etária, no entanto é um risco que se deve assumir quanto ao critério de legibilidade, tendo em vista que *alunos em nível mais avançado tenham supostamente uma maior capacidade ou probabilidade de acertos*.

A parte da Química que trata a transição do elétron entre os níveis e subníveis de energia muitas vezes está associado à temática sobre *“absorção/emissão de radiação”*

e também tende a se fundir com a categoria *“energia elétrica fornecida por um painel fotovoltaico”* quando envolve uma compreensão sobre o efeito fotoelétrico, seja na absorção e emissão da radiação. A temática *“rendimento no aquecimento”* também funde com a categoria *“absorção e emissão da radiação”* quando o processo de aquecimento é a radiação, assim como também funde com *“energia elétrica fornecida por um painel fotovoltaico”* pelo fato de que o rendimento ocorre via ondas de calor (radiação térmica ou eletromagnética). Entendendo que *“rendimento no aquecimento”* está relacionado a potência calorífica ou fluxo de calor por unidade de tempo, também se incorpora questões relacionadas a *Lei de Fourier* ou da condução térmica, pois a potência pode ser de absorção/emissão (via radiação) ou condução do calor, tendo o seu rendimento variando em função da espessura do material, condutividade térmica, etc. A *“capacidade térmica mássica”* está relacionada com o calor específico podendo ser sensível ou latente, em outros casos relaciona-se com problemas de trocas de calor, com isso foram consideradas todas as questões que envolvem essa grandeza, seja pra calculá-la ou não. Foi considerado *“balanço energético num sistema termodinâmico”* sempre que envolve uma compreensão sobre o processo de transformação de energia de acordo com a *1ª Lei da Termodinâmica*, isso pode acontecer em processos de transferências de calor estando o sistema fechado ou aberto, portanto não se restringe apenas ao funcionamento de máquinas térmicas, mas também na condução de calor entre corpos em contato com diferentes temperaturas. De modo geral, essas coincidências podem justificar uma mesma questão estar relacionada a duas temáticas ou categorias dentro da termodinâmica.

A incidência da temática sobre radiação térmica (ou eletromagnética) visando determinar comprimento de onda, velocidade de propagação, frequência ou refração enfatiza mais um domínio dentro da ondulatória (ótica geométrica) do que na termodinâmica, portanto não foram consideradas, a não ser que envolva dentro das causas variáveis relacionadas à física térmica (tipo, a frequência ou comprimento de onda *em função da temperatura*). Entretanto, quando remete a ótica física, na qual envolve a interpretação do campo magnético, espectros, e cores é algo mais pertinente por estar *associado ao modelo cinético molecular*, e isso compõe uma melhor compreensão

do que seja uma radiação. Vale lembrar que a radiação e termodinâmica fazem parte de naturezas epistemológicas distintas, sendo a radiação mais centrada no domínio da física quântica.

D1.3.2: Comentários sobre os Conteúdos Térmicos em Destaque

Naturalmente, a identificação do *padrão de conteúdo* da termodinâmica pode (e deve) ser encontrado nos programas curriculares e livros didáticos de Física recomendados pelo Ministério da Educação. Entende-se que o conteúdo nos itens que mais incidem nos exames nacionais são evidências mais concretas daquilo que o sistema exige, ainda que de forma intencional ou não, representam os conteúdos em destaque, o conhecimento considerado relevante diante da capacidade de resolver problemas.

Com isso foi realizado um mapeamento dos itens nos exames nacionais na última década (período de 2006 a 2016), e assim foram identificadas as incidências de **temáticas obrigatórias** a serem desenvolvidas pelos professores através de atividades práticas laboratoriais no ensino experimental, considerados assim como conhecimentos imprescindíveis na Física Térmica, que são: *rendimento no aquecimento; absorção e emissão da radiação; energia elétrica fornecida por um painel fotovoltaico; capacidade térmica mássica; balanço energético num sistema termodinâmico*. A obrigatoriedade dessas temáticas sugere que sejam padrões estabelecidos a serem ensinados dentro da termodinâmica.

Desde 2001, houve duas novas propostas de reforma para o programa curricular de Física em Portugal, porém os conteúdos permanecem com poucas mudanças, assim como a sua forma de abordagem nos livros didáticos, que apresentam uma forma estritamente orientada com as descrições nos programas, aparentemente com pouca variação de um para outro. A mais recente proposta (Fiolhais et al., 2014) reduziu (fundiu) para três atividades laboratoriais: *Radiação e potência elétrica de um painel fotovoltaico; Capacidade térmica mássica; Balanço energético num sistema termodinâmico*. Essa mudança entrou em vigor em 2015.

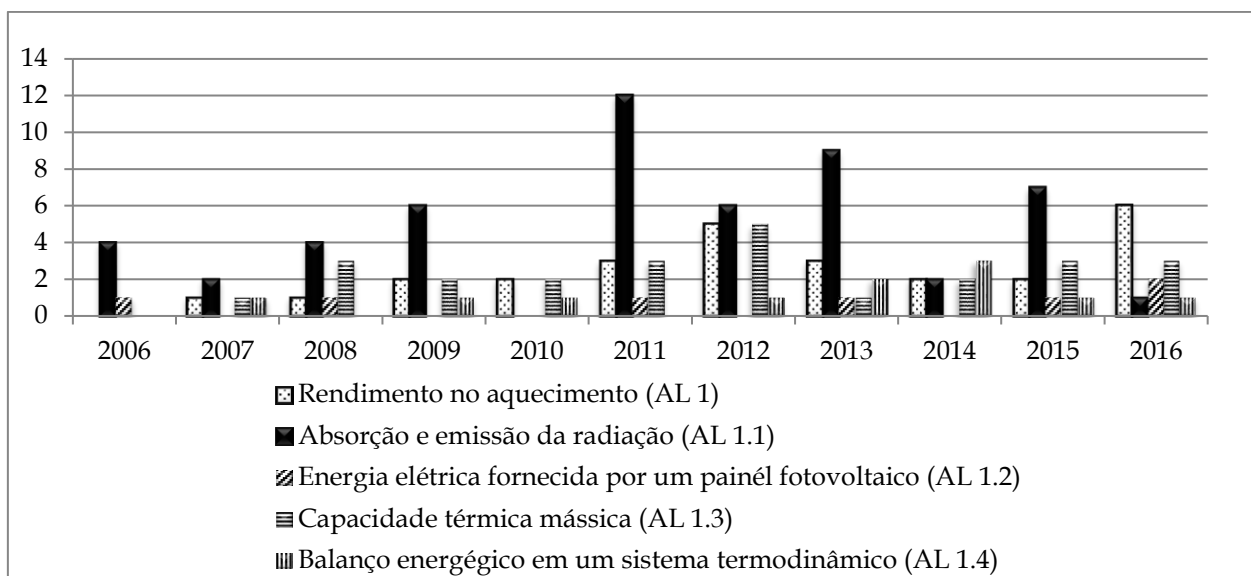


Figura 93: Quantidade de itens relacionados à temática em *Termodinâmica* nos exames nacionais em Portugal.

De acordo com o site do *Instituto de Avaliação Educativa* (IAVE), o conteúdo da disciplina de Física em nível de ensino secundário (Decreto-Lei n.º 286/89) de 1997 até 2006 se concentrava no 12º ano, e não foi encontrado nenhum item dentro da termodinâmica. Nesse mesmo ano, o conteúdo da disciplina Física e Química A estavam restritos ao nível de escolarização (Decreto-Lei n.º 74/2004) do 11º ano, sendo de 2007 a 2009 conteúdos do 11º e 12º ano, de 2010 a 2013 conteúdos do 10º e 11º ano, e a partir de 2014, somente o 11º ano. Em 2011 houve duas provas na 1ª fase e duas na 2ª fase, mais uma prova em época especial. De modo geral, tanto na 1ª fase quanto na 2ª, a quantidade de itens permanece com 28 itens e com duração de 120 min (+ 30 min de tolerância). Pode-se constatar uma maior incidência na temática sobre **“absorção/emissão de radiação” (45%)** ao longo dos últimos 10 anos, assim como na temática de **“rendimento no aquecimento” (22%)** e **“capacidade térmica mássica” (20%)**.

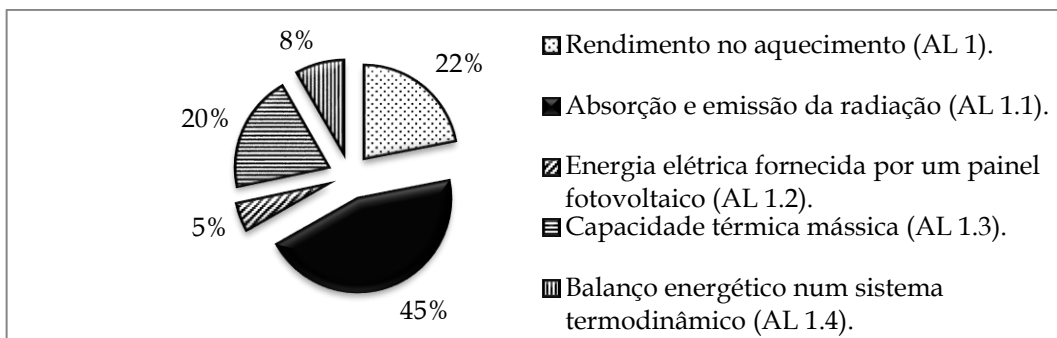


Figura 94: Incidências nas temáticas em termodinâmica entre 2006 e 2016.

Todos esses conteúdos na realidade são temáticas associadas a «processos termodinâmicos», e uma preocupação especial com a “*capacidade térmica mássica*”, grandeza física que se configura como uma propriedade térmica do corpo. A capacidade de resolver *problemas práticos* envolve a necessidade de compreender diversos conceitos e de mobilizar vários saberes, entretanto é importante verificar *competências de base conceitual* antes de *competências de aperfeiçoamento* que necessita de um domínio mais estruturado do conhecimento. Questões que explorassem conceitos básicos foram escassas e não foram consideradas representativas, subentende-se que para realizar a maioria das atividades necessite de conhecimentos sobre os processos de térmicos, e obviamente os conceitos básicos estão inclusos e necessariamente sendo mobilizados e aplicados nas diversas situações. A capacidade avaliada sem verificar a aquisição desses conceitos se torna um risco de afirmar sobre a progressão da aprendizagem em termos de aquisição em níveis de proficiências mais complexas e abstratas.

D2: Brasil

D2.1: PCN+ Ensino Médio (2002) - Orientações Educacionais Complementares aos Parâmetros Curriculares Nacionais: Ciências da Natureza, Matemática e suas Tecnologias (habilidades relativamente pertinentes ao TCE).

Quadro 86: Sugestões Temáticas do PCN+/2002 para Desenvolver Competências na Física Térmica.

N	Tema 2. Calor, ambiente e usos de energia
	Unidade temática: 1 - Fontes e trocas de calor
1	Identificar fenômenos, fontes e sistemas que envolvem calor para a escolha de materiais apropriados a diferentes situações ou para explicar a participação do calor nos processos naturais ou tecnológicos.
2	Reconhecer as propriedades térmicas dos materiais e os diferentes processos de troca de calor, identificando a importância da condução, convecção e irradiação em sistemas naturais e tecnológicos.
3	Utilizar o modelo cinético das moléculas para explicar as propriedades térmicas das substâncias, associando-o ao conceito de temperatura e à sua escala absoluta.
N	Unidade temática: 2 - Tecnologias que usam calor: motores e refrigeradores •
4	Compreender a relação entre variação de energia térmica e temperatura para avaliar mudanças na temperatura e/ou mudanças de estado da matéria em fenômenos naturais ou processos tecnológicos.
5	Identificar a participação do calor e os processos envolvidos no funcionamento de máquinas térmicas de uso doméstico ou para outros fins, tais como geladeiras, motores de carro etc., visando à sua utilização adequada.
6	Identificar o calor como forma de dissipação de energia e a irreversibilidade de certas transformações para avaliar o significado da eficiência em máquinas térmicas
N	Unidade temática: 3 - O calor na vida e no ambiente
7	Compreender o papel do calor na origem e manutenção da vida.
8	Reconhecer os diferentes processos envolvendo calor e suas dinâmicas nos fenômenos climáticos para avaliar a intervenção humana sobre o clima.
9	Identificar e avaliar os elementos que propiciam conforto térmico em ambientes fechados como sala de aula, cozinha, quarto etc., para utilizar e instalar adequadamente os aparelhos e equipamentos de uso corrente.
N	Unidade temática: 4 - Energia: produção para uso social
10	Identificar as diferentes fontes de energia (lenha e outros combustíveis, energia solar etc.) e processos de transformação presentes na produção de energia para uso social.
11	Identificar os diferentes sistemas de produção de energia elétrica, os processos de transformação envolvidos e seus respectivos impactos ambientais, visando às escolhas ou análises de balanços energéticos.
12	Acompanhar a evolução da produção, do uso social e do consumo de energia, relacionando-os ao desenvolvimento econômico, tecnológico e à qualidade de vida ao longo do tempo.

Fonte: Brasil (2002).

D2.2: Matriz de Referência de Ciências da Natureza e suas Tecnologias (campos considerados pertinentes ao TCE).

Quadro 87: Habilidades da Matriz de Referência de Ciências da Natureza e suas Tecnologias consideradas pertinentes ao TCE.

Competência de área 1 – Compreender as ciências naturais e as tecnologias a elas associadas como construções humanas, percebendo seus papéis nos processos de produção e no desenvolvimento econômico e social da humanidade.
H1 – Reconhecer características ou propriedades de fenômenos ondulatórios ou oscilatórios, relacionando-os a seus usos em diferentes contextos.
H2 – Associar a solução de problemas de comunicação, transporte, saúde ou outro com o correspondente desenvolvimento científico e tecnológico.
H3 – Confrontar interpretações científicas com interpretações baseadas no senso comum, ao longo do tempo ou em diferentes culturas.
H4 – Avaliar propostas de intervenção no ambiente, considerando a qualidade da vida humana ou medidas de conservação, recuperação ou utilização sustentável da biodiversidade.
Competência de área 2 – Identificar a presença e aplicar as tecnologias associadas às ciências naturais em diferentes contextos.
H5 – Dimensionar circuitos ou dispositivos elétricos de uso cotidiano.

H6 – Relacionar informações para compreender manuais de instalação ou utilização de aparelhos, ou sistemas tecnológicos de uso comum.
H7 – Selecionar testes de controle, parâmetros ou critérios para a comparação de materiais e produtos, tendo em vista a defesa do consumidor, a saúde do trabalhador ou a qualidade de vida.
Competência de área 6 – Apropriar-se de conhecimentos da física para, em situações-problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.
H20 – Caracterizar causas ou efeitos dos movimentos de partículas, substâncias, objetos ou corpos celestes.
H21 – Utilizar leis físicas e/ou químicas para interpretar processos naturais ou tecnológicos inseridos no contexto da termodinâmica e/ou do eletromagnetismo.
H22 – Compreender fenômenos decorrentes da interação entre a radiação e a matéria em suas manifestações em processos naturais ou tecnológicos, ou em suas implicações biológicas, sociais, econômicas ou ambientais.
H23 – Avaliar possibilidades de geração, uso ou transformação de energia em ambientes específicos, considerando implicações éticas, ambientais, sociais e/ou econômicas.

Fonte: Adaptado de MEC/INEP (2016).

D2.3: Brasil: Padrão de Conteúdos e Habilidades da Física Térmica segundo os documentos oficiais

“Tal como a compartimentação das disciplinas desintegra o tecido natural da complexidade, também a visão transdisciplinar é capaz de restituí-lo”. (Morin & Moigne, 2007) (EDGAR MORIN; JEAN-LOIUS LE MOIGNE, 2007, p. 55)

Segundo o PCNEM: “[...] o currículo deve ser articulado em torno de eixos básicos orientadores da seleção de conteúdos significativos, tendo em vista as competências e habilidades que se pretende desenvolver no Ensino Médio”. Segundo o PCN+ de Física, foram privilegiados seis temas geradores com abrangência para organizarem o ensino de Física, que são: (1) *movimentos: variações e conservações*; (2) **calor, ambiente, fontes e usos de energia**; (3) *equipamentos eletromagnéticos e telecomunicações*; (4) *som, imagem e informação*; (5) *matéria e radiação*; (6) *universo, terra e vida* (Brasil, 2002, p. 19). Diante disso, parece ser adequado escolher o paradigma da relação entre Ciência, Tecnologia, Sociedade e Ambiente¹⁰⁷ (CTSA) na articulação do tema gerador com outros subtemas transversais integradores.

O problema central passa a ser, então, o de **identificar as competências em Física desejadas**. Mas ainda que uma reflexão mais aprofundada nos permita listá-las, essas listas serão sempre parciais, dada a abrangência das habilidades envolvidas. Caberá sempre ao professor, dentro das condições específicas nas quais desenvolve seu trabalho, em função do perfil de sua escola e do projeto pedagógico em andamento, selecionar, priorizar, redefinir e organizar os objetivos em torno dos quais faz mais sentido trabalhar. (Brasil, 2002, *grifo nosso*)

¹⁰⁷ Esse é o paradigma vigente em Portugal para os conteúdos de Física, que por sua vez se adequam aos mais diversos temas geradores propostos pelos PCN no Brasil.

Dentro do tema “**Calor, ambiente e usos de energia**” encontram-se 4 unidades temáticas com a sugestão de 12 habilidades distribuídas, na qual 5 habilidades foram identificadas como sendo relativamente pertinentes ao TCE, que são as (ver no apêndice): 1, 2, 4, 5 e 9. A unidade 4, denominada “**Energia: produção para uso social**”, envolve uma relação da ciência com a tecnologia e sociedade, aspecto que o TCE não explora.

Os objetos de conhecimento da Física Térmica associados à Matriz de Referência do ENEM, se refere ao tópico “O calor e os fenômenos térmicos” e os conteúdos ou objetos do conhecimento que são:

Conceitos de calor e de temperatura. Escalas termométricas. **Transferência de calor e equilíbrio térmico.** Capacidade calorífica e calor específico. **Condução do calor.** Dilatação térmica. **Mudanças de estado físico e calor latente de transformação.** Comportamento de Gases ideais. Máquinas térmicas. Ciclo de Carnot. **Leis da Termodinâmica.** **Aplicações e fenômenos térmicos de uso cotidiano.** **Compreensão de fenômenos climáticos relacionados ao ciclo da água.** (Brasil, 2009; Brasil, 2016, p. 52)

As competências de referência para o ensino médio também encontram-se nas matrizes dos exames nacionais do ENCEEJA. Dentro da *Matriz de Referência de Ciências da Natureza e suas Tecnologias* (MEC/INEP, 2016), foram identificadas 3 habilidades entre 3 de 8 competências estabelecidas, sendo consideradas pertinentes a área da Física e que incide sobre o conteúdo explorado no teste TCE, que são (ver o apêndice): H3, H6 e H21.

Atualmente, o processo de transição das referências está sendo realizada através da *Base Nacional Comum Curricular*, com a 3ª versão recentemente finalizada, que por enquanto se restringe a educação infantil e fundamental, enquanto tramita o processo de reforma do ensino médio (Brasil, 2016). É muito provável que as habilidades que envolvem o conhecimento térmico na BNCC seja incluída nas matrizes de referência do SAEB, e reflita na do ENEM e ENCEEJA.

Na BNCC esta previsto uma distribuição de 10 competências gerais (Brasil, 2017, pp. 18-19): *conhecimento; pensamento científico, crítico e criativo; senso estético; comunicação; argumentação; cultura digital; autogestão; autoconhecimento e autocuidado;*

empatia e cooperação; autonomia. No documento, o conteúdo de termodinâmica nas séries iniciais sugere noções a serem desenvolvidas sobre as propriedades dos materiais e de sua relação com o meio ambiente, bem como da *formação sobre o conceito de calor*. Esse objeto de conhecimento encontra-se incluso na componente de **Ciências do 2º ano fundamental** dentro do eixo temático norteador denominado “Terra e Universo”, na qual explora a relação da energia solar como fonte de luz e calor (Brasil, 2017, p. 286).

Nos anos iniciais, as crianças já se envolvem com uma série de objetos, materiais e fenômenos, em sua vivência diária e na relação com o entorno. Tais experiências são o ponto de partida para possibilitar a **construção das primeiras noções** sobre os materiais, seus usos e propriedades, bem como suas interações com luz, som, **calor**, eletricidade e umidade, entre outros elementos, estimulando a construção de hábitos saudáveis e sustentáveis por meio da preservação da saúde a partir dos cuidados e riscos associados à integridade física e à qualidade auditiva e visual e da construção coletiva de propostas de reciclagem e reutilização de materiais. (Brasil, 2017, p. 277)

Considera-se a formação sobre o conceito de calor perpassa pela necessidade compreender as propriedades dos materiais, na qual se inclui as relacionadas aos fenômenos térmicos, bem como na comparação do aquecimento de diversas substância por meio da luz solar, com isso, consideram-se a aquisição das seguintes habilidades (Ibid, 2017, p. 287):

- EF02CI02: *Justificar o uso de diferentes materiais em objetos de uso cotidiano, tendo em vista algumas propriedades desses materiais (flexibilidade, dureza, transparência etc.);*
- EF02CI08: *Comparar e registrar o efeito da radiação solar (aquecimento) em diferentes tipos de superfície (água, areia, solo, superfície escura, superfície clara etc.).*

Também surge na componente de **Ciências do 7º ano fundamental** de forma mais abrangente na unidade temática “Matéria e Energia”, tendo como objetos de conhecimento (Brasil, 2017, p. 298): *máquinas simples; formas de propagação do calor; equilíbrio termodinâmico e vida na Terra; história dos combustíveis e das máquinas térmicas.* Algumas habilidades declaradas fazem correspondência de forma direta com o TCE (e outras de forma indireta quando envolve a relação da Ciência com questões econômicas e socioambientais), e são identificadas como (Ibid, 2017, p. 299):

- EF07CI02: *Diferenciar temperatura, calor e sensação térmica nas diferentes situações de equilíbrio termodinâmico cotidianas;*
- EF07CI03: *Utilizar o conhecimento das formas de propagação do calor para justificar a utilização de determinados materiais (condutores e isolantes) na vida cotidiana e explicar o princípio de funcionamento de alguns equipamentos (garrafa térmica, coletor solar etc.);*
- EF07CI04: *Avaliar o papel do equilíbrio termodinâmico para a manutenção da vida na Terra, para o funcionamento de máquinas térmicas e em outras situações cotidianas;*
- EF07CI05: *Discutir o uso de diferentes tipos de combustível e máquinas térmicas ao longo do tempo, para avaliar avanços, questões econômicas e problemas socioambientais causados pela produção e uso desses materiais e máquinas.*

Também foi verificada uma habilidade incidente com termo “calor” na componente de **Geografia no 1º ano fundamental** em sentido climatológico ao invés de energia térmica em processo de transferência, sendo assim um uso considerado equivocado, apesar de compreendido culturalmente. As habilidade relacionadas são (Ibid, 2017, p. 323):

- EF01GE05: *Observar e descrever ritmos naturais (dia e noite, variação de temperatura e umidade etc.) em diferentes escalas espaciais e temporais, comparando a sua realidade com outras;*
- EF01GE10: *Descrever características de seus lugares de vivência relacionadas aos ritmos da natureza (chuva, vento, calor etc.);*
- EF01GE11: *Associar mudanças de vestuário e hábitos alimentares em sua comunidade ao longo do ano, decorrentes da variação de temperatura e umidade no ambiente.*

Um equívoco semelhante ocorre na componente de **Geografia para a série do 6º ano fundamental**:

- EF06GE13: *Analisar consequências, vantagens e desvantagens das práticas humanas na dinâmica climática (ilha de calor etc.).*

Na componente de **Matemática do 4º ano fundamental**, dentro da unidade temática de geometria, tem-se como objeto de conhecimento “*Medidas de temperatura em grau Celsius: construção de gráficos para indicar a variação da temperatura (mínima e máxima) medida em um dado dia ou em uma semana*” (Ibid, 2017, p. 248). As habilidades

associadas buscam uma aproximação com o contexto cotidiano e sofrem o uso de recursos eletrônicos no processo de aprendizagem (Ibid, 2017, p. 249):

- EF04MA23: *Reconhecer temperatura como grandeza e o grau Celsius como unidade de medida a ela associada e utilizá-lo em comparações de temperaturas em diferentes regiões do Brasil ou no exterior ou, ainda, em discussões que envolvam problemas relacionados ao aquecimento global;*
- EF04MA24: *Determinar as temperaturas máxima e mínima diárias, em locais do seu cotidiano, e elaborar gráficos de colunas com as variações diárias da temperatura, utilizando, inclusive, planilhas eletrônicas.*

Na **Matemática do 5º ano fundamental** também se explora a grandeza de temperatura, buscando assim dar sentido real as relações matemáticas envolvidas na unidade temática “Grandezas e medidas”, e objeto de conhecimento “*Medidas de comprimento, área, massa, tempo, temperatura e capacidade: utilização de unidades convencionais e relações entre as unidades de medida mais usuais*” (Ibid, 2017, p. 252). A habilidade associada é (Ibid, 2017, p. 253):

- EF05MA19: *Resolver e elaborar problemas envolvendo medidas das grandezas comprimento, área, massa, tempo, temperatura e capacidade, recorrendo a transformações entre as unidades mais usuais em contextos socioculturais.*

De forma semelhante na **Matemática do 6º ano fundamental**, na mesma unidade temática do 5º ano, porém sendo denominado no objeto do conhecimento “*Problemas sobre medidas envolvendo grandezas como comprimento, massa, tempo, temperatura, área, capacidade e volume*”, ou seja, avançando da perspectiva de compreender as medidas para tentar resolver problemas relacionados (Ibid, 2017, p. 256). Nisso a habilidade relacionada é (Ibid, 2017, p. 259):

- EF06MA22: *Resolver e elaborar problemas que envolvam as grandezas comprimento, massa, tempo, temperatura, área (triângulos e retângulos), capacidade e volume (sólidos formados por blocos retangulares), sem uso de fórmulas, inseridos, sempre que possível, em contextos oriundos de situações reais e/ou relacionadas às outras áreas do conhecimento.*

Na análise da BNCC, constata-se dessa forma que hajam cerca de 14 habilidades que contribuam para o desenvolvimento de competências gerais dentro do contexto

do conhecimento técnico, estando distribuídas nas componentes de Ciências, Matemática e Geografia.

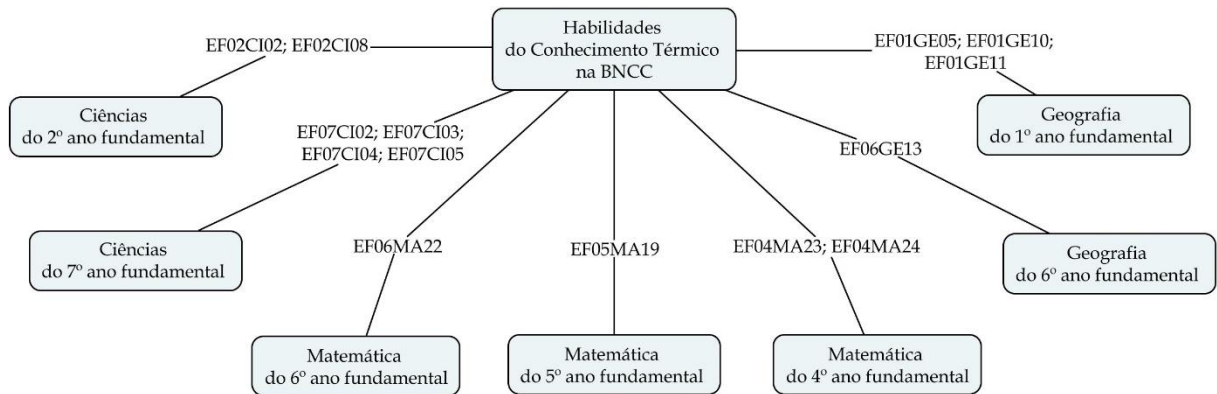


Figura 95: Habilidades consideradas pertinentes ao desenvolvimento de competências científicas dentro do conhecimento técnico, de acordo com a 3ª versão da BNCC.

E - Detalhes nos Resultados Obtidos

E1: Teste-Piloto (quatro escolas secundárias em Portugal)

E1.1: Quanto às etapas e a descrição nos procedimentos.

Tendo como referência as etapas e procedimentos para desenvolver a aplicação de um exame nacional, é possível obter informações preciosas de como delimitar e adaptar esses processos para outros contextos, na qual se aplicam outros diversos testes em larga escala. Pode-se observar que as etapas preliminares de testagem e calibração dos itens são fundamentais para os ajustes, bem como o uso de *Blocos Incompletos Balanceados (BIB)*, possibilitando uma maior abrangência de uma matriz sem cansar o respondente, e ainda assim obter níveis satisfatórios de consistência. Esses processos foram elaborados com base nos processos realizados no *Sistema de Avaliação da Educação Básica (SAEB)* no Brasil (atualmente, sendo o SINAEB), e servem de referência na adaptação para esse e outros estudos semelhantes. Foram estabelecidos e descritos os procedimentos em 3 fases: *concepção, aplicação e avaliação*.

- **Fase de concepção** com: *determinação da população a ser avaliada; elaboração da matriz ou utilização de uma matriz de referência já elaborada (p. ex., SAEB); construção dos itens; pré-testagem dos itens; Análise estatística e pedagógica dos itens (calibração e interpretação).*
- **Fase de aplicação:** *montagem dos cadernos de testes (blocos incompletos balanceados – BIB); elaboração dos testes padronizados; elaboração dos questionários contextuais; aplicação dos instrumentos; processamento e constituição da base de dados da avaliação.*
- **Fase de avaliação:** *análise dos resultados: utilização da TCT e TRI; produção dos resultados; construção e interpretação das escalas de proficiência; divulgação dos resultados.*

Quadro 88: Sugestões para Etapas e Procedimentos na Aplicação de Testes ou Exames em Larga Escala.

DESCRICHÃO		Procedimentos Recomendados em Exames de Larga Escala
CONCEPÇÃO	1. Determinação da população a ser avaliada.	Realiza-se um estudo sobre o tipo de amostragem (censitária ou probabilística) e do <i>design</i> . No <i>design experimental</i> se definem os procedimentos que atendam aos critérios de <i>validade externa</i> ,
	2. Elaboração da Matriz ou Utilização de uma Matriz de Referência já elaborada (p. ex. SAEB)	Elabora-se a matriz de referência <i>a priori</i> , e nela contém <i>descritores</i> que determinam as competências essenciais a serem avaliadas pelo exame. Faz-se uma análise taxonômica sobre <i>objetivos específicos</i> para determinar diferentes níveis abstração e complexidade. Deve-se ainda realizar a primeira etapa do processo de <i>validade do construto</i> , na qual é a <i>validade do conteúdo</i> . Para isso, os conteúdos a serem considerados padrões dentro de determinada área do conhecimento atende minimamente às finalidades pedagógicas de formação do nível de escolaridade. Pressupõe-se que a escolha do conteúdo deve estar relacionada a temáticas transversais, tendo a interdisciplinaridade e a contextualização como fundamentos construtivistas. Este processo é realizado por meios da opinião de peritos/especialistas no assunto quanto à sua adequação e exaustividade.
	3. Construção dos itens.	Utiliza-se o conhecimento sobre a “engenharia de construção de itens” tendo como base a matriz de referência. As competências definidas nos descritores por sua vez devem conter informações sobre <i>capacidades</i> (habilidades, fazendo a relação aos <i>saberes necessários</i>), <i>conteúdos</i> com base na noção dos <i>objetivos específicos</i> preestabelecidos. Na elaboração do item a competência deve ser desenvolvida dentro de uma <i>família de situações-problema</i> , ou seja, deve-se fazer referência ao contexto na qual a competência deverá agir. O processo de <i>validade de construto</i> também está presente, buscando identificar a uniformidade entre determinados grupos de itens na qual supostamente representam um determinado construto.
	4. Pré-testagem dos itens.	Aplica-se em uma amostra probabilística (representativa a população estabelecida) exercendo a função de <i>avaliação diagnóstica</i> quanto à qualidade dos itens e como <i>avaliação preditiva</i> quanto ao suposto desempenho dos respondentes e quanto à eficiência ou <i>fidedignidade</i> do teste.
	5. Análise Estatística e pedagógica dos itens (calibração e interpretação)	Avaliam-se os indicadores estatísticos dos itens e do teste quanto à <i>validade de construto</i> , nomeadamente a <i>validade de critério</i> . A validade de critério é de natureza estatística se baseado na <i>validade concorrente</i> (comparativa a outros testes semelhantes) e a <i>validade preditiva</i> (se os itens de fato irão avaliar futuramente os níveis de proficiência desejados). A <i>validade de construto</i> também está presente na interpretação pedagógica de determinados grupos de itens, podendo ser realizados conjuntamente métodos estatísticos de análise fatorial confirmatória de eventuais construtos preestabelecidos na matriz de referência. O conjunto dos procedimentos visa garantir uma <i>validade interna do instrumento</i> .
APLICÇÃO	6. Montagem dos cadernos de testes (<i>Blocos Incompletos Balanceados – BIB</i>).	Realização de procedimentos matemáticos exercendo a função de distribuição adequada dos descritores em diversas provas, mantadas através de blocos incompletos em relação à matriz de referência. Entretanto, o total de provas busca atender a matriz de referência em sua completude. Este procedimento também está associado à <i>validação do construto</i> , na qual resolve o problema de fadiga do respondente no sentido de avaliar todas as competências e ao mesmo tempo contribui para uma elevada consistência nas análises psicométricas. Este procedimento contribui para a <i>validade externa</i> durante no processo de montagem do teste, na qual deve estar ligado ou corresponder a critérios intrínsecos no momento da aplicação.
	7. Elaboração dos testes padronizados.	Confeccionam-se os testes padronizados já com os itens calibrados e testados, e já com os números de testes preestabelecidos pela análise anterior de BIB. Deve atender a critério de <i>validade externa</i> , nomeadamente ao rigor técnico e operacional para que não haja vazamento de informação das provas.
	8. Elaboração dos questionários contextuais	Confeccionam-se os questionários de forma mais imediata, tendo em vista que uma vez elaborados (utilizam-se geralmente os modelos anteriores) podem ser replicados naturalmente, pois não medem traços latentes, e, portanto, não são analisados com a Psicometria, mas sim através da estatística descritiva. A <i>validade externa</i> fica por parte dos <i>avaliadores externos</i> no sentido de conferir as informações <i>in locu</i> durante a visita técnica. A escolha dos avaliadores também faz parte desse processo e é de incumbência da entidade responsável pela administração do exame.
	9. Aplicação dos instrumentos.	Aplicam-se utilizando a técnica BIB-espiral, distribuindo as provas em cada sala de aplicação de modo uniforme e sistemático, sendo bem balanceada a amostra. Este procedimento contribui para a <i>validade externa</i> durante no processo <i>aplicação do teste</i> , na qual deve estar ligado ou corresponder aos critérios intrínsecos do momento da montagem dos testes.
	10. Processamento e constituição da base de dados da avaliação	Submetem-se os dados das provas dos respondentes recolhidos por meio de um cartão-reposta a um <i>Scantron</i> para que seja constituída uma base de dados. <i>Scantron</i> é um dispositivo utilizado para escanear formulários (do tipo cartão-resposta) utilizados em avaliações de larga escala, podendo ser provas de opção de múltiplas escolhas (politômico) ou de verdadeiro o falso (dicotômico). No caso do teste dicotômico, a <i>matriz-resposta</i> constitui de letras correspondentes as opções assinaladas, que por sua vez se convertem numa <i>matriz-binária</i> , sendo a alternativa correta ou correspondente à chave-de-correção com a numeração “1” e a errada “0”. De modo análogo ocorre no teste de múltiplas escolhas, entretanto, cada <i>ranking</i> da unidade de medida será considerado um item ou uma

		resposta a ser avaliada na <i>matriz-resposta</i> . Os dispositivos e mecanismos responsáveis pela constituição da base de dados devem ser rigorosos para garantir a <i>validade externa</i> do instrumento.
AVALIAÇÃO	11. Análise dos resultados: utilização da TCT e TRI.	Analisam-se a base de dados (ou <i>matriz-resposta</i>) utilizando os fundamentos da TCT e da TRI, que por sua vez utiliza a técnica estatística denominada <i>análise de multigrupos</i> , possibilitando a reintegração das diferentes provas (diferentes grupos de provas) numa só, e assim se reconstitui uma base de dados que contempla toda a <i>Matriz de Referência</i> . Para a realização de análises comparativas e a montagem das séries históricas de desempenho, são utilizados <i>itens-âncoras</i> pertencentes à <i>escala de proficiência de referência</i> , em outras palavras, itens pertencentes ao primeiro teste ou exame aplicado, na qual é tida como referência comparativa das séries dos anos posteriores. Por esse motivo não são divulgadas as provas que possuam escalas de proficiência previamente estabelecidas, pelo fato que são utilizados itens-âncora de provas anteriores, e elas se repetem nas provas seguintes. No caso, por exemplo, do SAEB, ENEM e ENCCEJA, a base de dados é enviada para análise a três equipes de analistas (estatísticos e psicometristas) independentes, visando maior fiabilidade através da comparação e conferência dos dados das três equipes. As equipes avaliadoras ou entidades de acreditação são definidas via edital de licitação ou por escolha da entidade responsável pela administração do exame. Os procedimentos de análise seguem critérios de <i>validade estatística</i> na qual estão contidos no processo de <i>validade interna</i> . Dessa forma, pode-se dizer que perpassam todas as etapas da <i>validade de construto</i> , ou seja, a <i>validade de conteúdo</i> e de <i>critério</i> (concorrente e preditiva). Dentro da <i>validade estatística</i> é importante evitar o erro denominado de <i>p-hacking</i> ou falso-positivo, na qual indicadores estatísticos são superestimados equivocadamente ou negligenciados por não adotarem os critérios de rigor estatístico.
	12. Produção dos resultados.	Definem-se as escores usando com rigor estatístico com base no modelo logístico de três parâmetros (3LP) tanto no SAEB quanto no ENEM, apesar de que possam ser utilizados outros modelos, como o de Rasch. Para uma melhor eficiência nos resultados adota-se o método <i>Expected a Posteriori</i> (EAP) com o auxílio de resolução do <i>método da quadratura Gaussiana</i> , atendendo a critérios de maior fiabilidade, garantindo aos respondentes um resultado mais confiável e passível de discriminação entre os classificados. Os riscos de <i>validade</i> são equivalentes à fase anterior.
	13. Construção e Interpretação das Escalas de Proficiência.	Constroem-se as escalas submetendo os itens a critérios de consistência estatística visando identificar itens-âncora, na qual são aqueles que forma a escala de proficiência. Dentro de cada desvio-padrão preestabelecido são agrupados grupos de itens-âncora que correspondem a níveis de proficiência ou podendo ser interpretados pedagogicamente como patamares de competência. Após ser construída a escala de proficiência, ela poderá ser utilizada posteriormente em outros testes. Se assim for, pode-se dizer no teste posteriormente que a escala está definida <i>a priori</i> como no caso do SAEB, fato que não ocorre no caso do ENEM e do ENCCEJA onde a escala é construída <i>a posteriori</i> . A escala sendo definida <i>a priori</i> implica a não divulgação da prova para que não se conheça ou identifique os itens-âncora que se repetem historicamente. Tanto no ENEM quanto no ENCCEJA sempre são disponibilizadas as provas, porém suas escalas de proficiência são reconstruídas a cada edição do exame, e as séries históricas de desempenho são possíveis por utilizarem itens-âncora não disponíveis ao público, mantidas em sigilo pela entidade responsável na administração do exame. Os riscos de <i>validade</i> são equivalentes à fase anterior.
	14. Divulgação dos resultados.	Divulga-se por meio de plataformas <i>online</i> da entidade aplicadora e por outros meios de divulgação como jornais, televisão, rádio, etc. Geralmente há um pronunciamento de entidades governamentais apresentando um relatório sobre todo o processo do exame, apontando eventuais falhas ou limitações, assim como dos procedimentos adotados para que se garanta a isonomia do processo. A <i>validade</i> dos fatos expostos deve corresponder aos resultados obtidos na análise.

Fonte: Elaborado partindo de ideias iniciais de Candian (n.d.).

E1.2: Quanto ao tempo estimado para a aplicação

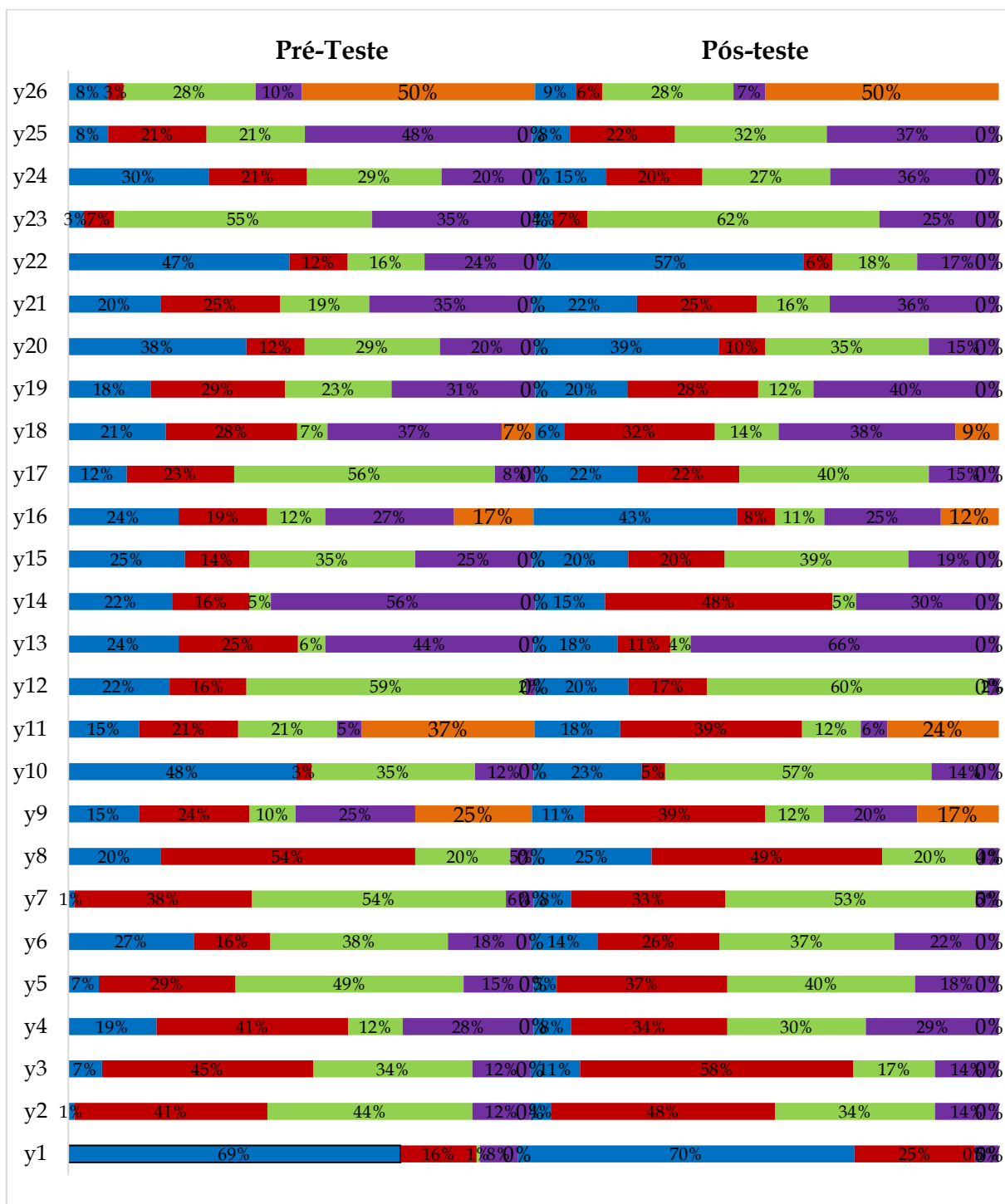
A sugestão é que após cerca de 80% dos alunos terminarem o teste, oferecer uma tolerância de ± 10 min para o restante, mas definir **50 min sendo o limite** para a realização do teste. Entretanto tempo estimado é bem inferior a 50 min (entorno de 30 min) pela natureza do teste, pois são questões intuitivas e os alunos não necessitam de um longo período assinalar a alínea correta, sendo estimado um tempo para responder (após a leitura) de apenas 1 min.

Quadro 89: Tempo estimado para a realização do teste TCE baseados em experiência prévia.

-
- O tempo médio para ler e preencher a parte de introdução do teste: 2 min;
 - O tempo médio de ler todas as questões, incluídos as respostas: 13 min (média de 30 s por questão);
 - A estimativa total para reflexão, reler e tomada de juízo são mais 30 s para cada questão (que no total se estima 1 min/questão). Dessa forma, se acrescentam um tempo extra de mais 13 min, onde a soma total ficaria com **28 min para total realização no teste** (± 30 min). Sendo o tempo disponível de uma aula entorno de 50 min, tirando os contratempos iniciais da aula, o tempo para aplicação do teste é considerado suficiente.
-

E1.3: Índice de Dificuldades dos Itens para o Pré-teste e o Pós-teste em Portugal.

Figura 96: Comparativo do percentual de respostas para o pré-teste e o pós-teste em Portugal.



E1.4: Fator de Concentração nas Análises dos Itens - Teste-Piloto

Tabela 55: Valores dos Fatores de Concentração nas Análises dos Itens da amostra portuguesa no teste-piloto, nomeadamente as percentagens de acerto e concentração dos grupos do 10º ano no pré-teste e pós-teste, conforme o *design* estabelecido.

ITEM	ID_pré	C_pré	ID/C	ID_pós	C_pós	ID/C
y1	0,74	0,53	AA	0,70	0,49	MM
y2	0,42	0,24	MM	0,48	0,21	MM
y3	0,46	0,18	MB	0,58	0,27	MM
y4	0,41	0,09	MB	0,34	0,08	BB
y5	0,29	0,19	BB	0,37	0,15	BB
y6	0,17	0,06	BB	0,26	0,05	BB
y7	0,55	0,34	MM	0,54	0,28	MM
y8	0,20	0,24	BM	0,26	0,19	BB
Y9*	0,24	0,04	BB	0,39	0,09	BB
y10	0,35	0,23	BM	0,57	0,27	MM
y11*	0,21	0,10	BB	0,39	0,12	BB
y12	0,60	0,32	MM	0,60	0,32	MM
y13	0,45	0,14	MB	0,66	0,39	MM
y14	0,16	0,26	BM	0,49	0,20	MB
y15	0,14	0,05	BB	0,21	0,05	BB
y16*	0,24	0,03	BB	0,44	0,16	MB
y17	0,13	0,25	BM	0,22	0,07	BB
y18*	0,37	0,13	BB	0,38	0,15	BB
y19	0,18	0,02	BB	0,20	0,08	BB
y20	0,29	0,07	BB	0,35	0,12	BB
y21	0,35	0,03	BB	0,36	0,04	BB
y22	0,47	0,13	MB	0,58	0,27	MM
y23	0,55	0,32	MM	0,63	0,37	MM
y24	0,20	0,02	BB	0,36	0,05	BB
y25	0,21	0,16	BB	0,33	0,09	BB
y26*	0,50	0,26	MM	0,50	0,26	MM

Nota: ID=índice de dificuldade; C=fator de concentração de análise.

E2: Índice de Dificuldade dos Itens nas amostras do Brasil e de Portugal.

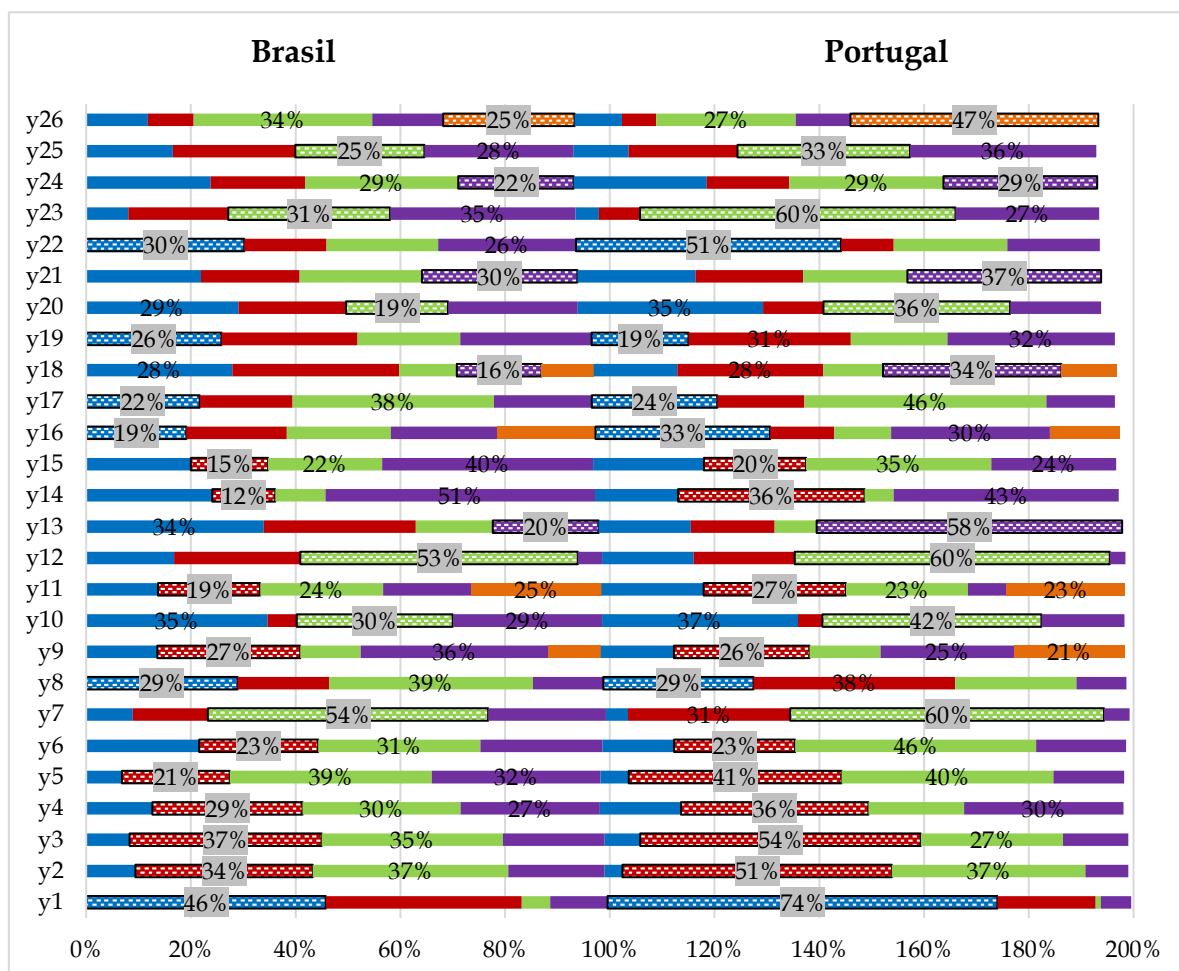


Figura 97: Comparativo do índice de dificuldade e do percentual dos distratores assinalado entre os respondentes de Brasil e Portugal.

*Itens y9, y11, y16, y18 e y26 possuem 5 alternativas (letra "e" na cor alaranjada, seguindo a ordem)

**O percentual de respostas em destaque (nas barras tracejadas) se refere a alternativa correta, ou seja, ao percentual de acertos. Do lado direito a leitura é equivalente, apenas muda o contexto.

E3: Parâmetros Clássicos da amostra no Brasil e em Portugal

Tabela 56: Parâmetros clássicos para cada item do conjunto dos alunos do 1º, 2º e 3º ano na amostra do contexto no Brasil/2015.

ITENS	PARÂMETROS CLÁSSICOS					PROPORÇÕES %					COEFICIENTES BISSERIAIS					MÉDIA DOS QUE ACERTARAM	DESVIO-PADRÃO	MÉDIA (de 0 a 10)
	ID	ACIM	ABAIX	DISCR	BISS*	A	B	C	D	E	A	B	C	D	E			
y1	0,46	0,28	0,62	0,35	0,34	0,46	0,38	0,05	0,11	-	0,34	-0,15	-0,23	-0,26	-	7,91	2,58	3,04
y2	0,34	0,17	0,52	0,34	0,39	0,09	0,34	0,38	0,19	-	-0,13	0,39	-0,16	-0,21	-	8,27	2,77	3,18
y3	0,37	0,20	0,56	0,36	0,39	0,08	0,37	0,35	0,20	-	-0,06	0,39	-0,17	-0,26	-	8,19	2,84	3,15
y4	0,29	0,17	0,39	0,22	0,27	0,13	0,29	0,31	0,27	-	-0,03	0,27	-0,15	-0,08	-	7,98	2,71	3,07
y5	0,21	0,12	0,33	0,21	0,32	0,07	0,21	0,39	0,33	-	-0,06	0,32	-0,03	-0,19	-	8,33	3,07	3,20
y6	0,23	0,16	0,31	0,15	0,22	0,22	0,23	0,31	0,24	-	-0,13	0,22	0,04	-0,11	-	7,89	2,92	3,04
y7	0,54	0,37	0,68	0,31	0,31	0,09	0,14	0,54	0,23	-	-0,13	-0,03	0,31	-0,30	-	7,74	2,65	2,98
y8	0,29	0,17	0,42	0,26	0,34	0,29	0,18	0,39	0,14	-	0,34	-0,04	-0,17	-0,15	-	8,19	2,92	3,15
y9	0,28	0,18	0,39	0,21	0,29	0,14	0,28	0,12	0,36	0,10	-0,10	0,29	-0,04	-0,09	-0,14	8,06	2,98	3,10
y10	0,30	0,13	0,50	0,36	0,43	0,35	0,06	0,30	0,29	-	-0,20	-0,08	0,43	-0,17	-	8,48	2,88	3,26
y11	0,20	0,11	0,29	0,18	0,26	0,14	0,20	0,24	0,17	0,25	-0,05	0,26	0,04	-0,15	-0,06	8,12	2,84	3,12
y12	0,54	0,43	0,60	0,16	0,18	0,17	0,24	0,54	0,05	-	-0,07	-0,09	0,18	-0,13	-	7,50	2,60	2,88
y13	0,21	0,09	0,37	0,28	0,43	0,35	0,30	0,15	0,21	-	-0,03	-0,16	-0,19	0,43	-	8,73	3,00	3,36
y14	0,12	0,06	0,20	0,14	0,34	0,25	0,12	0,10	0,53	-	-0,09	0,34	-0,09	0,00	-	8,64	3,17	3,32
y15	0,15	0,06	0,22	0,16	0,28	0,21	0,15	0,22	0,42	-	-0,02	0,28	0,08	-0,14	-	8,28	2,72	3,19
y16	0,20	0,09	0,33	0,24	0,39	0,20	0,20	0,20	0,21	0,19	0,39	-0,14	-0,09	0,03	-0,11	8,63	3,03	3,32
y17	0,22	0,13	0,31	0,17	0,23	0,22	0,18	0,40	0,19	-	0,23	-0,09	-0,02	-0,03	-	7,98	2,67	3,07
y18	0,17	0,08	0,26	0,18	0,34	0,29	0,33	0,11	0,17	0,10	-0,20	0,10	-0,09	0,34	-0,05	8,51	2,94	3,27
y19	0,27	0,15	0,36	0,22	0,29	0,27	0,27	0,20	0,26	-	0,29	-0,07	-0,05	-0,08	-	8,11	2,79	3,12
y20	0,21	0,10	0,28	0,18	0,28	0,31	0,22	0,21	0,26	-	-0,02	-0,08	0,28	-0,05	-	8,19	2,85	3,15
y21	0,32	0,19	0,41	0,22	0,23	0,23	0,20	0,25	0,32	-	0,03	-0,09	-0,06	0,23	-	7,85	2,61	3,02
y22	0,32	0,15	0,48	0,34	0,39	0,32	0,17	0,23	0,28	-	0,39	-0,13	-0,05	-0,14	-	8,32	2,78	3,20
y23	0,33	0,12	0,54	0,42	0,49	0,09	0,20	0,33	0,38	-	-0,10	-0,15	0,49	-0,19	-	8,62	2,83	3,32
y24	0,24	0,12	0,32	0,20	0,29	0,25	0,19	0,31	0,24	-	0,04	-0,12	-0,05	0,29	-	8,16	2,86	3,14
y25	0,27	0,14	0,40	0,26	0,36	0,18	0,25	0,27	0,31	-	-0,08	-0,10	0,36	-0,06	-	8,37	2,94	3,22
y26	0,27	0,13	0,40	0,27	0,38	0,13	0,09	0,37	0,15	0,27	-0,07	-0,01	-0,14	-0,06	0,38	8,44	2,93	3,24

Legenda: DPA – Desvio-Padrão Amostral; Em verde estão os valores para as alternativas corretas do teste.

*Coeficiente bisserial e ponto-bisserial *corrigidos* foram obtidos através do BILOG-MG e encontram-se no apêndice.

Tabela 57: Parâmetros clássicos para cada item do conjunto dos alunos do 10º, 11º e 12º ano na amostra do contexto em Portugal/2014.

ITENS	PARÂMETROS CLÁSSICOS					PROPORÇÕES %					COEFICIENTES BISSERIAIS					MÉDIA DOS QUE ACERTARAM	DESVIO-PADRÃO	MÉDIA (de 0 a 10)
	ID	ACIM	ABAIX	DISCR	BISS*	A	B	C	D	E	A	B	C	D	E			
y1	0,74	0,82	0,64	0,18	0,26	0,74	0,19	0,01	0,06	-	0,26	-0,18	-0,17	-0,28	-	8,72	3,83	3,35
y2	0,51	0,66	0,36	0,30	0,36	0,03	0,51	0,37	0,08	-	-0,28	0,36	-0,24	-0,21	-	11,38	3,90	4,38
y3	0,54	0,74	0,34	0,40	0,47	0,07	0,54	0,27	0,12	-	-0,17	0,47	-0,28	-0,33	-	11,64	3,93	4,48
y4	0,36	0,48	0,23	0,25	0,01	0,15	0,36	0,18	0,30	-	-0,06	0,01	-0,11	-0,13	-	11,51	4,01	4,43
y5	0,41	0,58	0,26	0,33	0,40	0,05	0,41	0,40	0,14	-	-0,24	0,40	-0,20	-0,24	-	11,77	4,01	4,53
y6	0,23	0,28	0,17	0,10	0,18	0,14	0,23	0,46	0,17	-	-0,27	0,18	0,13	-0,16	-	11,19	4,12	4,30
y7	0,60	0,66	0,51	0,15	0,20	0,04	0,31	0,60	0,05	-	-0,08	-0,10	0,20	-0,31	-	10,78	3,90	4,15
y8	0,29	0,44	0,16	0,28	0,40	0,29	0,39	0,23	0,10	-	0,40	-0,06	-0,27	-0,17	-	12,11	4,10	4,66
y9	0,26	0,40	0,14	0,26	0,40	0,14	0,26	0,14	0,25	0,21	-0,09	0,40	-0,01	-0,20	-0,15	12,21	4,18	4,70
y10	0,42	0,70	0,18	0,52	0,60	0,37	0,05	0,42	0,16	-	-0,44	-0,13	0,60	-0,22	-	12,45	3,86	4,79
y11	0,27	0,41	0,16	0,25	0,36	0,20	0,27	0,23	0,07	0,23	0,03	0,36	-0,05	-0,24	-0,25	11,99	4,02	4,61
y12	0,60	0,68	0,50	0,18	0,24	0,18	0,19	0,60	0,03	-	-0,26	-0,02	0,24	-0,25	-	10,88	3,86	4,19
y13	0,58	0,86	0,32	0,55	0,61	0,18	0,16	0,08	0,58	-	-0,29	-0,46	-0,35	0,61	-	11,88	3,76	4,57
y14	0,36	0,67	0,11	0,56	0,69	0,16	0,36	0,06	0,43	-	-0,34	0,69	-0,33	-0,34	-	13,11	3,78	5,04
y15	0,20	0,28	0,11	0,18	0,36	0,21	0,20	0,35	0,24	-	-0,18	0,36	0,05	-0,22	-	12,24	4,29	4,71
y16	0,33	0,59	0,14	0,45	0,57	0,33	0,12	0,11	0,30	0,13	0,57	-0,41	-0,24	-0,08	-0,21	12,68	3,94	4,88
y17	0,24	0,37	0,14	0,23	0,38	0,24	0,17	0,46	0,13	-	0,38	-0,12	-0,20	0,01	-	12,20	4,16	4,69
y18	0,34	0,55	0,16	0,39	0,48	0,16	0,28	0,11	0,34	0,11	-0,41	0,01	-0,25	0,48	-0,12	12,29	3,94	4,73
y19	0,19	0,27	0,12	0,15	0,32	0,19	0,31	0,18	0,32	-	0,32	-0,11	-0,01	-0,10	-	12,10	4,40	4,65
y20	0,36	0,53	0,21	0,32	0,41	0,35	0,12	0,36	0,17	-	-0,09	-0,27	0,41	-0,23	-	11,93	3,98	4,59
y21	0,37	0,45	0,26	0,19	0,22	0,23	0,21	0,20	0,37	-	-0,04	-0,22	0,04	0,22	-	11,18	3,77	4,30
y22	0,51	0,72	0,29	0,42	0,48	0,51	0,10	0,22	0,18	-	0,48	-0,29	-0,16	-0,32	-	11,77	3,85	4,53
y23	0,60	0,84	0,37	0,47	0,53	0,04	0,08	0,60	0,27	-	-0,27	-0,30	0,53	-0,38	-	11,63	3,86	4,47
y24	0,29	0,53	0,11	0,43	0,62	0,25	0,16	0,29	0,29	-	-0,21	-0,27	-0,19	0,62	-	13,14	4,07	5,05
y25	0,33	0,50	0,18	0,32	0,43	0,11	0,21	0,33	0,36	-	-0,18	-0,09	0,43	-0,22	-	12,15	4,16	4,67
y26	0,47	0,66	0,30	0,36	0,41	0,09	0,07	0,27	0,10	0,47	-0,12	-0,16	-0,27	-0,11	0,41	11,65	4,00	4,48

Legenda: DPA – Desvio-Padrão Amostral; Em verde estão os valores para as alternativas corretas do teste.

*Coeficiente bisserial e ponto-bisserial *corrigidos* foram obtidos através do BILOG-MG e encontram-se no apêndice F.

Tabela 59: Escala de desempenho (escore bruto) dos alunos portugueses de acordo com sua distribuição nas unidades escolares e nível de escolaridade (em 2014).

Escola	1			7			10			18			36			37		38		39			40			41		42			43			44			45			46			47			50			54			100			101			TOT	
Nív. / Escore	10	11	12	10	11	12	10	11	12	10	11	12	10	11	12	10	11	10	11	12	10	11	12	10	11	12	10	11	10	11	12	10	11	12	10	11	12	10	11	12	10	11	12	10	11	12	10	11	12	10	11	12	10	11	12	10	11	12	TOT
1		1									1																																									3							
2																																																				7							
3																																																				30							
4																																																				63							
5																																																				108							
6																																																				207							
7																																																				218							
8																																																				285							
9																																																				309							
10																																																				260							
11																																																				221							
12																																																				210							
13																																																				150							
14																																																					134						
15																																																				108							
16																																																				71							
17																																																				70							
18																																																				47							
19																																																				31							
20																																																				29							
21																																																				12							
22																																																				9							
23																																																				2							
24																																																					1						
TOT	54	51	11	54	54	21	23	29	27	43	45	18	64	62	16	8	25	32	4	60	69	18	119	84	9	23	39	143	34	86	57	17	19	14	28	79	40	50	23	58	58	58	57	2	41	45	8	331	122	69	68	16	2585						

10º ano (sombreado verde); 11º ano (sombreado alaranjado); 12º ano ; As escolas estão codificadas em números; O corte tracejado em azul corresponde ao grupo de desempenho inferior de 27 % que alcançou 9 valores no escore bruto, o grupo superior (27% da amostra com melhor desempenho) teve um corte no escore 12.

E5: Teste-t *student* comparando diferenças significativas entre as médias do 1º, 2º e 3º ano com relação ao turno matutino.

Tabela 60: Teste-t *student* comparando diferenças significativas entre as médias do 1º, 2º e 3º ano com relação ao turno matutino.

Teste-t: duas amostras presumindo variâncias diferentes												
Parâmetros	1º M	1º V	1º M	1º N	2º M	2º V	2º M	2º N	3º M	3º V	3º M	3º N
Média	7,21	6,95	7,21	5,94	7,40	7,29	7,40	6,97	6,93	7,24	6,93	7,11
Variância	7,00	6,43	7,00	5,76	8,47	7,52	8,47	6,41	6,86	6,04	6,86	6,47
Observações	373	367	373	95	748	328	748	146	590	190	590	120
Hipótese da dif. de média	0,00	-	0,00	-	0,00	-	0,00	-	0,00	-	0,00	-
gl	737	-	157	-	659	-	227	-	338	-	174	-
Stat t	1,36	-	4,52	-	0,57	-	1,83	-	-1,48	-	-0,70	-
P(T<=t) uni-caudal	0,09	-	0,00	-	0,28	-	0,03	-	0,07	-	0,24	-
t crítico uni-caudal	1,65	-	1,65	-	1,65	-	1,65	-	-1,65	-	-1,65	-
P(T<=t) bi-caudal	0,18	-	0,00	-	0,57	-	0,07	-	0,14	-	0,48	-
t crítico bi-caudal	1,96	-	1,98	-	1,96	-	1,97	-	1,97	-	1,97	-

E6: Teste-t *student* comparando diferenças significativas entre as médias dos grupos de escolas conforme os Livros Didáticos de Física (ou Manuais Escolares) em comum, tendo como referência o grupo com maior média.

Tabela 61: Teste-t *student* comparando diferenças significativas entre os grupos de respondentes conforme o Livro Didático de Física adotado na escola.

Teste-t: duas amostras presumindo variâncias diferentes									
Parâmetros	LDF_8	LDF_11	LDF_8	LDF_9	LDF_8	LDF_5	LDF_8	LDF_10	
Média	10,51	10,36	10,51	10,26	10,51	10,25	10,51	9,28	
Variância	16,19	13,93	16,19	15,65	16,19	12,96	16,19	19,70	
Observações	368	1463	368	430	368	147	368	177	
Hipótese da diferença de média	0,00	-	0,00	-	0,00	-	0,00	-	
gl	537	-	773	-	299	-	319	-	
Stat t	0,66	-	0,91	-	0,72	-	3,12	-	
P(T<=t) uni-caudal	0,25	-	0,18	-	0,24	-	0,00	-	
t crítico uni-caudal	1,65	-	1,65	-	1,65	-	1,65	-	
P(T<=t) bi-caudal	0,51	-	0,36	-	0,47	-	0,00	-	
t crítico bi-caudal	1,96	-	1,96	-	1,97	-	1,97	-	

E7: Teste-t *student* comparando diferenças significativas entre as médias dos diferentes níveis escolares nos dois contextos, assim como entre as médias gerais entre Brasil e Portugal.

Tabela 62: Teste-t *student* comparando diferenças significativas entre os grupos de respondentes conforme os níveis de escolaridade e na perspectiva geral entre o contexto Brasil e de Portugal.

Teste-t: duas amostras presumindo variâncias diferentes										
Parâmetros	1º ano	2º ano	1º ano	3º ano	10º ano	11º ano	10º ano	12º ano	Brasil	Portugal
Média	6,95	7,32	6,95	7,02	8,83	11,25	8,83	12,97	7,12	10,28
Variância	6,74	7,97	6,74	6,64	9,83	14,36	9,83	16,91	7,24	14,93
Observações	835	1222	835	900	1313	876	1313	396	2957	2585
Hipótese da dif. de média	0,00	-	0,00	-	0,00	-	0,00	-	0,00	-
gl	1886	-	1721	-	1630	-	541	-	4528	-
Stat t	-3,00	-	-0,52	-	-15,67	-	-18,48	-	-34,86	-
P(T<=t) uni-caudal	0,00	-	0,30	-	0,00	-	0,00	-	0,00	-
t crítico uni-caudal	-1,65	-	-1,65	-	-1,65	-	-1,65	-	-1,65	-
P(T<=t) bi-caudal	0,00	-	0,60	-	0,00	-	0,00	-	0,00	-
t crítico bi-caudal	1,96	-	1,96	-	1,96	-	1,96	-	1,96	-

E8: Fator de Concentração nas Análises dos Itens - Brasil

Tabela 63: Valores dos Fatores de Concentração nas Análises dos Itens da amostra brasileira, nomeadamente as porcentagens de acerto e concentração dos grupos do 1º, 2º e 3º ano, respectivamente.

ITEM	ID_1º	C_1º	ID/C	ID_2º	C_2º	ID/C	ID_3º	C_3º	ID/C	ID_Brasil	C_Brasil	ID/C
y1	0,48	0,21	MM	0,44	0,21	MM	0,47	0,23	MM	0,46	0,21	MM
y2	0,34	0,08	BB	0,36	0,10	BB	0,31	0,13	BB	0,34	0,10	BB
y3	0,37	0,10	BB	0,36	0,11	BB	0,38	0,11	BB	0,37	0,10	BB
y4	0,29	0,03	BB	0,30	0,05	BB	0,29	0,04	BB	0,29	0,04	BB
y5	0,21	0,10	BB	0,21	0,12	BB	0,21	0,12	BB	0,21	0,11	BB
y6	0,24	0,01	BB	0,24	0,01	BB	0,20	0,02	BB	0,23	0,01	BB
y7	0,51	0,20	MB	0,55	0,23	MM	0,55	0,23	MM	0,54	0,22	MM
y8	0,30	0,07	BB	0,31	0,09	BB	0,27	0,07	BB	0,29	0,08	BB
Y9*	0,28	0,07	BB	0,28	0,11	BB	0,28	0,12	BB	0,28	0,10	BB
y10	0,27	0,10	BB	0,32	0,11	BB	0,31	0,09	BB	0,30	0,10	BB
y11*	0,20	0,01	BB	0,20	0,02	BB	0,20	0,02	BB	0,20	0,02	BB
y12	0,55	0,25	MM	0,55	0,25	MM	0,50	0,20	MB	0,54	0,23	MM
y13	0,19	0,03	BB	0,22	0,05	BB	0,20	0,07	BB	0,21	0,05	BB
y14	0,10	0,18	BB	0,13	0,22	BM	0,13	0,23	BM	0,12	0,21	BM
y15	0,17	0,05	BB	0,14	0,10	BB	0,15	0,07	BB	0,15	0,08	BB
y16*	0,16	0,01	BB	0,22	0,00	BB	0,20	0,01	BB	0,20	0,0004	BB

y17	0,22	0,05	BB	0,24	0,07	BB	0,21	0,06	BB	0,22	0,06	BB
y18*	0,16	0,07	BB	0,16	0,10	BB	0,18	0,08	BB	0,17	0,08	BB
y19	0,24	0,00	BB	0,28	0,01	BB	0,27	0,01	BB	0,27	0,01	BB
y20	0,20	0,02	BB	0,21	0,01	BB	0,21	0,02	BB	0,21	0,01	BB
y21	0,28	0,00	BB	0,32	0,01	BB	0,35	0,03	BB	0,32	0,01	BB
y22	0,33	0,02	BB	0,35	0,03	BB	0,29	0,03	BB	0,32	0,03	BB
y23	0,32	0,08	BB	0,33	0,10	BB	0,34	0,11	BB	0,33	0,10	BB
y24	0,23	0,02	BB	0,26	0,01	BB	0,21	0,02	BB	0,24	0,01	BB
y25	0,26	0,01	BB	0,27	0,02	BB	0,26	0,02	BB	0,27	0,02	BB
y26*	0,28	0,08	BB	0,26	0,11	BB	0,27	0,11	BB	0,27	0,10	BB

E9: Fator de Concentração nas Análises dos Itens - Portugal

Tabela 64: Valores dos Fatores de Concentração nas Análises dos Itens da amostra portuguesa, nomeadamente as percentagens de acerto e concentração dos grupos do 10º, 11º e 12º ano, respectivamente.

ITEM	ID_10º	C_10º	ID/C	ID_11º	C_11º	ID/C	ID_12º	C_12º	ID/C	ID_Port.	C_Port.	ID/C
y1	0,73	0,51	AA	0,75	0,56	AA	0,78	0,61	AA	0,74	0,54	AA
y2	0,47	0,24	MM	0,56	0,32	MM	0,58	0,36	MM	0,51	0,28	MM
y3	0,47	0,18	MB	0,57	0,27	MM	0,68	0,42	MM	0,54	0,23	MM
y4	0,38	0,07	BB	0,31	0,03	BB	0,42	0,09	MB	0,36	0,05	BB
y5	0,37	0,17	BB	0,44	0,21	MM	0,46	0,20	MB	0,41	0,18	MB
y6	0,24	0,08	BB	0,22	0,18	BB	0,23	0,17	BB	0,23	0,12	BB
y7	0,62	0,36	MM	0,58	0,35	MM	0,60	0,38	MM	0,60	0,36	MM
y8	0,30	0,05	BB	0,24	0,15	BB	0,35	0,11	BB	0,29	0,08	BB
y9*	0,19	0,03	BB	0,32	0,05	BB	0,35	0,06	BB	0,26	0,03	BB
y10	0,28	0,20	BB	0,53	0,24	MM	0,64	0,35	MM	0,42	0,17	MB
y11*	0,24	0,05	BB	0,30	0,06	BB	0,32	0,10	BB	0,27	0,05	BB
y12	0,60	0,31	MM	0,59	0,30	MM	0,63	0,37	MM	0,60	0,31	MM
y13	0,46	0,12	MB	0,68	0,42	MM	0,78	0,59	AA	0,58	0,27	MM
y14	0,14	0,24	BM	0,58	0,32	MM	0,59	0,33	MM	0,36	0,17	BB
y15	0,16	0,04	BB	0,19	0,03	BB	0,31	0,07	BB	0,20	0,03	BB
y16*	0,20	0,03	BB	0,42	0,19	MB	0,56	0,34	MM	0,33	0,09	BB
y17	0,19	0,17	BB	0,26	0,10	BB	0,34	0,10	BB	0,24	0,13	BB
y18*	0,27	0,06	BB	0,37	0,11	BB	0,51	0,25	MM	0,34	0,08	BB
y19	0,19	0,03	BB	0,17	0,03	BB	0,21	0,04	BB	0,19	0,03	BB
y20	0,29	0,06	BB	0,40	0,12	BB	0,48	0,20	MB	0,36	0,09	BB
y21	0,37	0,05	BB	0,35	0,04	BB	0,39	0,06	BB	0,37	0,04	BB
y22	0,43	0,09	MB	0,56	0,25	MM	0,64	0,36	MM	0,51	0,17	MB
y23	0,51	0,24	MM	0,67	0,43	MM	0,77	0,57	AA	0,60	0,34	MM
y24	0,16	0,04	BB	0,41	0,09	MB	0,50	0,17	MB	0,29	0,02	BB
y25	0,29	0,07	BB	0,35	0,09	BB	0,43	0,12	MB	0,33	0,08	BB
y26*	0,42	0,18	MB	0,51	0,24	MM	0,55	0,29	MM	0,47	0,21	MM

E10: Fator de Concentração nas Análises dos Itens (Visualização Geral) - Teste-Piloto, Portugal e Brasil

Tabela 65: Classificação do Índice de Dificuldades e do Fator de Concentração nas Análises de Portugal e Brasil, com relação aos níveis de escolaridade e geral em cada contexto.

ITEM	Teste-Piloto		Portugal				Brasil			
	Pré	Pós	10º ano	11º ano	12º ano	Geral	1º ano	2º ano	3º ano	Geral
y1	AA	MM	AA	AA	AA	AA	MM	MM	MM	MM
y2	MM	MM	MM	MM	MM	MM	BB	BB	BB	BB
y3	MB	MM	MB	MM	MM	MM	BB	BB	BB	BB
y4	MB	BB	BB	BB	MB	BB	BB	BB	BB	BB
y5	BB	BB	BB	MM	MB	MB	BB	BB	BB	BB
y6	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
y7	MM	MM	MM	MM	MM	MM	MB	MM	MM	MM
y8	BM	BB	BB	BB	BB	BB	BB	BB	BB	BB
y9*	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
y10	BM	MM	BB	MM	MM	MB	BB	BB	BB	BB
y11*	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
y12	MM	MM	MM	MM	MM	MM	MM	MM	MB	MM
y13	MB	MM	MB	MM	AA	MM	BB	BB	BB	BB
y14	BM	MB	BM	MM	MM	BB	BB	BM	BM	BM
y15	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
y16*	BB	MB	BB	MB	MM	BB	BB	BB	BB	BB
y17	BM	BB	BB	BB	BB	BB	BB	BB	BB	BB
y18*	BB	BB	BB	BB	MM	BB	BB	BB	BB	BB
y19	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
y20	BB	BB	BB	BB	MB	BB	BB	BB	BB	BB
y21	BB	BB	BB	BB	BB	BB	BB	BB	BB	BB
y22	MB	MM	MB	MM	MM	MB	BB	BB	BB	BB
y23	MM	MM	MM	MM	AA	MM	BB	BB	BB	BB
y24	BB	BB	BB	MB	MB	BB	BB	BB	BB	BB
y25	BB	BB	BB	BB	MB	BB	BB	BB	BB	BB
y26*	MM	MM	MB	MM	MM	MM	BB	BB	BB	BB

E11: Análise dos Distratores e as Respectivas Concepções Térmicas em Destaque dentro dos Grupos Superiores (ACIM)

Tabela 66: Distratores mais atrativos entre os contextos, com base na AGI dentro do grupo ACIM.

Itens (k)	Análise Gráfica do Item: Distratores mais Atrativos nos Contextos (BR e PT) para os Grupos Superiores (ACIM) dentro de Níveis Específicos de Escore Bruto						Concepções Alternativas Associadas (BR/PT)
	BR	%: >0,25 ⁽⁴⁾ e >0,20 ⁽⁵⁾	(picos sig. em escore ≥9)	PT	%: >0,25 ⁽⁴⁾ e >0,20 ⁽⁵⁾	(picos sig. em escore ≥12)	
y1 ⁽⁴⁾	b	0,37	17, 23	b*	0,19*	19*, 21*	D6/ D6*
y2 ⁽⁴⁾	c; a*	0,37; 0,09*	15; 18*	c; d*	0,37; 0,08*	19, 22; 21*	[D7], [C6]; C6*/ [D7], [C6]; D7*, C6*
y3 ⁽⁴⁾	c; a*	0,35; 0,08*	16; 18*	c; a*	0,27; 0,07*	21; 16*	[C1]; C6*, C7*/ [C1]; C6*, C7*
y4 ⁽⁴⁾	c; d	0,30; 0,27	18, 20; 17	c*; d	0,18*; 0,30	23*; 22	[C1]; [D5]/ [C1]; [D5]
y5 ⁽⁴⁾	c; d	0,39; 0,32	18, 20; 18	c; d*	0,40; 0,13*	21, 23; 18*	[B4], [C1]; [B4], [C1]/ [B4], [C1]; B4*, C1*
y6 ⁽⁴⁾	c; d*	0,31; 0,23*	18, 20; 18*	c; a*; d*	0,45; 0,13*; 0,17*	22, 23; 21*; 21*	[C6], [D8]; C6*, D8*/ [C6], [D8]; sug.*; C6*, D8*
y7 ⁽⁴⁾	b*; d*	0,14*; 0,22*	17*, 22*; 16*	b	0,31	16, 21	A5*; A6*, C5*/ A5
y8 ⁽⁴⁾	c; b*; d*	0,39; 0,17*; 0,13*	17; 17*; 16*	b; c*	0,38; 0,23*	21; 19*	[D5]; D5*; sug.* / [D5]; D5*
y9 ⁽⁵⁾	d; e*	0,36; 0,10*	17; 16*	d; e	0,25; 0,21	22; 21	[C6], [D1]; B7*/ [C6], [D1]; B7
y10 ⁽⁴⁾	a; b; d*	0,35; 0,06*; 0,29	16, 17; 17*; 15	a; d*	0,37; 0,16*	18; 18*, 20*	[A3], [A1], [C4]; (!?); A1*, B3*, C4*/ [A3], [A1], [C4]; A1*, B3*, C4*
y11 ⁽⁵⁾	c; e; a*	0,24; 0,25; 0,14*	16, 18; 15, 17; 16*, 18*	c; e; a*	0,23; 0,22; 0,19*	17, 19, 21; 15, 17, 22; 18*, 20*, 22*	[A5], [D4]; [D7]; sug.* / [A5], [D4]; [D7]; sug.*
y12 ⁽⁴⁾	b*; d*	0,24*; 0,05*	15*, 18*, 20*; 17*	b*; a*	0,19*; 0,17*	17*, 23*; 19*, 21*	D12*, D12* / D12*, D12*
y13 ⁽⁴⁾	a; b	0,34; 0,29	13, 14; 15	a*	0,17*	20*	C5; A3, C4/ C5*
y14 ⁽⁴⁾	d	0,51	17, 18	d	0,43	13	[D2]/ [D2]
y15 ⁽⁴⁾	d; a*; c*	0,40; 0,20*; 0,22*	17, 20; 15*, 18*, 17*, 20*	c; d*	0,35; 0,24	18, 22; 17*, 20*	[A4], [A5], [B1]; A4*, A5*, B1*; A3*, A4*, A5*, B1*/ [A3], [A4], [A5], [B1]; A4*, A5*, B1*
y16 ⁽⁵⁾	c; d	0,20; 0,20	16; 16, 18	d	0,30	17, 19	[A1], [D1], [D2]; [D2]/ [D2]
y17 ⁽⁴⁾	c; d*	0,38; 0,19*	15, 18; 16*	d*	0,46; 0,13*	16, 21; 19*; 22*	[C6], [D6]; [C6], [D6]/ [C6], [D6]
y18 ⁽⁵⁾	a; b	0,28; 0,32	16; 16, 18	b	0,28	16, 19	[A3], [B3]; [A3], [B3]/ [A3], [B3]
y19 ⁽⁴⁾	b; d	0,26; 0,25	14, 17; 13, 16	b; d	0,31; 0,32	14, 18; 13, 16, 22	[B5], [D5]; [B5], [D5]/ [B5], [D5]; [B5], [D5]
y20 ⁽⁴⁾	a; b*; d	0,29; 0,21*; 0,25	15, 17; 16*, 18*, 13, 16	a; d*	0,35; 0,17*	16, 19, 22; 17*, 21*	[C2], [C3]; C9, D2; sug./ [C2], [C3]; sug.
y21 ⁽⁴⁾	a*; c*	0,22*; 0,23*	14*, 16*	a*; c*	0,22*; 0,19*	19*, 21*	A1*, B3*; A1*, C9*/ A1*, B3*, A1*, C9*
y22 ⁽⁴⁾	c*; d	0,21*; 0,26	17*; 18	c*	0,21*	14*, 17*, 18*, 22*	A1*, A2*, D2*; B3, B6, D2/ A1*, A2*, D2*
y23 ⁽⁴⁾	d	0,35	12, 13	d	0,27	15, 18, 20	[A3], [D9]/ [A3], [D9]
y24 ⁽⁴⁾	a*; c	0,24*; 0,29	14*, 18*; 13, 15	a; c	0,25; 0,29	não há; 19	C6*, D1*; [A3], [D1]/ C6, [D1]; [A3], [D1]
y25 ⁽⁴⁾	b*; d	0,23*; 0,28	14*, 16*; 13, 16	b*; d	0,20*; 0,35	13*; 17, 22	B8*; [sug.] / B8*, [sug.]
y26 ⁽⁵⁾	c	0,34	13, 16, 17	c	0,26	14, 16, 17, 21	[B6]/ [B6]

Nota: Asterisco e itálico representa os grupos na zona de "guessing" do item; ";" separa os alternativas, as porcentagens, os escores de pico para cada contexto e as concepções para cada alternativa; em colchetes correspondências de concepções térmicas entre os dois contextos; "/" separa os dados dos dois contextos; k é o número de alternativas que o item possui.

E12: Matriz de Correlações Tetracóricas dos Itens - Amostra Geral (Brasil e Portugal).

Tabela 67: Matriz de correlações tetracóricas dos itens - Amostra Geral (Brasil e Portugal).

	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14	y15	y16	y17	y18	y19	y20	y21	y22	y23	y24	y25	y26	
y1	-																										
y2	0,37	-																									
y3	0,26	0,23	-																								
y4	0,07	0,07	0,04	-																							
y5	0,08	0,12	0,21	0,16	-																						
y6	-0,04	-0,02	0,02	0,01	0,02	-																					
y7	0,05	0,05	0,03	0,05	0,04	0,00	-																				
y8	0,04	0,04	0,06	0,10	0,11	0,00	0,02	-																			
y9	0,00	0,06	0,07	0,07	0,04	0,03	0,03	0,02	-																		
y10	0,10	0,13	0,21	0,02	0,17	0,02	0,04	0,14	0,08	-																	
y11	0,14	0,08	0,05	0,07	0,10	0,04	0,07	0,07	0,04	0,10	-																
y12	0,03	0,06	0,01	0,05	0,00	0,02	0,04	0,07	-0,04	0,00	-0,05	-															
y13	0,22	0,20	0,24	0,09	0,28	-0,04	0,04	0,12	0,08	0,40	0,17	0,05	-														
y14	0,16	0,21	0,26	0,05	0,21	-0,01	0,06	0,07	0,20	0,37	0,12	0,04	0,43	-													
y15	0,08	0,06	0,12	0,06	0,15	0,01	-0,02	0,08	0,03	0,14	0,01	-0,01	0,12	0,17	-												
y16	0,10	0,14	0,16	0,05	0,15	-0,02	-0,02	0,10	0,11	0,29	0,14	-0,04	0,33	0,45	0,15	-											
y17	0,03	0,01	0,05	-0,03	0,01	0,00	-0,02	0,10	0,06	0,12	0,05	0,06	0,09	0,11	0,01	0,10	-										
y18	0,09	0,09	0,16	0,05	0,15	0,02	0,01	0,06	0,00	0,23	0,13	0,01	0,31	0,29	0,12	0,38	0,08	-									
y19	-0,01	0,01	0,01	0,03	-0,02	0,03	-0,03	0,16	0,05	0,02	0,05	0,03	0,00	-0,03	0,05	0,09	-0,02	0,00	-								
y20	0,12	0,09	0,09	0,03	0,16	0,01	0,06	0,02	0,04	0,17	0,06	0,01	0,25	0,25	0,00	0,14	0,03	0,16	0,03	-							
y21	-0,03	0,02	0,01	0,00	0,03	-0,03	0,02	-0,03	-0,03	0,02	0,00	0,04	0,08	0,00	-0,07	-0,02	0,03	0,05	-0,05	0,04	-						
y22	0,11	0,08	0,13	0,09	0,16	0,01	0,08	0,09	0,07	0,16	0,07	0,02	0,32	0,28	0,08	0,18	0,05	0,17	0,05	0,12	0,07	-					
y23	0,15	0,14	0,24	0,06	0,22	-0,01	0,09	0,05	0,08	0,34	0,13	-0,01	0,41	0,38	0,10	0,24	0,09	0,22	-0,02	0,19	0,05	0,30	-				
y24	0,02	0,08	0,13	0,01	0,10	0,04	-0,04	0,03	0,24	0,17	0,04	0,01	0,22	0,47	0,10	0,26	0,11	0,20	0,02	0,14	0,03	0,10	0,20	-			
y25	0,01	0,07	0,13	0,04	0,15	0,02	-0,01	0,12	0,01	0,18	0,05	0,01	0,18	0,16	0,13	0,13	0,06	0,12	0,04	0,12	0,05	0,11	0,18	0,17	-		
y26	0,13	0,12	0,15	0,05	0,09	-0,01	0,07	0,03	0,08	0,18	0,04	0,00	0,26	0,27	0,10	0,18	0,13	0,20	0,01	0,06	0,08	0,17	0,22	0,18	0,10	-	

E13: Identificação dos Itens-âncoras na Escala de Proficiências Térmicas para o Modelo 4PL (20 itens) – Brasil e Portugal

Tabela 68: Identificação dos itens-âncoras e quase-âncoras, de acordo com os critérios preestabelecidos relacionados a consistência dos itens.

Itens	Parâmetros				Habilidades θ e probabilidades de acertos empíricos correspondentes											Itens- âncoras e quase-âncoras
	a	b	g	u	-5,00	-4,00	-3,00	-2,00	-1,00	0,00	1,00	2,00	3,00	4,00	5,00	
y1	6,61	-0,84	0,00	0,75	0,00	0,00	0,00	0,00	0,19	0,75	0,75	0,75	0,75	0,75	0,75	y1
y2	3,10	-0,68	0,00	0,58	0,00	0,00	0,00	0,01	0,16	0,52	0,58	0,58	0,58	0,58	0,58	y2
y3	0,87	0,19	0,00	0,96	0,01	0,02	0,06	0,12	0,25	0,44	0,64	0,80	0,88	0,93	0,95	y3
y13	8,69	0,56	0,18	0,87	0,18	0,18	0,18	0,18	0,18	0,19	0,86	0,87	0,87	0,87	0,87	y13
y5	3,45	0,61	0,20	0,55	0,20	0,20	0,20	0,20	0,20	0,24	0,48	0,55	0,55	0,55	0,55	y5
y23	4,30	0,62	0,29	0,90	0,29	0,29	0,29	0,29	0,29	0,33	0,80	0,90	0,90	0,90	0,90	y23
y10	6,42	0,75	0,23	0,77	0,23	0,23	0,23	0,23	0,23	0,23	0,68	0,77	0,77	0,77	0,77	y10
y22	4,74	0,76	0,30	0,75	0,30	0,30	0,30	0,30	0,30	0,31	0,64	0,75	0,75	0,75	0,75	y22
y18*	7,04	0,91	0,17	0,59	0,17	0,17	0,17	0,17	0,17	0,17	0,44	0,59	0,59	0,59	0,59	y18
y20	3,81	0,96	0,21	0,57	0,21	0,21	0,21	0,21	0,21	0,22	0,40	0,56	0,57	0,57	0,57	y20
y16	8,94	1,05	0,18	0,74	0,18	0,18	0,18	0,18	0,18	0,18	0,40	0,74	0,74	0,74	0,74	y16
y14	5,48	1,20	0,12	1,00	0,12	0,12	0,12	0,12	0,12	0,12	0,34	0,99	1,00	1,00	1,00	y14
y26*	1,67	1,43	0,24	1,00	0,24	0,24	0,24	0,24	0,25	0,30	0,49	0,79	0,95	0,99	1,00	y26
y24	14,01	1,45	0,20	1,00	0,20	0,20	0,20	0,20	0,20	0,20	0,20	1,00	1,00	1,00	1,00	y24
y17	4,70	1,55	0,21	0,51	0,21	0,21	0,21	0,21	0,21	0,21	0,23	0,48	0,51	0,51	0,51	y17
y9*	9,52	1,65	0,24	0,73	0,24	0,24	0,24	0,24	0,24	0,24	0,24	0,71	0,73	0,73	0,73	y9
y25	1,72	1,98	0,23	1,00	0,23	0,23	0,23	0,23	0,23	0,25	0,35	0,62	0,89	0,98	1,00	y25
y8	1,93	2,37	0,26	1,00	0,26	0,26	0,26	0,26	0,26	0,27	0,31	0,50	0,83	0,97	1,00	y8
y19	7,48	2,52	0,22	1,00	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,24	0,98	1,00	1,00	y19
y15	1,04	3,00	0,12	0,98	0,12	0,12	0,12	0,12	0,13	0,16	0,22	0,34	0,55	0,76	0,88	y15

Nota: Os itens encontram-se em ordem crescente de dificuldade, ou seja, conforme os valores estimados para o parâmetro b; em verde e negrito são os itens-âncoras (obedecem aos 3 critérios), o restante em destaque são quase-âncoras (obedecem a 2 critérios).

F - Sintaxes

F1: Alguns Resultados da Análise Clássica no BILOG-MG - Brasil/2015

TESE_BR.BLM - TRADITIONAL IRT ANALYSIS OF A FIFTEEN-ITEM PRETEST FROM
 A TWO-STAGE TEST OF MATHEMATICS AT THE EIGHTH-GRADE LEVEL
 >COMMENTS

This example illustrates how the BILOG-MG program can be used for traditional
 >GLOBAL DFNAME='TESE_2957_BR.TXT', NPARAM=3, LOGISTIC, SAVE;
 >SAVE PARM='TESE_2585_PT.PAR', SCORE='TESE_2957_BR.SCO';
 >LENGTH NITEMS=26;
 >INPUT NTOTAL=26, NALT=5, NIDCHAR=4,
 KFNAME='TESE_2957_BR.TXT';
 >ITEMS INAMES=(I01(1)I26);
 >TEST1 TNAME='TESE_BR', INUMBER=(1(1)26);
 (4A1,26A1)
 >CALIB NQPT=31, CYCLES=100, NEWTON=10, CRIT=0.001, ACCEL=0.0, CHI=15,
 PLOT=1;
 >SCORE NOPRINT, RSCTYPE=4, INFO=2, POP;

ITEM	NAME	#TRIED	#RIGHT	PCT	ITEM*TEST CORRELATION		
					LOGIT	PEARSON	BISERIAL
1	I01	2957.0	1353.0	45.8	0.17	0.085	0.107
2	I02	2957.0	1002.0	33.9	0.67	0.134	0.174
3	I03	2957.0	1085.0	36.7	0.55	0.127	0.162
4	I04	2957.0	848.0	28.7	0.91	0.033	0.044
5	I05	2957.0	609.0	20.6	1.35	0.079	0.113
6	I06	2957.0	672.0	22.7	1.22	-0.001	-0.001
7	I07	2957.0	1583.0	53.5	-0.14	0.062	0.078
8	I08	2957.0	855.0	28.9	0.90	0.088	0.116
9	I09	2957.0	807.0	27.3	0.98	0.049	0.065
10	I10	2957.0	881.0	29.8	0.86	0.165	0.218
11	I11	2957.0	576.0	19.5	1.42	0.035	0.051
12	I12	2957.0	1567.0	53.0	-0.12	-0.039	-0.049
13	I13	2957.0	598.0	20.2	1.37	0.156	0.223
14	I14	2957.0	356.0	12.0	1.99	0.089	0.144
15	I15	2957.0	436.0	14.7	1.75	0.049	0.075
16	I16	2957.0	566.0	19.1	1.44	0.131	0.189
17	I17	2957.0	641.0	21.7	1.28	0.014	0.020
18	I18	2957.0	479.0	16.2	1.64	0.092	0.138
19	I19	2957.0	764.0	25.8	1.05	0.055	0.075
20	I20	2957.0	576.0	19.5	1.42	0.049	0.070
21	I21	2957.0	878.0	29.7	0.86	0.005	0.007
22	I22	2957.0	893.0	30.2	0.84	0.126	0.166
23	I23	2957.0	914.0	30.9	0.80	0.211	0.277
24	I24	2957.0	652.0	22.0	1.26	0.051	0.072
25	I25	2957.0	729.0	24.7	1.12	0.108	0.147
26	I26	2957.0	742.0	25.1	1.09	0.126	0.171

F2: Alguns Resultados da Análise Clássica no BILOG-MG - Portugal/2014

TESE_PT.BLM - TRADITIONAL IRT ANALYSIS OF A FIFTEEN-ITEM PRETEST FROM
 A TWO-STAGE TEST OF MATHEMATICS AT THE EIGHTH-GRADE LEVEL
 >COMMENTS



This example illustrates how the BILOG-MG program can be used for traditional
 >GLOBAL DFNAME='TESE_2585_PT.TXT', NPARAM=3, LOGISTIC, SAVE;
 >SAVE PARM='TESE_2585_PT.PAR', SCORE='TESE_2585_PT.SCO';
 >LENGTH NITEMS=26;
 >INPUT NTOTAL=26, NALT=5, NIDCHAR=4,
 KFNAME='TESE_2585_PT.TXT';
 >ITEMS INAMES=(I01(1)I26);
 >TEST1 TNAME='TESE_PT', INUMBER=(1(1)26);
 (4A1,26A1)
 >CALIB NQPT=31, CYCLES=100, NEWTON=10, CRIT=0.001, ACCEL=0.0, CHI=15,
 PLOT=1;
 >SCORE NOPRINT, RSCTYPE=4, INFO=2, POP;



ITEM	NAME	ITEM*TEST CORRELATION					
		#TRIED	#RIGHT	PCT	LOGIT	PEARSON	BISERIAL
1	I01	2585.0	1919.0	74.2	-1.06	0.085	0.115
2	I02	2585.0	1321.0	51.1	-0.04	0.171	0.214
3	I03	2585.0	1373.0	53.1	-0.12	0.257	0.322
4	I04	2585.0	917.0	35.5	0.60	0.118	0.152
5	I05	2585.0	1045.0	40.4	0.39	0.194	0.246
6	I06	2585.0	592.0	22.9	1.21	0.016	0.023
7	I07	2585.0	1539.0	59.5	-0.39	0.037	0.047
8	I08	2585.0	725.0	28.0	0.94	0.194	0.259
9	I09	2585.0	666.0	25.8	1.06	0.185	0.251
10	I10	2585.0	1072.0	41.5	0.34	0.363	0.458
11	I11	2585.0	692.0	26.8	1.01	0.160	0.216
12	I12	2585.0	1531.0	59.2	-0.37	0.072	0.091
13	I13	2585.0	1498.0	57.9	-0.32	0.381	0.481
14	I14	2585.0	910.0	35.2	0.61	0.436	0.562
15	I15	2585.0	501.0	19.4	1.43	0.148	0.214
16	I16	2585.0	854.0	33.0	0.71	0.329	0.427
17	I17	2585.0	617.0	23.9	1.16	0.169	0.232
18	I18	2585.0	870.0	33.7	0.68	0.261	0.338
19	I19	2585.0	473.0	18.3	1.50	0.130	0.190
20	I20	2585.0	909.0	35.2	0.61	0.196	0.252
21	I21	2585.0	925.0	35.8	0.58	0.054	0.069
22	I22	2585.0	1283.0	49.6	0.01	0.268	0.336
23	I23	2585.0	1516.0	58.6	-0.35	0.311	0.394
24	I24	2585.0	744.0	28.8	0.91	0.369	0.490
25	I25	2585.0	840.0	32.5	0.73	0.208	0.271
26	I26	2585.0	1186.0	45.9	0.17	0.217	0.273

G - Documentos Formais

G1: Consentimento da Autora do Teste TCE via *e-mail*.

RE: ITCE

 Shelley Yeo <S.Yeo@curtin.edu.au>
dom 16/04, 02:21
Você 

 Responder | 

Sinalizar para acompanhamento. Início por volta de domingo, 16 de abril de 2017. Conclusão por volta de domingo, 16 de abril de 2017.

Você encaminhou esta mensagem em 16/04/2017 02:59

By all means you may use the TCE. However, it has only been tested on 15-19 year-old students. If used for younger students, some questions may be inappropriate. Please ensure that no-one is able to keep a copy of the test, and do not release the answers to anyone.
Regards
Shelley Yeo

From: Marcel Bruno Pereira Braga Braga [mailto:marcelbruno79@hotmail.com]
Sent: Friday, 14 April 2017 2:56 AM
To: Shelley Yeo <S.Yeo@curtin.edu.au>
Subject: Re: ITCE

Dear Shelley,

I would like to request your consent for use of the TCE test for academic purposes.

Appreciate your attention.

Greetings.

G2: Carta aos Diretores (Modelo Brasil)



Serviço Público Federal
UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA EDUCAÇÃO

Florianópolis, 14 de maio de 2014

À Diretoria

EEB

Senhor(a) Diretor(a),

Pela presente apresentamos o Prof. MARCEL BRUNO PEREIRA BRAGA, docente da SEDUC - AM e doutorando da UFSC e da UC – Universidade de Coimbra-Portugal, que investiga aspectos de ensino-aprendizagem de Física no ensino médio - EM. Dentre as etapas de seus estudos, consta no projeto a realização de testes e questionários com alunos do EM das 2ª e 3ª séries, associado a entrevistas com professores.

Neste sentido, solicitamos a devida autorização para Marcel entrar em contato com os professores de Física desta EEB, para a exposição e discussão de seu projeto e possibilidade de agendar trabalho com os estudantes em sessões e horários a negociar, preferencialmente no contraturno.

Solicitamos também que sejam colocados à disposição do doutorando os recursos e meios atualmente disponíveis nesta escola, para a concretização desta etapa da pesquisa.

Contando com a sua compreensão e apoio, estamos à disposição para demais esclarecimentos.

Atenciosamente,

A handwritten signature in black ink, appearing to read 'J. Peres'.

José André Peres Angotti
Orientador

Programa de PG em Educação Científica e Tecnológica
Depto de Metodologia de Ensino - CED

G3: Carta de Apresentação aos Professores (Modelo Portugal)



DEPARTAMENTO
DE FÍSICA
UNIVERSIDADE DE COIMBRA
FACULDADE DE CIÊNCIAS
E TECNOLOGIA

CARTA DE APRESENTAÇÃO E ESCLARECIMENTO DE INVESTIGAÇÃO DE DOUTORAMENTO

Coimbra, 09 de Janeiro de 2014.

Caro Senhor(a) Professor(a) da componente disciplinar de Física:

Marcel Bruno Pereira Braga, Mestre em Ensino de Ciências pela Universidade do Estado do Amazonas (UEA), com pós-graduação *Lato Sensu* na área de Gestão Escolar pela Universidade do Estado do Amazonas (UEA), actualmente estudante de doutoramento em Ensino das Ciências da Faculdade de Ciência e Tecnologia da Universidade de Coimbra (FCTUC), está a desenvolver uma investigação no âmbito da sua tese, subordinada ao tema “*Uma Análise do Conteúdo do Livro Didáctico de Física e suas Implicações no Ensino-Aprendizagem da Termodinâmica: Um Estudo Comparativo Portugal-Brasil*” para a qual muito gostaria de poder contar com a sua valiosa colaboração.

O objectivo geral do estudo é analisar as principais características e identificar as contribuições dos manuais escolares para a aprendizagem da Física. Entre os elementos de análise estão as opiniões de docentes de Física do ensino secundário de Portugal e do Brasil quanto forma do uso do manual no ensino e as suas percepções sobre manuais escolares. Também serão objecto de estudo as percepções dos alunos. O uso das informações será exclusivamente para fins científicos.

A sua colaboração será um elemento indispensável para a pesquisa, e caso concorde em colaborar, será assinado um termo de consentimento livre e esclarecido, para o que serão prestados todos os esclarecidos considerados indispensáveis sobre os procedimentos relativos à metodologia de recolha de dados para a investigação.

O plano da investigação foi submetido à Comissão Nacional de Protecção de Dados (CNPD) N° de Processo 784/2013, e obtida Autorização N° 4277/2013 garantindo-se o cumprimento de todas as normas de utilização de documentos e de confidencialidade estabelecidas por este organismo. Todos os procedimentos formais necessários para viabilizar esta investigação serão realizados, partindo da autorização da direcção da escola, assim como dos consentimentos dos professores e encarregados de educação. Estará assegurado o total anonimato e confidencialidade dos dados envolvidos.

Antecipadamente grato sua atenção e colaboração.

Com a mais elevada estima e os melhores cumprimentos,

O Doutorando

Marcel Bruno P. Braga.

Rua Larga, P-3004 516 Coimbra, Portugal
Tel: +351 239 410 600, Fax: +351 239 829 158

G4: Carta de Apresentação aos Professores (Modelo Brasil)



CARTA DE APRESENTAÇÃO E ESCLARECIMENTO DE INVESTIGAÇÃO DE DOUTORAMENTO

Florianópolis, 18 de junho de 2014.

Caro Senhor(a) Professor(a) da componente disciplinar de Física:

Eu, Marcel Bruno Pereira Braga, Mestre em Ensino de Ciências pela Universidade do Estado do Amazonas (UEA), com pós-graduação *Lato Sensu* na área de Gestão Escolar pela Universidade do Estado do Amazonas (UEA), atualmente estudante de doutoramento em Ensino das Ciências da Faculdade de Ciência e Tecnologia da Universidade de Coimbra¹⁰⁸ (FCTUC), em regime de cotutela com o doutoramento em Educação Científica e Tecnológica da Universidade Federal de Santa Catarina¹⁰⁹ (UFSC), estou a desenvolver uma investigação no âmbito de minha tese, subordinada ao tema “*Uma Análise do Conteúdo do Livro Didático de Física e suas Implicações no Ensino-Aprendizagem da Termodinâmica: Um Estudo Comparativo Portugal-Brasil*” para a qual muito gostaria de poder contar com a sua valiosa colaboração.

O objetivo geral do estudo é analisar as principais características e identificar as contribuições dos livros didáticos para a aprendizagem da Física. Entre os elementos de análise estão as opiniões de docentes de Física do ensino secundário de Portugal e do Brasil quanto forma do uso do livro no ensino e as suas percepções sobre manuais escolares. Também serão objeto de estudo as percepções dos alunos. O uso das informações será exclusivamente para fins científicos.

A sua colaboração será um elemento indispensável para a pesquisa, e caso concorde em colaborar, será assinado um termo de consentimento livre e esclarecido, para o que serão prestados todos os esclarecidos considerados indispensáveis sobre os procedimentos relativos à metodologia de recolha de dados para a investigação.

O plano da investigação possui o consentimento e autorização da 18ª GERED de Florianópolis, e possui autorização (Nº 4277/2013) da Comissão Nacional de Proteção de Dados de Portugal, garantindo-se o cumprimento de todas as normas de utilização de documentos e de confidencialidade estabelecidas por este organismo. Todos os procedimentos formais necessários para viabilizar esta investigação serão realizados, partindo da autorização da direção da escola, assim como dos consentimentos dos professores e encarregados de educação. Estará assegurado o total anonimato e confidencialidade dos dados envolvidos.

Antecipadamente grato sua atenção e colaboração.

Com a mais elevada estima e os melhores cumprimentos,
O Doutorando.

Marcel Bruno P. Braga.

¹⁰⁸ Orientação do Prof. Dr. Décio Ruivo Martins (decio@pollux.fis.uc.pt) e Coorientação de Maria Augusta do Nascimento (guy@mat.uc.pt).

¹⁰⁹ Coorientação do Prof. Dr. José Angotti (zeangotti@gmail.com)

G5: TCLE – Encarregado de Educação



DEPARTAMENTO
DE FÍSICA
UNIVERSIDADE DE COIMBRA
FACULDADE DE CIÊNCIAS
E TECNOLOGIA

Ex^{mo}. Senhor Encarregado de Educação

Assunto: Termo de Consentimento Livre e Esclarecido (TCLE) para colaboração numa investigação de doutoramento na área do ensino da Física.

Marcel Bruno Pereira Braga vem por este meio solicitar o apoio na realização de uma pesquisa no âmbito do seu Doutoramento em Ensino da Física.

A viabilidade desta investigação depende do consentimento dos encarregados de educação. Este estudo respeita todas as condições éticas em pesquisa científica, de acordo com as leis vigentes em Portugal¹¹⁰.

Termo de Consentimento de Livre e Esclarecido

Na qualidade de encarregado de educação do aluno _____ do _____ ano, turma _____, N^o _____, da Escola _____ declaro autorizar:

		Autorizo	Não autorizo
1	O preenchimento de questionários e testes pelo meu educando no âmbito do estudo <i>“Uma Análise do Conteúdo do Livro Didáctico de Física e suas Implicações no Ensino-Aprendizagem da Termodinâmica: Um Estudo Comparativo Portugal-Brasil”</i> .	<input type="radio"/>	<input type="radio"/>
2	A eventual gravação em áudio das aulas – com o objectivo de analisar as referências do professor sobre o manual escolar.	<input type="radio"/>	<input type="radio"/>

_____, _____ de _____ de 2014.

Assinatura do Encarregado de Educação.

Marcel Bruno Pereira Braga

Assinatura do Responsável pela Investigação.

¹¹⁰ Este estudo respeita as condições éticas de pesquisa, de acordo com a autorização 4277/2013 dada pela Comissão Nacional de Proteção de Dados, que representa a Autoridade Nacional de Controlo de Dados Pessoais (Decreto-Lei n^o 43/2004 de 18 de Agosto) e a Monitorização de Inquéritos em Meio Escolar, controlado pela Direção-Geral de Estatísticas da Educação e Ciência (Decreto-Lei n^o 125/2011 de 29 de Dezembro).

G6: TCLE – Professores (Modelo Portugal)



Ex^{mo}. Senhor Professor

Assunto: Termo de Consentimento Livre e Esclarecido (TCLE) para colaboração numa investigação de doutoramento na área do ensino da Física.

Marcel Bruno Pereira Braga vem por este meio solicitar o apoio na realização de uma pesquisa no âmbito do seu Doutoramento em Ensino da Física.

O objetivo deste projeto é identificar os aspetos conceptuais nos manuais escolares de Física relacionados com a Termodinâmica, procurando identificar fatores no processo ensino-aprendizagem que possam influenciar a aquisição dos conceitos científicos pelos estudantes.

A viabilidade desta investigação depende da participação dos docentes. Este estudo respeita todas as condições éticas em pesquisa científica, de acordo com as leis vigentes em Portugal¹¹¹.

Termo de Consentimento de Livre e Esclarecido

Eu, _____, professor da Escola _____ declaro aceitar colaborar no estudo *“Uma Análise do Conteúdo do Livro Didático de Física e suas Implicações no Ensino-Aprendizagem da Termodinâmica: Um Estudo Comparativo Portugal-Brasil”*.

_____, _____ de _____ de 2014.

ASSINATURA DO PROFESSOR

Marcel Bruno Pereira Braga
ASSINATURA DO RESPONSÁVEL PELA INVESTIGAÇÃO

¹¹¹ Este estudo respeita as condições éticas de pesquisa, de acordo com a autorização 4277/2013 dada pela Comissão Nacional de Proteção de Dados, que representa a Autoridade Nacional de Controlo de Dados Pessoais (Decreto-Lei nº 43/2004 de 18 de Agosto) e a Monitorização de Inquéritos em Meio Escolar, controlado pela Direção-Geral de Estatísticas da Educação e Ciência (Decreto-Lei nº 125/2011 de 29 de Dezembro).

G7: TCLE – Professores (Modelo Brasil)



Ex^{mo}. Senhor Professor

Assunto: Termo de Consentimento Livre e Esclarecido (TCLE) para colaboração numa investigação de doutoramento na área do ensino da Física.

Venho por este meio solicitar o apoio na continuação de minha pesquisa investigativa no âmbito do Doutoramento em Ensino da Física.

O objetivo deste projeto é identificar os aspetos conceptuais nos livros didáticos de Física relacionados com a Termodinâmica, procurando identificar fatores no processo ensino-aprendizagem que possam influenciar a aquisição dos conceitos científicos pelos estudantes.

A viabilidade desta investigação depende da participação dos docentes. Este estudo respeita todas as condições éticas em pesquisa científica, de acordo com as leis vigentes.

Termo de Consentimento de Livre e Esclarecido

Eu, _____, professor da Escola _____ declaro aceitar colaborar no estudo *“Uma Análise do Conteúdo do Livro Didático de Física e suas Implicações no Ensino-Aprendizagem da Termodinâmica: Um Estudo Comparativo Portugal-Brasil”*.

_____, _____ de _____ de 2014.

ASSINATURA DO PROFESSOR

Marcel Bruno Pereira Braga
ASSINATURA DO RESPONSÁVEL PELA INVESTIGAÇÃO



TESTE PADRONIZADO DE TERMODINÂMICA

ORIENTAÇÕES GERAIS AO PROFESSOR

Caro(a) Professor(a):

1. Pedimos a sua **colaboração** na aplicação deste teste, mantendo o anonimato, indicando apenas as suas iniciais, a(s) turma(s) que lecciona, e dos alunos a respetiva numeração. A sua contribuição será fundamental para esta pesquisa realizada no âmbito de Doutoramento em Ensino da Física.
2. Este teste incide sobre conhecimentos prévios dos conceitos de **calor e temperatura**.
3. O teste deve ser respondido **individualmente** pelos alunos.
4. **Não é permitido** consulta nem o uso de qualquer instrumento eletrónico, pois é desnecessário.
5. Caso seja de interesse, o **professor poderá utilizar este instrumento** como recurso avaliativo, nesse caso será fornecido **grelha de correção**, e/ou os resultados dos alunos aferidos pelo investigador.
A sugestão ao professor é que seja uma *avaliação de orientação* (predictiva – ao prever hipóteses de êxito; de previsão – ao emitir hipóteses futuras; de diagnóstico – ao identificar fraquezas do sistema ou da pessoa) ou *de regulação* (formativa – ao serviço do indivíduo; formadora – ao serviço do professor), e *não de certificação*, pressupondo que o assunto não tenha sido ainda abordado em sala de aula.
6. Segundo seus elaboradores Shelley Yeo e Marjan Zadnik (*Curtin University of Technology, Austrália*), este teste poderá ter a função de: (1) instrumento de **pré-teste/pós-teste**; (2) avaliar as **concepções alternativas** em qualquer ponto durante a ensino, ou; (3) **planear/reformular** o ensino, conforme necessário.
7. Com poucas exceções, devidamente assinaladas, as respostas devem ser dadas através da colocação de uma cruz (X) ou de um sinal semelhante a (✓), em cada círculo O.
8. Caso haja rasuras, marcação de duas alternativas ou omissão em uma questão, **será desconsiderada**.
9. O teste tem **26 questões** e a estimativa para se responder cada questão é inferior a 1 min.
10. Cada questão possui uma **única resposta** correcta.
11. Todos **os itens são intuitivos** e não necessitam de cálculos, exigindo apenas conhecimento qualitativo e prático sobre a compreensão de fenómenos termodinâmicos relacionados com o quotidiano dos alunos.

Este estudo respeita as condições éticas de pesquisa, de acordo com a autorização 4277/2013 dada pela Comissão Nacional de Proteção de Dados, que representa a Autoridade Nacional de Controlo de Dados Pessoais (Decreto-Lei nº 43/2004 de 18 de Agosto) e a Monitorização de Inquéritos em Meio Escolar, controlado pela Direção-Geral de Estatísticas da Educação e Ciência (Decreto-Lei nº 125/2011 de 29 de Dezembro).

H - Instrumentos

H1: Tradução do TCE - Versão Portugal



VAMOS APLICAR...OS TEUS CONHECIMENTOS DE TERMODINÂMICA NA COZINHA?!

ORIENTAÇÕES AOS ALUNOS

Caro(a) Aluno(a):

1. Pedimos a tua colaboração na realização deste teste, mantendo o anonimato, indicando apenas a turma e o número na lista de chamada. A tua contribuição será fundamental para esta investigação.
2. Este teste incide sobre os teus conhecimentos prévios sobre os conceitos de calor e temperatura.
3. Com poucas exceções, devidamente assinaladas, as respostas devem ser dadas através da colocação de uma cruz (X) a alínea correta (a, b, c, d, e).
4. O teste é individual.
5. Não é permitido consulta e nem o uso de qualquer instrumento eletrónico.
6. O teste tem 26 questões e a estimativa para se responder a cada questão é inferior a 1 min.
7. Para cada questão escolhe a resposta que está mais próxima do teu entendimento.
8. Cada questão possui uma única resposta correcta.
9. Todos os itens são intuitivos e não necessitam de cálculos, exigindo apenas um conhecimento qualitativo e prático sobre a compreensão de fenómenos termodinâmicos relacionados com o quotidiano.
10. Procura ser cuidadoso(a) ao marcar a alternativa considerada correcta. Algumas questões possuem cinco alternativas.

SEXO DATA DE NASCIMENTO DATA DO TESTE ANO LECTIVO

F M ___/___/___ ___/___/___ 10º 11º 12º

TURMA: _____ Nº: _____

ESCOLA: _____



Imagina que **que estás com um grupo de amigos numa cozinha. **Todos** estão atentos e interessados em compreender os fenómenos físicos do dia-a-dia que ocorrem nesse ambiente.**

QUESTÕES

1. Qual é a temperatura mais provável dos cubos de gelo que se encontram armazenados no congelador de um frigorífico?
- a) -10°C b. 0°C c. 5°C
d. A temperatura depende do tamanho dos cubos de gelo.
2. Francisco pega em seis cubos de gelo no frigorífico e coloca quatro deles dentro de um copo com água, deixando os outros dois sobre a mesa. Agita várias vezes o copo até que os cubos de gelo estejam muito pequenos e tenham parado de derreter. Qual é a temperatura mais provável da água nesta situação?
- a. -10°C c. 5°C
b. 0°C d. 10°C
3. Os cubos de gelo que Francisco deixou sobre o balcão quase derreteram por completo e repousam sobre uma poça de água, ficando ainda pequenos pedaços de gelo. Qual é a temperatura mais provável desses cubos de gelo menores?
- a. -10°C c. 5°C
b. 0°C d. 10°C
4. Sobre o fogão está uma panela cheia de água. A água começou a ferver rapidamente. A temperatura mais provável da água é:
- a. 88°C b. 98°C c. 110°C
d. Nenhuma das respostas está correta.
5. Cinco minutos mais tarde, a água na panela ainda ferve. A temperatura mais provável da água agora é:
- a. 88°C c. 110°C
b. 98°C d. 120°C
6. Qual será a temperatura previsível do vapor formado, acima da água na panela?
- a. 88°C c. 110°C
b. 98°C d. 120°C
7. Susana pega em dois copos de água a 40°C e mistura-os à água de um copo a 10°C . Qual é a temperatura mais provável da mistura?
- a. 20°C c. 30°C
b. 25°C d. 50°C
8. Pedro deve usar água fervendo para fazer uma chávena de chá. Ele diz aos seus amigos: "Eu não poderia fazer chá se estivesse acampado numa montanha alta, porque a água não ferve a grandes altitudes".
- a. João diz: "Sim, a água ferve, mas não a uma temperatura tão elevada como aqui".
b. Tânia diz: "Não é verdade. A água ferve sempre à mesma temperatura".
c. Luís diz: "O ponto de ebulição da água diminui, mas a água ainda ferve a 100°C ".
d. Sandra diz: "Eu concordo com Pedro. A água nunca atingirá o seu ponto de ebulição". Indica com quem concordas.
9. Samuel pega numa lata de Coca-Cola e uma garrafa plástica de Coca-Cola do frigorífico, onde tinham sido colocadas na noite anterior. Ele rapidamente coloca um termómetro na Coca-Cola contida na lata. A temperatura é de 7°C . Qual é a temperatura mais provável da garrafa de plástico e da Coca-Cola no seu interior?
- a. Estão ambas abaixo de 7°C .
b. Estão ambas a 7°C .
c. Estão ambas acima de 7°C .
d. A Coca-Cola está a 7°C , mas a garrafa está a mais de 7°C .
e. Depende da quantidade de Coca-Cola e/ou do tamanho da garrafa.
10. Pedro coloca a lata de Coca-Cola sobre a mesa. Poucos minutos mais tarde, Carlos levanta a lata de Coca-Cola e diz que a superfície da mesa que estava debaixo da lata está mais fria do que o resto da mesa.
- a. Gabriel diz: "O frio foi transferido da Coca-Cola para a mesa".
b. André diz: "Não há energia deixada na mesa pela lata."
c. Sabrina diz: "Um pouco de calor foi transferido da mesa para Coca-Cola".
d. Elisa diz: "A lata faz com que o calor debaixo dela se desloque através dela para a mesa". Indica qual das explicações achas que é a melhor.
11. Patrícia pergunta aos amigos: "Se eu colocar 100 gramas de gelo a 0°C e 100 gramas de água a 0°C dentro do frigorífico, qual dos dois perderá a maior quantidade de calor?"
- a. Cristina diz: "Os 100 gramas de gelo".
b. Bernardo diz: "Os 100 gramas de água".
c. Marcelo diz: "Nenhum dos dois, porque ambos contêm a mesma quantidade de calor".
d. Mateus diz: "Não há resposta, porque o gelo não contém nenhum calor".
e. Júlia diz: "Não há resposta, porque não podemos ter água a 0°C ". Indica com qual dos amigos de Patrícia concordas.

12. Marta está a ferver água numa panela sobre o fogão. O que está no interior das bolhas que se formam na água em ebulição? Principalmente:

- a. Ar.
- b. Gás - oxigénio e hidrogénio.
- c. Vapor de água.
- d. Não há nada nas bolhas.

13. Depois de cozinhar alguns ovos em água a ferver, Mariana arrefece-os colocando-os numa panela com água fria. Qual das seguintes afirmações explica o processo de arrefecimento?

- a. A temperatura é transferida do ovo para a água.
- b. O frio desloca-se da água para os ovos.
- c. Objetos quentes arrefecem naturalmente.
- d. A energia é transferida dos ovos para a água.

14. Jacinta diz que não gosta de se sentar em cadeiras de metal porque “elas são mais frias do que as de plástico”.

- a. Rafael concorda e diz: “Elas são mais frias porque o metal é naturalmente mais frio do que o plástico”.
- b. Mário diz: “Elas não são mais frias, elas estão à mesma temperatura”.
- c. Luís diz: “Elas não são mais frias, as de metal só parecem mais frias porque são mais pesadas”.
- d. Sandra diz: “Elas são mais frias porque o metal tem menos calor para perder do que o plástico”.

Indica quem achas que está correto.

15. Um grupo está a ouvir a previsão do tempo no rádio: “... à noite fará 5 °C, mais frio do que os 10 °C da noite passada”.

- a. Marta diz: “Isso significa que esta noite será duas vezes mais fria que a noite passada”.
- b. Alice diz: “Não está certo. 5 °C não é duas vezes mais frio que 10 °C”.
- c. Rui diz: “Está parcialmente correto, mas ela deveria dizer que 10 °C é duas vezes mais quente do que 5 °C”.
- d. Gabriel diz: “Está parcialmente correto, mas ela deveria ter dito que 5 °C é metade do frio de 10 °C”.

Indica com que afirmação concordas mais.

16. Tomás pega numa régua de metal e numa de madeira do seu estojo. Ele diz que sente a régua de metal mais fria do que a de madeira. Qual é explicação com que concordas mais?

- a. O metal retira energia da mão mais rapidamente do que a madeira.
- b. A madeira é naturalmente uma substância mais quente do que o metal.
- c. A régua de madeira contém mais calor do que a régua de metal.
- d. Os metais são melhores irradiadores de calor do que a madeira.
- e. O frio flui mais facilmente de um metal.

17. Alzira pegou em duas garrafas de vidro contendo água a 20 °C e embrulhou-as em toalhas. Uma das toalhas estava molhada e a outra estava seca. Vinte minutos mais tarde, ela mediu a temperatura da água em cada garrafa. A água na garrafa embrulhada na toalha molhada estava a 18 °C e a garrafa embrulhada na toalha seca estava a 22 °C. A temperatura mais provável da sala durante esse tempo era:

- a. 26 °C
- b. 21 °C
- c. 20 °C
- d. 18 °C

18. Daniel pega simultaneamente em duas embalagens de leite achocolatado, uma fria, retirada do frigorífico e outra morna, que tinha sido deixada sobre a mesa por algum tempo. Por que razão a embalagem retirada do frigorífico parece mais fria do que a da mesa? Comparando com a embalagem morna, a embalagem fria:

- a. Contém mais frio.
- b. Contém menos calor.
- c. É pior condutora de calor.
- d. Retira calor mais rapidamente da mão de Daniel.
- e. Cede calor mais rapidamente para a mão de Daniel.

19. Rui sabe que a sua mãe faz sopa na panela de pressão porque esta cozinha mais rápida do que uma panela normal, mas não sabe o porquê disso (panelas de pressão têm a tampa selada, o que faz com que a pressão no interior atinja um valor maior que o da pressão atmosférica).

- a. Armanda diz: “Isso acontece porque a pressão faz a água ferver acima de 100 °C”.
- b. Carlos diz: “Isso ocorre porque a pressão alta gera calor extra”.
- c. Fernanda diz: “É porque o vapor está numa temperatura maior do que a sopa fervente”.

d. Tomás diz: “É porque panelas de pressão propagam o calor mais uniformemente através dos alimentos”.

Indica com qual deles concordas mais.

20. Patrícia acredita que o seu pai coloca um bolo na prateleira superior do forno elétrico porque a parte de cima é mais quente do que a de baixo.

a. Paula diz que é mais quente porque o calor sobe.

b. Samuel diz que é mais quente porque bandejas de metal concentram mais calor.

c. Rúben diz é mais quente na parte superior do forno porque quanto mais quente o ar, menos denso ele é.

d. Tiago não concorda com eles e diz que não é possível que a parte superior do forno seja mais quente.

Indica qual deles achas que está correto.

21. Beatriz lê uma questão de múltipla escolha num manual: “Suar arrefece-te porque o suor sobre a tua pele:

a. Humedece a superfície, e superfícies húmidas retiram mais calor do que superfícies secas.

b. Drena o calor dos poros e espalha-o pela superfície da pele.

c. Está à mesma temperatura do que a tua pele, mas está a evaporar e assim afasta o calor dela.

d. É ligeiramente mais frio do que a tua pele devido à evaporação, e dessa forma o calor é transferido da tua pele para o suor”.

Indica que resposta dirias para ela escolher.

22. Quando o André usa uma bomba para encher o pneu da bicicleta, observa que a bomba fica bem quente. Qual das explicações abaixo parece ser a melhor?

a. Energia foi transferida para a bomba.

b. Temperatura foi transferida para a bomba.

c. Calor propagou-se das suas mãos para a bomba.

d. O metal da bomba provoca o aumento de temperatura.

23. Porque vestimos agasalhos quando está frio?

a. Para manter o frio fora da roupa.

b. Para gerar calor.

c. Para reduzir a perda de calor.

d. As três respostas anteriores estão corretas.

24. Vítor pega em alguns gelados do congelador, onde os tinha colocado no dia anterior, e diz que os palitos de madeira estão a uma temperatura mais alta do que os gelados.

a. Débora diz: “Está certo, porque os palitos de madeira não ficam tão frios quanto o sorvete”.

b. Ivan diz: “Está certo, porque o sorvete contém mais frio do que a parte de madeira”.

c. Márcio diz: “Está errado, eles apenas parecem diferentes porque o palito contém mais calor”.

d. Ana diz: “Eu acho que eles estão à mesma temperatura porque estão juntos.

Indica com qual deles concordas.

25. Gabriela fala sobre um programa de televisão que viu na noite anterior: “Eu vi físicos fazerem supercondutores magnéticos a uma temperatura de $-260\text{ }^{\circ}\text{C}$ ”.

a. Jorge duvida disso: “Deves estar errada. Não se podem ter temperaturas tão baixas”.

b. Cátia discorda de Jorge e diz: “Sim, podem. Não existe um limite inferior para baixas temperaturas”.

c. Leonardo acredita que Gabriela está certa: “Eu acho que o íman estava perto da temperatura mais baixa possível”.

d. Gabriela não está segura: “Eu acho que os supercondutores são bons condutores de calor, logo não se pode esfriá-los a tão baixas temperaturas”.

Indica quem achas que está correto.

26. Alguns amigos estão a lembrar brincadeiras de crianças. Patrícia diz: “Eu costumava envolver as minhas bonecas em cobertores mas nunca entendi porque elas não aqueciam”.

a. Marcelo afirma: “É porque os cobertores eram provavelmente maus isoladores térmicos.

b. Laura replica: “É porque os cobertores eram provavelmente maus condutores”.

c. Fátima replica: “É porque as bonecas eram feitas de um material que não mantém bem o calor”.

d. Cecília diz: “É porque as bonecas eram feitas de um material que levava muito tempo a aquecer”.

e. Joana afirma: “Vocês estão todos errados”. Indica com quem concordas.

H2: Tradução do TCE - Versão Brasil



FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA



VAMOS APLICAR...OS TEUS CONHECIMENTOS DE TERMODINÂMICA NA COZINHA?!

ORIENTAÇÕES AOS ALUNOS

Caro(a) Aluno(a):

1. Pedimos a tua colaboração na realização deste teste, mantendo o anonimato, indicando apenas a turma e o número na lista de chamada. A tua contribuição será fundamental para esta investigação.
2. Este teste incide sobre os teus conhecimentos prévios sobre os conceitos de calor e temperatura.
3. Com poucas exceções, devidamente assinaladas, as respostas devem ser dadas através da colocação de uma cruz (X) a alínea correta (a, b, c, d, e).
4. O teste é individual.
5. Não é permitido consulta e nem o uso de qualquer instrumento eletrónico.
6. O teste tem 26 questões e a estimativa para se responder a cada questão é inferior a 1 min.
7. Para cada questão escolhe a resposta que está mais próxima do teu entendimento.
8. Cada questão possui uma única resposta correta.
9. Todos os itens são intuitivos e não necessitam de cálculos, exigindo apenas um conhecimento qualitativo e prático sobre a compreensão de fenómenos termodinâmicos relacionados com o quotidiano.
10. Procura ser cuidadoso(a) ao marcar a alternativa considerada correta. Algumas questões possuem cinco alternativas.

SEXO DATA DE NASCIMENTO DATA DO TESTE ANO LETIVO

F M ___/___/___ ___/___/___ 1º 2º 3º

TURMA: _____ Nº: _____

ESCOLA: _____



Imagina que **que estás com um grupo de amigos numa cozinha. **Todos** estão atentos e interessados em compreender os fenómenos físicos do dia-a-dia que ocorrem nesse ambiente.**

1. Qual é a temperatura mais provável dos cubos de gelo que se encontram armazenados no congelador de uma geladeira?
- b) $-10\text{ }^{\circ}\text{C}$
 c) $0\text{ }^{\circ}\text{C}$
 d) $5\text{ }^{\circ}\text{C}$
 d. A temperatura depende do tamanho dos cubos de gelo.
2. Francisco pega seis cubos de gelo no congelador e coloca quatro deles dentro de um copo com água, deixando os outros dois sobre a mesa. Agita várias vezes o copo até que os cubos de gelo estejam bem pequenos e tenham parado de derreter. Qual é a temperatura mais provável da água nesta situação?
- a. $-10\text{ }^{\circ}\text{C}$ c. $5\text{ }^{\circ}\text{C}$
 b. $0\text{ }^{\circ}\text{C}$ d. $10\text{ }^{\circ}\text{C}$
3. Os cubos de gelo que Francisco deixou sobre o balcão quase derreteram por completo e repousam sobre uma poça de água, ficando ainda pequenos pedaços de gelo. Qual é a temperatura mais provável desses cubos de gelo menores?
- a. $-10\text{ }^{\circ}\text{C}$ c. $5\text{ }^{\circ}\text{C}$
 b. $0\text{ }^{\circ}\text{C}$ d. $10\text{ }^{\circ}\text{C}$
4. Sobre o fogão está uma panela cheia de água. A água começou a ferver rapidamente. A temperatura mais provável da água é:
- a. $88\text{ }^{\circ}\text{C}$
 b. $98\text{ }^{\circ}\text{C}$
 c. $110\text{ }^{\circ}\text{C}$
 d. Nenhuma das respostas acima é correta.
5. Cinco minutos mais tarde, a água na panela ainda ferve. A temperatura mais provável da água agora é:
- a. $88\text{ }^{\circ}\text{C}$ c. $110\text{ }^{\circ}\text{C}$
 b. $98\text{ }^{\circ}\text{C}$ d. $120\text{ }^{\circ}\text{C}$
6. Qual será a temperatura previsível do vapor formado, acima da água na panela?
- a. $88\text{ }^{\circ}\text{C}$ c. $110\text{ }^{\circ}\text{C}$
 b. $98\text{ }^{\circ}\text{C}$ d. $120\text{ }^{\circ}\text{C}$
7. Susana pega dois copos com água a $40\text{ }^{\circ}\text{C}$ e mistura-os à água de um copo a $10\text{ }^{\circ}\text{C}$. Qual é a temperatura mais provável da mistura?
- a. $20\text{ }^{\circ}\text{C}$ c. $30\text{ }^{\circ}\text{C}$
 b. $25\text{ }^{\circ}\text{C}$ d. $50\text{ }^{\circ}\text{C}$
8. Pedro acredita que deva usar água fervente para fazer uma xícara de chá. Ele diz aos seus amigos: "Eu não poderia fazer chá se estivesse acampado numa montanha alta, porque a água não ferve a grandes altitudes."
- a. João diz: "Sim, a água ferve, mas não a uma temperatura tão elevada como aqui."
 b. Tânia diz: "Não é verdade. A água ferve sempre à mesma temperatura."
 c. Luís diz: "O ponto de ebulição da água diminui, mas a água ainda ferve a $100\text{ }^{\circ}\text{C}$."
 d. Sandra diz: "Eu concordo com Pedro. A água nunca atingirá o seu ponto de ebulição." Indique com quem você concorda.
9. Samuel pega numa lata de refrigerante cola e uma garrafa plástica de refrigerante cola da geladeira, onde tinham sido colocados na noite anterior. Ele rapidamente coloca um termômetro no refrigerante contido na lata. A temperatura é de $7\text{ }^{\circ}\text{C}$. Quais são as temperaturas mais prováveis da garrafa de plástico e do refrigerante cola de seu interior?
- a. Estão ambas abaixo de $7\text{ }^{\circ}\text{C}$.
 b. Estão ambas a $7\text{ }^{\circ}\text{C}$.
 c. Estão ambas acima de $7\text{ }^{\circ}\text{C}$.
 d. O refrigerante cola está a $7\text{ }^{\circ}\text{C}$, mas a garrafa está a mais de $7\text{ }^{\circ}\text{C}$.
 e. Depende da quantidade de refrigerante cola e/ou do tamanho da garrafa.
10. Pedro coloca a lata de refrigerante cola sobre a mesa. Poucos minutos mais tarde, Carlos levanta a lata de refrigerante cola e diz que a superfície da mesa que estava debaixo da lata está mais fria do que o resto da mesa.
- a. Gabriel diz: "O frio foi transferido do refrigerante cola para a mesa."
 b. André diz: "Não há energia deixada na mesa pela lata."
 c. Sabrina diz: "Um pouco de calor foi transferido da mesa para o refrigerante cola."
 d. Elisa diz: "A lata faz com que o calor debaixo dela se desloque através dela para a mesa."

Indique qual das explicações você acredita ser a melhor.

11. Patrícia pergunta aos amigos: “Se eu colocar 100 gramas de gelo a 0 °C e 100 gramas de água a 0 °C dentro da geladeira, qual dos dois perderá a maior quantidade de calor?”

- a. Cristina diz: “Os 100 gramas de gelo.”
- b. Bernardo diz: “Os 100 gramas de água.”
- c. Marcelo diz: “Nenhum dos dois, porque ambos contêm a mesma quantidade de calor.”
- d. Mateus diz: “Não há resposta, porque o gelo não contém nenhum calor.”
- e. Júlia diz: “Não há resposta, porque não podemos ter água a 0 °C.”

Indique com qual dos amigos de Patrícia você concorda.

12. Marta está fervendo água numa panela sobre o fogão. O que você acha que está no interior das bolhas que se formam na água em ebulição? Na maior parte:

- a. Ar.
- b. Gás oxigênio e hidrogênio.
- c. Vapor de água.
- d. Não há nada nas bolhas.

13. Depois de cozinhar alguns ovos em água fervente, Mariana resfria-os ao colocá-los numa panela com água fria. Qual das seguintes afirmações explica o processo de resfriamento?

- a. A temperatura é transferida do ovo para a água.
- b. O frio desloca-se da água para os ovos.
- c. Objetos quentes resfriam naturalmente.
- d. A energia é transferida dos ovos para a água.

14. Janete diz que não gosta de se sentar em cadeiras de metal porque “elas são mais frias do que as de plástico.”

- a. Rafael concorda e diz: “Elas são mais frias porque o metal é naturalmente mais frio do que o plástico.”
- b. Mário diz: “Elas não são mais frias, elas estão à mesma temperatura.”
- c. Luís diz: “Elas não são mais frias, as de metal só parecem mais frias porque são mais pesadas.”

d. Sandra diz: “Elas são mais frias porque o metal tem menos calor para perder do que o plástico.”

Indique quem você acha que está correto.

15. Um grupo está a ouvir a previsão do tempo no rádio: “... à noite fará 5 °C, mais frio do que os 10 °C da noite passada.”

- a. Marta diz: “Isso significa que esta noite será duas vezes mais fria que a noite passada.”
- b. Alice diz: “Não está certo. 5 °C não é duas vezes mais frio que 10 °C.”
- c. Rui diz: “Está parcialmente correto, mas ela deveria dizer que 10 °C é duas vezes mais quente do que 5 °C.”
- d. Gabriel diz: “Está parcialmente correto, mas ela deveria ter dito que 5 °C é metade do frio de 10 °C.”

Indique qual a afirmação que você concorda mais.

16. Tomás pega numa régua de metal e numa de madeira do seu estojo. Ele diz que sente a régua de metal mais fria do que a de madeira. Qual é a explicação com que você concorda mais?

- a. O metal retira energia da mão mais rapidamente do que a madeira.
- b. A madeira é naturalmente uma substância mais quente do que o metal.
- c. A régua de madeira contém mais calor do que a régua de metal.
- d. Os metais são melhores irradiadores de calor do que a madeira.
- e. O frio flui mais facilmente de um metal.

17. Alzira pegou em duas garrafas de vidro contendo água a 20 °C e embrulhou-as em toalhas. Uma das toalhas estava molhada e a outra estava seca. Vinte minutos mais tarde, ela mediu a temperatura da água em cada garrafa. A água na garrafa embrulhada na toalha molhada estava a 18 °C e a água na garrafa com a toalha seca estava a 22 °C. A temperatura ambiente mais provável da sala durante esse tempo era:

- a. 26 °C
- b. 21 °C
- c. 20 °C
- d. 18 °C

18. Daniel pega simultaneamente em duas embalagens de leite achocolatado, uma fria, retirada da geladeira e outra morna, que tinha

sido deixada sobre a mesa por algum tempo. Por que razão a embalagem retirada da geladeira parece mais fria do que a da mesa? Comparando com a embalagem morna, a embalagem fria:

- a. Contém mais frio.
- b. Contém menos calor.
- c. É pior condutora de calor.
- d. Retira calor mais rapidamente da mão de Daniel.
- e. Cede calor mais rapidamente para a mão de Daniel.

19. Rui sabe que a sua mãe faz sopa na panela de pressão porque esta cozinha mais rápido do que uma panela normal, mas não sabe o porquê disso (panelas de pressão têm a tampa selada, o que faz com que a pressão no interior atinja um valor maior que o da pressão atmosférica).

- a. Armanda diz: "Isso acontece porque a pressão faz a água ferver acima de 100 °C."
- b. Carlos diz: "Isso ocorre porque a pressão alta gera calor extra."
- c. Fernanda diz: "É porque o vapor está numa temperatura maior do que a sopa fervente."
- d. Tomás diz: "É porque panelas de pressão propagam o calor mais uniformemente através dos alimentos."

Indique com qual deles você concorda mais.

20. Patrícia acredita que o seu pai coloca um bolo na prateleira superior do forno elétrico porque a parte de cima é mais quente do que a de baixo.

- a. Paula diz que é mais quente porque o calor sobe.
- b. Samuel diz que é mais quente porque bandejas de metal concentram mais calor.
- c. Rubens diz é mais quente na parte superior do forno porque quanto mais quente o ar, menos denso ele é.
- d. Tiago não concorda com eles e diz que não é possível que a parte superior do forno seja mais quente.

Indique qual deles você acha que está correto.

21. Beatriz lê uma questão de múltipla escolha num manual: "Suar resfria-te porque o suor sobre a tua pele:

- a. Umedece a superfície, e superfícies úmidas retiram mais calor do que superfícies secas."

b. Drena o calor dos poros e espalha-o pela superfície da pele."

c. Está à mesma temperatura do que a tua pele, mas está evaporando e assim afasta o calor dela."

d. É ligeiramente mais frio do que a tua pele devido à evaporação, e dessa forma o calor é transferido da tua pele para o suor."

Indique a resposta que você diria para ela escolher.

22. Quando o André usa uma bomba para encher o pneu da bicicleta, observa que a bomba fica bem quente. Qual das explicações abaixo parece ser a melhor?

- a. Energia foi transferida para a bomba.
- b. Temperatura foi transferida para a bomba.
- c. Calor propagou-se das suas mãos para a bomba.
- d. O metal da bomba provoca o aumento de temperatura.

23. Porque vestimos agasalhos quando está frio?

- a. Para manter o frio fora da roupa.
- b. Para gerar calor.
- c. Para reduzir a perda de calor.
- d. As três respostas anteriores estão corretas.

24. Vítor pega alguns picolés do congelador, onde os tinha colocado no dia anterior, e diz que os palitos de madeira estão a uma temperatura mais alta do que a parte do "sumo gelado" que têm o sabor.

a. Débora diz: "Está certo, porque os palitos de madeira não ficam tão frios quanto o sumo gelado."

b. Ivan diz: "Está certo, porque o sumo gelado contém mais frio do que a parte de madeira."

c. Márcio diz: "Está errado, eles apenas parecem diferentes porque o palito contém mais calor."

d. Ana diz: "Eu acho que eles estão à mesma temperatura porque estão juntos."

Indique com qual deles você concorda.

25. Gabriela fala sobre um programa de televisão que viu na noite anterior: "Eu vi físicos fazerem supercondutores magnéticos a uma temperatura de -260 °C."

a. Jorge duvida disso: “Deves estar errada. Não se podem ter temperaturas tão baixas.”

b. Cátia discorda de Jorge e diz: “Sim, podem. Não existe um limite inferior para baixas temperaturas.”

c. Leonardo acredita que Gabriela está certa: “Eu acho que o íman estava perto da temperatura mais baixa possível.”

d. Gabriela não está segura: “Eu acho que os supercondutores são bons condutores de calor, logo não se pode esfriá-los a tão baixas temperaturas.”

Indique quem você acha que está correto.

26. Alguns amigos lembram-se das brincadeiras de crianças. Patrícia diz: “Eu costumava envolver as minhas bonecas em cobertores, mas nunca entendi porque elas não aqueciam.”

a. Marcelo afirma: “É porque os cobertores eram provavelmente maus isoladores térmicos.”

b. Laura replica: “É porque os cobertores eram provavelmente maus condutores.”

c. Fátima replica: “É porque as bonecas eram feitas de um material que não mantém bem o calor.”

d. Cecília diz: “É porque as bonecas eram feitas de um material que levava muito tempo a aquecer.”

e. Joana afirma: “Vocês estão todos errados.”

Indique com quem você concorda.

ANEXOS

1: CNPD – Autorização Formal para a Realização da Pesquisa

Proc. N.º: 784/2013 | 1

COMISSÃO NACIONAL
DE PROTECÇÃO DE DADOS

AUTORIZAÇÃO N 4277 /2013

I. Do Pedido

Marcel Bruno Pereira Braga, no âmbito da sua Tese de Doutoramento, notificou à CNPD um tratamento de dados pessoais com a finalidade de realizar um estudo observacional intitulado “Uma Análise do Conteúdo do Livro Didático de Física e suas implicações no Ensino-Aprendizagem da Termodinâmica: Um Estudo Comparativo Portugal-Brasil”.

O estudo pretende recolher informações sobre a aprendizagem dos alunos a partir de um diagnóstico sobre os manuais escolares adotados pelas escolas para a disciplina de Física. Para o efeito, serão recolhidas cópias de materiais de planeamento das aulas, registo áudio das aulas e de entrevistas, e serão aplicados questionários e testes aos alunos.

De acordo com as explicações do investigador, o estudo será dividido em três fases: uma fase pré-descritiva, uma fase descritiva e uma fase comparativa.

Serão incluídos no estudo os alunos de Física a frequentar o 10.º ano nas escolas participantes em Coimbra (E. S. Infanta Dona Maria, E.S. Quinta das Flores e E.S. de Avelar Brotero).

O investigador recolhe o consentimento informado dos participantes através dos respetivos professores, que conservará em focal de acesso reservado.

Os destinatários serão ainda informados sobre a natureza facultativa da sua participação e garantida confidencialidade no tratamento.

II. Da Análise

Porque em grande parte referentes à vida privada, os dados recolhidos pela requerente têm a natureza de sensíveis, razão pela qual o respetivo tratamento só pode basear-se

no consentimento expresso, esclarecido e livre dos titulares dos dados, ou dos seus legais representantes nos termos do disposto no n.º 2 do artigo 7.º da Lei n.º 67/98, de 26 de outubro {Lei de Proteção de Dados - LPD).

Por esta razão é necessário o «consentimento expresso do titular», entendendo-se por consentimento qualquer manifestação de vontade, livre, específica e informada, nos termos da qual o titular aceita que os seus dados sejam objeto de tratamento, o qual deve ser obtido através de urna “declaração de consentimento informado” onde seja utilizada uma linguagem clara e acessível.

Nos termos do artigo 10.º da LPD, a declaração de consentimento tem de conter a identificação do responsável pelo tratamento e a finalidade do tratamento, devendo ainda conter informação sobre a existência e as condições do direito de acesso e de retificação por parte do respetivo titular.

Os titulares dos dados, de acordo com a declaração de consentimento informado junta aos autos, apõem as suas assinaturas na mesma, deste modo satisfazendo as exigências legais.

O fundamento de legitimidade é o consentimento dos titulares dos dados.

A informação tratada é recolhida de forma lícita (cfr. alínea a) do n.º 1 do artigo 5.º da LPD), para finalidades determinadas, explícitas e legítimas (cfr. alínea b) do mesmo artigo).

III. Da Conclusão

Em face do exposto, a Comissão Nacional de Protecção de Dados (CNPd) autoriza o tratamento de dados pessoais supra apreciado, nos termos do n.º 2 do artigo 7.º, da alínea a) do n.º1 do artigo 28.º e do n.º 1 do artigo 30.º da LPD, consignando-se o seguinte:

Responsável pelo tratamento: Marcel Bruno Pereira Braga

*COMISSÃO NACIONAL
DE PROTECÇÃO DE DADOS*

Finalidade: Estudo observacional intitulado “Uma Análise do Conteúdo do Livro Didático de Física e suas implicações no Ensino-Aprendizagem da Termodinâmica: Um Estudo Comparativo Portugal-Brasil”.

Categoria de Dados pessoais tratados:

- **dos professores:** nome, contacto telefónico, e-mail, horário, manual adotado e material de suporte, planos de atividades, relação do professor com a turma, questionários sobre o ensino da física e voz;
- **dos alunos:** dados sociodemográficos (género, idade, ano de escolaridade), testes sobre a termodinâmica e a aprendizagem da física e voz.

Entidades a quem podem ser comunicados: Não há.

Formas de exercício do direito de acesso e retificação: Junto do responsável pelo tratamento.

Interconexões de tratamentos: Não há.

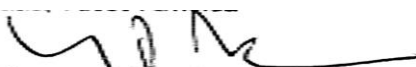
Transferências de dados para países terceiros: Não há.

Prazo de conservação: Os dados pessoais deverão ser destruídos um mês após a defesa da Tese. As gravações das entrevistas e das aulas devem ser destruídas imediatamente após a sua transcrição.

Da presente Autorização decorrem obrigações que o responsável deve cumprir. Deve, igualmente, dar conhecimento dessas condições a todos os intervenientes no circuito de informação.

Lisboa, 04 de junho de 2013.

Ana Roque (Relatora), Helena Delgado António, Carlos Campos Lobo, Luís Paiva de Andrade, Vasco Almeida


Luís Barroso (o Vogal que presidiu)