



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM MÉTODOS E GESTÃO EM AVALIAÇÃO

Kauê Tortato Alves

**ANÁLISE DAS PROPRIEDADES PSICOMÉTRICAS DA PROVA DE
CONHECIMENTOS ESPECÍFICOS DE LICENCIATURA EM CIÊNCIAS
BIOLÓGICAS NO ENADE 2017**

FLORIANÓPOLIS

2020

Kauê Tortato Alves

**ANÁLISE DAS PROPRIEDADES PSICOMÉTRICAS DA PROVA DE
CONHECIMENTOS ESPECÍFICOS DE LICENCIATURA EM CIÊNCIAS
BIOLÓGICAS NO ENADE 2017**

Dissertação submetida ao Programa de Pós-Graduação em Métodos e Gestão em Avaliação da Universidade Federal de Santa Catarina Para a obtenção do Grau de Mestre em Métodos e Gestão em Avaliação.

Orientador: Prof. Dr. Adriano Ferreti Borgatto

Florianópolis

2020

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

Alves, Kauê Tortato

Análise das Propriedades Psicométricas da Prova de
Conhecimentos Específicos de Licenciatura em Ciências
Biológicas no Enade 2017 / Kauê Tortato Alves ; orientador,
Adriano Ferreti Borgatto, 2020.

227 p.

Dissertação (mestrado profissional) - Universidade
Federal de Santa Catarina, Centro Tecnológico, Programa de
Pós-Graduação em , Florianópolis, 2020.

Inclui referências.

1. . 2. Teoria de Resposta ao Item. 3. Ancoragem de
Escala de Proficiência. 4. Avaliação em Larga Escala do
Ensino Superior. 5. Validação de Testes de Desempenho
Educativo. I. Borgatto, Adriano Ferreti . II.
Universidade Federal de Santa Catarina. Programa de Pós
Graduação em . III. Título.

Kauê Tortato Alves

**ANÁLISE DAS PROPRIEDADES PSICOMÉTRICAS DA PROVA DE
CONHECIMENTOS ESPECÍFICOS DE CIÊNCIAS BIOLÓGICAS NO ENADE 2017**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca
examinadora composta pelos seguintes membros:

Prof. Dr. Adriano Ferreti Borgatto
Universidade Federal de Santa Catarina CTC

Prof.^a Dr.^a Andréa Cristina Konrath
Universidade Federal de Santa Catarina - CTC

Prof.^a Dr.^a Sonia Purin da Cruz
Universidade Federal de Santa Catarina - CCR

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi
julgado adequado para obtenção do título de mestre em Métodos e Gestão em
Avaliação.

Prof. Dr. Marcelo Menezes Reis
Coordenador(a) do Programa

Prof. Dr. Adriano Ferreti Borgatto
Orientador(a)

Florianópolis, 19 de fevereiro de 2020.

Este trabalho é dedicado à minha mãe.

AGRADECIMENTOS

Agradeço:

Aos professores Dalton Francisco de Andrade, Pedro Alberto Barbeta e a meu orientador Adriano Ferreti Borgatto, pela paciência com minhas inquietudes, compreensão e cuidado em refrear meus ímpetos digressores, prestatividade na troca de ideias e indicação de caminhos.

Aos contribuintes, amigos e colegas de trabalho por permitirem e suportarem as consequências do afastamento que a pesquisa e a escrita necessitam.

Aos amigos AHJPZ, por contribuírem com momentos de catarse, laboratórios de criação, motivação e na revisão de ideias e textos.

À minha mãe, por suportar as ausências e por seu esforço para fornecer-me um ambiente propício à escrita.

os modelos são úteis apenas na medida em que eles
proporcionam aproximações razoáveis para relações do mundo
real....

(Coelho, 2014)

perfection is not the usual standard for testing statistical
models. Instead, we generally expect that a model should
closely approximate some phenomenon, but not perfectly
reproduce it.

(Kline, 2005)

RESUMO

O ENADE é um exame de desempenho que desde 2004 vem sendo aplicado a concluintes dos cursos de graduação de todo Brasil, tendo avaliado mais de 5 milhões de estudantes de mais de 150 áreas de conhecimento. Porém tal teste utiliza como medida a simples soma de acertos de questões objetivas e das notas de questões discursivas para compor o escore de cada indivíduo, o que se mostra estatisticamente frágil já que não há como calcular o erro padrão da medida. O objetivo geral desse trabalho foi avaliar a qualidade da prova de conhecimentos específicos da área de Licenciatura em Ciências Biológicas - do ENADE 2017, por meio de análise de conteúdo, da Teoria Clássica dos Testes (TCT), de Análise Fatorial e da Teoria de Resposta ao Item (TRI). Por meio de pesquisa documental e análise de conteúdo, foi possível constatar que as “diretrizes de avaliação”, que orientam a elaboração da matriz de referência, diferem das Diretrizes Curriculares Nacionais (DCN) e sua textualização não é estável ao longo das edições, o mesmo pode ser dito do número de questões por subcomponente (pedagógicas e biológicas) e por tipo (discursivas e objetivas). Além disso, constata-se uma superestimação da gama de conhecimentos que cada questão é capaz de avaliar. Com base nos índices de facilidade, discriminação, concluiu-se que os resultados obtidos na análise psicométrica pela TCT e pela TRI são complementares, sendo que a TRI fornece melhor fundamentação para a seleção dos itens a serem mantidos no instrumento. Dessa forma, pela análise da TCT conduzida pelo INEP, alguns itens pouco informativos para a população avaliada foram mantidos e outros itens com boa discriminação foram eliminados. Por meio da análise das cargas fatoriais, comunalidades e índices de ajustes, é possível concluir que a hipótese de unidimensionalidade da prova de conhecimentos específicos é de difícil sustentação, pois o modelo unidimensional apresentou índices de ajustes em valores limítrofes ao aceitável e o índice de consistência interna (*alpha* de Cronbach - α) indicou pouca consistência interna dos dados ($\alpha_{inicial}=0,566$; $\alpha_{final}=0,602$). Conjectura-se que alguns formatos de itens contribuem para fragilizar a validade das interpretações dos escores, pois tratados dicotomicamente não avaliam os sujeitos de menor proficiência. Os itens objetivos da prova de Conhecimentos Específicos apresentavam média de dificuldade de 2,2436 (escala 0,1) e erro padrão (EP) = 1,1175. Após a exclusão dos itens com parâmetros inadequados, o instrumento passou a ter uma dificuldade mais condizente com a população avaliada 1,2931 (EP = 0,1573) e com estimativas mais precisas. Apesar disso, pode-se considerar que a prova tem um nível difícil para a população, por 94,33% da população (12.616) estar abaixo do nível de dificuldade médio da prova. Por isso, salienta-se que a aplicação de modelos da TRI (especialmente o pré-teste e a equalização) e de procedimentos de ancoragem na escala pode ser um passo importante em direção ao aprimoramento da validade dos usos e interpretações dos escores do ENADE.

Palavras-chave: Teoria de Resposta ao Item. Ancoragem de Escala de Proficiência. Avaliação em Larga Escala do Ensino Superior. Validação de Testes de Desempenho Educacional

ABSTRACT

The goal of this work was to evaluate the quality of a multiple-choice test of teacher training in Biological Sciences – the ENADE 2017, this aim was intended through content analysis, TCT, factor analysis and TRI analysis. Through documentary research and content analysis it was found that the "evaluation guidelines", which guide the elaboration of the blue print, differ from the National Curriculum Guidelines (which should follow), and it was verified that "evaluation guidelines" aren't stable throughout the editions. The same temporal inconsistency can be observed in the number of items per subcomponent (pedagogical and biological) and by item type (open vs. close ended). In addition, there is an overestimation of the knowledge range that each item can evaluate. Based on the ease and discrimination indices, it was concluded that the results obtained in the psychometric analysis by the TCT and the IRT are complementary, and the IRT provides a better basis for the selection of the items to be kept in the instrument. Thus, by INEP's TCT analysis, some non-informative items for the evaluated population were kept and others with good discrimination were discarded. Through the analysis of factor loadings, commonalities and adjustment indices, it is possible to conclude that the hypothesis of the test unidimensionality is difficult to sustain, since the one-dimensional model presented adjustment indices bordering acceptable values, and the internal consistency index (Cronbach's α - α) indicated poor internal consistency of the data ($\alpha_{\text{first}}=0,566$; $\alpha_{\text{last}}=0,602$). It is conjectured that some item formats contribute to decrease the score's validity, since they are artificially dichotomized and thus do not evaluate the less proficient subjects. The standardized average difficulty of specific knowledge items was 2.2436 (SE: 1.1175). After the elimination of items with inadequate parameters, the instrument had an average difficulty more consistent with the population evaluated 1.2931 (SE: 0.1573) and with more accurate estimates. Nevertheless, it can be considered that it was a hard test for that population, for 94.33% of the population (12.616) are below the average difficulty level of the test. Therefore, it must be emphasized that the application of IRT models (especially pretesting and equalization) and scale anchoring procedures can be an important step towards improving the validity of the ENADE's scores interpretations and uses.

Keywords: Item Response Theory. Proficiency Scale Trough Anchoring. Large Scale Assessment in Higher Education. Learning Outcomes Test Validation

LISTA DE FIGURAS

Figura 1: Definição das dimensões teóricas empregadas pelo INEP para a construção da Matriz de Referência Tridimensional. Acima o quadro conforme adotado até 2014 e abaixo o novo referencial embasado em Marinho-Araujo e Rabelo (2015).....	46
Figura 2: Fluxo de trabalho da elaboração das provas pelo INEP.	52
Figura 3: Histograma da distribuição de acertos (escore) obtidos pela amostra Concluintes de Licenciatura em Ciências Biológicas no ENADE 2017.	104
Figura 4: Distribuição de frequências de escolha das alternativas em quintis de participantes agrupados em níveis crescentes de escore dos itens eliminados pelo INEP da prova de conhecimentos específicos da área de Licenciatura em Ciências Biológicas. Itens com conteúdo relativos à biologia: Q16 e Q18.....	110
Figura 5: Distribuição de frequências de escolha das alternativas em quintis de participantes agrupados em níveis crescentes de escore dos itens eliminados pelo INEP da prova de conhecimentos específicos da área de Licenciatura em Ciências Biológicas. Itens com conteúdo relativos à formação de professores: Q24L, Q25L e Q30L	111
Figura 6: Gráfico ilustrativo do Alfa de Conbach do instrumento (linha tracejada) e estimativas do Alfa para as hipóteses de retirada de cada um dos 27 itens objetivos de Conhecimentos da área de Licenciatura em Ciências Biológicas no ENADE 2017.	114
Figura 7: Gráfico de autovalores após a extração de componentes principais e fatores (<i>scree plot</i>) da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.	115
Figura 8: Excerto da questão Q14 da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.....	124
Figura 9: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q14 por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita).....	125
Figura 10: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q23 por quintis (abcissas) agrupando os participantes em níveis crescentes de	

escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita).....	127
Figura 11: Excerto da questão Q23 da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.....	128
Figura 12:Excerto da questão Q11 da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.....	129
Figura 13: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q11 por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita).....	131
Figura 14: Excerto da questão Q16 da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.....	133
Figura 15: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q16 por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita).....	134
Figura 16: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q18 por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita).....	135
Figura 17: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q25L por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita).....	136

LISTA DE QUADROS

Quadro 1: Testes de larga escala em âmbito internacional.	34
Quadro 2: Quadro comparativo das características de perfil conforme as DCN (à esquerda) e as diretrizes de prova (à direita).	60
Quadro 3: Quadro comparativo das competências conforme as DCN (à esquerda) e as diretrizes de prova (à direita).	61
Quadro 4: Quadro comparativo dos conteúdos conforme as DCN (a cima) e as diretrizes de prova (abaixo).	63
Quadro 5: Principais indicadores de informação restrita da adequação do modelo aos dados.	137
Quadro 6: Interpretação da escala de proficiência da licenciatura construída sobre o mapeamento de itens âncoras calibrados pelo ML3P.	142

LISTA DE TABELAS

Tabela 1: Quantitativos de estudantes, áreas IES e cursos avaliados ao longo das edições do ENADE.....	29
Tabela 2: Evolução do número de participantes ao longo dos cinco ciclos do ENADE classificados quanto ao tipo de curso, estágio do curso em que realizou a prova, e tipo de presença na prova.....	56
Tabela 3: Número de questões em cada componente da prova da área de Ciências Biológicas (Licenciatura e Bacharelado) ao longo dos cinco ciclos do ENADE.	57
Tabela 4: Sumário estatístico da amostra de Concluintes em Licenciatura em Ciências Biológicas e de sua interação com o ENADE 2017.....	104
Tabela 5: Índices de discriminação e dificuldade dos itens do ENADE de Licenciatura em ciências Biológicas de 2017 calculados pela TCT.....	109
Tabela 6: Análise fatorial de informação completa com simulações para um, dois e três fatores/dimensões explicativas para a variância da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.....	117
Tabela 7: Somatório quadráticos das cargas fatoriais de cada fator/dimensão e correlação entre os fatores nos modelos uni, bi e tridimensional.....	119
Tabela 8: Parâmetros de discriminação, dificuldade e probabilidade de acerto ao acaso, e respectivos erros padrões calibrados para o modelo logístico unidimensional da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.....	122
Tabela 9: Parâmetros finais dos itens com respectivo erro padrão e posicionamento na escala (0,1).....	140
Tabela 10: Itens ordenados por nível crescente de dificuldade, e tipificados quanto ao formato dos itens, parâmetros, carga fatorial, comunalidade e seu posicionamento na escala de proficiência.....	141
Tabela 11: Simulação do estabelecimento de pontos de ancoragem através da metodologia de padronização e escalonamento do INEP.....	145

LISTA DE ABREVIATURAS E SIGLAS

ACER	Conselho Australiano de Pesquisa Educacional (Australian Council for Educational Research)
ACP	Análise de Componentes Principais
ACT	<i>American College Testing</i> (teste de admissão ao Ensino superior organizado pela ACT Inc.)
AEE	Auxílio da Avaliação Educacional
AERA	<i>American Educational Research Association</i>
AFC	Análise Fatorial Confirmatória
AFE	Análise Fatorial Exploratória
AHELO	<i>Assessment of Higher Education Learning Outcomes</i>
AIC	Critério de Informação de Akaike
APA	<i>American Psychological Association</i>
BNI	Banco Nacional de Itens
BNI-ES	Banco Nacional de Itens do Ensino Superior
CAA	Comissão Assessora de Área
CAAP	<i>Collegiate Assessment of Academic Proficiency</i>
CAE	<i>Council for Aid to Education</i>
CAT	Comissão de Assessoramento Técnico
CC	Competências Cidadãs
CCI	Curva Característica do Item
CE	Conhecimentos Específicos
CENEVAL	<i>Centro Nacional de Evaluación para la Educación Superior</i>
CEQ	<i>Course Experience Questionnaire</i>
CERES	Cadastro de Elaboradores e Revisores de Itens da Educação Superior
CES	Câmara de Ensino Superior
CESGRANRIO	Fundação Centro de Seleção de Candidatos ao Ensino Superior do Grande Rio
CESPE/UnB	Centro de Seleção e de Promoção de Eventos Universidade de Brasília
CFI	Índice de ajuste comparativo (Comparative Fit Index)
CH	Carga horária
CLA	<i>Collegiate Learning Assessment</i>

CNE	Conselho Nacional de Educação
CPC	Conceito Preliminar de Curso
DEAES	Diretoria de Estatística e Avaliação da Educação Superior
DIF	Funcionamento Diferencial do Item
DLHE	<i>Destinations of Leavers from Higher Education</i>
dp	Desvio padrão (sd em inglês)
EaD	Educação a Distância
EAP	Esperança a Posteriori (<i>Expected a Posteriori</i>)
EGEL	<i>Exámenes Generales para el Egreso de Licenciatura</i>
EGETSU	<i>Examen General de Egreso del Técnico Superior</i>
EM	Esperança e maximização
ENADE	Exame Nacional de Desempenho dos Estudantes
ENC	Exame Nacional de Cursos
ENEM	Exame Nacional do Ensino Médio
EP	Erro padrão (SE em inglês)
ERIC	<i>Education Resources Information Center</i>
ETS	<i>Educational Testing Service</i>
EXAMINI-III	<i>Exámen Nacional de Ingreso al Posgrado</i>
FG	Formação Geral
FIFA	Análise Fatorial de Informação Completa (<i>Full Information Factor Analysis</i>)
ForGRAD	Fórum dos Pró-reitores de Graduação
GDS	Pesquisa de Destino dos Egressos (<i>Graduate Destination Survey</i>)
GPA	Média ponderada das notas da graduação (<i>Grade Average Point</i>)
GRE	<i>Graduate Record Examination</i>
GSA	<i>Graduate Skills Assessment</i>
GT	Grupo de Trabalho
ICFES	<i>Instituto Colombiano para el Fomento de la Educación Superior</i>
IES	Instituição de Ensino Superior
IGC	Índice Geral de Cursos
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
KMO	Critério de Kaiser-Meyer-Olkin

KOKOHS	Modelos e ferramentas para avaliação de competências no setor universitário (<i>Kompetenzmodelle und Instrumente der Kompetenzerfassung im Hochschulsektor</i>)
KR20	Kuder–Richardson Formula 20
LC	Leitura Crítica (em inglês: CR)
MADPDA	Modelo de Avaliação Dinâmica do Potencial de Domínio de Área
MAPP	<i>Measure of Academic Proficiency and Progress</i>
MBA	<i>Master of Business Administration</i>
MEC	Ministério da Educação
ML3P	Modelo Logístico de Três Parâmetros
MML	Máxima Verossimilhança Marginal (<i>Marginal Maximum Likelihood</i>)
MyGSI	<i>Malaysian Generic Skills Instrument</i>
NCME	<i>National Council on Measurement in Education</i>
NSSE	Pesquisa Nacional de Engajamento Estudantil (<i>National Survey of Student Engagement</i>)
OC	Objeto de Conhecimento
OCDE	Organização para a Cooperação e o Desenvolvimento Econômico
PISA	Programa Internacional de Avaliação de Estudantes (<i>Programme for International Student Assessment</i>)
PLD	Descrição do Nível de Proficiência (<i>Performance Level Description</i>)
RMSEA	Raiz quadrada da média dos erros de aproximação (<i>Root Mean Square Error of Approximation</i>)
RQ	Raciocínio quantitativo
SABIC	Critério e Informação Bayesiano ajustado para o tamanho da amostra (<i>Sample-size Adjusted Bayesian Information Criterion</i>)
SAEB	Sistema de Avaliação da Educação Básica
SAT	<i>Scholastic Aptitude Test</i> (Teste de Aptidão Acadêmica (teste de admissão ao Ensino superior organizado pelo College Board do Educational Testing Service))
SEDAP	Serviço de Acesso a Dados Protegidos
SESU	Secretaria de Educação Superior
SINAES	Sistema Nacional de Avaliação do Ensino Superior

SRMR Raiz quadrada padronizada da média do quadrado dos resíduos
(*Standardized Root Mean Squared Residual*)

TCT Teoria Clássica dos Testes

TI Tecnologia da Informação

TIC Tecnologias de Informação e Comunicação

TLI Tucker Lewis Index

TRI Teoria de Resposta ao Item

LISTA DE SÍMBOLOS

- θ *Theta* ou nível de proficiência
- ∞ Infinito
- a Índice ou Parâmetro de Discriminação
- b Índice ou Parâmetro de Dificuldade
- c Índice ou Parâmetro de Acertos ao Acaso
- Cf Carga Fatorial
- F_n Fator n, onde n é um número inteiro correspondente a uma das dimensões do modelo
- gl Graus de Liberdade
- H₀ Hipótese Nula
- h² Comunalidade
- I Número de Itens
- M2 Índice de adequação de modelos de Maydeu-Olivares e Joe (2006)
- n Número de sujeitos da amostra
- p Probabilidade
- r_{pb} Correlação Ponto Bisserial
- α *Alfa* de Cronbach, uma medida de consistência interna e fiabilidade
- Σ Somatório
- \bar{x} Média amostral
- χ^2 qui-quadrado

SUMÁRIO

1	INTRODUÇÃO	21
1.1	JUSTIFICATIVAS E LIMITAÇÕES	25
1.2	OBJETIVOS	26
1.2.1	Objetivo Geral	26
1.2.2	Objetivos Específicos	26
2	REVISÃO DE LITERATURA	28
2.1	HISTÓRIA E ENQUADRAMENTO TIPOLÓGICO DO ENADE	28
2.2	TESTES DE LARGA ESCALA NO CONTEXTO INTERNACIONAL	32
2.3	MODELOS TEÓRICOS E DEFINIÇÃO DO CONSTRUCTO.....	39
2.4	A ÁREA DE CIÊNCIAS BIOLÓGICAS E O ENADE	53
2.5	PSICOMETRIA: seus modelos, pressupostos e possibilidades	64
2.5.1	Teoria Clássica dos Testes	65
2.5.2	Teoria de Resposta ao Item: ou Teoria do Traço Latente	66
2.5.3	Dimensionalidade	70
2.6	Estado da arte da TRI e o ENADE	75
3	MÉTODOS	88
3.1	Coleta de dados.....	88
3.2	Tratamento e análise dos dados	89
3.3	Recursos Computacionais.....	101
4	RESULTADOS E DISCUSSÃO	102
4.1	Caracterização dos sujeitos da amostra e da população	102
4.2	Análise de conteúdo do teste	105
4.3	Análise dos itens pela Teoria Clássica dos Testes.....	108
4.3.1	Consistência interna	112
4.4	Dimensionalidade	114
4.5	O modelo unidimensional e suas propriedades.....	121
4.5.1.1	Análise e decisão sobre a eliminação ou manutenção de itens	123
4.5.2	Ajuste do modelo	137
4.6	A escala e sua interpretação	138
5	CONCLUSÕES	146
	REFERÊNCIAS	150

1 INTRODUÇÃO

Internacionalmente a educação superior ainda é pouco representada na pesquisa educacional empírica (ZLATKIN-TROITSCHANSKAIA; SHAVELSON, KUHN, 2015), principalmente para avaliar as mudanças na qualidade da educação superior resultantes de implementação de políticas, tais como as novas orientações para competências, que correm muito à frente da pesquisa existente.

Dentre as metodologias de pesquisa mais utilizadas para investigar a qualidade do Ensino Superior, estão os testes de larga escala. Os testes padronizados de desempenho diferem dos testes utilizados em sala de aula de muitas maneiras: quanto à natureza do aprendizado e do conteúdo avaliado; a qualidade dos itens do teste; a confiabilidade do teste; os procedimentos de administração e apuração da pontuação; a interpretação da pontuação (BANGERT, 2006)

Narrando a história dos testes padronizados, Bangert (2006) relata que, até 1926, muitas Instituições de Ensino Superior (IES) americanas utilizavam testes discursivos desenvolvidos localmente para avaliar o preparo dos candidatos para ingressar no ensino superior. Respondendo à demanda por métodos mais padronizados e eficientes para avaliar os aspirantes a vagas no ensino superior, o *College Entrance Examination Board* (CEEB) criou o *Scholastic Aptitude Test* (SAT) provendo as IES com resultados comparáveis das performances dos candidatos. Os militares estadunidenses, ao retornarem no fim da segunda guerra obtiveram, com a lei de reajustamento, benefícios de ingresso no ensino superior, e com isso surge uma necessidade ainda maior de avaliar a preparação de um ampliado número de candidatos ao ingresso no sistema de graduação dos Estados Unidos. Com isso, surge uma profusão de testes de admissão, tais como: o "*American College Testing Program*" (ACT), o "*Scholastic Aptitude Test*" (SAT), o "*Graduate Record Exam*" (GRE), o "*Graduate Management Admissions Test*" (GMAT), o "*Law School Admissions Test*" (LSAT) e o "*Medical College Admissions Test*" (MCAT).

No Brasil, a história dos desdobramentos do processo de avaliação do ensino superior tem sido tratada em várias publicações recentes (POLIDORI, 2009; RISTOFF; GIOLO, 2006 e o polêmico relatório da Organização para a

Cooperação e o Desenvolvimento Econômico, OCDE, 2018). Conforme Landeira-Fernandez e Primi (2002) o “Provão”, formalmente denominado de Exame Nacional de Cursos de Graduação (ENC), foi criado em 1995 por meio da Lei n.º 9.131 (BRASIL, 1995) e fazia parte de um sistema de avaliação da qualidade das Instituições de Ensino Superior do país. Ao apresentar os princípios da proposta do Sistema Nacional de Avaliação da Educação Superior (SINAES), Sobrinho (2004) afirma (de maneira figurada) que o sistema:

Entende que a avaliação não é neutra, não se limita a dimensões técnicas, produz importantes efeitos, tem a ver com valores, culturas e interesses. Avaliação, nesta concepção, não se identifica com o controle, com a mensuração e tampouco com os seus próprios instrumentos. Avaliação requer juízos de valor e mérito. No caso da avaliação da educação, é de sua natureza ser educativa; vale dizer, ação necessariamente social, pedagógica, formativa. (SOBRINHO, 2004, p.114, 115)

A Lei n.º 10.861 de 2004 instituiu SINAES que visa a assegurar o processo nacional de avaliação 1) institucional das IES, 2) de cursos de graduação e 3) do desempenho acadêmico dos estudantes. O terceiro enfoque é o objeto de pesquisa desta dissertação, e atualmente, é avaliado pela realização do Exame Nacional de Desempenho dos Estudantes (ENADE).

Segundo Barbosa, Freire e Crisóstomo (2011), as métricas de mensuração de desempenho discente, atualmente em uso no Brasil, são instrumentos de avaliação externa (i.e. ENADE) que têm sido muito criticados e questionados. Por conseguinte, essas avaliações também necessitam dos rigores científicos para que sejam consideradas a integridade e a qualidade dos instrumentos adotados.

O ENADE visa a “inferir a proficiência dos estudantes em relação aos conteúdos programáticos previstos nas diretrizes curriculares dos respectivos cursos de graduação, desenvolvimento de competências e habilidades necessárias ao aprofundamento da formação geral e profissional e o nível de atualização dos estudantes com relação à realidade brasileira e mundial” (BRASIL, 2018a) e tem como instrumentos complementares ao teste: o “Questionário do Estudante” (com 81 questões para Licenciatura, preenchido on-line), o “Questionário dos Coordenadores de Curso” (com 74 questões, preenchido on-line pelo coordenador), as “questões de avaliação da prova”

(nove questões de avaliação de reação respondidas pelo estudante ao final da prova). O objeto de estudo deste trabalho são os 27 itens objetivos do componente de conhecimentos específicos do ano de 2017. Esse estudo irá focar na prova de ciências biológicas na edição de 2017, por ser a última edição disponível.

Conforme a Portaria n.º 40 do Ministério da Educação, de 12 de dezembro de 2007, o componente de Conhecimentos Específicos (CE) tem como objetivo “aferir as competências, habilidades e conteúdos agregados durante a formação” (BRASIL, 2007), sendo essas especificadas nas diretrizes curriculares nacionais de cada curso.

Da Mata (2016) expõe que existem dois tipos principais de questões emergentes inerentes a qualquer processo de avaliação: as que tratam da eficácia da avaliação; e outras dizem respeito às implicações que surgem a partir dos resultados da avaliação. Essa dissertação aborda questões de eficácia relacionadas à validade do ENADE. Conforme o “*Standards for educational and psychological testing*” das instituições estadunidenses *American Educational Research Association (AERA)*, *American Psychological Association (APA)*, e o *National Council on Measurement in Education (NCME)*:

Validade refere-se ao grau em que as evidências e teorias suportam a interpretação dos escores de um teste, para um dado propósito de uso. Validade é, então, a consideração mais fundamental ao desenvolver e avaliar testes. O processo de validação envolve acúmulo de evidências relevantes para prover uma base científica forte (*soud*) para uma proposta de interpretação do escore. É a interpretação dos escores de um teste para um determinado uso que é avaliada, não o teste em si. Afirmações sobre a validade deve se referir a uma interpretação particular para usos especificados. É incorreto usar a frase não qualificada “a validade do teste”. (AERA; APA; NCME, 2014, p.11, tradução livre)

A validade é o grau em que uma determinada prova efetivamente atinge os objetivos da avaliação (questão de acuidade do instrumento). Validade não é uma característica da prova, e sim das inferências feitas a partir dos resultados. Isto é, uma prova pode ser válida para um objetivo específico, mas inválida para outros tipos de validade, como por exemplo a validade de conteúdo (trabalho da CAA sobre a matriz de referência), de critério e de constructo (CAMPOS,2013).

É preciso fornecer suporte para proposições subjacentes a cada interpretação de um determinado uso [de um teste]. As evidências que suportam a interpretação dos escores de um teste de desempenho em matemática para selecionar estudantes para um curso de um nível subsequente [...] não são as mesmas que permitem inferir a validade para outros propósitos (ex: avaliação de professores para promoção funcional). (AERA; APA; NCME, 2014, p.11, tradução livre)

O arcabouço teórico/metodológico utilizado pelo INEP para avaliar a prova ENADE em 2017 à luz da psicometria foi a Teoria Clássica dos Testes (TCT). A TCT é empregada para explicar o resultado do teste como um todo pela soma dos acertos atingidos em uma série de itens (escore total) por uma amostra dos sujeitos do teste. De acordo com Pasquali e Primi (2003), uma grande preocupação dos psicometristas é sobre o erro contido em tal procedimento de soma de acertos nos itens.

Sustenta-se a hipótese de que a adoção da TCT como principal método de análise de dados das provas do ENADE tem limitado as inferências e interpretações possíveis sobre o desempenho dos estudantes do Ensino Superior (especificamente os estudantes de Licenciatura em Ciências Biológicas). Por exemplo, o acompanhamento do desempenho dos cursos ao longo do tempo e a comparação entre as notas obtidas em diferentes edições é comprometida na TCT, pois não seria possível garantir que as variações nos parâmetros populacionais seriam decorrentes de variações da distribuição de proficiência da população ou de variação na dificuldade de um dos testes. Outra limitação diz respeito à comparação entre estudantes ou cursos em uma mesma edição, pois as notas pela TCT são compostas pelos somatórios dos acertos, sem ponderar se os acertos foram, por um lado, resultado de maior proficiência aos responder itens difíceis, ou por outro lado, respondendo-se itens fáceis e/ou acumulando acertos ao acaso.

O problema central de pesquisa dessa dissertação se refere à questão da validade do teste, isto é: **a prova do ENADE – aplicada em 2017 para os concluintes dos cursos de Licenciatura em Ciências Biológicas – atingiu os objetivos definidos nas diretrizes da prova?** A resposta a essa questão é possível por meio do emprego de ferramentas analíticas da Psicometria (TRI), pois, por meio de seus modelos é possível realizar a análise e verificação de evidências de validade e fidedignidade dos itens utilizados para a avaliação do

desempenho discente. Como desdobramento do problema central desta pesquisa encontram-se duas questões:

P1) Quão legítimo e justificável é que procedimentos e operações empíricas utilizados nas respostas dadas à prova objetiva de conhecimentos específicos de licenciatura em Ciências Biológicas sejam convertidos em uma representação numérica, ou score?

Por outro lado, sabe-se que o objetivo da avaliação do processo educacional não é apenas obter escores numéricos, os escores correspondem apenas a uma componente na avaliação quantitativa da performance do educando. Em psicometria, a análise dos resultados dedica-se a interpretar a realidade captada pelos instrumentos de avaliação, portanto, é dependente de como (e em consequente “para que”) o teste foi previamente estruturado, inclusive da qualidade dos parâmetros psicométricos dos seus itens. Sendo assim: **P2) Como as análises das respostas dos avaliados às situações-problema apresentadas nos testes podem resultar na identificação de outros componentes do processo de avaliação?** Dentre os resultados esperados a partir das tentativas de resposta a essas questões, pretende-se elaborar interpretações para uma escala de proficiência capazes de elucidar alguns dos motivos e obstáculos epistemológicos capazes de fornecer explicações plausíveis sobre os principais padrões de respostas e erros apresentados. Logo, esse estudo, na classificação dos *Standards* (AERA; APA; NCME, 2014), enquadra-se como pesquisa de evidências de validade pela análise do processo de resposta.

1.1 JUSTIFICATIVAS E LIMITAÇÕES

Esta pesquisa enfocará em uma das quarenta e quatro áreas que foram submetidas ao ENADE em 2017. Pode-se afirmar que foi feita uma escolha por conveniência, pois contribuíram para a decisão a formação do autor em Ciências Biológicas (bacharel e licenciado), e sua experiência como pesquisador sobre o processo recente de interiorização da educação superior pública e na elaboração de um currículo de licenciatura.

Os cursos de ciências biológicas já participaram de cinco edições do ENADE, porém justifica-se a escolha de apenas a edição mais recente (2017), pois, caso fossem realizadas paralelamente análises quantitativas para mais de

uma edição, sem a realização de testes de ligação para equalização, a comparação de escores não seria confiável, visto que as provas de anos distintos podem ter dificuldades diferentes e populações podem ter médias e distribuições diferentes em cada edição. Nesse sentido, a construção de uma escala envolvendo mais de uma edição demandaria a aplicação de um teste de ligação, procedimento inviável para o escopo desta pesquisa, considerando-se o tamanho da amostra necessária, a duração da pesquisa e os trâmites no comitê de ética.

1.2 OBJETIVOS

Nas seções a seguir estão descritos o objetivo geral e os objetivos específicos.

1.2.1 Objetivo Geral

O objetivo geral desse trabalho é avaliar a qualidade da prova de conhecimentos específicos da área de Licenciatura em Ciências Biológicas - do ENADE 2017, por meio da Teoria Clássica dos Testes, de Análise Fatorial e da Teoria de Resposta ao Item.

1.2.2 Objetivos Específicos

Para a consecução do objetivo acima descrito foi preciso desenvolver os seguintes objetivos específicos:

- I. Descrever e discutir as competências (os conteúdos) abordados nos itens de Conhecimentos Específicos de Ciências Biológicas - Licenciatura no ENADE 2017.
- II. Descrever as características da população pesquisada visando a compreender sua composição.
- III. Replicar a estimação dos parâmetros psicométricos dos itens por meio da TCT, de modo a compreender e discutir a metodologia atualmente empregada pelo ENADE.
- IV. Descrever e discutir as alternativas e os padrões de respostas dos estudantes conforme sua pontuação (TCT) de modo a aprofundar a compressão acerca das possibilidades e limitações dos atuais formatos dos itens.

- V. Avaliar a fiabilidade e dimensionalidade do instrumento do ENADE 2017 (prova de Conhecimentos Específicos de Ciências Biológicas - Licenciatura) para verificar a plausibilidade de sumarizar-se em um único valor a escala de proficiência dessas provas, conforme realizado atualmente pelo INEP.
- VI. Avaliar a adequação do ajuste do modelo selecionado aos dados empíricos.
- VII. Construir uma escala interpretativa para a dimensão dominante na prova de conhecimentos específicos (ou suas frações) por meio do modelo logístico de 3 parâmetros.

2 REVISÃO DE LITERATURA

Neste capítulo serão expostas as bases teóricas que levaram as análises realizadas nesse trabalho e uma breve revisão de estudos que contêm análises quantitativas de resultados do componente de Conhecimentos Específicos da Ciências Biológicas do ENADE 2017.

2.1 HISTÓRIA E ENQUADRAMENTO TIPOLOGICO DO ENADE

Cabe ressaltar que, ao longo de seus 15 anos de história, o ENADE passou por modificações no que concerne ao *modus operandi*, modelo teórico e objetivos pretendidos por suas provas.

Dentre as inovações mais recentes, estão o tempo mínimo de permanência do estudante na sala de aplicação da prova (por uma hora), adotado em 2013, a obrigatoriedade de resposta ao Questionário do Estudante e a publicação do Manual do Estudante, adotadas em 2014, e o curso como unidade de análise em 2015. Até 2015, a unidade de análise era a combinação de Área, IES e município, ou seja, se a IES oferecesse curso na Área em vários *campus* [sic] na mesma cidade, a nota era calculada de forma agregada. (BRASIL, 2018a)

Até o ano de 2009, a prova do ENADE era elaborada por uma empresa contratada. Conforme Campos (2013), essa empresa era responsável pela elaboração da prova, montagem, aplicação, divulgação de resultados e elaboração dos relatórios. Atualmente, tais empresas são responsáveis apenas pela aplicação e correção das provas. A partir de 2010, os itens de prova passaram a ser criados por professores do ensino superior interessados em elaborar e revisar itens, que foram selecionados em processos de chamada pública, dentre os corpos docentes dos cursos superiores avaliados pelo ENADE. Uma vez selecionados, tais professores passam a participar de capacitações e integram o Cadastro de Elaboradores e Revisores de Itens da Educação Superior (CERES)¹.

¹ Os elaboradores e revisores de itens são remunerados por esse serviço por meio do Auxílio da Avaliação Educacional (AAE). Conforme informantes, cada elaborador pode elaborar e ser remunerado por até seis itens. A seleção de integrantes do CERES obedece aos critérios

Conforme Marinho-Araujo e Rabelo (2016), desde 2010, a base para elaboração das provas do ENADE tem sido as Matrizes de Referência para avaliação do perfil e das competências.

Temporalmente, o exame tem um caráter amostral, pois desde 2004 vem sendo aplicado trienalmente a cada área de conhecimento (três grupos de áreas de conhecimento identificados por cores), mas ao mesmo tempo, desde 2009, pretende-se censitário a todos os concluintes do ano de aplicação (estudantes com mais de 80% da carga horária total cursada). Por isso a participação no ENADE é componente curricular obrigatório e requisito para a obtenção do título, sendo que os estudantes selecionados que não compareceram ao Exame ficam em situação irregular e são impedidos de colar grau. Para fornecer um panorama da dimensão do ENADE desde sua criação, estão compilados na Tabela 1 os totais obtidos com base nos microdados disponibilizados no portal do INEP:

Tabela 1: Quantitativos de estudantes, áreas IES e cursos avaliados ao longo das edições do ENADE.

Edição	Grupo	Convocados	Presentes	Abstenção	Áreas	IES	Cursos
2004	Azul	155.653	140.340	10.91%	13	554	2.279
2005	Verde	323.336	277.476	16.53%	20	980	5.898
2006	Vermelho	463.142	386.860	19.72%	15	1.600	5.896
2007	Azul	215.419	189.614	13.61%	16	753	3.454
2008	Verde	461.776	382.313	20.78%	30	1.406	8.059
2009	Vermelho	994.270	761.084	30.64%	22	1.696	7.379
2010	Azul	422.896	346.526	22.04%	19	951	4.221
2011	Verde	376.180	302.098	24.52%	84	1.389	8.814
2012	Vermelho	587.351	469.460	25.11%	17	1.646	7.228
2013	Azul	195.525	167.787	16.53%	17	903	3.678
2014	Verde	481.720	396.862	21.38%	43	1.486	9.963
2015	Vermelho	549.487	447.056	22.91%	26	1.759	8.121
2016	Azul	216.044	195.423	10.55%	18	997	4.300
2017	Verde	537.436	450.061	19.41%	44	1.478	10.571
2018	Vermelho	548.127	462.241	18.58%	27	1.762	8.813
Total	Total	6.528.362	5.375.201	21.45%	166	2.847	35.529

Fonte: O autor, com base nos microdados.

de vinculação a instituições e cursos com elevado IGC e CPC, além de cuidados com a distribuição geográfica e representatividade dos participantes, para que haja uma distribuição justa e equilibrada (não hegemônica) dentre as regiões do Brasil ou da organização acadêmica da IES a que se vinculam.

Para enquadrar o ENADE em um panorama tipológico da “pesquisa em avaliação da educação superior em larga escala”, expõe-se a seguir uma categorização presente na literatura (COHEN; SWERDLIK; STURMAN, 2014, NUSCHE, 2008).

Em uma revisão comparativa internacional sobre avaliação dos resultados educacionais do ensino superior em larga escala, a qual foi fomentada e publicada pela OCDE, Nusche (2008) distingue as iniciativas nacionais, quanto aos instrumentos empregados, em dois tipos: Questionários e Testes. Conforme a autora, os estudos que empregam questionários (*surveys*) avaliam indiretamente os resultados de aprendizado, ora por meio de autoavaliações: CEQ (Austrália), NSSE (USA); ora por questionários de acompanhamento da inserção laboral dos egressos e/ou na carreira acadêmica: DLHE (Reino Unido) GDS (Austrália).

Diferentemente dos *surveys*, aos testes de desempenho é atribuída a capacidade de avaliação direta dos resultados de aprendizado. Os testes educacionais e exames de larga escala do ensino superior (doravante chamados simplificada e de testes) tipificam-se quanto aos fins, à natureza do constructo, e à temporalidade de seus modelos teóricos. Quanto aos fins, os testes têm sido empregados para: acreditação e avaliação de cursos e sistemas de ensino, ingresso na pós-graduação, e/ou habilitação para o exercício de profissões específicas. Da finalidade do uso dos testes, surge a distinção quanto ao impacto do teste sobre os sujeitos da avaliação, sendo que quando o teste não resulta em benefícios diretos aos sujeitos, servindo apenas para avaliação e acreditação dos cursos, instituições e sistemas de ensino diz-se que é um teste *lowstakes*, e quando o impacto do desempenho incide diretamente sobre os sujeitos diz-se que é um *highstakes*.

Ocasionalmente as médias de testes admissionais da pós-graduação (GRE) ou de conselhos profissionais (*highstakes*) são utilizadas como indicadores da qualidade de cursos, mas nenhum desses testes foi desenvolvido ou é adaptado para avaliar programas educacionais, seu conteúdo é muito abrangente (não considera especialidades e ênfases), e o público que realiza tais testes não é uma amostra aleatória ou representativa. (WARREN, 2006)

Os testes também variam quanto à temporalidade dos seus modelos teóricos, assim, Cohen, Swerdlik e Sturman (2014) explicam que os testes educacionais podem ser de realização: quando visam a diagnósticos baseados no domínio de aprendizados formais/curricular, os quais estão formalizados em uma matriz de referência socialmente legitimada (bases e diretrizes curriculares); ou podem ser de aptidão: quando visam a prognósticos da aptidão dos examinandos baseados em aprendizados informais/tácitos, habilidades, atitudes e maturação psicossocial; geralmente selecionam para ingresso ou desempenho em uma experiência/nível educacional futuro. Em classificação similar, Nusche (2008) distingue os testes quanto à natureza do constructo avaliado, podendo visar a aspectos cognitivos (domínio de conteúdos), e/ou a aspectos não cognitivos (competências, atitudes e valores). Dentre os aspectos cognitivos mais empregados nos testes de larga escala está o domínio de áreas específicas e/ou também conhecimento gerais, supostamente comuns às diversas carreiras superiores. Também os resultados de aprendizado não cognitivos podem ser classificados em genéricos ou específicos de determinadas formações. Conforme Nusche (2008), tais aspectos não cognitivos devem ser avaliados por meio de aplicação, em vez de recuperação/memorização do conhecimento, pois assim enfocam na habilidade dos sujeitos ao resolver problemas.

A partir dessa tipologia, podemos enquadrar o ENADE, tema da presente dissertação, como um instrumento de mensuração direta dos resultados de aprendizado, com fins em acreditação e avaliação de cursos e sistemas de ensino, e portanto, como não resulta em benefícios ou punições diretas aos sujeitos da avaliação², poderia ser enquadrado como *lowstake* ao nível dos indivíduos. Sendo ele um teste de realização (i.e. com enfoque diagnóstico sobre o aprendizado passado), e segue uma linha de avaliação com enfoque em aspectos cognitivos do aprendizado (domínio de campo)

² Cabe fazer a ressalva de que o parágrafo 10º do artigo 5º da Lei 10.861 prevê a concessão “estímulos” aos estudantes com melhor desempenho: “§ 10. Aos estudantes de melhor desempenho no ENADE o Ministério da Educação concederá estímulo, na forma de bolsa de estudos, ou auxílio específico, ou ainda alguma outra forma de distinção com objetivo similar, destinado a favorecer a excelência e a continuidade dos estudos, em nível de graduação ou de pós-graduação, conforme estabelecido em regulamento.”

gerais e específicos, e não cognitivos (habilidades), e como o ENADE segue a corrente teórica da avaliação de competências, propõe que mais de um traço seja avaliado concomitantemente em cada item. Nesse sentido, pode ficar obscuro avaliar quando os erros dos avaliados em uma questão ocorreram por falta da habilidade, de uma atitude ou do domínio cognitivo do campo. Primi, Hutz e Silva (2011) e Silva (2011) ressaltam tal paradoxo em relação a esse ponto, uma vez que as comissões de especialistas de área produzem definições multidimensionais do que deve ser avaliado nas provas, enquanto que os especialistas em medida precisam construir provas unidimensionais para poder responder aos problemas de medida que enfrentam.

2.2 TESTES DE LARGA ESCALA NO CONTEXTO INTERNACIONAL

Nusche (2008) descreve que no México existem dois exames de larga escala, um aplicado aos bacharelados (EGEL) e outro aplicado a cursos tecnológicos e de engenharia (EGETSU) que assim como o ENADE avaliam majoritariamente conhecimentos disciplinares (de domínio específico de 39 e 36 áreas, respectivamente, cf. ZLATKIN-TROITSCHANSKAIA; SHAVELSON; KUHN 2015), porém o EGETSU ainda dispõe de uma seção que avalia habilidades genéticas (como inglês e TI). Nos Estados Unidos o *Major Field Test* possui essa estrutura com dois constructos (um generalista e um especializado), já o *Collegiate Learning Assessment* (CLA) avalia a performance em tarefas construídas sobre temáticas disciplinares amplas (Ciências naturais, sociais, humanidades e artes). Outros testes como o *Graduate Skills Assessment* (GSA) e o *Course Experience Questionnaire* (CEQ) da Austrália interessam-se por habilidades genéricas como o pensamento crítico, capacidade de resolução de problemas e escrita. Também alguns testes de ingresso à pós-graduação como o EXANI-III (México) e o GRE (USA) examinam em busca de competências genéticas como raciocínio verbal e matemático e capacidade de tratamento de informações. De todos esses testes, apenas o ENADE é censitário e obrigatório para indivíduos e cursos/IES (em ciclos trienais aplicados alternadamente a três conjuntos de áreas de conhecimento).

Segundo Zlatkin-Troitschanskaia, Shavelson e Kuhn (2015), a AHELO é a única avaliação internacional de ensino superior que apresenta uma amostra representativa em nível nacional e possibilita comparabilidade internacional no nível das IES.

Para licenciados, há nos EUA o *Test of Teaching Knowledge*, o qual avalia conhecimento de conteúdos e conhecimentos pedagógicos de professores recém-formados de várias disciplinas, incluindo matemática e ciências (ZLATKIN-TROITSCHANSKAIA; SHAVELSON; KUHN, 2015).

Conforme Zlatkin-Troitschanskaia, Shavelson e Kuhn (2015), na Alemanha, um programa de pesquisa nacional intitulado “Modelando e Medindo Competências na Educação Superior” (KoKoHs) iniciou em 2011. O programa compreende 70 projetos da Alemanha e da Áustria que modela e avalia competências genéricas e específicas em várias áreas, incluindo ciência da educação, formação de professores, psicologia, ciências sociais, engenharia, administração e economia. Segundo Zlatkin-Troitschanskaia, Shavelson e Kuhn (2015) o programa KoKoHs leva em conta requisitos curriculares e laborais, transforma modelos teóricos de competências em instrumentos de mensuração e valida a interpretação dos escores de testes.

Quadro 1: Testes de larga escala em âmbito internacional.

Teste	País (Instituição desenvolvedora)	Período de Existência e Público	Constructos (declarados)	Questões/ Duração e Referência e critérios	Tamanho	Periodicidade e Estimações de valor agregado
<i>Collegiate Assessment of Academic Proficiency</i> (CAAP)	Estados Unidos (<i>American College Testing Program's - ACT</i>)	1988	Habilidades gerais e Conhecimentos gerais (comuns a todos os currículos) Leitura, Pensamento crítico Escrita, Ciências e Matemática. ³	Objetivas e Discursivas (>72 por módulo) IES podem escolher 6 módulos de distintas habilidades para compor o teste/ (40 minutos por módulo)	Entre 1988 e 2001 foi usado por mais de 600 IES e 450.000 sujeitos	Em fluxo contínuo sob demanda. Não tem datas e periodicidade estabelecidas, e conta com testes com múltiplas formas equivalentes. Os estudos de valor agregado que empregam esse teste podem assumir desenhos transversais e horizontais e até ser usados para monitoramento com aplicações intensivas e periódicas.
<i>Collegiate Learning Assessment</i> (CLA)	Estados Unidos (<i>Council for Aid to Education- CAE</i>)	2002	Conhecimentos gerais e específicos Algumas competências são avaliadas por meio de tarefas (análise de políticas, redação de documentos...) com conteúdo pertinentes a grandes áreas (ciências naturais, sociais, humanidades e artes)	Discursivas /90 minutos	134 colleges	Anual, conta com testes com múltiplas formas equivalentes os estudos de valor agregado que empregam esse teste podem assumir desenhos transversais e horizontais e até serem empregados para monitoramento com aplicação intensiva (não apenas pré e pós-teste)

Continua na próxima página

³ Para mais informações sobre CAAP, CLA e MAPP ver Klein, et al. (2009)

Continuação do Quadro 1

Teste	País (Instituição desenvolvedora)	Período de Existência	Constructos (declarados)	Questões/ Duração e Referência a critérios	Tamanho	Periodicidade e Estimações de valor agregado
Exame Nacional de Cursos (ENC)	Brasil	1995-2003	Conhecimentos específicos de até 26 áreas	Objetivas/ 4 horas	460.000 sujeitos 6,500 cursos	Anual (Apenas para concluintes, não possibilita estimativas de valor agregado)
Exame Nacional de Desempenho dos Estudantes (ENADE)	Brasil (INEP + Cesgranrio ou Fundação Carlos Chagas (FCC) ou Centro de Seleção e Promoção de Eventos (Cespe-UnB) até 2009, a partir de 2010: INEP (BNIES))	2004-presente (Amostral até 2008 Ingressantes até 2011) Público: Concluintes (>80%CH)	Conhecimentos Específicos de até 52 áreas* Conhecimentos gerais (temas atuais sobre questões ambientais, sociais, de cidadania, políticas públicas) Habilidades genéticas e específicas*.	Objetivas (35[8 gerais, 27 específicas]) e Discursivas (5) / 4 horas	~5.4 bilhões de sujeitos em 15 anos (med. anual ~ 350.000) 2.184 IES 35.529 cursos 166 áreas	Trienal para cada grupo/ciclo que alternadamente (em rodízio) são convocados (compulsoriamente) a aplicações anuais. Desenho transversal até 2011. Atualmente utiliza medidas de constructos diferentes (ENEM X ENADE) para conjecturar valor agregado.
<i>Exámen General Para el Egreso de la Licenciatura (EGEL)</i>	México (CENEVAL)	1994-presente	Conhecimentos específicos e Habilidades específicas	Objetivas(100 a 300 dependendo da área)/ (8 a 10 horas)/Referenciado em critérios	200.000 anuais	Adesão voluntária de estudantes e IES. Anualmente Aplicação única (não possibilita estudos de valor agregado)

Continua na próxima página

Continuação do Quadro 1

Teste	País (Instituição desenvolvedora)	Período de Existência	Constructos (declarados)	Questões/ Duração e Referência a critérios	Tamanho	Periodicidade e Estimações de valor agregado
<i>Exámen Nacional de Ingreso al Posgrado (EXANI-III)</i>	México (<i>Centro Nacional de Evaluación para la Educación Superior - CENEVAL</i>)	1997-presente	Habilidades genéricas *Seleção (Foco nos resultados individuais).	Objetivas (120) / (4 horas)/	13.604	Adesão voluntária de estudantes e IES. Anualmente Aplicação única (não possibilita estudos de valor agregado)
<i>Exámenes Generales para el Egreso del Técnico Superior Universitario (EGETSU)</i>	México (CENEVAL)	2000-presente (mudou para EGATSU?)	Conhecimentos específicos e generalistas (questões sociais e de economia, TI e inglês) Habilidades específicas e generalistas	Objetivas(250[87 Gerais, 163 específicas])/ (6 horas)/Referenciado em critérios	21.000 sujeitos 48 IES	Adesão voluntária de IES. Anualmente Aplicação única (não possibilita estudos de valor agregado)
<i>Graduate Record Examination (GRE)</i>	Estados Unidos, Canadá e outros (ETS)	1966-presente Público: ingressantes na pós-graduação	Habilidades genéricas (raciocínios verbal e quantitativo e escrita analítica) *Objetiva seleção (Foco nos resultados individuais).	Mais de 3 horas		Em fluxo contínuo sob demanda. Aplicação única (não possibilita estudos de valor agregado)
<i>Graduate Skills Assessment (GSA)</i>	Australia (<i>Australian Council for Educational Research -ACER</i>)	2000-presente	Habilidades genéricas	Objetivas e Discursivas (2)/ 3 horas	Aprox. 2000 sujeitos Aprox. 20 IES	2 vezes ao ano (início e fim do ano letivo) Transversal Modelo de Rasch

Continua na próxima página

Continuação do Quadro 1

Teste	País (Instituição desenvolvedora)	Período de Existência	Constructos (declarados)	Questões/ Duração e Referência a critérios	Tamanho	Periodicidade e Estimações de valor agregado
Modelando e Medindo Competências na Educação Superior (KoKoHs)	Alemanha e Áustria (<i>German Federal Ministry of Education and Research - BMBF</i>)	1ª fase 2011-2015 2ª fase 2016- Presente	Competências genéticas (autodisciplina e competências gerais de pesquisa) e específicas das áreas de: engenharia, ciências sociais e econômicas, ciências da educação (e psicologia) e licenciaturas em ciências.	Objetivas (150)	650 (piloto) 3.500 sujeitos de licenciaturas	Projetos de pesquisa avaliam estudantes de diferentes cursos e fases
<i>Major Field Tests</i>	Estados Unidos (ETS)	1990	Conhecimentos e habilidades específicos (para graduação e MBAs). *Seleção (Foco nos resultados individuais).	Objetivas /2 horas graduação 3 horas para MBAs	Quase 70 mil sujeitos 606 IES	Em fluxo contínuo sob demanda. Aplicação única (não possibilita estudos de valor agregado)
<i>Malaysian Generic Skills Instrument (MyGSI)</i>	Malásia (ACER My)		Conhecimentos disciplinares, habilidades práticas, habilidades e responsabilidades sociais, valores, atitudes e profissionalismo, comunicação e liderança no trabalho, Pensamentos crítico, resolução de problemas e habilidades científicas, gestão de informações e habilidades de aprendizado ao longo da vida, habilidades gerenciais e empreendedoras.	13 constructos (Baseados no <i>Malaysian Qualification Framework - MQF</i>) 102 itens	1.262	

Continua na próxima página

Continuação do Quadro 1

Teste	País (Instituição desenvolvedora)	Período de Existência	Constructos (declarados)	Questões/ Duração e Referência a critérios	Tamanho	Periodicidade e Estimações de valor agregado
<i>Measure of Academic Proficiency and Progress</i> (MAPP)	Estados Unidos (<i>Educational Test Service-ETS</i>)	2006-presente Sua história remonta ao " <i>Academic Profile</i> " (1992-2006) ao qual veio substituir	Habilidades genéricas e específicas (Além de escrita e matemática, a leitura e pensamento crítico são respondidos no contexto de grandes áreas – humanidades, ciências sociais ou ciências naturais)	Objetivas e Discursivas (2 formas longas [108 questões], 6 curtas [36]) / duas horas ou 40 minutos	375 IES 1 milhão de sujeitos	Idem CAAP
SABER PRO (Ensino Superior) SABER TyT (cursos técnicos e tecnológicos)	Colômbia (<i>Instituto Colombiano para el Fomento de la Educación Superior - ICFES</i>)	Público: Concluintes (>75%CH)	1ª fase: 5 módulos gerais: Raciocínio quantitativo (RQ), Leitura Crítica (CR), Competências Cidadãs (CC), Inglês, Comunicação Escrita e na 2ª fase do teste: 1 módulo de competências específicas (40 áreas)	Objetivas (160 [35RQ+35LC+35CC+55Ing]) 1 Aberta (comunicação escrita) /4 horas e 40 minutos	106.189 em 2012	Anual Transversal? (Saber 11 X Saber PRO) Pontuação pelo ML2P

Fonte: O autor, adaptado de (NUSCHE, 2008)

2.3 MODELOS TEÓRICOS E DEFINIÇÃO DO CONSTRUCTO

Embora o ENADE seja um teste que avalia o desempenho dos estudantes, sua finalidade é fornecer uma estimativa de qualidade dos cursos de graduação e sua contribuição ao aprendizado dos estudantes. Conforme Primi, Da Silva e Bartholomeu (2018, p. 131), nesse modelo de avaliação há duas suposições implícitas: 1) que qualquer processo pedagógico executado ao nível do curso impactará em todos os estudantes a ele vinculados; e 2) que os “cursos com diferentes níveis de qualidade terão impacto diferenciado, sendo que cursos de alta qualidade provocarão desenvolvimento em nível mais elevado do que cursos de baixa qualidade” (*ibidem*). Em síntese: Presume-se uma relação direta entre a qualidade do curso e o desempenho dos estudantes. “Porém, existem muitos questionamentos sobre se o impacto de cursos com diferenciados níveis de qualidade é substancialmente forte e nítido a tal ponto de ser possível avaliar a qualidade de um curso a partir da aprendizagem dos alunos”(*ibidem*).

De acordo com Da Mata (2016, p.92), para elaboração de um instrumento de avaliação de aprendizagem, faz-se necessário que a aprendizagem seja primeiramente definida por algum fundamento científico. Entretanto, muitos dos organizadores das avaliações do ensino superior não demonstraram até o presente terem tido tal cuidado. Paraphraseando Da Mata (2016), ao investigar-se a natureza da aprendizagem e quais os fatores envolvidos no processo de ensino-aprendizagem, também se desenvolvem oportunidades de avaliar tais processos, mas é necessário partir de algum modelo teórico que fundamente o constructo.

Na literatura internacional, encontra-se no *thesaurus* da base de dados Education Resources Information Center (ERIC mantida pelo Departamento de Educação dos Estados Unidos) que o termo “resultados de aprendizado” (*learning outcomes*) é o mais relacionado à avaliação educacional por testes de larga-escala no Ensino Superior. Conforme Nusche (2008), o termo tem sua origem na “educação

baseada em resultados”, um modelo educacional estruturado na identificação clara e explícita de resultados, para definir e avaliar a aprendizagem do estudante.⁴

Conforme Nusche (2008), vários autores definem resultados de aprendizagem como algo que pode ser observado, demonstrado e mensurado e define resultados de aprendizagem (*learning outcomes*) como “aquilo que um aprendiz sabe ou pode fazer como resultado de um aprendizado” (OTTER, 1992 *apud* NUSCHE, 2008). Uma definição operacional semelhante é fornecida por Cumming e Ewell (2017) e Eisner (1979, *apud* NUSCHE, 2008), que afirmam que resultados de aprendizagem são essencialmente o nível de conhecimento, habilidade e competências que alguém carrega consigo, intencionalmente ou não, após ter se comprometido em alguma situação de ensino-aprendizado.

Da Mata (2016, p.29), buscando referência em autores da psicologia da aprendizagem explícita uma distinção fundamental ao afirmar que a aprendizagem gera consequências em duas direções: a primeira refere-se a um processo interno do organismo e a segunda está relacionada a um processo externo denominado de resposta emitida que pode ser observada e medida. Alguns teóricos defendem o emprego do termo aprendizagem apenas a essas mudanças internas e em relação às respostas emitidas, o mais adequado seria a utilização do termo desempenho.

⁴ Na cultura acadêmica estadunidense, a média das notas (GPA - *grade point average*) tem sua forma de cálculo uniformizada e costuma ser usada para comparações institucionais. Na revisão sistemática de YORK; GIBSON; RANKIN (2015) o GPA lidera a lista das medidas de sucesso acadêmico, respondendo por 54,8% das publicações sobre o assunto. Estudos que utilizam mensurações de pensamento crítico e retenção são as segundas mais prevalentes com 19,4%, e habilidades acadêmicas e engajamento vêm em terceiro com 16,1% das publicações.

Entretanto, YORK; GIBSON; RANKIN (2015) afirmam que o GPA é problemático para pesquisadores e gestores porque carece de padronização entre as instituições, acurácia, evidências de validade, estabilidade, e porque os autores, que referindo-se ao GPA, ao utilizar o termo sucesso acadêmico, omitem na verdade que falam de uma restrita porção do fenômeno, desenvolvendo resultados não generalizáveis. Na mesma linha, Warren (2006) afirma que o GPA é um composto de avaliações baseadas em comparações de notas entre estudantes e, portanto, é um constructo não referenciado em critérios. Considerando o exposto, o autor deduz que o GPA não traz informações sobre o que foi aprendido, exemplificando com o fato notório que estudantes de ciências exatas apresentam menor GPA em comparação com as ciências humanas, mas, uma vez que não há razões para acreditar que estudantes dessas duas áreas sejam inerentemente diferentes (uns menos ou mais proficientes que os outros). Por esse motivo, o GPA também não possibilita comparações entre cursos.

Em suma, “a aprendizagem é um evento interno, não observável, inferido através do desempenho” (*Ibidem*).

Pode-se deduzir, em consonância com o Artigo 5º da Lei n.º 10.861 de 14 de abril de 2004, que o desempenho é o objeto de mensuração do ENADE:

Art. 5º A avaliação do **desempenho** dos estudantes dos cursos de graduação será realizada mediante aplicação do Exame Nacional de Desempenho dos Estudantes - ENADE.

§ 1º O ENADE aferirá o **desempenho** dos estudantes em relação aos conteúdos programáticos previstos nas diretrizes curriculares do respectivo curso de graduação, suas **habilidades** para ajustamento às exigências decorrentes da evolução do conhecimento e suas **competências** para compreender temas exteriores ao âmbito específico de sua profissão, ligados à realidade brasileira e mundial e a outras áreas do conhecimento. (BRASIL, 2004)

Cabe explicitar ao leitor que o tratamento dado à fundamentação teórica do ENADE nessa seção visará a descrever conforme descrito em trabalhos de referência como ela está sendo operacionalizada pelo INEP, e não apontar a visão dos autores de como deveria ser. Essa revisão foi elaborada buscando descrever os trabalhos basilares (possivelmente) utilizados como referência teórica para elaboração dos modelos teóricos do teste, bem como através de comentadores que tenham citado tais trabalhos basilares ao analisarem empiricamente resultados do ENADE.

Isto posto, foram encontrados dois trabalhos basilares que tratam dos modelos teóricos adotados no ENADE: o primeiro é Limana e de Brito (2005), o segundo é Marinho-Araujo e Rabelo (2015) esses últimos trabalhos foram textualmente mencionados no Relatório de área de 2017 (BRASIL, 2018a), mas não nos anteriores.

Conforme da Mata (2016, p. 96) “o modelo de avaliação concebido e empregado no ENADE é o modelo de avaliação dinâmica das competências e do potencial de aprendizagem descrito por Limana e De Brito (2005)”. Porém não se encontrou menções a tal modelo teórico em textos oficiais do INEP, havendo apenas evidências indiretas de que ele tenha constituído uma das fundamentações iniciais do ENADE. Limana e de Brito dão a entender que “A prova usa o modelo de avaliação dinâmica [...]”(idem, 2005, p.23). Além disso o autor Amir Limana foi

Coordenador-Geral do ENADE junto ao INEP/MEC, desde a sua implementação até fevereiro de 2008; e a autora “Márcia Regina Ferreira de Brito Dias” foi consultora no INEP de 2004 a 2008 junto à Diretoria de Avaliação da Educação Superior, e compôs a Comissão Assessora de Avaliação da Formação Geral dos ENADE 2005 até 2013.

Eles sintetizam sua proposta da seguinte forma:

[...]o Modelo de Avaliação Dinâmica do Potencial de Domínio de Área (MADPDA) busca avaliar as mudanças decorrentes da passagem dos estudantes da Educação Superior pelas IES onde buscam a formação. É um modelo que busca avaliar o processo e não o produto, sendo essa avaliação basicamente, uma situação de teste-intervenção-reteste. (LIMANA; DE BRITO, 2005, p.23)

Sumarizando, Limana e de Brito (2005) buscam fundamentar as escolhas metodológicas do SINAES, norteados pelo objetivo de “verificar o efeito do curso sobre o aprendizado do estudante” e para tanto investem na explicitação de um quadro teórico que tem como peça central o conceito de Avaliação Dinâmica elaborado em 2002 por Sternberg e Grigorenko. O modelo de avaliação dinâmica é baseado em abordagens desenvolvidas a partir da oposição aos modelos da psicometria clássica, tem uma longa história que pode remeter aos escritos de Alfred Binet (1857-1911), e conforme Da Mata (2016), também tem por base a teoria da modificação cognitiva de Feurstein, avaliação do potencial de aprendizagem de Budoff e a avaliação dinâmica do efeito acadêmico (ver: STERNBERG; GRIGORENKO, 2002) apoiada na teoria triádica de inteligência (capacidades analíticas, criativas e práticas. Ver: STERNBERG, 1985), além da teoria do domínio de área (*Mastery Learning* de BLOOM, 1968).

Logo, a partir do modelo explicitado por Limana e de Brito (2005), é possível concluir que para esses autores o objetivo primordial do ENADE seria verificar o valor agregado pela passagem dos estudantes pelos cursos de graduação, conforme pode ser notado nas frases: “a prova precisa ser formulada de forma a permitir comparabilidade, tendo em vista a intenção de se medir a mudança” (LIMANA; De BRITO, 2005, p. 19) e “Mesmo que apenas uma pequena parcela de estudantes seja a mesma que já realizou o exame quando ingressante, será possível extrair indicadores de ganho, que é um conceito central no novo modelo”

(op. cit., p.27)⁵. Entretanto, tais estudos de valor agregado possivelmente têm validade limitada, pois não se encontrou na literatura nada que aponte a condução pelo INEP de estudos com a utilização da equalização de provas de diferentes ciclos, nem com a TCT, nem através da TRI, portanto tais estudos de valor agregado estão apenas restritos aos de regressão hierárquica que tem sido empregada no cálculo do IDD, e que faz uso dos escores brutos (da TCT) do ENADE e proficiência do ENEM (TRI), portanto ainda não há como ter a certeza de que os resultados das diferentes provas são comparáveis em uma mesma escala.

Segundo Da Mata (2016), dentre os quatro enfoques teóricos empregados por Limana e de Brito (2005) para propor o modelo de avaliação dinâmica supostamente empregado pelo ENADE (até 2010), o que mais relaciona as concepções de aprendizagem com as concepções sobre avaliação é a chamada teoria do domínio de área, desenvolvida por Benjamin Bloom (1913 - 1999) com base nas concepções teóricas de John Carroll (1916 - 2003). Uma breve digressão sobre a teoria do domínio de um aprendizado faz-se oportuna, pois sua fundamentação teórica possui forte aderência e assemelha-se com os pressupostos da psicometria moderna. Em seu primeiro postulado (citando Carroll, 1963) o autor lança a suposição de que se a aptidão para uma determinada disciplina está normalmente distribuída em uma população, e se a todos os estudantes são providas as mesmas instruções o resultado final será uma distribuição normal em uma mensuração de desempenho apropriada, e a correlação entre a aptidão (prévia) e o desempenho será relativamente alta. Por outro lado, se a mesma distribuição de aptidão é tratada com o tipo e qualidade de instrução, e com o tempo apropriado para as características e necessidades de cada estudante, o esperado é que grande parte dos estudantes atinjam o domínio do aprendizado, reduzindo as correlações

⁵ Observa-se algumas inconsistências nas conjecturas das possibilidades do ENADE, as quais parecem decorrentes de desconhecimento dos autores acerca das limitações metodológicas inerentes à equalização (pois mesmo que haja no INEP a disponibilidade de “dados referentes ao ENEM e resultados dos vestibulares” sobre a qual nos falam os autores, dificilmente serão comparáveis os instrumentos dos vestibulares com ENADE), e amostragem aleatória (pois raros e de proficiências incomuns são os concluintes em 3 anos de graduação). Ver também as considerações sobre a parte comum da prova que trata do conhecimento tácito que o indivíduo necessita para ser bem sucedido, mas que não é ensinada (op. cit., p.27).

entre aptidão e desempenho a níveis próximos de zero. Então, Bloom (1968), citando Carroll (1963), postula que se estudantes com alta aptidão podem compreender ideias complexas, enquanto que, ao mesmo tempo os estudantes com baixa aptidão podem aprender apenas as ideias mais simples; nessa lógica a aptidão seria inversamente proporcional ao montante de tempo necessário para que um estudante completasse uma tarefa relacionada ao domínio do aprendizado de uma matéria. Nesse postulado de que a “aptidão é um preditor da taxa de aprendizagem e não propriamente do nível (ou complexidade) do aprendizado” está implícita a hipótese de que: não havendo limitações intelectuais, concedido o tempo necessário e auxílio do tipo apropriado, cerca de 90% dos estudantes de uma sala podem dominar o aprendizado de uma área. Assume-se que para alguns estudantes isso demandará mais tempo, esforço e ajuda que para outros, e, em alguns casos, essas demandas de tempo e esforço podem tornar algumas tarefas de aprendizado proibitivas.

Carroll e Spearritt (1967) também postulam sobre a qualidade da instrução como sendo o grau em que a apresentação, explicação e organização dos elementos de um aprendizado aproximam o aprendizado ao ótimo para um determinado indivíduo. Em termos genéricos, Carroll e Bloom defendem que a qualidade da instrução deve ser avaliada quanto a seu “efeito ao nível individual” e não quanto ao resultado em nível de grupo (conjuntos aleatórios de estudantes). E propõe agrupamentos de classes em estilos de aprendizado, uma vez que é improvável que a instrução ocorra individualmente para cada estudante. Por fim, empregando desenvolvimentos teóricos criados em 1965 por Gagné, e elementos de sua própria taxonomia (i.e. Bloom, 1956), Bloom (1968) defende a estruturação de uma hierarquia de complexidade nas tarefas de aprendizagem, sendo também cumulativa a organização de tais tarefas, partindo da compreensão e memorização de fatos, passando por conceitos complexos e abstratos até a aplicação de princípios e análises de proposições teóricas completas. Essa última proposta encontra equivalente teórico na unidimensionalidade dos constructos da teoria psicométrica e na monotonicidade crescente das curvas características dos itens, o que se relaciona com uma visão cumulativa do processo de aprendizagem.

Para compreensão de como a concepção avaliativa do ENADE vinculou-se ao modelo de competências, Marinho-Araujo e Rabelo (2015) apontam uma explicação histórica (intertextual), visto que as Diretrizes Curriculares Nacionais (DCN) da maioria dos cursos foram promulgadas em 2004 tendo como meta orientadora o “desenvolvimento de competências vinculado a conhecimentos científicos, saberes e práticas próprias da área profissional”. Sobre esse contexto histórico, o Fórum de Pró-Reitores de Graduação das Universidades Brasileiras fornece elementos para a compreensão de como a categoria das Competências foi inserida em todas as DCN dos cursos de graduação:

Cabe ressaltar que tanto as IES ao elaborarem suas propostas de diretrizes curriculares como o GT seguiram as determinações expressas no Edital n.º 4/97 da SESu/MEC. Este Edital define o conjunto de elementos constitutivos que as diretrizes curriculares deveriam contemplar: perfil, competências e habilidades, conteúdos curriculares, duração dos cursos, estrutura modular, estágios e atividades complementares e conexão com a avaliação institucional. A apresentação desses elementos no Edital, ao mesmo tempo que procurou ordenar o debate nacional, também pode ter contribuído para limitar a criatividade e a flexibilidade necessárias nesse processo de construção coletiva. (ForGRAD, 2000)

A categoria “competências” já estava presente nas orientações contidas no Parecer n.º 776/97/CES/CNE, de 03 de dezembro de 1997 (BRASIL, 1997), (“Orientação para as diretrizes curriculares dos cursos de graduação”), porém no contexto frasal de “reconhecimento de competências adquiridas fora do ambiente escolar”.

Sobre a adoção dessa categoria no ENADE, fato digno de menção é a alteração das expressões utilizadas nos Relatórios de área do INEP que até 2016 mencionava, especificamente, nos anexos “Concepção e elaboração das provas do ENADE”, uma categorização das diretrizes de prova em “características de perfil, recurso e objetos de conhecimento”, e em 2017 passa a empregar nova nomenclatura “Perfil, Competência e Conteúdo” conforme pode constatar-se na Figura 1:

Figura 1: Definição das dimensões teóricas empregadas pelo INEP para a construção da Matriz de Referência Tridimensional. Acima o quadro conforme adotado até 2014 e abaixo o novo referencial embasado em Marinho-Araujo e Rabelo (2015)

CARACTERÍSTICA(S) DE PERFIL Conjunto de características do egresso do curso.	
CONJUNTO DE RECURSOS Uma expansão do termo competências, que compreende a mobilização de conhecimentos, saberes, escolhas éticas e estéticas, habilidades, posturas, entre outros, para permitir agir eficazmente em um determinado tipo de situação, apoiado em conhecimentos, mas sem limitar-se a eles.	
OBJETO(S) DE CONHECIMENTO São os conteúdos que devem ser mobilizados por meio dos recursos (competências e habilidades) pelo profissional dotado do perfil esperado. Algumas vezes o item solicita a utilização de dois ou três objetos de conhecimento. Neste caso, o item deve ser capaz de articular todos os conteúdos.	
PERFIL	Conjunto de características esperadas do egresso da Educação Superior, construído na articulação entre uma base teórica e uma prática real, e que contempla a identidade pessoal e a identidade profissional.
COMPETÊNCIA	Mobilização reflexiva e intencional de diferentes recursos (conhecimento, saberes, habilidades, esquemas mentais, afetos, crenças, princípios, funções psicológicas, posturas e outros) necessários para o enfrentamento de uma situação-problema específica.
CONTEÚDO	Conteúdos curriculares estabelecidos pelas Diretrizes Curriculares Nacionais dos cursos de graduação ou pelo Catálogo Nacional de Cursos Superiores de Tecnologia.

Fonte: BRASIL (2016, 2018a)

Para Marinho-Araujo e Rabelo (2015), competência é compreendida como a **mobilização** reflexiva e intencional de diferentes recursos (conhecimentos, saberes, habilidades, esquemas mentais, afetos, crenças, princípios e outros) os quais são necessários para o enfrentamento de uma situação-problema específica. Mas advertem que as competências não podem ser definidas somente em termos de prontidão para o desempenho de uma ação, mas antes, exigem “dos sujeitos a identificação dos conhecimentos pertinentes e a mobilização autônoma de outros recursos, configurados em uma postura ativa diante de desafios ou problemas contextualizados”. E distinguem (com base em Wittorski) entre competências individuais e coletivas (com o objetivo de colaboração na interação interpessoal), e entre competências técnicas e transversais.

Esses recursos técnicos (conceituais, teóricos, metodológicos, práticos) e transversais (pessoais, interpessoais, éticos, estéticos) deveriam ser mediados intencionalmente na educação superior para o enfrentamento de cenários profissionais ou situações de trabalho. A ampliação, integração e complementação desses recursos, considerando sua transversalidade em situações análogas, porém não idênticas, e o **valor de uso que eles possam apresentar em determinada ação**, constituem-se em indicadores individuais ou coletivos de competências pessoais [...] (MARINHO-ARAÚJO; RABELO, 2015, p.13)

Já Zlatkin-Troitschanskaia, Shavelson e Kuhn (2015) acrescentam à noção de competência outros componentes (todos latentes): cognitivos (conhecimentos e habilidades), metacognitivos (autorregulação) e não cognitivos (motivação, afetivos, volitivos e disposições sociais), além de incluir performances observáveis pela realização de tarefas à luz de critérios que variam com os contextos e situações.

Sobre a avaliação de aspectos não cognitivos, cabe ponderar que Pike (1989) constata uma alta correlação entre a idade e a performance em testes de habilidades genéticas (ACT-COMP), e sugere que tais aprendizados estão mais relacionados à maturação psicossocial do que à proficiência e portanto, o provável relacionamento da maturação com fatores externos à universidade (acontecimentos da vida) pode impedir o controle estatístico necessário à avaliação do impacto das IES.

Em via de síntese, Nusche (2008) menciona modelos teóricos sobre a mensuração de resultados de aprendizado que partem de críticas à distinção entre aprendizados cognitivos e não cognitivos e se ela é apropriada para captar e discutir os resultados obtidos na experiência real de mensuração dos resultados de aprendizagem. E em substituição, Rycher (2004 *apud* NUSCHE, 2008) propõe a adoção da categoria “competências” que [supostamente] contemplaria e transcenderia tal distinção.

Citando Deluiz (2001), Marinho-Araujo e Rabelo (2015, p.18) explicam que os modelos deram origem a diversas concepções orientadoras para a identificação, construção e avaliação de competências: Condutivista/behaviorista; funcionalista; construtivista e crítico-emancipatória. Conforme Marinho-Araujo e Rabelo (2015), sobre as matrizes condutivista e funcionalista pesam críticas à sua submissão à lógica do mercado, e de que estariam limitando-se à descrição de funções e tarefas dos processos produtivos. A Construtivista revela a dimensão processual coletiva e contextual na investigação das competências, mas não enfatiza o papel do contexto social, minimizando sua dimensão política. A crítico-emancipatória concebe a noção e competência como multidimensional, envolvendo facetas que vão do individual ao sociocultural, situacional e processual e entende que o desenvolvimento de competências pelo sujeito passa por um processo sociocultural, pela construção da

consciência crítica e do compromisso com a transformação social, articulando as dimensões educacional, profissional e ético-política.

A norma que regeu as questões logísticas do ENADE de 2017, teste objeto desse estudo, foi a Portaria Normativa n.º 8/MEC/2017, de 26 de abril de 2017, e o Edital n.º 26/2017/INEP de 16 de junho de 2017 que determinaram: A prova daquele ano deveria ser aplicada no dia 26 de novembro de 2017 e ter duração mínima de 1 hora e meia (para saída da sala de aplicação) e máxima de 4 horas, que seria composta por um total de 40 questões, sendo 10 questões de formação geral (2 discursivas) idênticas para todas as áreas e 30 questões de componente específico, ou seja, relativas a cada área (curso). As questões objetivas apresentam cinco alternativas, devendo haver apenas uma correta.

No atual processo de formulação do ENADE, é designada uma Comissão Assessora da Área (CAA) que tem suas atribuições expressas em sua portaria de designação, dentre elas, criar as recomendações, em termos de diretrizes, objetivos e especificações do exame, as quais serão necessárias para os integrantes do CERES elaborarem e revisarem a prova. Conforme Campos, dentre as atribuições da CAA estão:

“a elaboração das diretrizes e matrizes de prova para a avaliação, acompanhamento do processo de capacitação de elaboradores e revisores técnico-pedagógicos de itens, acompanhamento da revisão técnico-pedagógica de itens elaborados para o BNI-ENADE, indicação da homologação dos itens que integrarão o BNI e análise após a aplicação do ENADE do gabarito preliminar dos itens de múltipla-escolha e os padrões de respostas dos itens discursivos.” (CAMPOS, 2013, p.31)

Inicialmente, a comissão assessora discute as diretrizes para elaboração da prova, subsidiados pelas Diretrizes Curriculares Nacionais dos cursos de graduação. Isso posto, é publicada uma portaria com as especificações gerais da prova, como número e formato das questões, duração, e explicitação das próprias diretrizes, que nessa edição foram categorizadas em “características do perfil do concluinte” no Art. 5º, “competências desenvolvidas” no Art.6º, e “conteúdos” no Art.7º (BRASIL, 2017d). As diretrizes para o componente de Formação Geral são publicadas em outra portaria específica. Na seção 2.4 será apresentado um panorama comparativo entre as Diretrizes Curriculares Nacionais e as Diretrizes de Avaliação.

Em 2017, pela primeira vez foram publicadas portarias distintas para o Bacharelado (Portaria INEP n.º 471 de 6 de junho de 2017, BRASIL, 2017c) e a Licenciatura (Portaria INEP n.º 472 de 6 de junho de 2017, BRASIL, 2017d). Além disso, a portaria da Licenciatura, em seu Art. 8º, §2º, define que para áreas que conferem diploma de Licenciatura, cinco questões referenciadas na Portaria INEP n.º 510 de 6 de junho de 2017 (Diretrizes de Pedagogia, BRASIL, 2017e). Porém, não consta na referida portaria o número de itens dedicados a temáticas pedagógicas; sendo essa uma informação importante para as licenciaturas e variável⁶ entre as edições. No relatório e área constata-se que a prova de Conhecimentos Específicos da edição de 2017 apresentou 15 objetivas e 2 discursivas comuns ao bacharelado e à licenciatura e relativas a temáticas biológicas; 1 questão discursiva e as 12 últimas questões objetivas foram exclusivas de cada área. Parte da história dessas modificações está documentada nas portarias normativas, cadernos de prova e relatórios de área do ENADE, editadas anualmente e disponibilizadas no portal eletrônico do INEP (<http://www.inep.gov.br>).

Posteriormente, os integrantes da CAA, “sob a orientação dos servidores da Daes/Inep, constroem a matriz de prova, em que cada item é definido a partir da articulação entre uma característica de perfil, uma competência e até três conteúdos” (BRASIL, 2018a). A matriz tridimensional é formada por colunas e linhas que formam células ao se entrecruzarem. Nas linhas são relacionadas as características definidoras do perfil profissional esperado para o estudante a ser avaliado. As colunas apresentam os recursos (componente das competências), que serão avaliados. Nas células está descrita a dimensão do objeto do conhecimento, bem como apontamentos sobre a quantidade de questões e dificuldade esperada de cada questão.

Para Marinho-Araujo e Rabelo (2015), a matriz de referência para avaliação do perfil e das competências surge como instrumento metodológico e uma tecnologia psicoeducacional importante para a formação e o acompanhamento

⁶ Observa-se uma alteração na composição de prova, pois no ciclo anterior (2014), além das 3 discursivas (todas comuns ao bacharelado) havia 17 questões objetivas de conteúdos biológicos comuns ao bacharelado e apenas 10 exclusivas de cada área.

intencional de trajetórias profissionais, e expõe que desde 2010 o ENADE fundamenta-se na Matriz de Referência. Os autores mencionam funções e possibilidades para as matrizes como servir de fundamentação teórica para procedimentos e instrumentos de avaliação, orientação de indicadores de desempenho em processos de intervenção, mapeamento e trajetória de formação profissional.

Mata (2016) recomenda que ao aplicar uma avaliação é necessário identificar para cada item de prova qual ou quais competências são exigidas para a sua exata solução. Entretanto cabe considerar que os sujeitos e as IES avaliadas não têm acesso prévio à(s) matriz(es) de avaliação, outrossim, aproximadamente 4 meses antes da prova são publicadas as portarias com as diretrizes de área, e só posteriormente à prova, quando da publicação do relatório de área (“Anexo IX Concepção e Elaboração das Provas do ENADE”, BRASIL, 2018a), é que serão disponibilizadas as relações dos recursos (característica de perfil, competências e objetos de conhecimento) abrangidos em cada item, o que não substitui a matriz supracitada, mas permite reconstruí-la de maneira dedutiva (ver seção 4.2). Percebe-se, ao analisar as portarias dos ciclos anteriores, que houve modificações no texto das diretrizes de prova, portanto após três anos do ingresso de um estudante, não há garantia de que as diretrizes de prova vigentes durante sua formação serão as mesmas utilizadas em seu ENADE.

No caso da prova de 2017, (ver anexo A) as combinações uma a uma de 7 características de perfil, com 10 competências e 18 conteúdos podem resultar uma diversidade potencial de até 1.260 combinações possíveis para cada um dos 22 itens, considerando-se o emprego de apenas 1 conteúdo por item (sendo que são permitidos até três conteúdos por item). Há, ainda, no instrumento da licenciatura, a presença de cinco itens que empregam as diretrizes de prova descritas na portaria da Pedagogia (Port. 510 de 6 de junho de 2017) com 6 características de perfil, 13 competências e 23 conteúdos (1.794 combinações possíveis para cada um dos 5 itens objetivos).

Tal matriz de avaliação serve à elaboração da encomenda de itens de prova, dirigida aos integrantes do CERES-BNIES, os quais:

Após a capacitação os professores receberão a encomenda, a especificação do item a ser elaborado por eles e deve contemplar, o perfil, o recurso a ser utilizado e o objeto de conhecimento. A encomenda é montada com base na matriz de referência elaborada pela Comissão Assessora de Área e entregue de acordo com a área de interesse do conteúdo do colaborador e de acordo com suas escolhas. (CAMPOS, 2013)

“O revisor tem a prerrogativa de devolver este item para o elaborador, com orientação do que deve ser modificado, ou descartá-lo, ou encaminhá-lo para a próxima etapa, a revisão da comissão assessora de área do ENADE” (CAMPOS, 2013). Após a revisão da CAA e montagem da prova, esta é encaminhada à Comissão de área técnica (CAT), formada por especialistas em elaboração e revisão de itens e por servidores da Coordenação-Geral do ENADE.

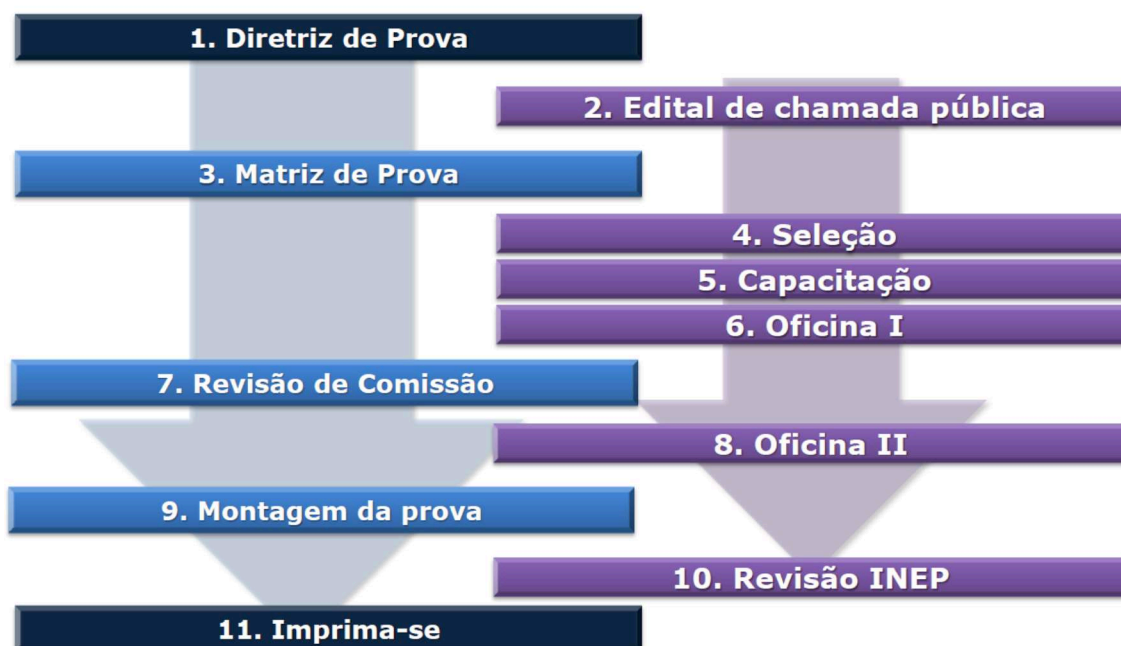
Segundo Campos (2013), como o BNIES do ENADE é uma ação muito recente e precisa superar alguns obstáculos que prejudicam a sua validade, pois, [pelo que se sabe até o presente] os itens que compõem o ENADE não são pré-testados. “O que se realiza é um intenso trabalho de “lapidação” do item, onde ele é preparado, por meio de constantes revisões, para que possa realmente compor a prova” (CAMPOS, 2013). Ainda conforme a autora, atualmente a Coordenação-Geral do ENADE somente consegue produzir o instrumento a ser aplicado naquele ano, não havendo a possibilidade de se verificar estatisticamente a validade do item produzido durante as oficinas.

Os itens elaborados e aprovados por uma revisão de pelo menos três revisores passam a integrar o Banco Nacional de Itens (BNIES), que:

tem por objetivo armazenar itens de qualidade técnica que permitam a montagem de provas capazes de estimar com maior precisão a proficiência dos estudantes com referência às competências, habilidades e características do perfil do egresso previstas nas diretrizes de prova. (CAMPOS, 2013)

Conseqüentemente, na construção do ENADE tenta-se basicamente operacionalizar a medida dos constructos (proficiência em determinadas competências) selecionadas e definidos nas diretrizes de prova de cada área, que podem ser editadas pela CAA a cada triênio, e tornados mensuráveis na criação e revisão das questões pelos integrantes do CERES-BNIES. A síntese desse processo produtivo está representada a seguir na Figura 2:

Figura 2: Fluxo de trabalho da elaboração das provas pelo INEP.



Fonte: Andrade (2017)

A análise dos resultados deve dedicar-se a interpretar a realidade captada pelos instrumentos de avaliação, portanto, é dependente de como (e em conseqüente “para que”) o teste e seus itens foram previamente estruturados.

Campos (2013) descreve os três principais tipos de estruturações de itens: o primeiro tipo é o “item de Interpretação” que é composto de uma situação problema para o qual deseja-se que o estudante mobilize determinados recursos, as habilidades e saberes capazes de evidenciar sua competência. O segundo tipo denominado de “múltipla escolha” [também chamado “múltipla escola complexa”] que disponibiliza três a cinco afirmações e pede ao estudante a análise de quais são verdadeiras ou falsas. O terceiro tipo denominado “asserção-razão” “apresenta duas proposições que são ligadas pela palavra PORQUE, sendo que a segunda proposição deve ser avaliada se constitui razão ou justificativa da primeira” (CAMPOS, 2013). No ENADE, as alternativas deste formato de item seguem um padrão único em que as duas primeiras alternativas indicam a veracidade das duas sentenças, mas a segunda alternativa considera inverídica a relação entre asserção e razão (causa-efeito, justificativa da primeira sentença pela segunda), a terceira e a quarta alternativa consideram que uma das sentenças não está correta

(respectivamente a asserção ou a razão) e a quinta indica a hipótese de nenhuma das sentenças estarem corretas.

As estatísticas básicas dos resultados do ENADE, calculadas à luz da TCT, foram divulgadas em relatórios referentes a cada curso (área) e disponibilizados na página eletrônica do INEP. No relatório de área são reportados aspectos do perfil dos respondentes e coordenadores de curso (agrupados por Região Geográfica, Categoria Administrativa, Organização Acadêmica), medidas de tendência central e histogramas ilustrando a dispersão de cada componente de prova (CE, FG) e de cada questão discursiva, também são apresentados os índices de facilidade (proporção de acertos) e discriminação (ponto-bisserial não corrigido) relativos a cada item da prova.

Conforme Campos (2013), a alta demanda de trabalho da Coordenação-Geral do ENADE, responsável pelo banco, e a inexistência de um fluxo de trabalho que possa contribuir para a constante revisão e renovação do banco, soma-se à não aplicação de uma teoria de medição da qualidade do item como principais problemas do BNIES.

2.4 A ÁREA DE CIÊNCIAS BIOLÓGICAS E O ENADE

Conforme Marandino, Selles e Ferreira (2009), no início do século XX, os conhecimentos das Ciências Biológicas organizavam-se em ramos com epistemologias bem variadas, ora por ramos mais descritivos da história natural, ora por estudos com tradições experimentais como a fisiologia humana, a embriologia e a biologia celular. Esse contexto fragmentado reforçava concepções sobre um menor status e cientificidade dos conhecimentos biológicos, quando contrastados à física. Segundo as autoras, a ideia de uma unificação das ciências biológicas surgiu muito depois da palavra “biologia” cunhada por Lamarck e Treviranus, no início do século XIX. Até o final do século XIX, não havia uma interpretação que fundamentasse nas bases do positivismo lógico e de axiomas matemáticos as relações entre as diferentes espécies de seres vivos, e apesar da publicação da Origem das Espécies, em 1859, a teoria da evolução padecia com lacunas teóricas,

dentre as principais uma teoria da hereditariedade, já que a teoria das gêmulas havia sido experimentalmente refutada. Com isso, enquanto as ciências experimentais desenvolviam-se paralelamente, fundamentadas em bases empíricas, de outro lado as disciplinas naturalísticas (zoologia, botânica, e a incipiente ecologia) eram reconhecidas como especulações metafísicas por não empregarem teorias quantificáveis. Apenas com a redescoberta dos trabalhos de Mendel 34 anos após sua publicação, e com a resolução dos conflitos entre biometristas (Darwinistas que defendiam uma herança fluída e miscível transmitida por gêmulas) e mendelistas (que defendiam uma herança particulada) é que, na primeira década do século XX, desenvolveu-se a genética e a genética de populações (e todo ferramental estatístico) com as teóricas capazes de fornecer modelos matemáticos e a base empírica para a teoria da seleção natural. Esse arcabouço teórico desenvolvido no intervalo entre as guerras mundiais deu origem ao movimento da “moderna síntese” com sua teoria sintética da evolução, que mesmo longe do consenso (como a oposição de microbiologistas e empiristas à síntese neodarwinista) e com lacunas empíricas (sobre a natureza bioquímica da hereditariedade, questão só consensuada alguns anos após a descrição do modelo do DNA e sua função nas décadas de 50 e 60) resultou em uma estrutura disciplinar arquitetada sob discursos unificadores com pretensões à hegemonia. Fato é que, mesmo sem a comprovação experimental da origem de espécies, nenhuma disciplina biológica nega a existência da evolução biológica e seu papel central na junção das ciências biológicas.

Sobre a origem da disciplina escolar de biologia, Marandino, Selles e Ferreira (2009, p.51) explicam que as disciplinas surgem como amálgamas mutáveis de subgrupos e tradições, sendo ministradas por professores não especialistas que frequentemente obtêm um lugar no currículo mediante justificativas como pertinência e utilidade. A partir disso, elas desenvolvem-se para uma consolidação acadêmica, vinculando-se às universidades por conjunto de especialistas formados nessas tradições. No Brasil, esse processo não foi diferente, e conforme Pedroso e Ferreira (2015) houve uma certa discrepância entre o ensino secundário e o ensino superior, já que o primeiro, a partir de 1942, passou a oferecer a disciplina Biologia (e não mais História Natural), enquanto que o segundo continuou a oferecer

sistematicamente o curso de História Natural. Segundo Ivor Goodson (1995, *apud* MARANDINO; SELLES; FERREIRA, 2009) esse movimento faz as disciplinas escolares afastarem-se dos objetos do cotidiano em direção a conteúdos mais abstratos e distantes da realidade dos alunos, o que constitui uma virada academicista em busca de um status científico, recursos materiais e prestígio e território profissional. Nesse sentido Bizzo esclarece outro aspecto histórico da ampliação do campo das Ciências Biológicas:

A formação de professores em área multidisciplinar, como é o caso das Ciências Biológicas, irá trazer dificuldades adicionais, em especial a partir da reforma universitária de 1968, com a Lei 5.692, de 1971, e a normatização que lhe seguiu, visando a formação de professores para o I e II Graus. Hoje o ensino de ciências biológicas, tanto no Ensino Fundamental como no Ensino Médio, se ressentem das deficiências introduzidas pelo modelo de formação daquela época, o qual leis e normas posteriores não modificaram essencialmente. (BIZZO, 2004, p. 152)

No Brasil, desde a fundação do Colégio Pedro II, em 1837, a disciplina de História Natural esteve presente nos currículos (com conteúdos de zoologia, botânica, geologia e mineralogia). Os cursos de Ciências Biológicas sucederam os de História Natural, inicialmente implantados quando da criação da Universidade de São Paulo (USP), em 1934. Após quase cinco décadas da criação do primeiro curso superior, em 3 de setembro de 1979, foi promulgada a Lei n.º 6.684, que regulamentou a profissão de Biólogo. Vinte e dois anos depois, a partir de exigência da Lei n.º 9.394/96 (BRASIL, 1996), foi aprovado o Parecer CNE/CES no 1.301 de 2001, que estabeleceu as Diretrizes Curriculares Nacionais para os Cursos de Ciências Biológicas, homologado pelo Ministério da Educação, gerando a Resolução do CNE/CES n.º 7, de 11 de março de 2002, e contemplando a formação de professores de biologia (CORTELAZZO; RIBEIRO, 2013). Desde 2002, as diretrizes para formação de inicial de professores (Resolução CNE/CP n.º 1, de 18 de fevereiro de 2002; BRASIL, 2002a), em seu artigo 7º, determinam que os cursos de licenciatura devem ter identidades e matrizes curriculares próprias e distintas dos Bacharelados, por mais que possam estar articuladas.

A prova do ENADE tem como público os concluintes dos cursos superiores avaliados no ciclo, e também era aplicada a uma amostra dos ingressantes até

2010, sendo que 2008 foi a última edição em que ingressantes de Ciências Biológicas participaram. Ao longo dos últimos cinco ciclos do ENADE, a Área de Licenciatura em Ciências Biológicas totalizou 140.528 convocados para realizar o exame; desses aproximadamente 117.000 (~82,9%) compareceram e responderam a prova objetiva de conhecimentos específicos. Em 2008, o exame atingiu o auge do número de estudantes convocados (~36.000), porém, tratando-se apenas de concluintes e cursos de graduação (atual enfoque do SINAES), o auge do número de participantes concluintes dos cursos de ciências biológicas foi a prova de 2014, com 21.539 concluintes vinculados a mais de 770 cursos de 361 instituições diferentes, sendo que em 2017 houve uma pequena redução (ver Tabela 2).

Tabela 2: Evolução do número de participantes ao longo dos cinco ciclos do ENADE classificados quanto ao tipo de curso, estágio do curso em que realizou a prova, e tipo de presença na prova.

Tema	Categoria	2005	2008	2011	2014	2017	Total
Tipo de Curso	Bacharelado			7.230	7.422	7.971	22.623
	Licenciatura			17.208	19.704	17.381	54.293
Estágio do Participante	Ingressante	15.728	19.395	0	0	0	35.123
	Concluinte	117.90	16.699	24.438	27.126	25.352	105.405
Presença	Ausente	3.043	5.457	4.362	5.418	4.602	22.882
	Presente	24.475	30.637	19.987	21.690	20.745	117.534
	Desconsiderados (problema adm. ou part. indevida)			89	18	5	112
Total	Total de convocados	27.518	36.094	24.438	27.126	25.352	140.528

Fonte: O autor, com base nos microdados.

Conforme Silva (2011), é necessário investigar o quanto a prova conseguiu operacionalizar a medida do constructo, conforme planejado previamente, essa questão se refere à abrangência do constructo avaliado pela prova. Por meio dessa comparação, é possível compreender as ênfases e omissões atualizadas na construção da prova de 2017.

Também foram utilizados como fonte de dados os cadernos de questões das provas de licenciatura, para, por meio de análise de conteúdo (quali-quantitativa), avaliar a coerência entre os seguintes elementos: 1) as provas, 2) a portaria com as diretrizes de elaboração da prova, 3) as diretrizes curriculares nacionais, 4) os descritores das relações propostas entre os atributos de “perfil”, “competência”,

“conteúdos” para cada questão, que são expostos nos apêndices do relatório de área (uma vez que não foi possível ter acesso à matriz de referência, consideradas pelo INEP como dado sigiloso).

Na Tabela 3, apresentada a seguir, observa-se a evolução temporal do número de questões de cada componente do ENADE, note-se que apenas de 2011 para 2014 houve a manutenção na proporção dos componentes de prova das licenciaturas, o que pode gerar insegurança por parte das coordenações de curso sobre o que será exigido no próximo ciclo avaliativo.

Tabela 3: Número de questões em cada componente da prova da área de Ciências Biológicas (Licenciatura e Bacharelado) ao longo dos cinco ciclos do ENADE.

Componente	Categoria	2005	2008	2011	2014	2017
FG	Discursiva	3	2	2	2	2
	Objetiva	7	8	8	8	8
CE	Discursiva comum ⁷	2	1	3	3	2
	Objetiva comum	18	19	17	17	15
	Discursiva exclusiva	2	2	0	0	1
	Objetiva exclusiva	8	8	10	10(5) ⁸	12(5)

Fonte: O autor, com base nas provas e portarias das diretrizes de prova. FG=Formação Geral; CE=Conhecimentos Específicos

Cabe considerar que às diretrizes curriculares nacionais é atribuída a função de subsídio para a elaboração das diretrizes de prova, o que deixa margem à reelaboração textual das diretrizes de prova por parte das comissões assessoras de área. Sendo assim, cabe comparar o texto presente nas DCN dos cursos de licenciatura e bacharelado em ciências biológicas (Parecer CNE/CES n.º 1.301/2001, aprovado em 6 de novembro de 2001, BRASIL, 2001b) e com a redação das diretrizes de prova de 2017 expostas nas Portarias do INEP n.º 472 (Licenciatura) de 6 de junho de 2017(BRASIL, 2017d).

Para facilitar a análise textual, foram utilizadas marcações de destaque em cores variadas para ilustrar os espaços ocupados por conjuntos de expressões com semântica semelhante ou equivalente, os textos exclusivos da licenciatura estão

⁷ Por comum, diz-se das questões que estão presentes nas provas da Licenciatura e também no Bacharelado no mesmo formato e mesma ordem no caderno de provas.

⁸ Apenas a partir de 2014, 5 questões dentre as exclusivas da licenciatura tiveram como referência a Matriz de Avaliação da área de Pedagogia.

demarcados em **fonte vermelha**, sendo as demais cores de fonte indicativas de textos comuns às portarias de licenciatura e bacharelado.

Quanto ao perfil, observa-se nas diretrizes de prova uma reformulação quase completa das DCN, sendo que as diretrizes de 2002 investem maior detalhamento em relação aos **conteúdos biológicos** e **ambientais**, enquanto as diretrizes de prova, apesar de mais coesas, apresentam ênfases em temáticas de **cidadania e diversidade social** e **epistemologia e popularização da ciência**. Apesar dessa diferença, as diretrizes de prova limitam a ênfase sobre responsabilidade social à característica exclusiva das licenciaturas (VII) e regidas por um adjetivo de sensibilização, sendo que nas DCN (característica d) há um verbo mais forte que exige do formando o comprometimento. O item V das diretrizes de prova (**fonte azul**) não encontra correspondente nas DCN, sendo a única inserção de conteúdo estranho às DCN realizada deliberadamente pela CAA, trata-se de recomendações comportamentais e atitudinais cuja intenção é gerar à docilidade dos avaliados ao **mundo do trabalho**. No

Quadro 2, é possível comparar os “espaços” e qualificações que cada assunto/característica de perfil assume nos dois textos:

Quadro 2: Quadro comparativo das características de perfil conforme as DCN (à esquerda) e as diretrizes de prova (à direita).

Perfil dos Formados (DCN)	Perfil dos Concluintes (ENADE)
<p>a) generalista, crítico, ético, e cidadão com espírito de solidariedade;</p> <p>b) detentor de adequada fundamentação teórica, como base para uma ação competente, que inclua o conhecimento profundo da diversidade dos seres vivos, bem como sua organização e funcionamento em diferentes níveis, suas relações filogenéticas e evolutivas, suas respectivas distribuições e relações com o meio em que vivem;</p> <p>c) consciente da necessidade de atuar com qualidade e responsabilidade em prol da conservação e manejo da biodiversidade, políticas de saúde, meio ambiente, biotecnologia, bioprospecção, biossegurança, na gestão ambiental, tanto nos aspectos técnicos-científicos, quanto na formulação de políticas, e de se tornar agente transformador da realidade presente, na busca de melhoria da qualidade de vida;</p> <p>d) comprometido com os resultados de sua atuação, pautando sua conduta profissional por critério humanísticos, compromisso com a cidadania e rigor científico, bem como por referenciais éticos legais;</p> <p>e) consciente de sua responsabilidade como educador, nos vários contextos de atuação profissional;</p> <p>f) apto a atuar multi e interdisciplinarmente, adaptável à dinâmica do mercado de trabalho e às situações de mudança contínua do mesmo;</p> <p>g) preparado para desenvolver idéias inovadoras e ações estratégicas, capazes de ampliar e aperfeiçoar sua área de atuação.</p>	<p>I. observador, crítico e integrador ao interpretar e avaliar os padrões e processos biológicos e suas interfaces com outras áreas do saber;</p> <p>II. comprometido com a produção de conhecimento, a transformação social e a educação emancipatória;</p> <p>III. ético, com responsabilidade social e ambiental, comprometido com a sua contínua atualização profissional e com a divulgação científica;</p> <p>IV. consciente de sua responsabilidade como educador frente à comunidade, nos vários contextos de atuação profissional[e], compreendendo a ciência como uma atividade social com potencialidades e limitações;</p> <p>V. empático, propositivo e colaborativo nas relações interpessoais que envolvem o mundo do trabalho;</p> <p>VI. criativo e empreendedor na concepção de ideias inovadoras para o desenvolvimento humano e de sociedades sustentáveis;</p> <p>VII. sensível às questões ligadas aos direitos humanos, à diversidade sociocultural e ambiental e à identidade de gênero no contexto escolar.</p>

Fonte: O autor, adaptado de Parecer CNE/CES n.º 1.301/2001 (BRASIL, 2001b) e Portaria 472/2017 (BRASIL, 2017d).

Quanto às competências, novamente o texto das diretrizes de prova é mais sucinto, porém nas portarias há omissões a temáticas presentes (apenas) nas competências das DCN, tais como: trabalho e mercado, formação continuada, cidadania e diversidade social, senso crítico, bem como há uma restrição das temáticas educacionais e crítico-sociais apenas às diretrizes de prova da licenciatura. No Quadro 3, é possível comparar os “espaços” e discursos que cada assunto/competência assume nos dois textos:

Quadro 3: Quadro comparativo das competências conforme as DCN (à esquerda) e as diretrizes de prova (à direita).

Competências e Habilidades (DCN)	Competências (ENADE)
<p>a) Pautar-se por princípios da ética democrática: responsabilidade social e ambiental, dignidade humana, direito à vida, justiça, respeito mútuo, participação, responsabilidade, diálogo e solidariedade;</p> <p>b) Reconhecer formas de discriminação racial, social, de gênero, etc. que se fundem inclusive em alegados pressupostos biológicos, posicionando-se diante delas de forma crítica, com respaldo em pressupostos epistemológicos coerentes e na bibliografia de referência;</p> <p>c) Atuar em pesquisa básica e aplicada nas diferentes áreas das Ciências Biológicas, comprometendo-se com a divulgação dos resultados das pesquisas em veículos adequados para ampliar a difusão e ampliação do conhecimento;</p> <p>d) Portar-se como educador, consciente de seu papel na formação de cidadãos, inclusive na perspectiva sócio-ambiental;</p> <p>e) utilizar o conhecimento sobre organização, gestão e financiamento da pesquisa e sobre a legislação e políticas públicas referentes à área;</p> <p>f) Entender o processo histórico de produção do conhecimento das ciências biológicas referente a conceitos/princípios/teorias;</p> <p>g) Estabelecer relações entre ciência, tecnologia e sociedade;</p> <p>h) Aplicar a metodologia científica para o planejamento, gerenciamento e execução de processos e técnicas visando o desenvolvimento de projetos, perícias, consultorias, emissão de laudos, pareceres etc. em diferentes contextos;</p> <p>i) Utilizar os conhecimentos das ciências biológicas para compreender e transformar o contexto sócio-político e as relações nas quais está inserida a prática profissional, conhecendo a legislação pertinente;</p> <p>j) desenvolver ações estratégicas capazes de ampliar e aperfeiçoar as formas de atuação profissional, preparando-se para a inserção no mercado de trabalho em contínua transformação;</p> <p>k) Orientar escolhas e decisões em valores e pressupostos metodológicos alinhados com a democracia, com o respeito à diversidade étnica e cultural, às culturas autóctones e à biodiversidade;</p> <p>l) atuar multi e interdisciplinarmente, interagindo com diferentes especialidades e diversos profissionais, de modo a estar preparado a contínua mudança do mundo produtivo;</p> <p>m) avaliar o impacto potencial ou real de novos conhecimentos/tecnologias/serviços e produtos resultantes da atividade profissional, considerando os aspectos éticos, sociais e epistemológicos;</p> <p>n) comprometer-se com o desenvolvimento profissional constante, assumindo uma postura de flexibilidade e disponibilidade para mudanças contínuas, esclarecido quanto às opções sindicais e corporativas inerentes ao exercício profissional.</p>	<p>I. elaborar, executar e gerenciar programas e projetos nas áreas de Biodiversidade e Ambiente, Saúde e Biotecnologia;</p> <p>II. identificar e problematizar questões inerentes às Ciências Biológicas, incluindo seus aspectos científicos, históricos e filosóficos;</p> <p>III. realizar consultorias educacionais e outras atividades afins definidas na legislação e em políticas públicas;</p> <p>IV. executar técnicas básicas e aplicadas, em laboratório e em campo, no contexto educacional;</p> <p>V. realizar atividades de gestão educacional nos diversos espaços institucionais;</p> <p>VI. instrumentalizar a comunidade escolar para a prática social crítica, emancipatória e transformadora;</p> <p>VII. comunicar e divulgar o conhecimento científico;</p> <p>VIII. desenvolver produtos e processos de inovação científica, tecnológica e educacional;</p> <p>IX. elaborar, aplicar e avaliar estratégias pedagógicas e materiais didáticos;</p> <p>X. desenvolver e aplicar estratégias de avaliação na prática docente.</p>

Fonte: O autor, adaptado de Parecer CNE/CES n.º 1.301/2001 (BRASIL, 2001b) e Portaria 472/2017 (BRASIL, 2017d).

É importante notar que o parecer que fundamenta as DCN não faz distinção entre bacharelado e licenciaturas nem em relação às características de perfil, nem em relação às competências e habilidades. Em relação aos conteúdos, há uma distinção entre Bachareis e Licenciados onde aos primeiros os conteúdos deveriam sofrer adaptações com relação aos biomas e ecossistemas regionais e aos segundos deveriam ser acrescentadas noções de física, química e matemática e ciências da saúde para atender ao ensino fundamental. Tal recomendação insere uma miríade de possíveis dimensões e conteúdos os quais quase foram tratados pelas diretrizes de prova. O texto das DCN novamente demonstra-se mais detalhado e prolixo, comparado à estrutura de tópicos expressa nas diretrizes de prova, conforme observa-se no Quadro 4:

Quadro 4: Quadro comparativo dos conteúdos conforme as DCN (a cima) e as diretrizes de prova (abaixo).

Conteúdos curriculares (DCN)	
<p>BIOLOGIA CELULAR, MOLECULAR E EVOLUÇÃO: Visão ampla da organização e interações biológicas, construída a partir do estudo da <u>estrutura molecular</u>[II] e <u>celular</u>[IV], <u>função e mecanismos fisiológicos da regulação</u>[I] em <u>modelos eucariontes</u>[III, IV], <u>procariontes</u>[III] e de <u>partículas virais</u>[III], fundamentados pela informação <u>bioquímica</u>, <u>biofísica</u>[II], <u>genética</u>[V] e <u>imunológica</u>[III]. Compreensão dos <u>mecanismos de transmissão da informação genética</u>[V], em <u>nível molecular</u>[II], <u>celular</u>[IV] e <u>evolutivo</u>[VI].</p> <p>DIVERSIDADE BIOLÓGICA: Conhecimento da classificação, <u>filogenia</u>[VII, VIII], <u>organização</u>[I], <u>biogeografia</u>[XIII], <u>etologia</u>[VII], <u>fisiologia</u>[I] e estratégias adaptativas <u>morfo-funcionais</u>[I] dos seres vivos.</p> <p>ECOLOGIA: <u>Relações entre os seres vivos e destes com o ambiente</u>[IX] ao longo do tempo <u>geológico</u>[IX e/ou XIII]. Conhecimento da <u>dinâmica das populações</u>[IX], <u>comunidades e ecossistemas</u>[IX], da <u>conservação e manejo da fauna e flora</u>[IX] e da <u>relação saúde</u>[XI], <u>educação e ambiente</u>[X].</p> <p>FUNDAMENTOS DAS CIÊNCIAS EXATAS E DA TERRA: Conhecimentos matemáticos, físicos, químicos, estatísticos[XII], geológicos e outros fundamentais para o entendimento dos processos e padrões biológicos.</p> <p>FUNDAMENTOS FILOSÓFICOS E SOCIAIS: Reflexão e discussão dos <u>aspectos éticos</u>[XV] e <u>legais relacionados ao exercício profissional</u>[XVI]. Conhecimentos básicos de: História, Filosofia e Metodologia da Ciência[XVII], Sociologia e Antropologia, para dar suporte à sua atuação profissional na sociedade, com a consciência de seu papel na formação de cidadãos.</p> <p>CONTEÚDOS ESPECÍFICOS</p> <p>A modalidade Licenciatura deverá contemplar, além dos conteúdos próprios das Ciências Biológicas, conteúdos nas áreas de Química, Física e da <u>Saúde</u>[XI], para atender ao ensino fundamental e médio. A <u>formação pedagógica</u>[XV-XVIII], além de suas especificidades, deverá contemplar uma visão geral da educação e dos processos formativos dos educandos. Deverá também enfatizar a instrumentação para o ensino de Ciências no nível fundamental e para o ensino da Biologia, no nível médio.</p>	
Conteúdos ENADE Licenciatura	
<p>I. Morfofisiologia animal; II. Bioquímica e Biofísica; III. Microbiologia, Imunologia e Parasitologia; IV. Biologia Celular; V. Genética; VI. Evolução; VII. Zoologia; VIII. Botânica; IX. Ecologia; X. Educação Ambiental; XI. Ambiente e Saúde;</p>	<p>XII. Bioestatística; XIII. Biogeografia e Paleontologia; XIV. Bioética; XV. Legislação e políticas públicas educacionais; XVI. Planejamento e métodos em pesquisa educacional; XVII. Didática para o ensino de Ciências e de Biologia; XVIII. Currículo no ensino de Ciências e de Biologia.</p>

Fonte: O autor, adaptado de Parecer CNE/CES n.º 1.301/2001 (BRASIL, 2001b) e Portaria 472/2017 (BRASIL, 2017d). Obs: Utilizou-se a formatação “sublinhado” para marcar o início e término das expressões que encontram correspondência com os tópicos listados nas diretrizes de prova. A formatação em negrito representa conteúdos presentes apenas nas DCN e as palavras em vermelho destacam expressões presentes apenas nas diretrizes de prova da licenciatura (contrastando com os tópicos em preto que representam expressões comuns às das habilitações).

Nos conteúdos, as redundâncias presentes nas DCN, (i.e. fisiologia[I] e ecologia[IX] repetidas 4 vezes, “bioquímica e biofísica”[II], e “Microbiologia, Imunologia e Parasitologia”[III] repetidas 3 vezes) contrastam com as omissões nas diretrizes de provas de assuntos presentes nas DCN (marcados em **negrito**). Os conteúdos negligenciados ou preteridos pela CAA são relacionados a fundamentos das ciências exatas (matemáticos, físicos, químicos, geológicos), e fundamentos filosóficos e das ciências sociais (sociologia e antropologia, história e filosofia da ciência). Além disso do conteúdo específico das DCN de licenciatura não foram contemplados pela CAA na elaboração das diretrizes de prova os conteúdos relacionados à psicologia da educação e instrumentação para o ensino.

Nota-se também que os objetos de conhecimento II, III e XIII das diretrizes de prova são na verdade objetos compostos por miscelâneas de mais de uma área do conhecimento, e de outra forma, tratadas como independentes resultariam em sete conteúdos diferentes.

Conforme o Parágrafo único do Art. 11 das Resolução CNE/CP 1, de 18 de fevereiro de 2002 (Diretrizes Curriculares Nacionais para Licenciaturas; BRASIL, 2002a), o tempo dedicado às dimensões pedagógicas não será inferior à quinta parte da carga horária total do curso. Além disso, o parecer 009/2001/CNE/CP (BRASIL, 2001a) que deu origem à resolução supracitada enfatiza a necessidade de articular os conteúdos pedagógicos aos conteúdos disciplinares desde as primeiras fases, opondo-se a tratamentos dissociados entre os campos (modelo 3+1 de formação de professores).

2.5 PSICOMETRIA: seus modelos, pressupostos e possibilidades

De um modo geral, a psicometria procura explicar o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas, tipicamente chamadas de itens. A psicometria moderna tem duas vertentes: a teoria clássica dos testes (TCT) e a teoria de resposta ao item (TRI). Nas seções a seguir serão apresentados os fundamentos dessas duas teorias complementares em sua tarefa de assegurar a validade das avaliações.

2.5.1 Teoria Clássica dos Testes

A TCT começou a se estruturar, como é conhecida hoje, após os trabalhos de Spearman no início do século XXI, que possibilitaram o desenvolvimento da Psicometria Clássica em bases estatísticas e empíricas com representações lineares capazes de gerar modelos (SILVEIRA, 2015). Segundo Muñiz (2003, *apud* SILVEIRA, 2015), o propósito de Spearman era “encontrar um modelo estatístico que fundamentasse adequadamente as pontuações dos testes e permitisse estimar os erros de medida associados a todo processo de medição”. Também não se pode deixar de mencionar Thurstone, que, na década de 1930, desenvolveu a análise fatorial múltipla e teorias relacionadas ao escalonamento psicométrico.

Conforme Pasquali (2013), a análise empírica dos itens é feita através de uma série de parâmetros que as questões devem possuir para conseguir medir adequadamente o que se propõem. Dentre os parâmetros, os principais são a discriminação e a dificuldade, essa última definida na TCT como a percentagem de acertos (representada pela letra “*b*”). Quanto mais próxima a taxa de acertos estiver de 100%, mais fácil é o item (PASQUALI, 2013). Para ser considerada apta a avaliar os sujeitos de uma população (ex.: alunos de um curso), uma questão deve ser mais acertada por alunos que tiveram bom desempenho do que pelos que tiveram desempenho ruim, a essa capacidade de distinguir quem sabe de quem não sabe dá-se o nome de discriminação (parâmetro representado pela letra “*a*”).

Outro índice da TCT que mede essa capacidade de discriminação das questões, e que foi escolhido pelo INEP para ser utilizado no ENADE 2017, é o denominado coeficiente de correlação ponto-bisserial, o qual é derivado do coeficiente de correlação de Pearson. Ou seja, sua principal suposição é que a relação entre o resultado da resposta em um determinado item (erro 0 ou acerto 1) e o escore total do teste (soma dos acertos no restante dos itens) é linear (correlação positiva). Valores menores para correlações ponto-bisserial (item-total) revelaram uma tendência de escolha da opção errada pelos participantes que obtiveram os maiores escores, um forte indicativo de que o item deve ser retirado do instrumento

(LIRA, 2004). Quanto maior a proporção de itens de prova com elevados índices de discriminação, mais precisa será a avaliação.

Pasquali e Primi (2003) explicam que a grande crítica à teoria clássica da psicometria (TCT) já havia sido apontada desde a década de 1920 por Thurstone, e diz respeito à variação na capacidade de mensuração que o ferramental estatístico dessa teoria sofre, pois, as medidas inferidas alteram-se conforme os diferentes itens (questões) usados para compor o instrumento (teste). Além disso, os autores explicam que os parâmetros dos itens (dificuldade e discriminação) de um teste dependem da amostra de sujeitos em que eles foram calculados, isto é, este parâmetro é dependente da amostra (*group-dependent* ou *sample-dependent*) e, se ela não for rigorosamente representativa da população, os parâmetros dos itens não podem ser considerados válidos. Há também uma dependência em relação ao teste utilizado, não permitindo comparações entre edições, pois testes com dificuldades diferentes produzem escores diferentes. Conforme Pasquali e Primi (2003), pela TCT as comparações e o acompanhamento dos cursos ao longo do tempo dependem muito de que as amostras dos sujeitos apresentem os parâmetros invariáveis entre as edições, já pela TRI o cálculo dos parâmetros dos itens (dificuldade e discriminação) independe da amostra de sujeitos utilizada - diz-se que os parâmetros são independentes dos sujeitos (*not group-dependent*).

2.5.2 Teoria de Resposta ao Item: ou Teoria do Traço Latente

Conforme Pasquali e Primi (2003), foi somente após os anos 50 que os psicometristas começaram a descobrir a solução para o problema, baseados na teoria do traço latente de Lazarsfeld (1959), nos trabalhos de Lord (1952) e do dinamarquês Rasch (1960), formaram-se as bases da moderna Teoria da Resposta ao Item, (também chamada Teoria do Traço Latente). Conforme Coelho (2014), essa teoria foi originalmente desenvolvida por Lord em 1952, mas só foi axiomatizada por Birnbaum em 1968 e por Lord em 1980. Birnbaum implementou, em 1968, um terceiro parâmetro no modelo de Rasch, sendo criado assim o Modelo Logístico de Três Parâmetros (ML3P). Deve-se também a Birnbaum a descrição da Função de Informação do Item (utilizando um modelo de mensuração de Fisher) e também a

primeira técnica utilizando o método da máxima verossimilhança para estimar os parâmetros dos itens.

Em 1970, Bock e Lieberman (1970) introduziram o método da máxima verossimilhança marginal para a estimação dos parâmetros dos itens. Tal processo da estimação dos parâmetros do modelo é realizado em duas etapas. Na primeira etapa, estimam-se os parâmetros dos itens, portanto, assume-se certa distribuição para as proficiências. Na segunda etapa, assumem-se os parâmetros dos itens conhecidos, os quais serão utilizados para estimar as proficiências por outro processo de estimação, sendo o método EAP (Esperança a posteriori), o mais utilizado em avaliação educacional no Brasil. Apesar do avanço que esse procedimento trouxe para a solução do problema, ele requeria que todos os parâmetros dos itens fossem estimados simultaneamente. Bock e Aitkin (1981) propuseram uma modificação no método acima e utilizaram o algoritmo Esperança e Maximização (EM) de Dempster, Laird e Rubin (1977), de modo a permitir que os itens pudessem ter seus parâmetros estimados em separado, a fim de reduzir o custo computacional (Coelho, 2014. p. 57).

A Teoria de Resposta ao Item, conforme Andrade, Tavares e Valle (2000), compreende modelos matemáticos que apresentam formas de representar a relação entre a probabilidade de um indivíduo acertar um item de prova, e seu traço latente (proficiência) na área de conhecimento avaliada. Tal proficiência não pode ser diretamente observada, em outras palavras, conforme Pasquali (2013), a proficiência é uma entidade latente que intenta-se definir e medir para poder explicar um determinado comportamento de acertar ou errar as respostas aos itens de um teste. De acordo com Pasquali e Primi (2003), as aptidões ou traços latentes, servem como preditores de desempenho do sujeito em uma determinada tarefa. Tal técnica é conhecida por modelagem latente, em que o traço latente é a causa e o desempenho o efeito (LOPES, 2011, p. 41).

Hambleton, Swaminathan e Rogers (1991 *apud* COELHO, 2014) salientam cinco avanços e impacto que apontam porque a TRI, supera as limitações na psicometria clássica:

- os cálculos dos parâmetros clássicos dos itens (dificuldade e discriminação);
- o cálculo do nível de aptidão do sujeito independente do teste;
- a possibilidade de emparelhar itens com a aptidão do sujeito;
- definição de modelo e não precisa fazer suposições que aparentam serem improváveis, tais como os erros de medida serem iguais para todos os sujeitos;
- não necessitam trabalhar com testes estritamente paralelos, como exigido pela TCT.

A razão do “atraso” na aplicação desta teoria está no fato da complexidade de manipulação de seus modelos matemáticos, que apresentavam algoritmos de tal dificuldade que a tecnologia computacional da época era incapaz de resolver de uma maneira útil e prática[...]. Este problema vem sendo solucionado com o avanço da tecnologia da informática, da disponibilidade de softwares apropriados que entraram no mercado nos anos de 1980 e, com os procedimentos de estimação dos parâmetros. Assim, a TRI vem se tornando a técnica predominante no campo de estudos para testes. O primeiro software para as análises da TRI surgiu somente em 1979 com o BICAL de Wright, Mead e Bell (1979), seguidos depois pelo LOGIST de Wingersky, Barton e Lord (1982) e pelo BILOG de Kim (1997). (COELHO, 2014)

No contexto internacional, a TRI vem sendo empregada amplamente por vários países: Estados Unidos, França, China, Holanda, Coreia do Sul e principalmente nos países participantes do Programa Internacional de Avaliação de Estudantes (PISA) que conta com a participação de 32 países, tais como Estados Unidos, Coreia do Sul, Canadá e Brasil. (SCHER, 2014)

No Brasil, a TRI foi usada pela primeira vez em 1993 pela Secretaria de Educação do Estado de São Paulo (conforme GATTI, 1996 *apud* ANDRADE; TAVARES; VALLE, 2000), porém seu uso mais completo na análise dos dados de uma avaliação em larga escala foi em 1995 quando foi aplicada aos dados do Sistema Nacional de Ensino Básico (SAEB), e mais recentemente, no Exame Nacional do Ensino Médio (ENEM).

“Embora, a TRI como um conjunto de procedimentos de avaliação não seja totalmente estranha nos estudos referentes à Psicologia e à Educação, considerando seu uso pelo INEP desde a década de

1990, poucos e isolados estudos se apresentam à comunidade científica, o que impede uma discussão mais pormenorizada dos resultados obtidos” (LOPES; VENDRAMINI, 2015, p.46).

Zlatkin-Troitschanskaia, Shavelson e Kuhn (2015) mencionam que alguns desafios específicos da educação superior como a diversidade nacional e internacional, bem como, a diversidade de cursos de graduação, programas de estudo e instituições que fazem da pesquisa em avaliação de resultados de aprendizado de nível superior, e especialmente nos domínios de um campo de conhecimentos específicos, uma tarefa complexa e multidimensional. No mesmo mote, Moreira Júnior (2014) afirma que uma implantação da TRI no ENADE seria mais complicada do que no ENEM, já que há dezenas de áreas diferentes, o que implicaria na criação, pré-teste e calibração de dezenas de bancos de itens.

Nesse sentido, a escassez de estudos psicométricos sobre a prova do ENADE sugere que há um valioso conjunto de dados disponíveis a ser investigado por futuras pesquisas abrangendo modelagem estatística para construção de escalas de proficiência e avaliação das propriedades psicométricas do Exame e de seus itens, dentre essas modelagens destaca-se a TRI.

Conforme Primi, Hutz e Silva (2011) o modelo da TRI é empregado para agregar dados de respostas aos itens calibrando-se em uma mesma escala medidas intervalares para os itens (nível de dificuldade) e para os estudantes (nível de proficiência), tal escala, frequentemente denominada escala (0,1) representa a média amostral em zero e desvio padrão em um. É o estabelecimento dessa relação (posicionamentos estabelecidos probabilisticamente) entre os itens e os respondentes “que proporciona ao modelo o termo teoria, uma vez que, diante de dados empíricos, que são as respostas dadas por um grupo de pessoas a um conjunto de itens, é construído um modelo que tenta explicar a relação entre eles”. (NUNES; PRIMI 2009, p. 42 *apud* LOPES, 2011, p. 38)

De acordo com Andrade, Tavares e Valle (2000, p. 7), a escolha de qual modelo se adotar depende da natureza dos itens (dicotômicos ou politômicos), do número de populações envolvidas e da quantidade de traços latentes (uni ou multidimensional).

Segundo Embretson e Reise (2000 *apud* COELHO, 2014), os modelos utilizados na TRI partem de dois pressupostos relevantes: 1) a plausibilidade de uma Curva Característica do Item (CCI), e a independência local e/ou unidimensionalidade. A forma de uma CCI é uma função matemática monotônica crescente (geralmente logarítmica ou ogiva normal), e é usada para descrever como a alteração do traço latente relaciona-se à mudança na probabilidade de acerto de uma resposta específica. 2) A independência local é obtida quando, condicionada aos níveis do traço latente, os itens do teste são independentes. Assim, a probabilidade de responder a um item é precisamente determinada pelo nível do traço latente do respondente e não por suas respostas aos outros itens do conjunto. Entretanto, para Andrade, Tavares e Valle (2000), a independência local é vista como consequência da correta determinação da dimensionalidade dos dados.

Conforme Pasquali (2018, p. 27), a métrica do parâmetro que representa a dificuldade do item, denominado parâmetro b_i , pode variar em teoria de $-\infty$ a $+\infty$, mas na prática, vai de -3 a +3 desvios padrões. Já o parâmetro de discriminação que representa a inclinação da CCI, chamado de a_i , na prática deve ser positivo (varia de 0 a 3), onde quanto mais próximo de zero significa baixa a discriminação. Conforme Pasquali (2018), Baker (2001) sustenta que valores de “a” abaixo de 0,65 são medíocres, e acima desse valor são considerados moderados. Já Oliveira (2006 citando LINACRE; WRIGHT, 1994) e Vendramini e Dias (2005) estabelecem como valores de referência aceitáveis: para discriminação “a” maiores que 0,30; para a dificuldade do item (parâmetro b) entre -2,95 e 2,95 e para a probabilidade de acerto ao acaso (parâmetro c) valores menores que 0,40 (para itens com 5 alternativas). Outrossim, segundo Nojosa (2001), para a escala normal o intervalo dos valores plausíveis para “a” é de 0,6 a 1,6, para “b” o intervalo seria de -2 a 2 e para “c” (com 5 alternativas) o esperado seria valores entre 0,1 e 0,3.

2.5.3 Dimensionalidade

Quando um teste é analisado pela TRI ou pela TCT, tem-se como pressuposto a unidimensionalidade do teste (PASQUALI, 2013), e caso tal

pressuposto não seja atendido elevam-se as magnitudes dos erros padrões dos escores calculados, e pode-se dizer que o instrumento padece de um vício na validade estrutural.

De um ponto de vista estrito, conforme apontado por Bock, Gibbons e Muraki (1988 *apud* OLIVEIRA, 2017, p. 81), caso se deseje resumir o desempenho de um conjunto de itens em uma nota simples, esses itens devem se referir a uma única dimensão, assim, conforme o autor, é inadequado produzir só uma nota para o componente de prova se os itens dessa prova não são claramente relativos à mesma dimensão. A ideia de unidimensionalidade pode ser compreendida a partir da organização dos itens, ordenando-os em uma linha em função de sua dificuldade. Os itens têm diferentes níveis de dificuldade e isto ocorre porque há diferenças, que podem ser ordenadas em uma hierarquia de complexidade da aquisição dessas competências/ habilidades/ conteúdos avaliados (PRIMI, HUTZ, SILVA, 2011).

Pasquali e Primi (2003) esclarecem que na maioria dos modelos da TRI há apenas uma aptidão (θ) responsável pela realização de um conjunto de tarefas (itens do teste). Dadas as diversas habilidades desenvolvidas durante uma formação, pode parecer mais coerente associar o desempenho do respondente a mais de um traço latente. Contudo, para satisfazer o pressuposto da unidimensionalidade, é suficiente admitir que haja uma aptidão dominante (um traço latente dominante) responsável pelo desempenho no conjunto de itens do teste. Logo, “a presença de um fator geral subjacente, nessa perspectiva, indicaria que os padrões esperados são mais frequentes [sic] do que os inesperados” (SILVA, 2011, p. 98).

Pasquali (1996, p.79) explica que responder um item é um comportamento multimotivado, como qualquer outro comportamento humano, daí, a rigor é impossível falar em unidimensionalidade estrita. Todavia a análise da covariação entre o item e o traço latente que se deseja medir é possível. Essa covariação é expressa pela carga fatorial do item, o que significa que itens com pouca carga são representações inadequadas do traço e devem ser removidos do teste. Assim a exclusão de itens tomando como critério a análise fatorial é mais razoável que

baseá-la no índice de correlação ponto-bisserial, que esse pressupõe tanto a unidimensionalidade, quanto a adequação dos demais itens do instrumento.

A unidimensionalidade, de modo geral e prático, “consiste em aferir a consistência interna entre os itens do teste. Isso significa tipicamente avaliar quanto os itens se entendem entre si, ou em termos estatísticos, verificar as intercorrelações ou a covariância entre os itens” (Pasquali, 2018). Ainda segundo Pasquali (*op. cit.*) existem duas principais categorias de técnicas estatísticas multivariadas para avaliar a unidimensionalidade dos itens de testes de aptidão: a primeira, que ele chama Análise Fatorial Comum (ACF), trabalha matrizes de correlações que sumarizam as informações das respostas pelas intercorrelações entre os itens. A segunda, chamada *Full-Information Factor Analysis* (FIFA) e desenvolvida por Bock, Gibbons e Muraki em 1988, também é uma técnica de análise fatorial, mas é baseada em todos os dados empíricos agrupados em frequências de padrões de respostas.

Conforme Damásio (2012), o primeiro passo durante a implementação de um análise fatorial exploratória é observar se a matriz de dados é passível de fatoração, isto é, analisar se os dados podem ser submetidos ao processo de análise fatorial e para isso ele sugere o cálculo do índice de Kaiser- Meyer-Olkin (KMO), que é um teste estatístico que sugere a proporção de variância dos itens que pode estar sendo explicada por uma variável latente (LORENZO-SEVA; TIMMERMAN; KIERS, 2011). Assim, tal índice indica quão adequada é a aplicação da análise fatorial ao conjunto de dados.

O KMO é calculado por meio do quadrado das correlações totais dividido pelo quadrado das correlações parciais, das variáveis analisadas (Field, 2005). Seu valor pode variar de zero a um. Valores iguais ou próximos a zero indicam que a soma das correlações parciais dos itens avaliados é bastante alta em relação à soma das correlações totais. Nesses casos, possivelmente a análise fatorial será inapropriada (Pasquali, 1999). Como regra para interpretação dos índices de KMO, valores menores que 0,5 são considerados inaceitáveis, valores entre 0,5 e 0,7 são considerados medíocres; valores entre 0,7 e 0,8 são considerados bons; valores maiores que 0,8 e 0,9 são considerados ótimos e excelentes, respectivamente (Hutcheson & Sofroniou, 1999). (DAMÁSIO, 2012, p. 215)

Laros explica que a análise fatorial (AF) é um dos procedimentos psicométricos mais utilizados na construção, revisão e avaliação de instrumentos psicológicos.

O ponto central de partida na análise fatorial é o princípio da parcimônia: um grande número de variáveis observadas pode ser explicado por um número menor de variáveis hipotéticas, não-observadas. Estas variáveis hipotéticas, também chamadas de fatores, são responsáveis [princípio da causalidade] pelo fato de as variáveis observadas correlacionarem entre si. (LAROS, 2012, p. 164)

Assim, a análise fatorial é utilizada para descobrir as variáveis latentes que estão subjacentes à escala. Para alcançar este fim, a análise fatorial usa a matriz de correlações ou a de covariâncias entre as variáveis mensuradas. Na teoria, essas variáveis latentes são as causas subjacentes das variáveis mensuradas. A análise fatorial produz cargas fatoriais, as quais podem ser consideradas pesos de regressão das variáveis mensuradas para predizer o constructo subjacente. “Nos casos onde existe mais de um fator [dimensão] subjacente aos dados, a análise fatorial também produz correlações entre os fatores” (LAROS, 2012, p. 166).

“A análise fatorial tem como lógica precisamente verificar quantos constructos comuns são necessários para explicar as covariâncias (as intercorrelações) dos itens. As correlações entre os itens são explicadas, pela análise fatorial, como resultantes de variáveis-fonte que seriam as causas destas covariâncias”. (PASQUALI, et al, 1996, p. 96). A premissa é que não existe um verdadeiro número de dimensões, mas sim qual é o número de dimensões necessário e suficiente para representar fielmente as principais relações dos dados gerados pelas respostas aos itens (RECKASE, 2009).

A lógica da análise fatorial confirmatória (AFC) adequa-se muito bem ao problema da unidimensionalidade. Contudo há dificuldades com alguns postulados básicos da análise fatorial, em especial com a questão das intercorrelações e da linearidade. A base da análise fatorial são as medidas de associação entre os itens (covariância e correlação), e existe uma série de índices estatísticos para estimar essa associação, não devendo ser arbitrária a escolha desses índices. A correlação tetracórica é a preferida para estabelecer a relação entre as variáveis dicotômicas ou

dicotomizáveis, como são as respostas ao ENADE, entretanto quando há itens muito fáceis ou muito difíceis esse evento faz surgir na matriz tetracórica dois fenômenos: 1º quando a correlação positiva é indefinida (divisão por zero) ou 2º casos de Heywood em que as comunalidades extrapolam o domínio dos valores possíveis (entre 0 e 1) apresentando coeficientes negativos ou superiores a 1. Já o pressuposto da linearidade suposto na AFC, defende que a relação entre o traço latente e as respostas dos sujeitos é linear, porém empiricamente nem sempre tal relação linear é corroborada.

A análise paralela é uma técnica que compara uma matriz de covariância dos dados observados com a de uma matriz de covariância gerada a partir de uma matriz de dados aleatórios com as mesmas dimensões dos dados observados (COELHO, 2014). A hipótese testada (H_0) é de que os dados observados não diferem de dados aleatórios, ou seja, comportam-se de modo pouco correlacionado. A cada extração de fator explicativo de covariâncias não aleatórias essa hipótese nula é testada novamente com os resíduos de variância não explicados (pelos componentes ou fatores extraídos). Empregou-se para a análise paralela tanto a extração de componentes principais (ACP) quanto de fatores análise fatorial confirmatória (AFC). Sobre a diferença entre essas metodologias Laros explica:

Uma diferença fundamental entre os dois métodos é que a ACP trabalha com a variância total observada, enquanto a AFC trabalha somente com a variância partilhada dos itens (variância erro e variância única são excluídas). Em outras palavras, a ACP tem como ponto de partida a variância observada das variáveis [respostas aos itens] e a AFC a covariância observada entre as variáveis. Na AFC, os fatores são estimados para explicar as covariâncias entre as variáveis observadas, portanto os fatores são considerados como as causas das variáveis observadas. Em contraste, na ACP, os componentes são estimados para representar a variância das variáveis observadas de uma maneira tão econômica quanto possível. Os componentes principais são somas otimamente ponderadas das variáveis observadas, neste sentido, as variáveis observadas são consideradas as causas dos componentes principais. (LAROS, 2012, p. 167)

Em outras palavras, enquanto para a análise fatorial as respostas são o efeito/resultado da manifestação das proficiências (fatores/dimensões); para a análise de componentes principais as respostas são as causas das dimensões latentes.

Conforme Pasquali (2018, p. 121), a análise fatorial de Informação completa foi introduzida em 1981 por Bock e Aitkin, e detalhada por Bock, Gibbons e Muraki em 1988, e tem esse nome porque não utiliza parâmetros estatísticos que sumarizam a informação dos dados empíricos (tais como as correlações tetracóricas), outrossim trabalham com a frequência absoluta dos padrões de resposta tratados por Máxima Verossimilhança Marginal (*Marginal maximum likelihood* - MML) com auxílio do algoritmo de Esperança Maximização (EM). Ela emprega em cada padrão de resposta um produto da probabilidade de acertos multiplicada pela probabilidade de erro de cada item individual, que resulta na probabilidade de um dado sujeito com um dado *theta* apresentar tal padrão de resposta. Calcula-se o produto dos padrões de resposta para os diversos níveis de *theta* da população e multiplica-se pela probabilidade de ocorrência desse nível (probabilidade esperada pela curva normal). A quantidade de níveis de *theta* é equivalente ao número de pontos de quadratura definido para análise, quanto maior esse número, mais precisas as estimativas, porém mais lento o processamento. O somatório dos produtos descritos calculados para cada nível de *theta* da frase anterior resultará na probabilidade de ocorrência daquele padrão de resposta na população. Conforme Bock, Gibbons e Muraki (1988 *apud* OLIVEIRA, 2017, p. 32) a FIFA é mais vantajosa em relação às matrizes tetracóricas por: evitar coeficientes tetracóricos indeterminados para itens com dificuldades muito altas ou baixas, acomodar melhor o efeito de itens omitidos ou não alcançados, possibilitar o teste da significância estatística de fatores adicionais, além de ser uma técnica mais robusta contra a falta de normalidade dos dados (PASQUALI, 2018, p. 123). No entanto, vale manter uma ressalva apontada por Pasquali:

a análise fatorial se fundamenta em equações lineares e, ao que parece, as correlações entre os itens seriam normalmente de caráter curvilíneo. Conseqüentemente, a prova da unidimensionalidade do teste, que mede um e um só traço, fica difícil de ser cabalmente demonstrada. (PASQUALI, 1996, p.88)

2.6 Estado da arte da TRI e o ENADE

Foi realizada uma pesquisa sistematizada da Literatura pertinente às contribuições sobre o Exame Nacional de Desempenho dos Estudantes e especialmente sobre a aplicação da TRI às questões de conhecimento (geral e específico) do exame. Anteriormente à sistematização da pesquisa bibliográfica, foi realizada uma sondagem no mecanismo de busca “Periódicos Capes” (Primo da ExLibris) com o objetivo de selecionar as bases de dados com maior probabilidade de ocorrerem trabalhos que atendessem aos seguintes critérios de busca: Assunto OU Título: “ ENADE “ OR "Exame Nacional de Desempenho de Estudantes" OR "*National Student Performance Exam*". Com base nesta primeira sondagem, foram selecionadas as das bases de dados que apresentaram a maior quantidade de trabalhos, a saber: *Directory of Open Access Journals* (DOAJ, com 57 ocorrências na sondagem); SciELO (*CrossRef*, com 28). Além dessas bases, para escrutinar a publicação internacional sobre a temática, também se realizaram buscas nas bases Scopus (Elsevier), e *Web of Science* (WoS). Na continuidade, procedeu-se a consulta diretamente em cada uma das bases, em busca de artigos indexados, utilizando a seguinte estratégia de busca: “ ENADE “ OR "Exame Nacional de Desempenho de Estudantes" OR "*National Student Performance Exam*". A fim de complementar essa revisão, acessou-se a Biblioteca Digital Brasileira de Teses e Dissertações pelo sítio <http://bdtd.ibict.br/>, aplicando-se os mesmos critérios de busca.

A primeira classificação/seleção dos estudos foi realizada através da avaliação dos títulos e dos resumos (abstracts), e nessa etapa, foram coletados todos os resumos e dados bibliográficos disponíveis, identificados quanto ao método, área abordada número de cursos, instituições, anos (edições do ENADE), e dados tratados (foco/objeto/indicadores de estudo) e por fim selecionados os trabalhos cuja leitura do documento completo seria realizada, aplicando-se como critério de inclusão/exclusão foi a aplicação da TRI às questões da prova.

A partir da leitura dos resumos desses trabalhos realizou-se a seleção dos artigos sumarizados a seguir no estado da arte, utilizando os seguintes critérios:

1. Tratam do ENADE;
2. Empregam técnicas de psicometria para tratamento dos dados.

A análise do texto integral dos artigos incluiu aqueles resgatados pelas bases e que se enquadravam no escopo (critérios de inclusão e exclusão), e a partir desses foram levantados 26 trabalhos, após leitura completa desses, concluiu-se que apenas 19 atendiam aos critérios de inclusão-exclusão.

A Literatura que discute sobre o tema (aplicações da Teoria de Resposta ao Item aos resultados do ENADE) é constituída por artigos, teses, dissertações, capítulos de livros e resumos de eventos, todos com cunho quantitativo, a maioria publicado em periódicos, bases ou anais de eventos nacionais. Pelo escopo desse trabalho, esse resultado já era esperado, pois a maior parte do público interessado no ENADE é nacional. A fim de sintetizar as informações, serão apontados os objetivos do estudo, quais as metodologias de calibração utilizadas, qual a área de conhecimento e qual foi a(s) seção(ões) e as edição(ões) da prova analisada, o tipo de delineamento e principais resultados.

Landeira-Fernandez e Primi (2002) realizaram um estudo de correlação entre os conceitos de 113 dos 1.173 cursos de Psicologia que participaram do ENC em 2000 com as suas respectivas taxas candidato/vaga de seus vestibulares. Eles revelaram que 46% da variabilidade dos conceitos atribuídos pelo ENC pode ser explicada pela relação candidato/vaga, isto é, podem ser explicadas pelas diferenças de habilidades e competências específicas que os alunos apresentam antes de terem iniciado o curso e estarem sob influência deste. Também realizara a aplicação da prova do ENC de 2000 a ingressantes e concluintes para averiguar a validade preditiva dos itens. Empregando análise de regressão logística para identificar itens com boa validade (20 dos 39), ou seja, capazes de demonstrar uma interação significativa, e/ou aqueles que não demonstram diferenças significativas na variável “curso” entre os calouros. “A presença desse tipo de questões impede eventuais discriminações entre os alunos de diferentes cursos com demandas diferenciadas reduzindo assim a variância de constructo irrelevante e consequentemente aumentando a validade desse processo de avaliação” (LANDEIRA-FERNANDEZ; PRIMI, 2002).

Francisco (2005) realizou um estudo de caso por meio da aplicação da TRI, com a finalidade de verificar o desempenho dos concluintes do curso de Matemática

da Universidade Estadual do Centro-Oeste (UNICENTRO) no Exame Nacional de Cursos (ENC) no período de 2000 até 2003 (SCHER, et al, 2014). Por ser uma dissertação do programa de Métodos Numéricos em Engenharia, o autor desenvolve uma vasta dedução dos procedimentos matemáticos envolvidos na estimação dos parâmetros através dos modelos logísticos unidimensionais de um, dois e três parâmetros da TRI. Segundo o autor, o nível das proficiências/habilidades do grupo dos respondentes (UNICENTRO) é baixa, ($\theta \sim -0,3$), a dificuldade média dos testes é alta (de 1,2 a 2,2) e os índices de discriminação (“a”) médios nos três anos de prova estavam acima de 1. O autor propõe uma classificação dos itens quanto a dificuldade sem especificar o método ou os critérios ou referências para estabelecer tal classificação. E conclui que a dificuldade geral dos itens diminuiu no decorrer dos anos e a proficiências dos estudantes manteve-se constante, porém não realizou a equalização para garantir essa conclusão. Uma limitação desse estudo foi a amostra utilizada para a calibração dos itens. Foram analisadas quatro provas (de 2000 a 2003), com número de itens igual a 25, 40, 30 e 40, e com número de respondentes igual a 46, 59, 41 e 41, respectivamente (SCHER, et al, 2014). Não foi possível determinar se os dados usados na calibração são somente da amostra da IES estudada ou de todos os participantes do ENC.

Também utilizando o modelo de três parâmetros Pasquali (2005), realizou um estudo cujo objetivo era analisar os parâmetros psicométricos do ENC/Provão de Psicologia, aplicados nos anos de 2000 e 2001. Em 2000, participaram do exame 9.362 estudantes e em 2001 foram 9.201, os resultados indicaram um índice de discriminação (parâmetro a) médio de $a=0,518$ ($dp=0,273$) em 2000 e em 2001 a média do a das questões foi $a=0,510$ ($dp=0,227$), o que demonstra a precariedade do poder discriminativo dos itens. Já o parâmetro b, que refere-se à dificuldade dos itens, apresentou questões com índices muito discrepantes nos dois anos (ex: questões com b entre 8 e 24 dp). De acordo com os resultados obtidos, Pasquali (2005) conclui que as provas foram mal elaboradas, e apesar de uma significativa melhora em 2001, as questões não representavam a realidade do curso, ou seja, muito difíceis para a população avaliada.

Oliveira (2006) utilizou a TRI para proceder a uma análise das propriedades psicométricas da prova de Medicina do ENADE de 2004. A amostra utilizada foi de 8.124 sujeitos e as análises foram realizadas apenas com as questões objetivas tanto do componente de Formação Geral (FG) quanto do componente de Conhecimentos Específica (CE), totalizando 28 itens. Relata ter obtido os seguintes valores do coeficiente de fidedignidade KR-20 de Kuder e Richardson ($KR-20_{FG} = 0,391$, $KR-20_{CE} = 0,681$) e do Alfa de Cronbach ($\alpha_{FG} = 0,462$; $\alpha_{CE} = 0,834$). Para os modelos de 2 e 3 parâmetros a autora relata que muitos itens não atenderam aos critérios estipulados (resíduo inferior a 2,0 e limite crítico b (entre -2,95 e 2,95) e para a superior a 0,3) ML2P com 14 itens em desajuste, 9 itens no ML3P e apenas 3 no modelo de Rasch. A autora conclui que os dados demonstraram o melhor ajuste do Modelo Logístico Unidimensional de um parâmetro (Modelo de Rasch) nos dois componentes do ENADE pois os parâmetros de *infit* e *outfit* tinham valores críticos entre 0,70 e 1,30. Já

Nogueira (2008) aplicou a TRI para avaliar as questões da prova do componente de formação geral (FG) das edições do ENADE de 2004 e 2005, visando a estimar a proficiência dos estudantes, especialmente nos conteúdos especialmente relacionadas a conceitos estatísticos. A amostra utilizada foi de 403.512 estudantes de vários cursos que fizeram as provas de 2004 e 2005. Os resultados indicaram que as provas se ajustam ao modelo de Rasch, principalmente devido à baixa frequência de resultados inesperados de padrões de erros e acertos.

Em outra pesquisa, Primi, et al. (2010) buscaram detectar itens com funcionamento diferencial (DIF), ou seja, itens que apresentavam comportamento (CCI) muito diferentes entre dois grupos. O DIF pode ser definido como sendo a observação, em pessoas com a mesma habilidade, mas pertencente a grupos distintos (i.e. feminino vs. masculino, brancos e não brancos...), de uma chance diferenciada de acerto de um item. “Os parâmetros estimados pela TRI devem ser invariantes em relação a[*sic*] habilidade da amostra, mas quando isso não ocorre, constata-se a anomalia denominada DIF” (SCHER, et al, 2014). Para tanto, os autores utilizaram-se de um banco de dados composto por 26.613 estudantes

participantes do ENADE de Psicologia do ano de 2006, e comparou os grupos de ingressantes e concluintes, e também os grupos de IES privadas e públicas. Por meio da análise de Rasch e por meio da regressão logística, os autores puderam perceber que, quanto à distinção entre os estudantes ingressantes e concluintes, cinco itens apresentaram DIF, sendo que três itens favoreceram os ingressantes e dois itens favoreceram os concluintes. Já para os estudantes de universidades públicas e privadas, sete itens apresentaram DIF, sendo que 6 itens favoreceram estudantes de instituições públicas. Segundo os autores, um grupo menor de itens com DIF indicaram problemas em sua formulação, favorecendo os estudantes com menores níveis de habilidade. Contudo, a maior parte dos itens que apresentaram DIF também tinha por característica propriedades psicométricas adequadas. Dessa maneira, os autores puderam concluir a existência de um pequeno número de itens potencialmente capazes de produzir um viés à prova, haja vista que esses itens se referem mais aos sistemas de ensino do que ao desenvolvimento da formação do aluno (ingressantes ou concluintes) fator esse que apresenta relação direta e esperada com a proficiência. Para os autores o DIF pode indicar que o item está discriminando constructos secundários (ex: aspectos culturais não relacionados à proficiência, mas que gera distinções espúrias entre estudantes pertencentes a instituições públicas ou privadas de ensino) não relevantes para o objetivo primário do teste que é mensurar apenas a proficiência (constructo primário). Considerando pessoas com o mesmo nível de habilidade a partir do escore global na prova, os autores concluem que estudantes concluintes e os de universidade pública tendem a ter vantagem nos itens com DIF. Também discutem alternativas metodológicas para se tentar lidar com esse problema, como a resolução do DIF, e que o julgamento do viés precisa ser feito com base em uma discussão mais ampla do conteúdo e da análise da influência que esses itens têm na medida global obtida a partir das pontuações dos sujeitos (SILVA, 2011).

Com o objetivo de avaliar a prova de Formação Geral do ENADE por meio da TRI, Lopes (2011) analisou as respostas de 49.497 estudantes da área de Pedagogia que participaram do ENADE no ano de 2005 e equalizou com uma prova (simulado) aplicado a 260 estudantes de pedagogia de dois campi de uma IES. O

modelo empregado na análise dos dados foi o modelo de um parâmetro (Rasch), que permitiu indicar os itens com melhores propriedades psicométricas para comporem os itens de ligação do “simulado”. Também para esse estudo, o modelo de Rasch foi o que mais se ajustou aos padrões de resposta. (LOPES, 2011).

Silva (2011) investigou o ENADE de administração de 2006 com o objetivo de discutir a validade das interpretações e a metodologia aplicada pelo INEP para avaliar a qualidade dos cursos. Participaram 71.838 ingressantes e 58.428 concluintes da área de Administração que realizaram o ENADE em 2006. A autora verificou-se a predominância de um fator geral apesar das diretrizes multidimensionais. Aplicando o mapeamento de itens de Rasch e o julgamento de um painel com cinco juízes, ela identificou três níveis, e posteriormente realizou interpretações de escala baseadas em critério. Por fim, aplicou modelagem de regressão linear hierárquica múltipla com características de perfil sendo a variável dependente a proficiência individual dos concluintes, como variáveis independentes no nível individual: o tipo de escola que frequentou no ensino médio, a escolaridade dos pais, renda familiar, forma de sustento e trabalho e o desempenho, e como variáveis independentes no nível do curso a média do desempenho no CE e no FG dos ingressantes. A autora constata que apenas 4% da variabilidade do desempenho dos concluintes pode ser atribuída ao “efeito das IES” e conclui que a avaliação realizada pelo ENADE apresenta pouca variabilidade para dizer que o exame avalia a qualidade do curso e as diferenças de competência. A partir desse trabalho, foi publicado o artigo de Primi, Da Silva e Bartholomeu (2018) cujo objetivo é identificar qual parcela da variância da prova do ENADE está associada ao efeito do curso. Os autores concluem que as notas do ENADE decorrem de diferenças importantes do nível de habilidade dos estudantes no momento de entrada no curso, e que no interior dos cursos a maior parte da variabilidade das notas do ENADE está relacionada aos perfis socioeconômicos dos sujeitos.

Segundo Silva (2011), os pontos centrais das críticas ao ENADE estão relacionados:

ao uso de uma metodologia transversal; a quantidade e formato dos itens diante da especificação[sic] de conteúdos, habilidades e competências proposta nas diretrizes do exame; a mensuração de

ingressantes e concluintes a partir da mesma prova; a falta de uma medida de controle inicial do conhecimento do estudante antes da entrada na instituição; a inexistência de estudos consistentes de valor agregado. (SILVA, 2011, p.18)

Do ponto de vista técnico, outra crítica feita ao ENADE se refere à profundidade em que as habilidades e competências podem ser avaliadas, comparativamente ao número reduzido de itens.

O estudo de Primi, Hutz e Silva (2011) teve por objetivo estabelecer critérios de desempenho com base na análise dos itens da prova de Psicologia do ENADE 2006, por meio do procedimento de referência ao item da TRI, ou seja, para a determinação de pontos de corte, formando grupos de competências requeridas para a resolução de itens. Para tanto, os pesquisadores utilizaram um banco de dados contendo as respostas ao ENADE de 26.613 estudantes da área de Psicologia. O instrumento utilizado foi a prova do ENADE, contendo 30 questões que mediam o Conhecimento Específico. Para estabelecer os pontos de corte das notas, foi utilizado o mapa de itens, o índice de separação⁹, e indicou que a prova tem precisão para distinguir de dois a três estratos significativamente diferentes. As análises dos resultados permitiram observar que 72,9% dos estudantes concluintes da área de Psicologia estão na Faixa 3, faixa considerada de compreensão mínima dos conteúdos abarcados pelo exame – aquém do que se espera de um profissional graduado nesse curso (PRIMI; HUTZ; SILVA, 2011). Os resultados mostraram que, de maneira geral, os estudantes concluintes concentram-se na competência mínima, enquanto que os ingressantes se distribuem, em sua maioria, abaixo da proficiência mínima.

Lopes e Vendramini (2013) buscam equalizar uma prova acadêmica de Pedagogia aplicada a 260 estudantes (grupo de interesse) de uma universidade particular do interior paulista com o ENADE 2005 via Teoria de Resposta ao Item, utilizam para isso os microdados das questões objetivas de CE de 49.497

⁹ O índice de separação é calculado pela razão da variância verdadeira (corrigida pela estimativa de variância de erro) com o erro de medida, indicando, portanto, a dispersão das medidas dos estudantes em unidades de erro padrão, ou seja, quantas vezes a dispersão verdadeira entre as medidas supera o erro padrão da medida

estudantes (grupo de referência). Para isso, selecionou 5 itens de FG e 11 de CE da prova objetiva do ENADE, com propriedades de item âncora, para servirem de itens de ligação, que juntamente com outros 24 itens inéditos constituíram a prova aplicada ao grupo de interesse. Apenas os 30 itens objetivos de conhecimentos específicos aplicados ao grupo de interesse, pela prova nomeada “PAD”, foram analisados no artigo. Os estudantes do grupo de interesse (260) foram agrupados entre ingressantes (16,2%), concluintes (23,9%), e intermediários (59,6%) que é uma categoria não avaliada no ENADE. Após a equalização as autoras fizeram uma comparação entre as notas padronizadas e equalizadas obtidas pelos estudantes (do grupo de interesse e de referência), segmentando-os em relação ao pertencimento a cursos com conceito ENADE (de 1 a 5) e em relação à categoria de ingressantes e concluintes. A população do grupo de interesse apresentou ingressantes assemelhados aos ingressantes dos cursos conceito 2, e concluintes assemelhados aos estudantes concluintes dos cursos conceito 4. Nota-se uma ausência de detalhamento sobre qual o procedimento estatístico utilizado para equalização dos testes.

Coelho (2014) apresenta uma boa revisão das bases matemáticas empregadas na TRI e conduz testes de comparações entre diferentes modelos (1, 2 e 3 parâmetros) uni e multidimensionais aplicados a três bases de dados de diferentes tamanhos ($n=94, 198, 436$), sendo o terceiro banco de dados referente ao ENADE de Estatística aplicado em 2009. A análise dessa terceira base de dados também fundamenta o artigo de Coelho, Ribeiro Junior e Bonat (2014). Os autores tratam como uma única prova objetiva a junção dos itens de conhecimentos específicos e de formação geral e parte do pressuposto de que se trata de um instrumento multidimensional. Utiliza análise fatorial paralela para verificar a dimensionalidade do constructo (mas não apresentam KMO). Para comparar os modelos, emprega AIC e BIC, sem comparação com dados da literatura, nem fundamentação em índices de adequação (*goodness-of-fit*) com critérios de corte reconhecidos tradicionalmente aplicados à análises da TRI (i.e. CFI, GFI, RMSEA, SRMR), e apenas sugere a exclusão de questões sem apresentar o resultado da calibração do modelo após a exclusão de itens. Por fim, a autora compara a

classificação ranqueada dos respondentes e seus escores (TCT e do modelo da TRI previamente selecionado) patronizados. Conclui que “os conjuntos de dados, foram sugeridos e aceitáveis por análises de diagnóstico estatístico, conforme o modelo, por representarem práticas em estudo, com ajustes e estimativas de parâmetros precisos e adequados”. A partir desse trabalho surgem as publicações de Coelho, Ribeiro Junior e Bonat (2014a, 2014b) em que os autores relatam os resultados e conclusões obtidos mediante a análise da prova do ENADE aplicada à área de estatística 2009. Concluem pela exclusão de dois dentre os 32 itens analisados, e que em geral, a prova apresenta boa coerência interna com a maioria dos itens, bem como boa aderência ao instrumento de medida proposto.

Scher (2014) analisa a prova do ENADE 2009 da área de Administração, utilizando o modelo logístico de três parâmetros para tratar a junção dos itens de conhecimentos específicos e de formação geral (35 itens – 8 excluídos pelo coeficiente ponto-bisserial) como uma única prova objetiva. Os autores não apresentam resultados sobre a dimensionalidade e fiabilidade do teste, mas segundo eles “Os resultados apontaram a viabilidade da utilização da TRI nos itens avaliados na prova do ENADE e observaram também o ganho em relação à proficiência entre os ingressantes e concluintes”, a concentração da informação da prova estava no intervalo entre 0 e 3 desvios padrões.

Da Mata (2014) analisa as provas do ENADE de 2005 e 2008 da área de Engenharia Civil, utilizando os modelos uni e multidimensionais da TRI. Um diferencial de seu trabalho é a ampla fundamentação teórica que busca conceituar e distinguir desempenho, competências, habilidade, aprendizagem e assenta-se nas teorias da aprendizagem (especialmente as teorias de aprendizagem de Gagné, a taxonomia de Bloom (1956, 1983) e as operações de pensamento de Raths, et al. (1977), articuladas semanticamente à síntese CHC das habilidades humanas proposta por McGrew com base nos modelos teóricos desenvolvidos por Cattell, Horn e Carroll e Spearman). Com isso, o objetivo de fundo do trabalho da autora foi identificar o que as duas provas efetivamente medem (habilidades ou memória). Através das análises de conteúdo fundamentadas nas teorias acima descritas, a

autora questiona a supervalorização das questões de prova às quais são atribuídas múltiplas habilidades, mas para as quais não há avaliação da sua dimensionalidade pelo INEP. A conclusão de da Mata (2016) sugere que os resultados obtidos na análise psicométrica pela TCT e pela TRI não se mostraram complementares, tal como apontado na literatura sobre essas duas vertentes de análise estatística, indicando informações díspares nos índices de ajustes, índices de facilidade, discriminação, cargas fatoriais e comunalidades. Também, que as provas contêm pouca informação, os dados não apresentaram bons ajustes psicométricos na maioria dos modelos analisados, o índice de consistência interna (*alpha* de Cronbach) indicou pouca confiabilidade dos dados, e que a teoria CHC talvez ainda não apresente a amplitude necessária para mapear e explicar as habilidades relacionadas ao conceito de domínio de área.

Lopes e Vendramini (2015) buscam avaliar as propriedades psicométricas da prova de Pedagogia do ENADE 2005, por meio do modelo de Rasch (ML1P), utilizam para isso os microdados das questões objetivas de 49.497 estudantes, sendo 7 questões de FG e 26 questões do CE, não tendo avaliado dois itens excluídos pelo INEP. Os dois componentes da prova foram calibrados separadamente, os critérios avaliados foram a correlação item-total, as medidas de *infit* e *outfit*, o índice de dificuldade “b” e a carga residual. Foram obtidos pelas autoras 7 itens com correlação item-total abaixo de 0,3, mas nenhum abaixo de 0,2. Concluem que os estudantes concluintes apresentaram um desempenho melhor (média = 49,4) do que o desempenho dos estudantes ingressantes (média = 41,5), que a prova apresenta boas propriedades psicométricas e que apresentou um nível de dificuldade semelhante ao nível de habilidade dos avaliados.

Camargo, et al. (2016) analisam a prova do ENADE 2012 de Ciências Contábeis utilizando o modelo logístico de três parâmetros para tratar a como uma única prova objetiva, a junção dos itens de conhecimentos específicos e de formação geral (35 itens – 8 excluídos pelo r_{pb}). Compara as médias dos estudantes distinguidos conforme a origem escolar, escolaridade dos pais (superior), horas de estudo extraclasse e natureza administrativa da IES. Conclui que os itens contidos na prova ENADE representaram um alto grau de dificuldade para o grupo que

realizou a prova. Constrói subdivisão em níveis de proficiência baseado em ancoragem de itens (intervalos e 1 dp), não explicitando a consulta a especialistas dentre os métodos.

Oliveira (2017) analisa as provas de formação geral dos ENADE de 2010, 2011 e 2012, realiza análise de dimensionalidade por meio da análise fatorial de correlações tetracóricas com o método dos resíduos mínimos, e por ela obtém evidências de unidimensionalidade nos três instrumentos, também apresenta análise da consistência interna dos instrumentos, constatando valores pequenos para o α . O autor conclui que a nota referente às questões objetivas desse componente não deveria ser utilizada para a tomada de decisões quando calculada à luz da TCT. Também comparou o ajuste de modelos logísticos de três e quatro parâmetros (TRI) e verificou-se que o último teve melhor ajuste aos itens objetivos. Já quanto aos itens discursivos, propôs a calibração pelo modelo de resposta gradual, sugerindo ajustar a correção dos itens discursivos em bandas de competência, e ressalta a vantagens desse tipo de item, relacionada a cobertura de uma maior faixa de competências. Por fim, sugere a adoção de um modelo de calibração de formato misto, permitindo calibrar simultaneamente os itens objetivos (ML4P) e discursivos (Gradual) em uma mesma escala.

Houve ainda alguns trabalhos não lidos na íntegra por serem de difícil acesso (apenas em livros esgotados e revistas impressas não assinadas pela UFSC, ou em monografias não disponibilizadas digitalmente), mas sobre os quais há menções nos trabalhos de outros autores, e por isso, serão brevemente descritos a seguir:

Lopes e Vendramini (2015) citam Vendramini (2009) que investigou a dimensionalidade do componente de Formação Geral dos ENADE de 2004, 2005 e 2006 em seu trabalho intitulado “Análise Fatorial baseada na TRI do componente de Formação Geral do ENADE” publicado na Revista de Educação ANEC em edição não disponibilizada na internet. Como principal resultado, as autoras apontam que os coeficientes de Kuder-Richardson foram baixos (0,353, 0,393 e 0,255, respectivamente para os anos de 2004, 2005 e 2006), e uma estrutura fatorial

complexa com muitos fatores, resultados já esperados, haja vista a abrangência da prova de Formação Geral e a pequena quantidade de itens da prova de FG.

Molk e Calderon (2014) citam Veiga (2009) que em sua dissertação de mestrado investigou a prova de pedagógica do ENADE cuja edição não foi possível determinar. Conforme Molk e Calderon (2014), a autora empregou modelos de 1 parâmetro (Rasch) para analisar as respostas de 49.497 estudantes e relacionar as habilidades e competências supostamente por eles adquiridas. “Constatou que a prova do ENADE engloba todas as habilidades necessárias para a boa formação do pedagogo” (*Idem*).

Lopes (2011) cita Bartholomeu, et al. (2011) que analisaram a prova de direito do ENADE 2006 com os objetivos de estabelecer pontos de corte em face ao procedimento de referência ao item, com o emprego do modelo de Rasch, e avaliar os estudantes de cursos de Direito que responderam a uma prova com itens de ligação selecionados do referido ENADE. Os sujeitos da pesquisa foram 73.106 estudantes que realizaram o ENADE no ano de 2006 e 1.311 estudantes da Universidade particular agrupados em dois *campi* distintos.

Também Lopes (2011) cita Silva, et al. (2011) que conduziram um estudo sobre o ENADE 2006 aplicado à área de Administração (com 141.689 sujeitos) e para avaliar o desempenho dos cursos de uma IES particular do interior paulista, aplicaram uma prova com itens de ligação selecionados do referido ENADE a outros 916 estudantes desse mesmo curso, divididos em quatro *campi*. Os estudantes do grupo focal apresentaram média de habilidade inferior à média de dificuldade dos itens, indicando que a prova foi difícil para a população avaliada.

3 MÉTODOS

3.1 Coleta de dados

A coleta de dados foi realizada mediante a aplicação de um teste padronizado pela equipe de aplicação do ENADE 2017. Não houve necessidade de coleta de dados primários com os estudantes, logo, nenhuma forma direta de obtenção de informações (questionário, teste, entrevista) pessoais ou acadêmicas foi empregada.

O instrumento utilizado foi a prova de Conhecimentos Específicos do ENADE, apenas as questões objetivas foram analisadas, ou seja, as questões de 9 a 35, ou 27 das 40 questões do teste; tal escolha justifica-se pelo escopo de análise desse trabalho, que prevê a avaliação unidimensional do constructo (proficiência em Conhecimentos Específicos de Ciências Biológicas) por dicotomização (respostas certas ou erradas), tendo sido descartadas pela análise: as questões do componente de Formação Geral, as questões discursivas (estas, por não seguirem uma correção intrinsecamente dicotômica) e o questionário de percepção da prova.

Ao conjunto de dados no menor nível de desagregação (indivíduo) convencionou-se denominá-los microdados. Os microdados relativos ao ENADE, inclusive a edição 2017, estão disponíveis na página eletrônica do INEP, no endereço <http://portal.inep.gov.br/microdados>. Nesses arquivos estão as bases de dados de cada ano, que contêm os padrões de respostas às questões objetivas, as respostas dadas aos questionários complementares, além de dados pessoais e institucionais estão dispostos nas colunas para cada um dos participantes (dispostos nas linhas). Além disso, nos arquivos há um dicionário de variáveis (que descrevem as legendas das colunas e dos valores mais comuns das variáveis) e das sintaxes para leitura das bases de dados em alguns programas estatísticos.

Os microdados (BRASIL, 2018b) selecionados para tratamento quantitativo foram os vetores de respostas contendo a sequência das alternativas escolhidas pelos participantes nos itens de Conhecimentos Específicos da prova do ENADE 2017 (uma linha para cada indivíduo, 27 símbolos posicionados sequencialmente, sendo um para cada questão: “A”, “B”, “C”, “D”, “E”, “.”, ou “*”). As questões não

preenchidas (em branco) foram representadas pelo INEP como “.” (caractere de ponto final), o qual foi substituído pelo código “P” no pré-tratamento de dados (pois nesse trabalho o algoritmo empregado para correção dos padrões de resposta, as marcações com “P” são consideradas erradas); quando o participante realiza marcação ambígua (duas alternativas) o INEP representa pelo código “*” (recodificado para “P”). Na fase de pré-tratamento de dados, foram desmembrados os vetores de resposta de cada participante, de modo que a cada item correspondesse uma coluna separada por “;” (ponto e vírgula). Justifica-se tal recodificação e desmembramento dos vetores para atender à sintaxe da linguagem R (R Core Team, 2018) e do programa R Studio (Rstudio, 2018), empregados nas análises. Os sujeitos com respostas em branco (*missing*) em mais de 24 das 27 questões objetivas de Conhecimentos Específico não foram computados na calibração dos parâmetros, mas foram alvo de estatística descritiva para melhor caracterização amostral.

3.2 Tratamento e análise dos dados

A análise da consistência da correção de prova realizada com base nos vetores de respostas (ABCDE) presentes nos microdados foi realizada subtraindo-se a matriz de acertos e erros (0 e 1) gerada pelos autores, da matriz gerada pela desagregação do vetor de acertos e erros produzido pelo INEP nos microdados “DS_VT_ACE_OCE”. Caso os padrões de resposta forem idênticos, a subtração das matrizes deverá resultar em uma matriz nula (todos os elementos iguais a 0), sendo esse o resultado obtido, evidenciando a adequação do método de pré-tratamento de dados empregado nesse trabalho.

A análise do índice de discriminação de cada alternativa, calculados à luz da TCT, pode fornecer pistas sobre aprimoramentos para as diretrizes de construção dos itens. Conforme Borgatto e Andrade (2012), o índice de discriminação clássico da TCT codificado sob a letra “a” (também chamado Índice-D por PRICE, 2016, p. 184), analisa para determinado item, diferença da porcentagem de acertos do grupo de estudantes com melhor (27% melhores) subtraída da porcentagem de acertos do grupo pior desempenho (27% piores). Espera-se que, para um item com boa

qualidade, a porcentagem de acerto seja maior para o grupo com melhor desempenho, e 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. Por outro lado, constituem comportamentos anômalos e que demandam análise *ad hoc* de especialistas, casos em que os valores da discriminação dos distratores são positivos e maiores que a discriminação do gabarito, ou seja, quando as proporções de escolhas dos distratores pelos indivíduos pertencentes ao extrato (27%) de alto desempenho supera a proporção dos indivíduos do extrato (27%) de baixo desempenho. No entanto, conforme explica Vilarinho (2015), a discriminação considera os 27% superiores como um grupo homogêneo, sendo que para itens com proporção de acertos inferior a 27% é esperado que o coeficiente de discriminação não seja sensível, pois dilui-se o efeito da proficiência na heterogeneidade do grupo.

Também foi replicado o cálculo das propriedades psicométricas da TCT com os microdados de 2017 (BRASIL, 2018b) e os valores comparados com os resultados apresentados no Relatório de Área do ENADE 2017 de Ciências Biológicas – Licenciatura (BRASIL, 2018a).

O índice de discriminação da TCT, que é utilizado pelo INEP para estimar a capacidade de discriminação das questões do ENADE desde 2004, é o denominado coeficiente de correlação ponto-bisserial, o qual é derivado do coeficiente de correlação linear de Pearson. Sua principal suposição é que a relação entre o acerto de um determinado item e o escore total do teste é linear (correlação positiva). Valores menores para correlações ponto-bisserial (item-total) revelaram uma tendência de escolha da opção errada pelos participantes que obtiveram os maiores escores; ou seja, há uma baixa correlação entre as respostas dadas para o item e o escore total do componente avaliado (no caso Conhecimentos Específicos), um forte indicativo de que o item deve ser retirado do instrumento. De acordo com Lira (2004), este método é indicado quando uma das variáveis (Y) é dicotômica e a outra (X) é contínua [ou ao menos discreta]; onde a variável contínua X, representa a característica de interesse para o estudo [traço latente ou escore] e a variável dicotômica Y representa os grupos [de participantes que acertaram ou erraram o item]. (LIRA, 2004)

Um problema de ordem prática é que o INEP adota o coeficiente ponto-bisserial como critério de eliminação de itens, mas em um teste com pequeno número de itens como é a prova de conhecimentos específicos ($I < 30$), um viés de superestimação desse parâmetro poderia reduzir a confiabilidade global do teste pela manutenção de itens de fraca discriminação (ver seção 4.3). Nesse sentido, PRICE (2016, p. 187) recomenda que seja feita uma correção no caso de testes compostos de menos de 25 itens - o item analisado deve ser retirado do computo do escore total no teste, evitando um efeito espúrio devido à grande proporção de variância explicada pelo desempenho no próprio item sobre o escore total. Também Pasquali (2013) recomenda que para esse cálculo os pontos obtidos pelo acerto do próprio item, sob análise, devem ser retirados do cálculo dos escores, pois elevam espuriamente a correlação ponto-bisserial do item e a estimativa da consistência interna do teste como um todo. Nota-se a ausência de menção dessa correção na metodologia de cálculo descrita pelo INEP nos relatórios de área. Suspeitando-se da ausência dessa correção no relatório de área, replicou-se o cálculo do Coeficiente Ponto-Bisserial, e foi possível constatar, pela identidade entre os valores presentes no relatório e os valores obtidos pelo cálculo do coeficiente ponto-bisserial não corrigido (sem desconsiderar a auto correlação dos itens), que a equipe técnica do INEP não considera relevante tal viés de estimação. Prescindida tal correção, resulta que os coeficientes bisseriais (r_{pb}) divulgado nos relatórios síntese de área são superestimados, fato também reportado por Oliveira (2017, p. 36) para o componente de formação geral nos ENADE de 2010, 2011 e 2012.

Um outro coeficiente utilizado para tratar a discriminação de testes à luz da TCT é a correlação bisserial, que conforme Peng (2004) é extensivamente utilizado quando um item que não é naturalmente binário, ou seja, não é do tipo verdadeiro ou falso, foi artificialmente dicotomizado para correção da prova. Tanto a correlação ponto-bisserial quanto a correlação bisserial são medidas estatísticas que estimam a correlação do resultado de um item em particular com o resultado do teste (isto é, o escore bruto total), sendo, portanto, medidas da capacidade de discriminação do item em relação ao resultado do teste (SOARES, 2005). Conforme Santo (1978), a correlação bisserial tenta estimar qual seria a correlação fosse levado em conta o

conhecimento parcial sobre a questão, isto é, se houvesse probabilidades de proficiência em um item para além das alternativas 0% ou 100%. A explicação é que, quando se atribui zero para uma resposta errada e 1 para uma resposta certa, está-se simplificando o mundo real, portanto não se pode dizer que os participantes que assinalaram uma das alternativas incorretas tinham absoluto desconhecimento, ou não tinham informação alguma. A correlação bisserial pressupõe a existência de uma distribuição normal de conhecimentos sobre o item, que é artificialmente dicotomizado em 0 ou 1. Conforme Lira (2004), o que difere o “Coeficiente de Correlação Ponto-Bisserial” do “Coeficiente de Correlação Bisserial” é que, no primeiro, a variável Y é originalmente dicotômica, não necessitando ser contínua e nem normalmente distribuída (BUNCHAFT e KELLNER, 1999 *apud* LIRA, 2004). Em outras palavras, o Coeficiente Bisserial estima o Coeficiente de Pearson entre o escore total na prova X e a variável “latente”, subjacente ao item. (LIRA, 2004) utiliza-se então a hipótese de que, subjacente à resposta de cada item, existe uma variável “latente”, contínua e normal, que determina o resultado (certo ou errado, zero ou um) no item.

Também foi realizada a análise gráfica dos padrões de escolha às alternativas escrutinando-se as proporções de escolha (aos distratores e gabarito) para cada nível de desempenho (sujeitos agrupados conforme os seus escores brutos segmentados em quintis) objetivando-se induções sobre os efeitos que alguns padrões de construção de itens podem ter sobre os atributos psicométricos dos itens.

No que se refere à consistência interna da prova, utilizou-se como medida o coeficiente Alfa de Cronbach, que:

[...] estima quão uniformemente os itens contribuem para a soma não ponderada do instrumento, variando numa escala de 0 a 1. Esta propriedade é conhecida por consistência interna da escala, e assim, o α pode ser interpretado como coeficiente médio de todas as estimativas de consistência interna que se obteriam se todas as divisões possíveis da escala fossem feitas (CRONBACH, 1951 *apud* MAROCO; GARCIA-MARQUES, 2006)

O estudo da dimensionalidade está diretamente relacionado à validade do instrumento. Tabachnick e Fidel (2014) e Figueiredo Filho e Silva Junior (2010)

recomendam, previamente às análises fatoriais, a aplicação da verificação da medida de adequação amostral de Kaiser-Meyer-Olkin (KMO) que é um índice calculado pela razão entre “o somatório das correlações quadráticas” e o “somatório das correlações parciais quadráticas mais o somatório das correlações quadráticas”. Ainda conforme as autoras um critério de corte para dados fatoráveis seria $KMO > 0,6$. Sendo os dados compostos por respostas dicotômicas, a investigação da dimensionalidade inicia-se a partir da análise de componentes principais sobre a matriz de correlação tetracórica, análise paralela e, por fim, análise fatorial de informação completa.

Para a análise da dimensionalidade da prova de Conhecimentos Específico (CE), analisou-se inicialmente as correlações dos resíduos, e a coexistência de resíduos de variância inexplicados em um subconjunto de itens pode ser entendida como padrões de respostas inesperados, evidência de outras dimensões além da principal, e/ou mau funcionamento de alguns itens. Para isso, o programa efetua primeiramente a calibração da dificuldade dos itens e da habilidade dos sujeitos, e a partir disso são calculadas as respostas esperadas; em seguida, é feito o cálculo dos resíduos (resposta observada-esperada) e a preparação da matriz residual (em que se remove a influência da primeira dimensão); calculam-se as correlações entre esses resíduos para então permitir a análise de componentes principais dessa matriz de correlação entre resíduos.

Na análise fatorial convencional, a interpretação pode ser baseada apenas em cargas positivas. No julgamento dos resíduos da FIFA, a interpretação deve ser baseada no contraste entre cargas positivas e negativas. Como o programa mostra os resultados dos componentes de resíduos, o resultado do primeiro componente (dimensão) já foi removido e passa-se a analisar as dimensões secundárias, ou também chamados de componentes dos contrastes.

Ao interpretar o significado de um componente ou um fator, a abordagem convencional sugere olhar para as maiores cargas positivas, a fim de inferir o significado substantivo do componente. Na análise de resíduos, é necessário identificar os agrupamentos de resposta nos contrastes e interpretar o significado dos componentes. O primeiro contraste pode tanto representar outra(s) dimensão(ões) dependendo dos valores encontrados (se os autovalores forem

maiores que 2,0), ou mesmo uma mudança de intensidade do constructo principal (SILVA, 2011).

Nesse trabalho as análises de dimensionalidade foram realizadas com a utilização do pacote psych (REVELLE, 2017) do software R.

Ao processo de estimação dos parâmetros dos itens baseado nas respostas dos questionários para aferição do instrumento, dá-se o nome de “calibração dos itens” (ANDRADE; TAVARES; VALLE, 2000). Nesse estudo, a calibração dos parâmetros dos itens foi operacionalizada pelo modelo logístico de três parâmetros (3PLM) da Teoria de Resposta ao Item – TRI. Justifica-se essa escolha metodológica pois compreende-se que para além de estimar a dificuldade (b) de cada item (questão) – o que poderia ser realizado pela TCT ou por modelos de Rasch, mantendo-se os demais parâmetros constantes dentre todos os itens - objetiva-se também avaliar a importância de diferenças em outras características psicométricas dos itens, tais como: diferentes níveis de discriminação (a), ou seja, a capacidade de cada item distinguir os respondente dentre portadores ou não de uma dada competência (tais variações de discriminação são tornadas constantes ($a=1$) em modelos de um parâmetro); também pode haver dentre os itens diferentes probabilidades de acerto casual (c). Esse último parâmetro não é estimado por modelos de 2 parâmetros, mas possui relevância em se tratando de testes educacionais, nos quais a qualidade dos distratores de cada um dos itens pode interferir diferencialmente na probabilidade de os indivíduos acertarem casuisticamente, sem necessariamente deter a proficiência que se intentou estimar.

Francisco (2005) explica que para a estimação conjunta dos parâmetros dos itens e das proficiências dos sujeitos avaliados não havendo uma métrica definida a priori, e por isso, pode haver mais de um conjunto de parâmetros capaz de produzir as mesmas CCI do modelo de maior verossimilhança (falta uma identificabilidade). Conforme Francisco (2005), Birnbaum (1968) propôs o uso de um processo iterativo (algoritmo de Newton-Raphson) em que o passo inicial é dar como estabelecidas as estimativas das habilidades (escores padronizados com média = 0 e desvio padrão = 1) para geração de uma primeira estimativa dos parâmetros dos itens, em um segundo passo, com essa estimativa provisória dos parâmetros (dados por conhecidos), recalcula-se as proficiências dos sujeitos em uma métrica que

reduza as estimativas de erro padrão (maximizam a verossimilhança). Repetem-se alternadamente os dois passos, substituindo-se os parâmetros e a proficiência gerados na última rodada até o momento que um critério de parada seja alcançado (geralmente quando as estimativas deixam de variar de uma rodada para a outra). Os processos de estimação dos parâmetros e proficiências são operacionalizados por sistemas de equações cuja resolução computacional faz-se por cálculos matriciais envolvendo inversões de matrizes de ordem $(P \cdot I + n)$ sendo “P” o número de parâmetros do modelo, “I” o número de itens, e “n” o número de sujeitos da amostra; a título de ilustração, nesse estudo tal sistema de equações resultaria em uma matriz quadrada com mais de 1,2 milhões de linhas e colunas. A exigência computacional para a inversão de matrizes dessa ordem desafiou os teóricos da TRI. Por isso Birnbaum (1968) propôs partir-se de três suposições iniciais: 1) os parâmetros dos itens são localmente independentes dos demais itens; 2) os parâmetros dos itens são independentes dos indivíduos; 3) as proficiências dos indivíduos são independentes dos itens. Com isso, as covariâncias dos itens com os indivíduos e entre si não são relevantes para as estimações de parâmetros e proficiências, e segundo Francisco (2005) a grande vantagem desse método é que ele permite, a partir da propriedade da independência local, que os itens sejam estimados individualmente, isso exige operações com 27 matrizes 3x3 para o ML3P, o que reduz drasticamente a exigência computacional.

O algoritmo de calibração do modelo confirmatório utilizado nessa pesquisa empregou como *prioris* os seguintes valores e distribuições:

parâmetro “a”) assume distribuição log-normal com log da média 0 e log do desvio padrão 0,5;

parâmetro “c”) é atribuída uma distribuição a priori para o logito de “c” do tipo normal, com média de tal logito = -1,386294 (que resultaria em $c \sim 0,2$) e desvio padrão 0,5. Vale a pena salientar que o parâmetro “c” é denominado no software R como parâmetro “g”. Assim, na codificação e sintaxe do pacote mirt (CHALMERS, 2012) a representação dos *prioris* do modelo é a seguinte:

```
"F = Q09-Q35L \nPRIOR = (Q09-Q35L, a1, lnorm, 0, 0.5), (Q09-Q35L, g, norm, -1.386294, 0.5)"
```

Nesse estudo, a estimação dos parâmetros dos itens (dificuldade, discriminação, acerto ao acaso) foi feita via Máxima Verossimilhança Marginal; e as proficiências dos sujeitos foram estimadas pelo método da Esperança a Posteriori (EAP). Conforme Francisco (2005), essa forma de estimação das proficiências dos sujeitos tem como vantagem a possibilidade de ser estimada diretamente sem o uso de processos iterativos.

Para o aperfeiçoamento da validade das interpretações do ENADE, e atendimento ao sexto objetivo dessa dissertação, é necessário avaliar a adequação dos ajustes do modelo unidimensional aos dados empíricos. Conforme (KLINE, 2005), um teste estatístico de adequação de modelo é um teste em que a matriz de covariância gerada pelo modelo estimado é comparada para saber se aproxima-se o suficiente da matriz de covariância gerada a partir dos dados empíricos de tal forma que as diferenças são tão pequenas que podem ser razoavelmente atribuídas a erros de amostragem. Se não for possível, então 1) as covariâncias dos dados contêm informações que estão contradizendo o modelo, e 2) esses resultados demandam ao pesquisador que explique as discrepâncias entre o modelo e os dados que tenham excedido aquilo que seria esperado pela aleatoriedade.

Para avaliar a adequação do modelo que está sendo testado tradicionalmente o qui-quadrado (χ^2), ou algum derivado, é o tipo de índice estatístico que é mais utilizado. Apesar desta estatística ter a vantagem de ter uma distribuição conhecida, o χ^2 é fortemente influenciado pelo tamanho da amostra, pela normalidade dos dados e pela complexidade do modelo. Para sanar tal desvantagem, muitos índices foram desenvolvidos. Porém, ao invés de confiar em apenas uma forma de mensuração, Nye e Drasgow (2011) afirmam que é altamente recomendado que pesquisadores examinem conjuntamente mais de um índice ao avaliarem seus modelos.

Para essa pesquisa, a avaliação da qualidade dos ajustes dos modelos foi feita pelas estatísticas RMSEA (erro quadrático médio de aproximação, em inglês *Steiger-Lind root mean square error of approximation*: STEIGER, 1990), SRMR (resíduo quadrático médio padronizado, em inglês *Standardized Root Mean Square Residual*), CFI (índice de ajuste comparativo, em inglês *Bentler's Comparative Fit*

Index: BENTLER, 1990) e TLI (índice de Tucker-Lewis), usando a função M2 do pacote mirt do software R. Conforme argumentam Cai e Hansen (2013), essas estatísticas têm a vantagem de ser pouco sensíveis ao tamanho da amostra, ao contrário do teste qui-quadrado de comparação de modelos, que, usualmente, rejeita a adequação de modelos quando a amostra é muito grande, mesmo nos casos em que os modelos são adequados.

Conforme Kline (2005), RMSEA, SRMR, TLI e CFI são todos índices de ajuste aproximados, e não delimitam onde terminam as variações ao acaso nos desvios de covariância entre o esperado pelo modelo e observado na amostra e onde começam as evidências de um modelo mal ajustado. Para isso, χ^2 seria o índice a ser usado, segundo o autor. Ainda segundo Kline (2005), RMSEA e SRMR são medidas absolutas e o RMSEA seria uma medida de falha no ajuste, pois, à medida que seus valores aumentam, piores são os ajustes e, portanto, um valor de RMSEA = 0 indicaria o melhor ajuste possível para um dado modelo (não necessariamente um ajuste perfeito).

O CFI é mais usado quando se espera distribuições assimétricas, ele compara o aprimoramento do modelo de referência relativo a um modelo nulo (variáveis independentes), que é tipicamente um modelo com correlações nulas. Um CFI = 1 costuma não significar um modelo perfeito, mas apenas que o χ^2 do modelo é menor que seus graus de liberdade. Porém Kline (2005) adverte que os índices incrementais baseados em modelos nulos de independência, como o CFI e o TLI, dizem apenas “quão bem o modelo em análise está se portando, comparativamente ao pior modelo possível” para a mesma amostra, e que evidenciar que seu modelo é significativamente diferente do pior modelo possível não é uma boa evidência de validade. Ainda Kline (2005), baseado no trabalho de Hu e Bentler de 1999, sugere utilizar o CFI sempre em conjunto com outro índice baseado em resíduos de correlações (como o SRMR), pois enquanto o primeiro é mais sensível a falhas na calibração das cargas fatoriais, o segundo é mais sensível a erros na estimação das correlações entre os fatores. Mas o autor ressalta que, se o conjunto de dados é pouco correlacionado, um índice de adequação incremental pode não ser muito informativo.

Nesse ínterim, cabe revisar algumas considerações sobre o valor dado à parcimônia dos modelos e a penalidade para o aumento de complexidade dos índices supramencionados, SRMR não acrescenta penalidades à complexidade, para valorizar a parcimônia dos modelos mais simples, para o CFI a penalidade é sempre igual a 1 independente da complexidade do modelo, já para o TLI e o RMSEA a penalidade, quanto ao aumento da complexidade, é dada em função da razão entre o qui-quadrado e os graus de liberdade (χ^2/gf).

Os metodologistas consideram o ajuste adequado quando RMSEA < 0,05 e TLI e CFI são maiores que 0,90 (Brown, 2015, p. 72 e 74). Já os estudos clássicos de Hu e Bentler (1998, 1999 *apud* NYE; DRASGOW, 2011. p. 551), baseados em estudos empíricos e de simulação, forneceram recomendações de parâmetros mínimos para muitos desses índices: RMSEA < 0,06, SRMR < 0,08, TLI > 0,95, e CFI > 0,95.

Para atender ao sétimo objetivo específico da dissertação, foram aplicados métodos de escalonamento comportamental baseados na TRI, e assim fornecer um sentido mais amplo às interpretações feitas sobre os escores do teste e definir melhor os significados dos vários pontos da escala. Segundo Silva, os métodos de escalonamento:

têm sido considerados fundamentais na apresentação dos resultados comparativos em muitos sistemas educacionais ao redor do mundo (Cizek, 1996; Impara & Plake, 2000; Shen, 2001; MacCann & Stanley, 2006). O Standard Setting pode ser caracterizado pelo estabelecimento de um ou mais pontos de corte na escala de avaliação. (SILVA, 2011, p.70)

Segundo Ubriaco (2009, p. 35), “ponto de corte” é o conceito operacional ao passo que “padrão de desempenho” é a versão conceitual do nível de proficiência. A mesma autora explica que, a determinação desses níveis de desempenho se baseia em três etapas. Primeiramente, determina-se (dentre consensos estabelecidos por especialistas de área) o número de níveis e da localização dos pontos de corte, que dividem a escala em categorias. Posteriormente, elabora-se descrição dos níveis de desempenho (PLD), etapa que envolve a seleção dos nomes dos níveis, a alocação dos itens nos níveis e a descrição pedagógica dos itens para caracterizar os níveis, nessa etapa é importante que a escala de proficiência gerada a partir da TRI seja

empregada no ordenamento dos itens. Dessa forma, buscou-se descrever o que os diferentes níveis da escala significam em termos teóricos e qualitativos e, conseqüentemente, em termos dos constructos avaliados. E, por fim, alocam-se as proficiências dos alunos, delimitando tais resultados em categorias/níveis.

Conforme Silva (2011), a referência no desempenho do grupo normativo (também conhecida como “referência na norma”), apesar de ser o mais utilizado (inclusive no ENADE), visa à comparação do resultado de um indivíduo com os de seu grupo-referência, mas não permite ao desenvolvedor do teste estabelecer *a priori* o que cada subgrupo disposto em níveis realmente é capaz de fazer. A referência no critério permite estabelecer *a priori* parâmetros que diferenciem os sujeitos em níveis de proficiência. E, por fim, a referência ao conteúdo é utilizada quando o conjunto de problemas presente no instrumento, ou em um banco de itens calibrados e equalizados, pode ser considerado uma amostra representativa do universo de problemas de um determinado conteúdo ou domínio.

É necessário esclarecer que o objetivo desta dissertação não é o estabelecimento de padrões de performance (*Standard Setting*), pois tal propósito demandaria a análise de um painel de juízes para estabelecimentos de padrões baseadas em critérios com emprego de metodologias próprias para esse fim (i.e. Bookmark de LEWIS; MITZEL, 1995, ou Método de ANGOFF, 1971). Conforme Cizek, Bunch e Koons (2004), a pergunta a ser respondida no estabelecimento de padrões de performance é qual o conhecimento mínimo dominado por um sujeito com um nível limítrofe de proficiência entre dois níveis, ou mais especificamente para o estabelecimento de padrões e níveis baseados em critérios: com base nos conhecimentos de área dos juízes, onde melhor se posicionam os limites entre níveis (básico, intermediário, avançado). Em outras palavras, a dissertação visa à construção de uma interpretação para uma escala referenciada na norma (assim como é o conceito ENADE), e não a proposição de padrões de desempenhos baseados em critérios.

Para a construção e interpretação da escala, foi utilizado o método de ancoragem de escala por suavização proposto por Beaton e Allen (1992) e as adaptações de Haberman, Sinharay e Lee (2011). Para um item ser considerado

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. A razão subjacente a esse tipo de procedimento é a pressuposição de que o nível de dificuldade dos itens reflete o nível de complexidade do conhecimento e competências avaliadas (CIZEK; BUNCH; KOONS 2004, SILVA, 2011).

Andrade, Tavares e Valle (2000) explicam o que é um item âncora por meio da metodologia “*Smoothing Scale Anchoring*” aplicada ao “*National Assessment of Educational Progress*” NAEP de 1990 por Beaton e Allen (1992, p. 194). Segundo esses autores, “níveis âncoras são pontos selecionados pelo analista na escala de habilidade para serem interpretados pedagogicamente. Já os itens âncora são itens selecionados, segundo a definição dada abaixo, para cada um dos níveis âncora” (ANDRADE; TAVARES; VALLE, 2000, p.110).

Conforme a metodologia de ancoragem na escala por suavização, para ser caracterizado como “item âncora” em um determinado nível da escala, o item deve atender a três critérios:

- 1) ser respondido corretamente por uma grande proporção de indivíduos com este nível de habilidade (pelo menos 65%, mas os autores sugerem a aplicação de até 80%)

- 2) a diferença entre a proporção de acertos dos dois grupos de indivíduos posicionados exatamente nos pontos de ancoragem (níveis âncoras, adjacentes) deve ser grande. Ou seja, a diferença deve ser superior a 30% seguindo-se a recomendação de Beaton e Allen (1992, p.202) ou superior a 15% seguindo-se a recomendação de Haberman, Sinharay e Lee (2011, p. 10).

- 3) ser respondido corretamente por uma proporção minoritária de indivíduos do nível de habilidade imediatamente inferior (no máximo 50%).

Beaton e Allen (1992, p. 194) advertem que, de toda forma, é importante que os pontos de ancoragem sejam selecionados criteriosamente. Se a distância entre os pontos de ancoragem que delimitam os níveis âncoras é muito pequena, não haverá itens que atendam os critérios de ancoragem e sejam em número suficiente para servirem de base à generalização (da interpretação). Ainda segundo eles, os

pontos de ancoragem devem estar distanciados o suficiente para que as respostas aos itens possam ser diferenciadas entre eles, mas não devem estar tão distantes que as interpretações sejam triviais. Devido à dificuldade de atendimento simultâneo dos três critérios propostos por Beaton e Allen (1992), no mesmo trabalho os autores já apontam uma classificação intermediária denominada “item quase âncora”, que nessa pesquisa definimos como o item que atende a pelo menos dois dos critérios supramencionados (alguns autores, ex. TREVISAN, et al. 2019, ainda exigem que os itens tenham ≥ 1).

3.3 Recursos Computacionais

Na estimação dos parâmetros dos itens e na estimação do traço latente dos indivíduos, utilizou-se o programa R Studio (Rstudio, 2018) em sua versão livre, optou-se pela linguagem R por possuir códigos-abertos, de tal forma que é possível codificar os algoritmos necessários, também por possuir pacotes matemáticos e estatísticos específicos para as modelagens psicométricas. Os pacotes utilizados para análise de dados foram o mirt (CHALMERS, 2012), psych (REVELLE, 2017).

4 RESULTADOS E DISCUSSÃO

4.1 Caracterização dos sujeitos da amostra e da população

Em 2017, foram convocados para o ENADE 17.381 concluintes da área de Licenciatura em Ciências Biológicas os quais estavam vinculados a 481 cursos de 296 IES diferentes (BRASIL, 2018a). No entanto, com uma taxa de abstenção de (19,12%) e suprimindo-se as provas anuladas pelo INEP e as provas objetivas entregues com 25 ou mais itens em branco ou anulados (em outras palavras, considerando-se apenas as provas com 3 ou mais respostas válidas), restaram 14.053 testes com dados válidos para as análises.

A maioria dos concluintes cursou presencialmente (86,40% dos convocados, 88,22% dos presentes), correspondendo a 450 dos 481 cursos, ou seja, 93,55% dos cursos são presenciais. Dentre os inscritos, 3.444 frequentavam Educação a Distância – (19,86% dos licenciandos) – e nesse subgrupo o índice de abstenção foi um pouco maior, 28,9%, o que resultou em uma proporção de 17,35% das respostas válidas pertencentes a concluintes vinculados a um dos 31 cursos EaD.

Quanto à categoria acadêmica, 68% dos presentes estavam vinculados a IES do tipo Universidade, seguido por Centros Universitários 12,24%, Institutos Federais 11,21% e Faculdades 8,65%¹⁰. Quanto à categoria administrativa, a maior parte (63,2%) dos concluintes estava matriculada em Instituições de Ensino Superior Públicas (sendo 40,1% em IES Federais e 23,1% em Estaduais) e 18,1% em Instituições Privadas sem Fins Lucrativos e 16% em IES com Fins Lucrativos.

Conforme Cortelazzo e Ribeiro (2013), os cursos de Ciências Biológicas apresentam características peculiares em relação ao conjunto de cursos de graduação oferecidos no país: são preferencialmente oferecidos em instituições universitárias, especialmente os bacharelados, e há uma participação maior das instituições públicas nessa formação, quando comparada aos outros cursos.

¹⁰ Em 2014, as faculdades respondiam por 13.03% dos convocados (n=3535), sendo fato digno de estudos posteriores essa repentina redução no número de concluintes (apenas 1864 indivíduos curavam em faculdades em 2017, os dados apontam para uma redução do número de cursos em faculdades de 56 para 43, e uma ampliação da proporção de IES privadas com fins lucrativos).

Contrariando o panorama de 2013 descrito pelos autores supracitados e demonstrando a evolução demográfica do curso, encontra-se no relatório de área um panorama distinto:

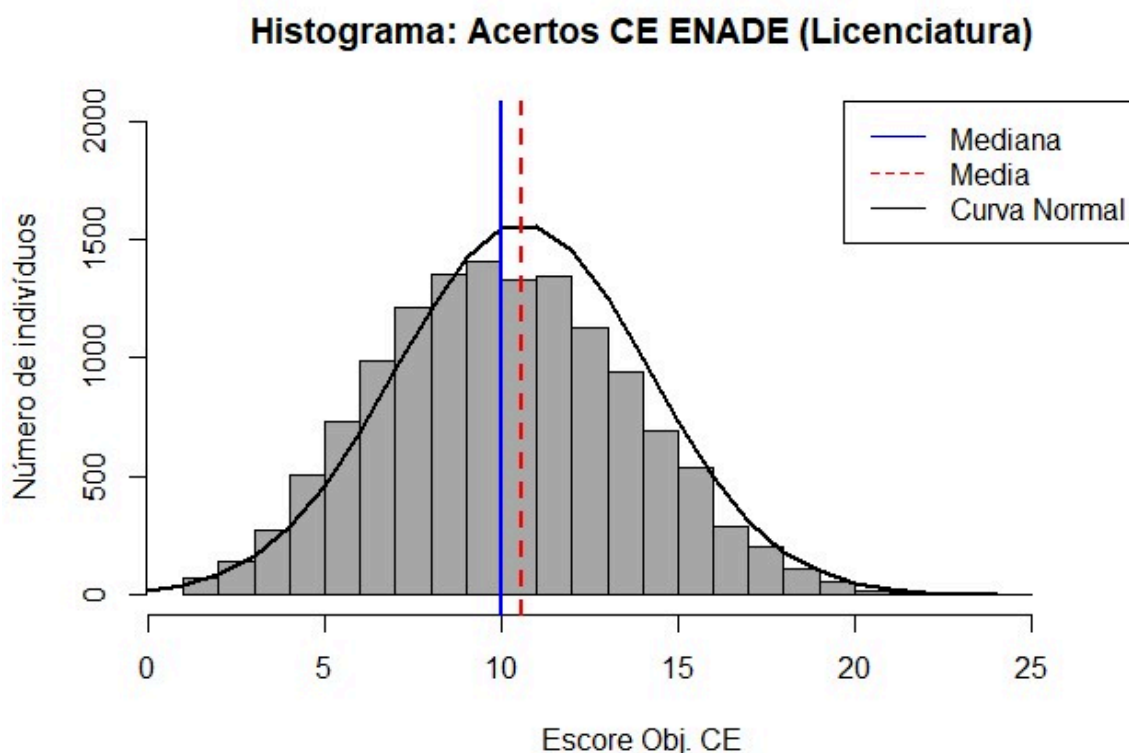
Considerando-se os cursos de Bacharelado e a Categoria Administrativa da IES, destaca-se a predominância das Instituições Privadas de ensino, que concentraram 153 dos 268 cursos de Ciências Biológicas, número correspondente a 57,1% dos cursos avaliados. Lembra-se que com respeito aos cursos de Licenciatura, a predominância é de Instituições Públicas, com 57,4% dos cursos. (BRASIL, 2018a)

As matrículas concentram-se em cursos localizados principalmente nas regiões brasileiras sudeste (34,7%), nordeste (30,7%) e sul (16,2%). Para um detalhamento maior sobre a demografia e distribuição geográfica dos convocados, consultar o relatório de Área (BRASIL, 2018a).

Quanto à avaliação dos estudantes sobre a dificuldade da prova, 51,5% (n=7.242) consideraram o nível de dificuldade mediano e 35,4% (n= 4.971) consideraram-na difícil. Quanto ao tempo de prova 47,8% consideraram adequado, 24,6% longa e 17,5% muito longa; ainda sobre o tempo 42% utilizou entre 3 e 4 horas para fazer a prova, 32% entre 2 e 3 horas 13% entre uma e duas horas, e apenas 8% afirmam que não ter conseguido terminar a prova em 4 horas. E a maioria (52,2%) dos respondentes considerou os enunciados das questões claros o suficiente para respondê-las. Quanto à percepção do estudante em relação ao domínio dos conteúdos para responder à prova de conhecimentos específicos: 54,2% afirma que estudou e aprendeu muitos desses conteúdos; 17,3% estudou a maioria desses conteúdos, mas não os aprendeu; e 13,9% estudou alguns desses conteúdos, mas não os aprendeu (BRASIL, 2018a).

Facultando aos leitores a análise gráfica da distribuição dos escores na amostra, apresenta-se, na Figura 3, um histograma contendo os dados empíricos e uma simulação de uma curva normal com média ($\bar{x} = 10,6$ dp = 3,58) calculados:

Figura 3: Histograma da distribuição de acertos (escore) obtidos pela amostra Concluintes de Licenciatura em Ciências Biológicas no ENADE 2017.



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

Com base no histograma apresentado na Figura 3, é possível inferir que a distribuição de frequências dos escores assemelha-se à normal, com uma ínfima assimetria nos valores superiores (cauda longa à direita).

A Tabela 4 apresenta um sumário de fácil consulta com as principais medidas de tendência central e dispersão, e parâmetros estatísticos da TCT calculados com base na amostra de concluintes da Licenciatura em Ciências Biológicas e sua interação com o teste de 2017:

Tabela 4: Sumário estatístico da amostra de Concluintes em Licenciatura em Ciências Biológicas e de sua interação com o ENADE 2017

Indicador	Valor (dp)	Indicador	Valor (dp ou EP)
N de itens	27	Escore máx. do grupo inf. (<27%)	8
N de examinandos	14.053	N indivíduos grupo inferior	3.935
Respostas válidas para análise da TRI	13.374	Escore mín. do grupo sup. (>73%)	13
Média (dp)	10,598 (3,578)	N indivíduos grupo superior	3.997
CV	33,76%	Dificuldade (b)% acertos	38,91%

		(\bar{x} e dp dos 27)	(14,74%)
Escore mín. CE	0	Discriminação pela TCT (<i>a</i>) (\bar{x} e dp dos 27)	0,1650 (0,1108)
Escore máx CE	25	Correlação ponto-bisserial (\bar{x} e dp dos 27 itens)	0,2783 (0,1042)
Quartis Q1 Q2(mediana) Q3	Escala (0- 27) 8 10 13	Correlação bisserial (\bar{x} e dp dos 27 itens)	0,2005 (0,1271)
		Alfa de Cronbach (dos 27 itens)	0,5633
<i>Missing values</i>	1,398 (0,37%)	KMO	0,8088

Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b). Os parâmetros em negrito são relativos ao modelo inicial.

4.2 Análise de conteúdo do teste

Conforme Pasquali (2013), a análise teórica de um item visa a verificar a compreensão das tarefas propostas no teste por parte dos sujeitos avaliados (análise semântica) e a avaliação da pertinência do item a tal ou tal unidade e avaliando tal ou tal processo cognitivo (análise de conteúdo pela CAA). Ainda segundo Pasquali (2013, p. 189), a validade de conteúdo segue passos bem definidos. Primeiramente, é preciso definir objetivos da avaliação, saber qual o domínio cognitivo que se quer avaliar, seja: conhecer, compreender, aplicar e analisar.

Para ter acesso à matriz de avaliação (*blue print*) elaborada previamente à “encomenda dos itens”¹¹ de 2017, solicitou-se tal documento junto ao INEP, o caminho indicado pelos servidores do instituto foi solicitar pelo portal de “acesso à informação”. Através desse canal de comunicação, um dos servidores do INEP respondeu que a informação está ao final do relatório de área, porém “a tabela está em outro formato”. Além disso, esclarece que “O que mais existir e não for visualizado no portal do INEP, ou na biblioteca local do INEP, faz parte do conteúdo de documento sigiloso” (INEP, 2019 Anexo B), porém, não menciona a entidade

¹¹ Forma como é denominada a demanda de elaboração de itens para os elaboradores do BNIES.

classificadora que considerou como documentos sigilosos a “matriz de referência” e os “manuais de treinamento de elaboração de itens”. Ainda em tal resposta, o agente público faculta a consulta presencial à biblioteca do INEP, ou mediante cadastro de projeto de pesquisa junto ao Serviço de Acesso a Dados Protegidos (SEDAP) para obtenção de acesso a tais dados pessoalmente, sugestão inviável a quem não reside próximo a Brasília.

Então, visando a contornar essa lacuna, a partir do relatório de área, foi possível deduzir e tentar reproduzir a matriz de referência empregada pelos membros da comissão assessora de área para a composição da prova de licenciatura. Como pode ser observado na reconstituição da matriz de referência 2017 (Apêndice A), observa-se a repetição da vinculação da característica de perfil “I. observador, crítico[a] e integrador ao interpretar e avaliar os padrões e processos biológicos e suas interfaces com outras áreas do saber” em nove itens (relacionou-se com 33% dos itens, 38% dos conhecimentos), e oito itens demandaram o emprego da competência “II. identificar e problematizar questões inerentes às Ciências Biológicas, incluindo seus aspectos científicos, históricos e filosóficos” (relacionou-se com 30% dos itens e 37% dos conhecimentos avaliados). Em contraste, a característica de perfil “II. comprometido com a produção de conhecimento, a transformação social e a educação emancipatória” e duas competências: “III. realizar consultorias educacionais e outras atividades afins definidas na legislação e em políticas públicas” e “V. realizar atividades de gestão educacional nos diversos espaços institucionais” não foram contempladas, ou ao menos identificadas pela CAA, em nenhum item da prova de Licenciatura.

Ainda conforme o relatório de área, a prova de licenciatura englobou 63 objetos de conhecimento, o que resulta em uma proporção de 2,33 conteúdos por item, e evidencia a intenção da CAA de vincular cada questão a mais de um objeto de conhecimento. Assim como Coelho (2014), cabe questionar a supervalorização das questões de prova, às quais são atribuídos múltiplos vínculos com objetos de conhecimento, sendo que muitos desses vínculos são circunstâncias, como, por exemplo, a identificação de que as questões Q23 e Q14 avaliariam conhecimentos de Zoologia, quando na verdade os poucos conhecimentos zoológicos servem

apenas de pano de fundo no texto base. Nesse ínterim, Troncon, Borges e Bollela (2018. p. 78) apontam “a necessidade de delimitação estreita e muito precisa do aspecto que vai ser abordado em cada questão” e ponderam que situações com múltiplos componentes podem dificultar a interpretação da avaliação. No mesmo sentido, mas contraditoriamente às diretrizes de prova que possibilitam tratar até três conhecimentos em uma mesma questão, o Guia de elaboração e revisão de itens do BNI do ENADE prescreve: “O item deve ser estruturado de modo que se configure uma unidade de proposição [...] explicitando uma única situação-problema e uma abordagem homogênea do conteúdo selecionado” (BRASIL, 2011. p. 9).

Sobre esse tema, a pesquisa reverbera os resultados já apontados em outras áreas e edições do ENADE por Primi, et al. (2010); Primi, Hutz e Silva (2011), Silva (2011), Verhine, Dantas e Soares (2006), sobre desproporções entre o número de itens da prova de CE (limitados pela tradição, normas e tempo disponível para o ENADE) e o número de competências e conteúdos que o teste se propõe mensurar. Isso tem resultado em itens em que há a pretensão de avaliar o desempenho em muitos conteúdos e lacunas, pois nem todos os conteúdos e suas nuances, profundidades e complexidades são devidamente distribuídos ao longo da prova. Também se constata a superestimação da gama de conhecimentos que as questões são capazes de individualmente avaliar, pois relacionamentos teóricos fugazes e circunstanciais são tratados no relatório de área (BRASIL, 2018a) com a mesma importância que os conteúdos centrais para o processo de resposta de cada item.

Conforme o relatório de área, no instrumento de conhecimentos específicos do ENADE 2017, a temática Ambiente e Saúde (OC11) foi contemplada em sete itens. Em contraste, dois objetos de conhecimento foram vinculados a apenas um item cada um “X. Educação Ambiental” - Q22, e “XII. Bioestatística” - Q24, sendo que o item 24 vinculado à Bioestatística foi eliminado do cômputo dos escores pelo critério do coeficiente ponto-bisserial (ver seção 4.3). Esses fatos relacionados à dificuldade de cobertura da abrangência de conteúdos presentes nas diretrizes de prova, mencionados nos parágrafos acima, são compreensíveis ao se considerar a pequena quantidade de itens (15 relativos a conteúdos biológicos, e 7 relativos a conteúdo de licenciatura) que em tese devem relacionar-se aos 18 objetos de

conhecimento previstos na matriz de referência das licenciaturas em ciências biológicas (potencialmente 22, considerando-se os três objetos de conhecimento constituídos de miscelâneas de duas ou mais áreas de conhecimento das ciências biológicas). A esses somam-se outros cinco itens relativos a conteúdo da matriz de referências da pedagogia, que tem outros 23 (objetos de) conhecimento. Mas o insucesso em abranger uma gama tão extensa de diretrizes pode comprometer a validade de conteúdo do teste devido à “sub-representação do constructo”, pela negligência a alguns objetos de conhecimento que contrasta com redundâncias e repetições de um mesmo (objeto de) conhecimento em mais de 5 itens, assim evidenciando falhas de amostragem de certos conteúdos, escolhas e ênfases diferenciais e contingentes, que mesmo compatíveis com o arbítrio cultural típico das decisões curriculares e avaliativas, pouco contribuem para a legitimidade da avaliação e resultam em uma interpretação restrita dos escores do teste.

4.3 Análise dos itens pela Teoria Clássica dos Testes

A importância dessa etapa do processamento, além da verificação de que a leitura dos dados foi feita corretamente, é que estas estatísticas são utilizadas posteriormente como valores iniciais para os processos de estimação realizados nas fases seguintes. Além disso, Andrade; Tavares; Valle (2000) afirmam que estatísticas como a correlação bisserial fornecem um diagnóstico preliminar dos itens, servindo, por exemplo, na identificação de itens com problemas no gabarito.

Inicialmente, nessa seção, serão discutidos os parâmetros utilizados para avaliar o poder de discriminação dos itens. O princípio implicado nesses parâmetros é o pressuposto de que para ser considerado apto a avaliar os alunos e os cursos, um item deve ser mais acertado por alunos que tiveram bom desempenho do que pelos que tiveram desempenho ruim (desempenhos inferidos pelos escores). Na Tabela 5, exposta a seguir, estão dispostos os principais parâmetros calculados pela TCT:

Tabela 5: Índices de discriminação e dificuldade dos itens do ENADE de Licenciatura em ciências Biológicas de 2017 calculados pela TCT.

Questão	Discriminação (a)	Dificuldade e (b)	Ponto-Bisseria I	P.Bisserial Corrigido	Bisserial	Alfa se excluído
Q09	0,254	62,97%	0,3624	0,2375	0,3049	0,5485
Q10	0,342	53,75%	0,4343	0,3113	0,3887	0,5383
Q11	0,179	43,26%	0,2902	0,1564	0,2003	0,5590
Q12	0,181	40,96%	0,3006	0,1685	0,2129	0,5578
Q13	0,145	33,79%	0,2749	0,1466	0,1823	0,5613
Q14	0,077	33,20%	0,2104	0,0801	0,1077	0,5685
Q15	0,164	29,97%	0,2913	0,1679	0,2219	0,5578
Q16	-0,033	24,90%	0,0906	-0,0305	-0,0378	0,5803
Q17	0,333	46,26%	0,4269	0,3029	0,3784	0,5395
Q18	-0,058	13,51%	0,0262	-0,0689	-0,1077	0,5801
Q19	0,164	42,82%	0,2862	0,1524	0,1926	0,5598
Q20	0,125	29,14%	0,2564	0,1324	0,1765	0,5621
Q21	0,358	54,95%	0,4378	0,3154	0,4050	0,5366
Q22	0,066	28,50%	0,1957	0,0706	0,0938	0,5696
Q23	0,117	46,84%	0,2460	0,1091	0,1435	0,5650
Q24L	0,017	21,89%	0,1450	0,0294	0,0406	0,5733
Q25L	0,038	19,92%	0,1500	0,0389	0,0541	0,5720
Q26L	0,090	26,55%	0,2219	0,1001	0,1363	0,5658
Q27L	0,111	25,69%	0,2308	0,1106	0,1548	0,5642
Q28L	0,263	61,42%	0,3637	0,2378	0,3109	0,5476
Q29L	0,214	37,54%	0,3310	0,2027	0,2606	0,5530
Q30L	0,068	19,07%	0,1836	0,0747	0,1098	0,5679
Q31L	0,253	43,79%	0,3539	0,2241	0,2869	0,5497
Q32L	0,239	44,96%	0,3463	0,2156	0,2756	0,5509
Q33L	0,267	53,39%	0,3638	0,2344	0,3043	0,5476
Q34L	0,226	39,49%	0,3387	0,2097	0,2724	0,5516
Q35L	0,254	72,10%	0,3578	0,2428	0,3349	0,5477

Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

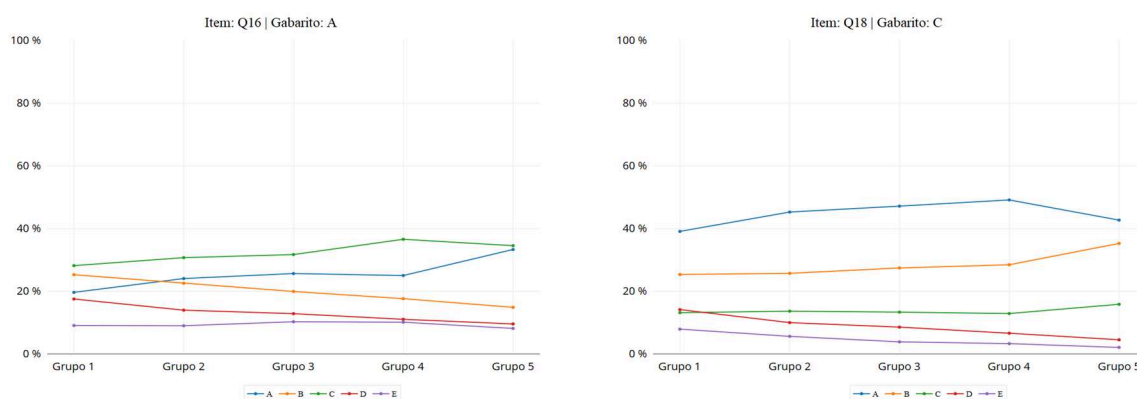
As questões que apresentaram maior índice de discriminação foram Q10 e Q21 e estão relacionadas, respectivamente às temáticas de “fisiologia” e “imunologia e genética”. Já as questões Q18 e Q16 apresentaram índices de discriminação negativos, ou seja, no grupo com maior escore os indivíduos acertaram menos aquela questão, proporcionalmente aos de menor escore, fato indicativo de item confuso ou com erro de gabarito.

Na seção 4.5.1.1 serão discutidos os critérios de eliminação e manutenção de itens pela TRI, e será traçado um paralelo com índices da TCT expostos na

Tabela 5, por hora basta saber que para o INEP “questões com índice de discriminação [ponto-bisserial] fraco, com valores $\leq 0,19$, são eliminadas do cômputo das notas” (BRASIL, 2018a), sendo assim, pelo critério do índice ponto-bisserial, foram eliminados do cômputo do escore final da prova de CE da licenciatura os itens 16, 18, 24, 25 e 30 no ENADE 2017 (ver Tabela 5).

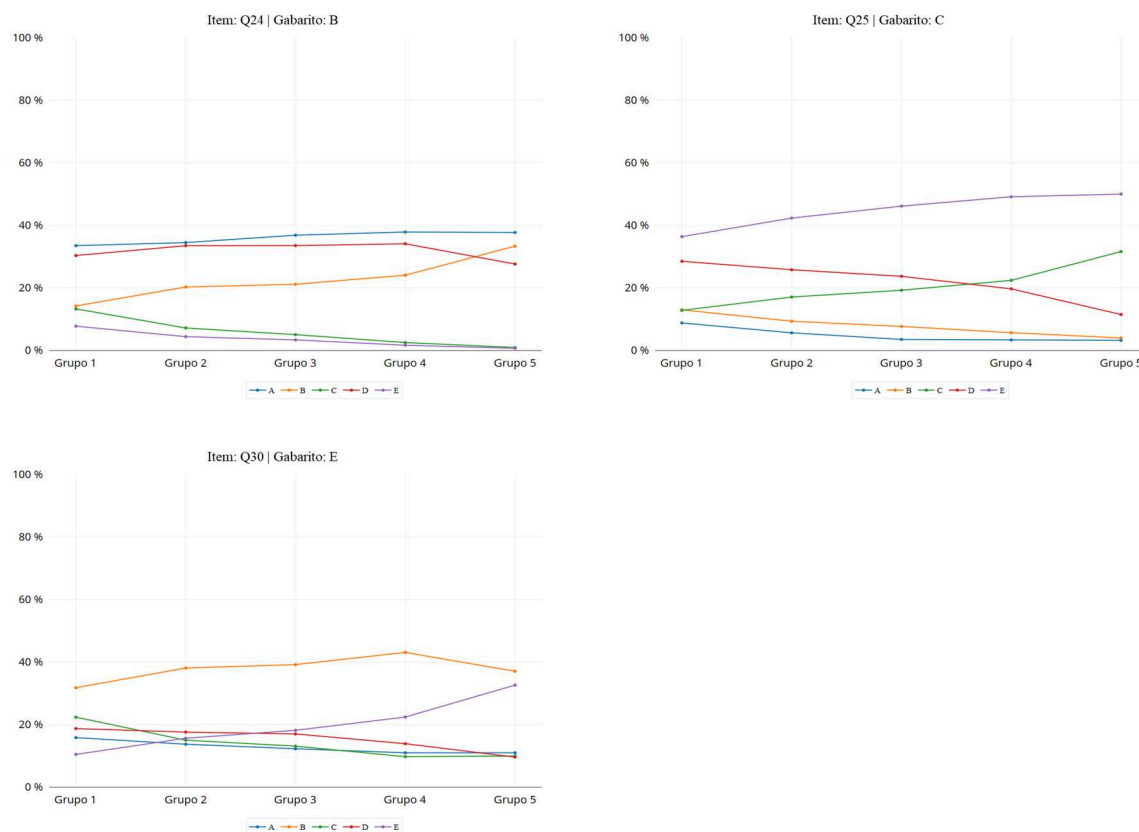
As curvas dos itens supracitados, geradas a partir da TCT pela proporção de escolha dos itens (eixo y) pelos indivíduos agrupados em quintis (eixo x) de escore crescente (esquerda) e a curva característica do item da TRI (direita), estão representadas nos gráficos (Figura 4 e Figura 5), expostos a seguir:

Figura 4: Distribuição de frequências de escolha das alternativas em quintis de participantes agrupados em níveis crescentes de escore dos itens eliminados pelo INEP da prova de conhecimentos específicos da área de Licenciatura em Ciências Biológicas. Itens com conteúdo relativos à biologia: Q16 e Q18



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

Figura 5: Distribuição de frequências de escolha das alternativas em quintis de participantes agrupados em níveis crescentes de escore dos itens eliminados pelo INEP da prova de conhecimentos específicos da área de Licenciatura em Ciências Biológicas. Itens com conteúdo relativos à formação de professores: Q24L, Q25L e Q30L



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

Observa-se, dentre os itens eliminados pelo INEP, dois padrões característicos: o primeiro padrão, representado pelos itens Q16, Q18, os quais apresentam pouca informação para a gama de proficiências da população avaliada (fato que será ilustrado na próxima seção pelo modelo unidimensional de 3 parâmetros pela CCI) e tem como característica identificável um padrão crescente em duas das alternativas (o gabarito e um distrator) na curva de frequência de escolha das alternativas pelos estudantes pertencentes aos quintis de escore (imagens à esquerda). O segundo padrão é composto por itens de elevada dificuldade (acima do máximo de proficiência da amostra), mas que ainda detém índices de discriminação meritórios, o padrão de “dupla atratividade crescente” apesar de ainda presente nos quintis de escore, exhibe no último quintil uma leve

inflexão evidenciando uma ampliação na frequência de escolhas do gabarito pelo grupo pertencente ao quintil superior.

Conforme o INEP (BRASIL, 2018a), a classificação dos itens, quanto ao índice de dificuldade, foi estabelecida da seguinte forma: das 27 questões, uma questão foi considerada muito difícil ($\leq 15\%$), 13 foram as difíceis (16% a 40% de acertos), outras dez de nível mediano (41% a 60%) e três fáceis (61% a 85%). Nota-se que, conforme o critério do INEP, uma questão muito difícil deve ser acertada por uma proporção de participantes inferior a 15%, sendo que, em itens com cinco alternativas, a proporção de acertos ao acaso frequentemente é superior ao previsto em tal critério, questiona-se se a existência de tal item é plausível na prática, pois nesse caso implica que os distratores possuem atratividade superior à alternativa certa, fato que, em conjunto com fracos índices de discriminação evidenciaria um potencial erro de gabarito ou item mal elaborado.

4.3.1 Consistência interna

Os valores do Alfa de Cronbach para todos os itens objetivos do componente de Conhecimentos Específicos foram 0,566 na Licenciatura, o que pode ser considerado ($0,5 \leq \alpha < 0,6$) moderado segundo preconizado por Landis e Koch (1977), mas inaceitável conforme Murphy e Davidsholder (1988 *apud* MAROCO; GARCIA-MARQUES, 2006) e pobre segundo George e Mallery (2003). Pelo critério do índice ponto-bisserial, conforme os critérios do INEP, foram eliminados os itens Q16, Q18, Q24L, Q25L, Q30L (ver anexo A) após tais eliminações o α atingiu 0,602.

Comparando esses valores com valores presentes na literatura, Primi, Hutz e Silva (2011) obtiveram um valor de Alfa de Cronbach igual a 0,76 para as 30 questões da prova de conhecimentos específicos do ENADE de psicologia em 2006. Vendramini (2009), analisando os itens de Formação Geral das provas do ENADE de 2004, 2005 e 2006, obteve valores de KR20 (análogo do Alfa) iguais a 0,353, 0,393 e 0,255, respectivamente. Mas quais as consequências dos baixos valores dos parâmetros de Consistência Interna dos Instrumentos do ENADE?

“São essencialmente duas as consequências directas[sic] de uso de dados com fraca fiabilidade: (a) Existe elevada probabilidade da medida não ser válida – O resultado pode nada dizer sobre o constructo que se pretendia medir. Mas mesmo se a medida for válida, (b) O erro de medida é elevado, pelo que a variabilidade observada afecta[sic] o poder de qualquer teste estatístico realizado, aumentando a probabilidade de resultados não-significativos.” (MAROCO; GARCIA-MARQUES, 2006)

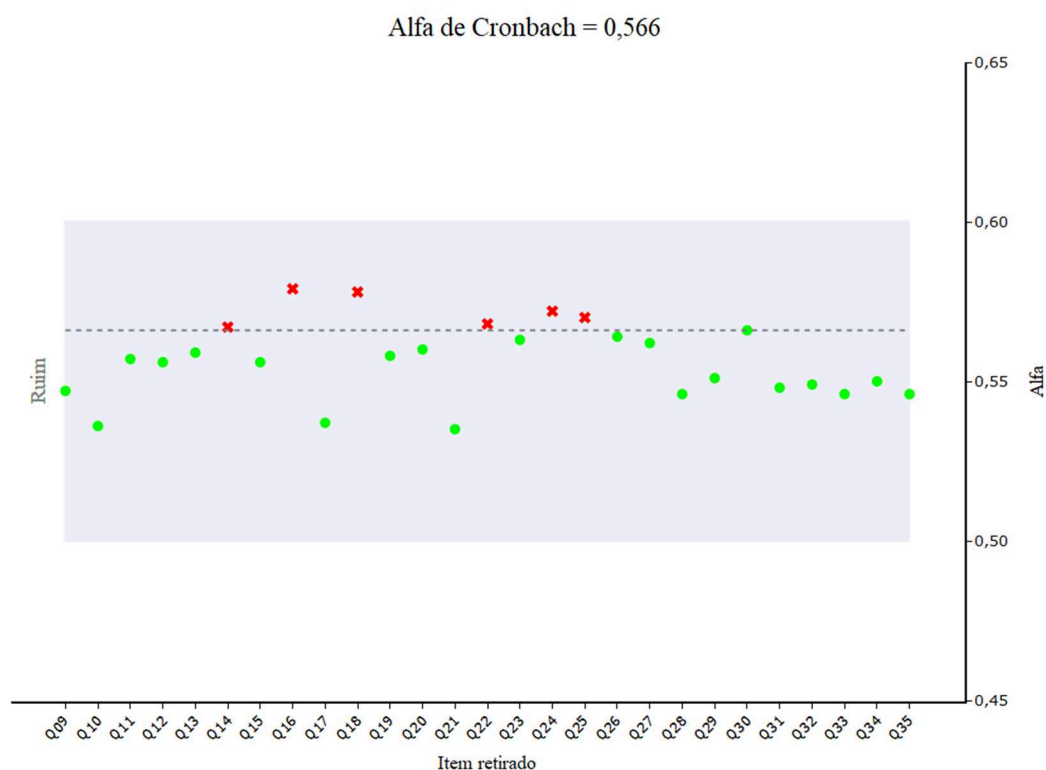
Como conciliar a baixa consistência interna e, ao mesmo tempo, a presença de um único fator geral indicando a unidimensionalidade do instrumento?

[...] Resultados como esse são frequentes em provas educacionais (*achievement*) e são bem diferentes dos padrões que os pesquisadores que lidam com testes e escalas psicológicas estão acostumados a ver. Nelas, um princípio frequentemente adotado é a maximização da consistência, isto é, a construção de várias questões medindo uma mesma dimensão subjacente. Já as provas educacionais são construídas com princípios fundamentalmente diferentes, pois elas têm como propósito maximizar a representação de um domínio multidimensional. Isso se reflete nos índices de consistência. (PRIMI; HUTZ; SILVA, 2011)

Tal resultado já era esperado, pois reflete não só a natureza complexa do constructo avaliado (proficiência em um teste que engloba a maior parte das áreas de uma disciplina científica, conforme as diretrizes do curso), mas também a história de composição do campo de conhecimento das ciências biológicas, que segundo Pedroso e Selles (2015) formou-se no início do século XIX pela unificação de campos científicos com origens epistemológicas distintas, em comunidades científicas próprias e inicialmente não inter-relacionadas (disciplinas morfofisiológicas [citologia, embriologia, fisiologia, anatomia] e naturalísticas [zoologia, botânica] unidas pela teoria sintética da evolução cuja hegemonia discursiva sempre esteve em disputa). Conforme as autoras, “a unificação das Ciências Biológicas não foi produzida de modo consensual nos meios acadêmicos, [mas na escolas] parece ter incorporado em grande parte essa idéia[sic] ao constituir uma nova disciplina escolar – a disciplina escolar Biologia – em substituição às disciplinas escolares separadas que estavam presentes pelo menos até a metade do século XX no país.” (FERREIRA; SELLES, 2015)

Na Figura 6, estão representados os valores de Alfa na hipótese de exclusão dos itens objetivos da prova de 2017, retirados hipoteticamente um a um:

Figura 6: Gráfico ilustrativo do Alfa de Cronbach do instrumento (linha tracejada) e estimativas do Alfa para as hipóteses de retirada de cada um dos 27 itens objetivos de Conhecimentos da área de Licenciatura em Ciências Biológicas no ENADE 2017.



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

Observa-se que os itens que contribuem para a redução da consistência interna, e que se hipoteticamente retirados do instrumento resultariam na elevação do Alfa de Cronbach, e da fiabilidade do instrumento da Licenciatura, foram as de código Q14, Q16, Q18, Q22, Q24L, Q25L e estão identificadas como “x” vermelho na Figura 6. Cabe lembrar, que por meio das análises da TCT, percebeu-se que tais itens são em geral ou muito difíceis ou pouco discriminativos.

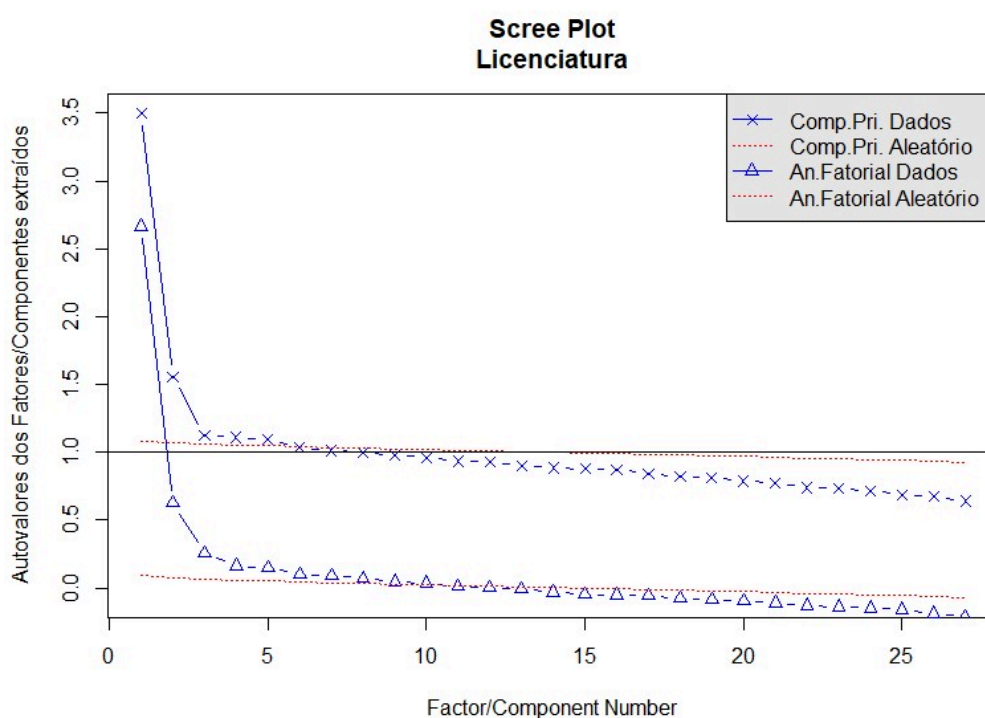
4.4 Dimensionalidade

Tabachnick e Fidel (2014) e Figueiredo Filho e Silva Junior (2010) recomendam, previamente às análises fatoriais, a aplicação da verificação da medida de adequação amostral de Kaiser-Meyer-Oiklin (KMO), tendo o valor de 0,6 ou superior com critério de corte para uma adequação da base de dados à análise

fatorial. Como para a amostra de estudantes de licenciatura do ENADE 2017 o KMO foi 0,8088, valor superior ao critério de corte e, portanto, considera-se que os dados da prova de conhecimentos específicos de 2017 são fatoráveis.

As análises paralelas fatorial e de componentes principais tiveram o objetivo de verificar a pertinência da utilização de modelos unidimensionais. Por essa técnica, todo autovalor que estiver acima da linha vermelha (simulação com dados aleatórios), indica que a dimensão é importante para explicar a variância, ou seja, conforme a Figura 7, a análise paralela indica de duas a três dimensões:

Figura 7: Gráfico de autovalores após a extração de componentes principais e fatores (*scree plot*) da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

No instrumento da licenciatura, o primeiro componente explicou 12,95% da variância total na análise de componentes principais, o que segundo Reckase (1979, p. 227), não indica unidimensionalidade, pois o primeiro fator deve responder por mais de 20% da variância total.

Para estimar o número adequado de fatores/dimensões que deveriam ser extraídos e testados na análise fatorial de informação completa, aplicou-se o teste Very Simple Structure (VSS)¹² do pacote Psych (Revelle, 2017) do software R. Tal teste apontou pelo critério de MAP de Velicer que os dados podem ser adequadamente representados por um único fator, já o critério de informação bayesiano ajustado pelo tamanho da amostra (SABIC), atingiu o mínimo de -1149,24 em três fatores, indicando ser esse o número de fatores mais eficiente a testar-se.

Portanto, considerando-se que a técnica do *scree plot* (Figura 7) aponta a existência de 2 fatores predominantes, e que o critério de BIC recomenda verificar a adequação de 3 fatores, foram realizadas análises fatoriais de informação completa testando-se modelos logísticos de 3 parâmetros com uma, duas e três dimensões, e os resultados das cargas fatoriais (F_n), com n sendo o número da dimensão e comunalidades (h^2) obtidos em cada um desses modelos pela rotação 'oblimin' está representado na Tabela 6, a seguir:

¹² Para mais informações ver REVELLE, William; ROCKLIN, Thomas. Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. **Multivariate Behavioral Research**, v. 14, n. 4, p. 403-414, 1979. Disponível em: <http://www.test.personality-project.org/revelle/publications/vss.pdf>

Tabela 6: Análise fatorial de informação completa com simulações para um, dois e três fatores/dimensões explicativas para a variância da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017

Item	Unidimensional		Bidimensional			Tridimensional			
	F1	h2	F1	F2	h2	F1	F2	F3	h2
Q09	0,388	0,150	-0,353	0,152	0,150	-0,029	0,214	-0,352	0,169
Q10	0,587	0,345	-0,586	-0,003	0,344	-0,046	0,077	-0,580	0,367
Q11	0,259	0,067	-0,227	0,120	0,067	-0,004	0,169	-0,238	0,078
Q12	0,282	0,080	-0,259	0,104	0,079	-0,079	0,114	-0,207	0,080
Q13	0,597	0,357	-0,619	-0,063	0,386	-0,026	0,033	-0,624	0,406
Q14	0,181	0,033	-0,185	-0,064	0,038	-0,021	-0,043	-0,163	0,034
Q15	0,314	0,099	-0,264	0,364	0,207	-0,065	0,381	-0,220	0,207
Q16*	-0,156	0,024	0,398	0,507	0,407	-0,118	0,411	0,516	0,442
Q17	0,675	0,455	-0,697	0,052	0,490	-0,104	0,126	-0,637	0,497
Q18*	-0,575	0,331	0,357	-0,378	0,277	0,108	-0,400	0,283	0,282
Q19	0,353	0,124	-0,322	0,014	0,104	-0,134	0,004	-0,230	0,107
Q20	0,448	0,200	-0,408	-0,034	0,167	-0,099	-0,010	-0,340	0,165
Q21	0,539	0,291	-0,490	0,205	0,286	-0,235	0,182	-0,319	0,282
Q22	0,519	0,270	-0,460	-0,246	0,267	-0,115	-0,227	-0,361	0,249
Q23	0,185	0,034	-0,206	-0,014	0,042	-0,022	0,013	-0,194	0,043
Q24L*	0,782	0,612	-0,510	-0,409	0,418	-0,194	-0,455	-0,448	0,552
Q25L*	0,428	0,183	-0,167	-0,190	0,063	-0,011	-0,169	-0,156	0,060
Q26L	0,378	0,143	-0,381	0,097	0,156	-0,195	0,067	-0,238	0,155
Q27L	0,207	0,043	-0,064	0,583	0,345	-0,048	0,562	-0,036	0,328
Q28L	0,398	0,159	-0,282	0,363	0,216	-0,280	0,281	-0,073	0,214
Q29L	0,350	0,123	-0,291	0,200	0,128	-0,267	0,122	-0,089	0,132
Q30L*	0,398	0,158	-0,039	0,633	0,404	0,011	0,602	-0,055	0,355
Q31L	0,363	0,132	-0,254	0,349	0,190	-0,225	0,291	-0,089	0,188
Q32L	0,362	0,131	-0,336	0,117	0,128	-0,370	-0,011	-0,053	0,161
Q33L	0,392	0,154	-0,353	0,151	0,150	-0,525	-0,050	0,047	0,241
Q34L	0,405	0,164	-0,334	0,284	0,197	-0,243	0,241	-0,155	0,201
Q35L	0,454	0,206	-0,410	0,175	0,202	-0,561	-0,039	0,017	0,296

Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018a) Nota: *=Itens excluídos do computo do escore pelo INEP.

Conforme Hair Júnior (2014, p.115), “um dos pressupostos da análise fatorial é a estrutura simples de seus componentes, ou seja, a mesma variável não deve contribuir para a construção de fatores distintos” (FIGUEIREDO FILHO; SILVA JUNIOR, 2010). Sendo assim, as cargas fatoriais consideradas significativas foram as superiores a $|0,3|$ e estão destacadas em negrito e vermelho, e os casos limítrofes (entre $|0,3|$ e $|0,2|$) apenas negritados.

Conforme Figueiredo Filho e Silva Junior (2010), as comunalidades representam a proporção da variância para cada variável incluída na análise que é explicada pelos componentes extraídos. Como critérios de corte para a comunalidade adotou-se a recomendação de Coelho (2014, p.97), que afirma que “para que um item se ajuste, é necessário que sua comunalidade seja maior que 0,3.” (aceitando-se como limítrofes até 0,2). Observa-se pelo resultado das análises fatoriais de informação completa apresentadas acima que as comunalidades dos itens são em geral bastante baixas, o que “é um indício de que os itens não estão linearmente correlacionados e, por isso, não devem ser incluídos na análise fatorial.” (FIGUEIREDO FILHO; SILVA JUNIOR, 2010). Porém a recomendação dos autores de exclusão de itens com baixa comunalidade não poderia ser levada em conta, caso contrário, observando-se o melhor dos cenários da análise fatorial, a validade de conteúdo ficaria comprometida, tendo em vista que restariam apenas quatro dos itens testados acima.

Ao realizar a análise fatorial de informação completa pressupondo o modelo unidimensional (existência de um fator predominante), os itens Q16 e Q18 apresentam cargas fatoriais negativas, uma evidência de que são itens com alguma anomalia não tratável por um modelo unidimensional, pois suas cargas fatoriais estariam inversamente proporcionais à proficiência, ou seja, discriminando positivamente os estudantes que não dominam a proficiência avaliada. Esses dois itens predominaram na extração de resíduos dos modelos unidimensionais, porém dado o exposto na seção anterior, não há motivos para acreditar que estejam avaliando adequadamente conhecimentos ou competência relevantes aos futuros professores, e por esse motivo recomenda-se a retirada desses itens do cômputo das proficiências.

As cargas fatoriais do ML3P bidimensional revelaram que a dimensão predominante F1 engloba dezesseis itens com conteúdo das ciências biológicas (Q9 a Q23) e de saberes pedagógicos (i.e. Q24L, Q26L, Q29L, Q32L, Q33L, Q34L, Q35L) os quais compartilham a característica comum de exigir precisão no uso da linguagem e atenção a detalhes (pegadinhas) especialmente em situações que exigem aplicação assertiva de princípios teóricos. A segunda dimensão latente é

composta por apenas oito itens (Q15, Q16, Q18, Q24L, Q26L, Q27L, Q30L, Q31L), sendo cinco questões relativas a saberes pedagógicos e três, a saberes biológicos. Em hipótese, a característica distintiva desse fator é demandar a “aplicação de raciocínio analógico na interpretação e compreensão do sentido global e informações implícitas no enunciado e nos textos bases”. Tal dimensão secundária mostrou-se quase independente da dimensão principal, pois a correlação entre os fatores F1 e F2 no modelo bidimensional foi de -2,3% (rotação ‘oblímin’, ver Tabela 7).

Tabela 7: Somatório quadráticos das cargas fatoriais de cada fator/dimensão e correlação entre os fatores nos modelos uni, bi e tridimensional

		Uni	Bidimensional		Tridimensional		
Rotação	Indicador	F1	F1	F2	F1	F2	F3
<i>Não rotacionado</i>	$\Sigma(Cf^2)$	5,068	3,922	1,986	2,795	1,612	1,883
<i>Não rotacionado</i>	%Variância	18,8%	14,5%	7,4%	10,4%	6%	7%
<i>Oblimin</i>	$\Sigma(Cf^2)$	5,068	3,795	2,091	1,212	1,828	2,55
<i>Oblimin</i>	Corr_F1	1	1	-0,023	1	-0,205	0,577
<i>Oblimin</i>	Corr_F2					1	0,104

Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b) Nota: $\Sigma(Cf^2)$ =Somatório quadráticos das cargas fatoriais de cada fator/dimensão; Corr_Fn Correlação entre o fator n e o fator identificado em cada coluna.

Já a análise do modelo tridimensional possibilitou inferir uma distinção mais contrastante entre os itens carregados em F3 com enfoque em conteúdos disciplinares das ciências biológicas e raciocínio lógico (Q09, Q10, Q13, Q17, Q20, Q21, Q22, Q24L) e os itens carregados em F1 que exigem mais de capacidades relacionadas à interpretação de textos por uso de pensamento analítico (Q29L, Q32L, Q33L, Q34L, Q35L), e os itens carregados em F2 que exigem a compreensão e interpretação de textos pelo uso de pensamento analógico (Q15, Q27L, Q28L, Q30L, Q31L).

Conforme o Padrão 1.25 da APA, AERA e NCME:

quando consequências indesejadas resultam das análises e usos de um teste, podem realizar-se tentativas de investigação como tais consequências surgiram a partir da sensibilidade do teste a características distintas daquelas às quais pretendia-se avaliar, ou dos motivos porque o teste falhou como um todo em representar o constructo. (AERA; APA; NCME, 2014, p.30, tradução livre)

Visando-se a iniciar tais investigações na próxima seção, serão analisados os itens com parâmetros, estimados pelo ML3P unidimensional, cujos valores são insuficientes segundo os critérios de corte discutidos na metodologia. No mesmo intento, também se realizou um exercício de análise discriminante, em que se avaliam as relações entre as variáveis do constructo pretendido e as de um constructo similar, mas do qual espera-se que o constructo enfocado apresente distinções¹³. Para isso, foram incluídos os itens objetivos de formação geral na análise fatorial de informação completa para verificar se na estimação das cargas fatoriais (ver Apêndice B) alguma das questões elaboradas com base na matriz de referência do ENEM (cf. BRASIL, 2007) apresentaram carga fatorial alta nos mesmos fatores do CE. Tal exercício evidenciou que os itens ímpares mais relacionados a raciocínio lógico-matemático, precisão no uso da linguagem e conhecimentos científicos estão associados ao fator 3, já itens pares de formação geral que demandam interpretação textual associam-se mais fortemente ao fator 1, não houve no componente de formação geral nenhum item que tenha carregado maior variância no fator 2 (compreensão de texto). Sendo assim, fazem-se pertinentes as palavras de Landeira-Fernandez e Primi:

O problema detectado nessa análise preliminar refere-se à questão de variância confiável de constructo irrelevante, significando que: “a avaliação é muito ampla, contendo excesso de variância confiável associada com constructos distintos... que afeta as respostas de maneira irrelevante ao constructo sendo interpretado” (Messick, 1995, p. 742). Resultados de uma avaliação de conteúdos específicos de psicologia devem ser influenciados prioritariamente pelo domínio do conteúdo. Variáveis auxiliares, embora necessárias para a realização da avaliação, tais como capacidade de leitura e escrita ou riqueza de vocabulário, não deveriam possuir uma influência muito grande. (LANDEIRA-FERNANDEZ; PRIMI, 2002)

Dado o exposto na análise fatorial, conclui-se que o exame dos modelos multidimensionais não apresentou uma delimitação unívoca nem dentre as matrizes de avaliação (saberes biológicos distintos dos saberes pedagógicos), nem dentre as

¹³ A título de informação, a correlação do escore parcial do FG apresentou correlação de 0.3577 com os itens (comuns ao bacharelado) de conhecimentos biológicos, e correlação de 0.2930 com os conhecimentos pedagógicos. Já a correlação entre o escore parcial da parte do CE referente a conhecimentos biológicos e a parte do CE com conhecimentos pedagógicos ficou em 0.3620. Por fim, a correlação entre os itens objetivos de CE (15 biológicos + 12 pedagógicos) e FG foi 0.3967.

características de perfil, conteúdo e competências descritas para os itens. Já pela análise discriminante, foi possível compreender que o fator principal F3 (que carrega a maior parte da carga fatorial) contém itens cujas características¹⁴ possibilitam a interpretação de um escore mais coerente com os conhecimentos e habilidades da matriz de referência.

Cabe lembrar o paradoxo em relação aos usos e expectativas dos diferentes grupos de interessados que compõem o público de uma avaliação, conforme exposto por Primi, Hutz e Silva (2011), e Silva (2011), pois enquanto as comissões de especialistas de área produzem definições multidimensionais do que deve ser avaliado nas provas, os psicometristas (especialistas em medida) têm que agir como mediadores dos processos avaliativos, e pautados pela parcimônia preferem testes unidimensionais, visto que geralmente o que lhes é demandado é um indicador único para tomada de decisões gerenciais. Sendo assim, dedicaremos a próxima seção à apresentação e análise do modelo unidimensional.

4.5 O modelo unidimensional e suas propriedades

Assumindo-se a hipótese de predomínio de apenas um fator na composição do constructo, realizou-se a calibração dos itens através de um modelo logístico de três parâmetros unidimensional obtendo-se os parâmetros iniciais (antes da exclusão dos itens), na escala (0,1), conforme exposto na Tabela 8, a seguir:

¹⁴ (i.e. precisão no uso da linguagem e atenção a detalhes (pegadinhas) especialmente em situações que exigem: “aplicação assertiva de princípios teóricos” e “matemática, raciocínio lógico-matemático e conhecimentos científicos”)

Tabela 8: Parâmetros de discriminação, dificuldade e probabilidade de acerto ao acaso, e respectivos erros padrões calibrados para o modelo logístico unidimensional da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017

ITENS	a	EP(a)	B	EP(b)	C	EP(c)
Q09	0,781	0,041	-0,477	0,120	0,115	0,037
Q10	1,233	0,077	0,126	0,073	0,126	0,029
Q11	0,535	0,042	1,084	0,146	0,104	0,031
Q12	0,601	0,053	1,234	0,132	0,112	0,032
Q13	1,200	0,146	1,891	0,079	0,231	0,014
Q14	0,450	0,160	3,679	0,394	0,196	0,058
Q15	0,711	0,079	1,985	0,096	0,105	0,024
Q16*	0,622	0,900	10,161	12,066	0,247	0,005
Q17	1,494	0,120	0,630	0,051	0,187	0,021
Q18*	0,642	0,819	11,643	12,939	0,135	0,003
Q19	0,661	0,088	1,338	0,154	0,173	0,043
Q20	0,818	0,132	2,375	0,131	0,166	0,023
Q21	1,194	0,054	-0,039	0,055	0,081	0,022
Q22	0,863	0,209	3,278	0,366	0,228	0,016
Q23	0,388	0,040	1,221	0,277	0,132	0,044
Q24L*	1,362	0,382	3,628	0,516	0,206	0,006
Q25L*	0,750	0,270	4,704	0,984	0,169	0,013
Q26L	0,713	0,155	3,044	0,261	0,165	0,025
Q27L	0,547	0,093	3,189	0,229	0,114	0,028
Q28L	0,815	0,041	-0,384	0,106	0,107	0,034
Q29L	0,749	0,051	1,128	0,078	0,082	0,022
Q30L*	0,833	0,225	3,718	0,500	0,142	0,015
Q31L	0,752	0,042	0,666	0,079	0,078	0,022
Q32L	0,756	0,044	0,608	0,084	0,084	0,024
Q33L	0,804	0,038	0,042	0,079	0,081	0,024
Q34L	0,803	0,072	1,118	0,091	0,120	0,029
Q35L	0,934	0,038	-1,013	0,086	0,099	0,032

Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018a). EP: Erro padrão

Observando-se os parâmetros iniciais estimados sem a exclusão de nenhum item, sete questões não atendem aos critérios de corte para o índice de discriminação recomendados por Baker (2001 *apud* PASQUALI, 2018) que considera medíocres itens com parâmetros “a” inferiores a 0,65. Porém, para esse momento em que as aplicações da TRI à elaboração dos itens do ENADE ainda podem ser consideradas experimentais, considera-se tal valor de corte muito extremado. Fazendo uma analogia com a carga fatorial, um item que apresenta valor do parâmetro “a” superior a 0,55, também apresenta carga fatorial superior a 0,3.

Quanto ao parâmetro de dificuldade, segundo Pasquali (2018), os valores plausíveis sob um critério prático estão compreendidos entre -3 e +3, o que resultaria em uma indicação para eliminação de nove itens. Porém, muitos trabalhos consideram admissíveis valores entre -4,5 e +4,5, sendo esse o critério adotado nessa pesquisa.

Outro aspecto implicado nesses “valores de corte” é que eles não deveriam determinar a eliminação acrítica dos itens, mas apenas apontar a necessidade de consultar especialistas na intenção de avaliar se a presença desses itens é essencial à validade de conteúdo (é um item exclusivo na cobertura daquele conteúdo/habilidade), e se à luz das distribuições de frequência dos padrões de resposta da TCT os itens não apresentam indícios de erros de gabarito (duplo ou trocado) e se aparentam estar verdadeiramente medindo o que se propõe na matriz de referência (validade de constructo), ou de outra forma, se a recomendação de eliminação do item é o encaminhamento mais parcimonioso e plausível.

Sendo assim, para essa população discutiremos a plausibilidade na eliminação dos itens Q11, Q14, Q23, e Q27L pelo critério da baixa discriminação e os itens Q16, Q18 e Q25L pelo parâmetro de dificuldade (nesses casos, muito elevados). Cabe lembrar que os itens Q16, Q18 e Q25 já haviam sido eliminados do cômputo do escore final pelo INEP seguindo-se o critério ponto-bisserial e que os itens Q14, Q16, Q18 e Q25L já eram apontados na análise do Alfa de Cronbach por contribuírem negativamente com a consistência interna da prova, reforçando-se a recomendação de desconsiderá-los.

4.5.1.1 Análise e decisão sobre a eliminação ou manutenção de itens

Itens indiciados por baixa discriminação:

A questão Q14 ilustra uma deficiência no formato de elaboração de questões do tipo asserção-razão, e está transcrita na Figura 8 apresentada a seguir:

Figura 8: Excerto da questão Q14 da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.

QUESTÃO 14

Acerca dos tentilhões de Galápagos, Darwin faz as colocações transcritas a seguir.

Observando essa transição gradual, essa diversidade, em um grupo de aves pequeno e intimamente aparentado, pode-se postular que, a partir de uma escassez original de aves, uma espécie tenha sido tomada e modificada para fins distintos. Infelizmente, a maior parte dos espécimes da tribo dos tentilhões mesclaram-se; mas tenho motivos fortes para suspeitar que algumas das espécies do subgrupo *Geospiza* se encontram confinadas a ilhas distintas. Se cada uma das ilhas tem representantes do *Geospiza*, isso pode ajudar a explicar o grande, e singular, número de espécies desse subgrupo neste pequeno arquipélago e, como consequência do seu número, a série graduada perfeita no tamanho dos seus bicos.

DARWIN, C. *Journal of Researches into the Natural History and Geology of the countries visited during the voyage round the world of H.M.S. Beagle*, Revised Edition, London: Henry Colburn, 1845, p. 403-420 (adaptado).

Estudos posteriores mostraram que as modificações na forma do bico dos tentilhões de Galápagos também alteraram a frequência do canto de atração sexual.

A partir dessas informações, avalie as asserções a seguir e a relação proposta entre elas.

- I. Todas as espécies de tentilhões de Galápagos são derivadas de um ancestral comum.

PORQUE

- II. Pressões seletivas que resultaram na modificação da forma dos bicos dos tentilhões, somadas ao isolamento geográfico, levaram ao isolamento reprodutivo pré-zigótico etológico.

A respeito dessas asserções, assinale a opção correta.

- A** As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
B As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
C A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
D A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
E As asserções I e II são proposições falsas.

Fonte: Brasil (2017e)

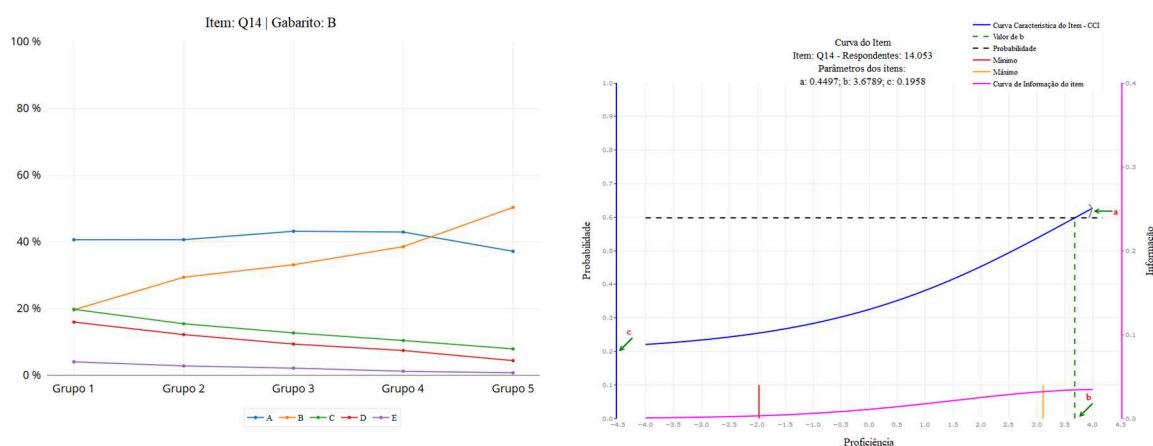
Nesse formato de elaboração, o respondente deve estabelecer um julgamento sobre a veracidade de cada uma das frases e, sendo ambas forem verdadeiras, ainda deverá demonstrar a habilidade de julgar se a relação estabelecida entre as sentenças é uma relação causal correta (razão ou justificativa). Dado o exposto, observa-se que uma resposta incorreta pode ter até três causas diferentes:

- Desconhecimento ou inabilidade para reconhecer a veracidade ou falsidade da asserção (primeira sentença);
- Desconhecimento ou inabilidade para reconhecer a veracidade ou falsidade da razão (segunda sentença) que, não necessariamente, está avaliando as mesmas características (conteúdo, competência...) da primeira sentença;
- Desconhecimento ou inabilidade para reconhecer a veracidade, plausibilidade ou falsidade da relação causal proposta entre as sentenças.

O que resulta em perda de informações sobre a proficiência dos respondentes, pois devido ao modo de correção dicotômico, são igualados sob a rubrica de “resposta errada” tanto o respondente que errou por não apresentar proficiência para superar nenhuma das capacidades avaliadas, quanto o respondente que demonstrou conhecer e ser competente para identificar com maestria a veracidade da asserção e da razão, mas esteve em dúvida sobre o sentido da expressão “Porque” que apresenta-se ambígua e polissêmica ao expressar causa, efeito, razão ou justificativa, sendo que esses sentidos podem apresentar-se complementares ou até contraditórios ou opostos dependendo do contexto e da ordem com que as sentenças são diagramadas.

Retomando o caso concreto da questão 14 pela TRI, é possível observar um baixo índice de discriminação, e a curva gerada com base na TCT (Figura 9) pode fornecer informações sobre os motivos desse resultado:

Figura 9: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q14 por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita)



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

Observa-se no gráfico de proporção de respostas às alternativas em cada quintil (Figura 9 à direita) que até o quarto quintil os respondentes (com bom desempenho) em proporções crescentes reconheciam como verdadeira a relação causal entre a hipótese de ancestralidade comum dos tentilhões e a explicação sobre os fatores e formas de especiação. Em suma, correção da prova dicotomizada

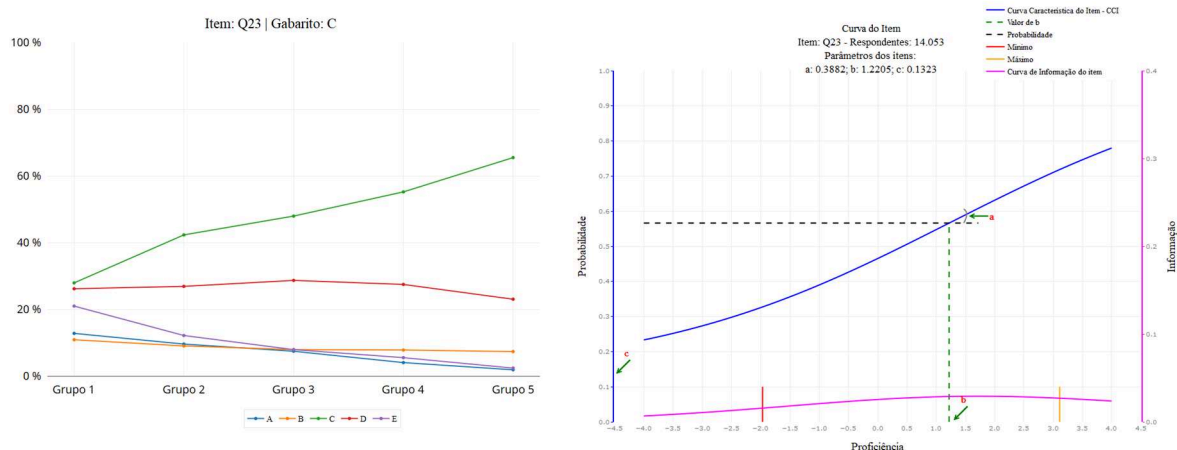
em questão asserção-razão resulta em perda de informações e subestimação de proficiências.

A principal evidência de ancestralidade comum está relacionada às homologias entre as espécies, porém caso não houvesse a formação de novas espécies, tal ancestralidade não seria cogitada por Darwin, logo a justificativa não está errada (à luz da história da ciência), apesar de constituir-se em uma falácia lógica (pois a especiação é o efeito da ancestralidade comum é não a causa). Nesse sentido, apenas a ordem das sentenças torna a relação causal falsa, e por isso, apenas o grupo de respondentes que reconheceu a lógica unidirecional do formato asserção-razão (efeito-causa), foi capaz de marcar a alternativa B. Em outras palavras, cabe considerar em alguns casos se a variabilidade captada pela aplicação de uma correção dicotômica a esse formato de item não está gerando uma variância de constructo irrelevante, conforme explicado nos *Standards*:

Se uma questão requer dos sujeitos uma habilidade de processamento linguístico para interpretá-la, mas não faz parte dos propósitos do teste a mensuração de habilidades linguísticas, a complexidade da questão pode resultar em variância irrelevante para o constructo nos escores do teste (AERA; APA; NCME, 2014, p.77, tradução livre)

A questão Q23 também apresentou baixo parâmetro de discriminação, porém nesse caso a habilidade em avaliação seria a interpretação de linguagem imagética em um contexto de divulgação científica (o que exigiria do respondente o julgamento acerca da melhor forma de realizar a transposição didática sem perder a essência do conteúdo comunicado- a “seleção natural” - pela simplificação da sua representação). Também nessa questão houve um padrão característico de duplo gabarito, conforme observa-se na Figura 10 apresentada a seguir:

Figura 10: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q23 por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita)



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

Nesse caso, observa-se uma confusão dos quintis de escore superior entre o distrator “D” e a alternativa correta “C”. Ilustramos o enunciado e as alternativas com elevada discriminação (TCT) para facultar ao leitor a possibilidade de interpretar e julgar a plausibilidade das hipóteses defendidas:

Figura 11: Excerto da questão Q23 da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.

QUESTÃO 23

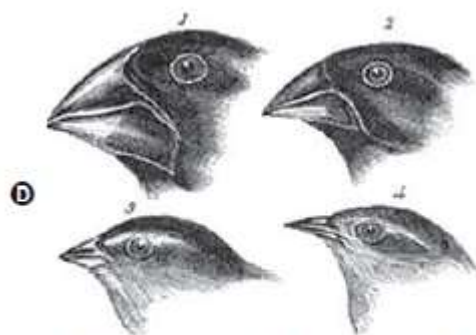
Os organismos produzem descendência em número muito maior em relação aos que serão capazes de sobreviver, o que resulta em uma luta e/ou competição pela sobrevivência. A seleção natural age sobre quaisquer entidades capazes de se reproduzir, de apresentar herança de suas características de uma geração para a outra e de variar em aptidão (isto é, número relativo de descendentes diretos que elas produzirem) de acordo com as características que possuem.

RIDLEY, M. *Evolução*. Porto Alegre: Artmed, 2006 (adaptado).

Qual das imagens a seguir melhor representa o processo de seleção natural no contexto da divulgação científica?



Disponível em: <<https://biologiacosmos.wordpress.com>>. Acesso em: 17 jul. 2017 (adaptado).



Disponível em: <<http://educador.brasilescola.uol.com.br>>. Acesso em: 17 jul. 2017 (adaptado).

Fonte: Brasil (2017e)

Como proposta de explicação para a confusão dos respondentes, lançamos a hipótese de que houve um relacionamento espúrio com o enunciado presente na questão 14 (que apresenta os tentilhões como uma evidência da hipótese de Darwin) quanto ao enunciado da própria questão 23 que menciona “capacidade de reprodução” o que relaciona-se ao número de descendentes viáveis, entretanto a imagem da alternativa correta apresenta no terceiro quadro (em uma noção de sucessão temporal de fatos) apenas um descendente vivo! Nesse ínterim deduz-se que o aspecto enfatizado na representação esquemática da alternativa C é a seleção negativa, ou seja, a sobrevivência do mais apto em detrimento (morte) do menos apto, mas que no enunciado da questão a reprodução diferencial do mais apto é o aspecto relevante.

A questão 11, exposta na Figura 12, apresenta um conflito menos evidente entre o gabarito (D) e o distrator (C), sendo um bom exemplo para discutirmos a inadequação da aplicação de formas de correção dicotômica a itens de “múltipla escola complexa”.

Figura 12: Excerto da questão Q11 da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.

QUESTÃO 11

Razanandrongobe sakalavae é um grande Archosauria predador do Jurássico médio da Ilha de Madagascar. Em estudo recente, pesquisadores analisaram novos fragmentos da mesma região e da mesma espécie e identificaram a espécie *R. sakalavae* como pertencente ao clado Crocodylomorpha. As profundas e gigantescas mandíbulas do crocodilo pré-histórico eram armadas com enormes dentes serrados e muito semelhantes em tamanho e forma às dos dinossauros terópodes e, especialmente, às do carnívoro *Tiranosaurus rex*. Essas mandíbulas são uma notável evidência de que esses répteis crocodylianos se alimentaram de tecidos orgânicos duros, como ossos e tendões.

DAL SASSO, C. et al. *Razanandrongobe sakalavae*, a gigantic mesoeucrocodylian from the Middle Jurassic of Madagascar, is the oldest known notosuchian. *PeerJ* 5:e3481, 2017 (adaptado).

A evolução das aves ratitas tem sido atribuída à especiação vicariante, impulsionada pela dissolução do supercontinente Gondwana no período Cretáceo. O isolamento inicial da África e de Madagascar é indicativo de que o avestruz e os pássaros-elefante-de-madagascar – esse último extinto – devem ser as linhagens de ratitas mais antigas. O sequenciamento do genoma mitocondrial de dois pássaros-elefante e análises filogenéticas revelaram que essas aves são os parentes mais próximos do kiwi da Nova Zelândia e estão distantes da linhagem basal de avestruzes. Essas evidências sugerem que a convergência para o gigantismo e a ausência de voo foi facilitada no Terciário, após a extinção dos dinossauros, com a exploração do nicho de herbivoria diurna.

MITCHELL, K. J. et al. Ancient DNA reveals elephant bird and kiwi are sister taxa and clarifies ratite bird Evolution. *Science*, v. 344, n. 898, 2014 (adaptado).

Considerando os textos apresentados, avalie as afirmações a seguir.

- I. A semelhança de denticões entre os dinossauros terópodes e o crocodilo fóssil exemplifica uma característica homoplásica.
- II. A **divergência** a respeito da evolução das linhagens de aves ratitas constitui exemplo de hipóteses de vicariância e de dispersão.
- III. A distribuição biogeográfica das espécies de aves ratitas é considerada cosmopolita.
- IV. A ilha de Madagascar representa um *hotspot* biogeográfico dada, entre outras características, a presença de endemismo de espécies.

É correto apenas o que se afirma em

- A I e III.
- B II e III.
- C II e IV.
- D I, II e IV.
- E I, III e IV.

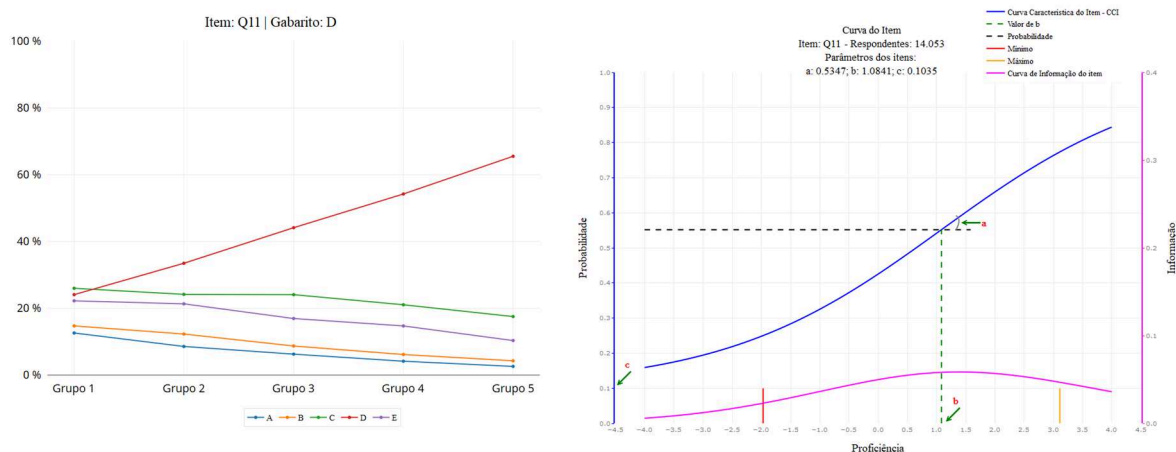
Área livre

Nessa forma de elaboração de item, o respondente que saiba mais da metade dos conteúdos propostos poderá ter o mesmo desempenho do estudante que não sabe nada, pois havendo um erro de julgamento toda a questão será considerada errada. No caso concreto do item acima, o estudante deveria:

- I. Saber o conceito de homoplasia e que dinossauros e crocodilomorpha não constituem grupo monofilético.
- II. Saber e diferenciar os conceitos de vicariância e de especiação alopátrica (que o elaborador denominou pela expressão não usual “dispersão”) e compreender que o enunciado apresenta subliminarmente uma controvérsia científica relacionada à filogenia das ratitas (kiwi sendo grupo irmão de pássaro-elefante e não de avestruzes como anteriormente ao sequenciamento era a hipótese aceita).
- III. Saber o conceito de cosmopolita e que as aves artistas possuem distribuição restrita ao hemisfério sul.
- IV. Saber o que é um *hotspot* biogeográfico, sua relação com os endemismos e que Madagascar enquadra-se nessa categoria.

No geral, os estudantes com elevado desempenho assinalaram a alternativa correta (D), porém independentemente de pertencerem quintil de elevado ou de baixo desempenho uma fração de aproximadamente 20% dos estudantes foram atraídos pela alternativa C. Tal fato, ilustrado na Figura 13 apresentada a seguir, pode estar relacionado ou à confusão de termos de difícil memorização (homoplasia e homologia), ou à exigência de um conhecimento muito específico (filogenia de répteis extintos).

Figura 13: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q11 por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita)



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

De toda forma, considerando-se temerário a eliminação das questões Q11, Q14 e Q23, pois elas são apontadas na matriz de referência como as únicas relacionadas aos conteúdos de Evolução (6) e Biogeografia (13), compartilhando também a tarefa de tratar os conteúdos de zoologia com o item Q13. Por esse motivo, e tendo-se em vista que o item Q11 apresentou um índice de discriminação limítrofe ao critério de corte (0,535~0,55), que não apresentou um padrão de duplo gabarito, nem resultou em redução do Alfa de Cronbach, optou-se por mantê-lo no modelo final, pois de outra forma a sub-representação dos conteúdos ligados a evolução seria uma falha irreparável.

O item Q27L também é um caso limítrofe, pois seu índice de discriminação foi 0,547 e considerando-se o erro padrão de estimação desse parâmetro 0,0928 é possível que atinja o critério de corte (0,55). Outro fato que corrobora para sua manutenção é que apesar de ter sido descrito como um dos três itens que tratam a temática do “Currículo no ensino de Ciências e de Biologia” (OC18), é o único dentre os três que invés de tratar temas transversais, trata o currículo sobre a perspectiva da sociologia do currículo.

Itens indicados pelo critério do índice de dificuldade

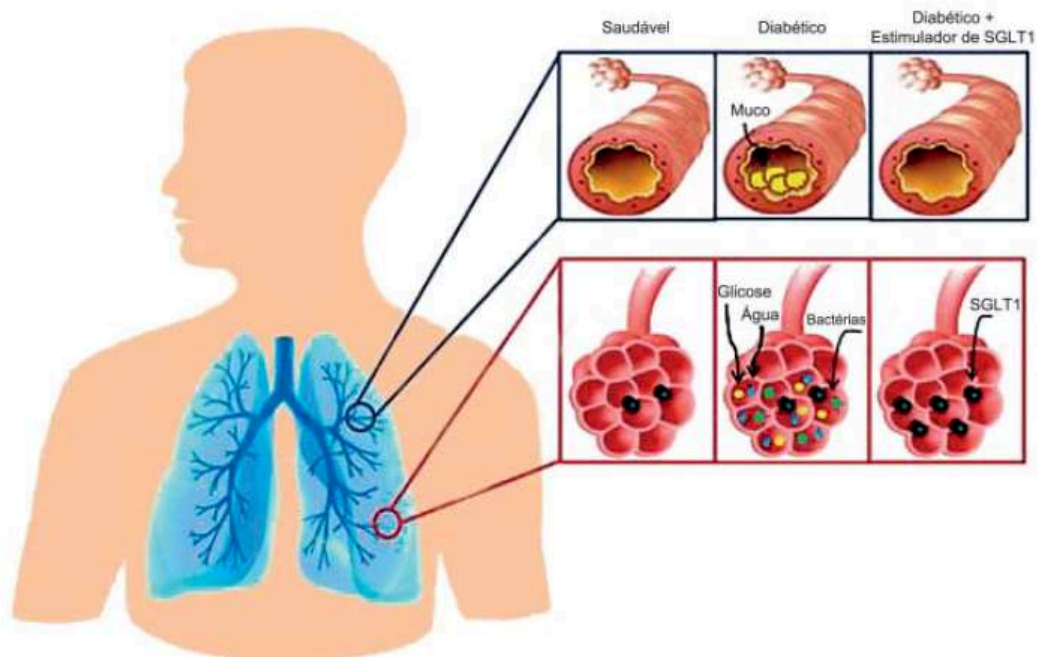
A questão Q16 foi apontada como passível de eliminação na estimação dos parâmetros iniciais com um índice de dificuldade muito elevado e apresentando um erro padrão elevado para os parâmetros de discriminação e dificuldade. Trata-se de uma questão de asserção-razão que emprega em seu “texto base” uma imagem com uma representação de escala confusa (ver Figura 14), pois bactérias aparecem representadas com tamanhos inferiores às moléculas de proteína e em proporções semelhantes às moléculas de água e a glicose. Além disso, a cor verde utilizada para representar as bactérias, em meio ao fundo vermelho dos alvéolos pode ter confundido indivíduos com algum grau de discromopsia (daltonismo). A questão está relacionada a fisiologia, bioquímica e saúde-educação, objetiva na asserção pretende avaliar a compreensão do texto base (texto e imagem), na razão (2ª sentença) o objetivo é avaliar se os respondentes compreendem os mecanismos de osmoregulação, a característica da glicose como soluto osmoticamente ativo e o conceito de “potencial hídrico”, por fim na relação (“porque”) o objetivo é avaliar se o estudante sabe que a composição do muco presente nos brônquios é de polímeros de glicose (mucopolissacarídeos) e consegue relacionar essa informação.

Figura 14: Excerto da questão Q16 da prova de Conhecimentos Específicos da área de Licenciatura em Ciências Biológicas no ENADE 2017.

QUESTÃO 16

A proteína transportadora de glicose (SGLT1) acoplada ao íon sódio está naturalmente presente em diversos órgãos, sendo responsável pela reabsorção de glicose e pelo transporte contra o gradiente de concentração de glicose e a favor do gradiente de concentração do sódio.

Pesquisas na área de biotecnologia têm demonstrado que estimuladores dessa proteína podem auxiliar no combate a infecções bacterianas em pacientes diabéticos, conforme ilustrado na figura abaixo.



Disponível em: <<http://www.comunica.ufu.br>>. Acesso em: 18 jul. 2017 (adaptado).

A respeito da atuação da SGLT1 nos alvéolos pulmonares de diabéticos, avalie as asserções a seguir e a relação proposta entre elas.

- I. O uso terapêutico de fármacos estimuladores da SGLT1 em pacientes diabéticos promove a redução do volume de muco e, conseqüentemente, da proliferação bacteriana.

PORQUE

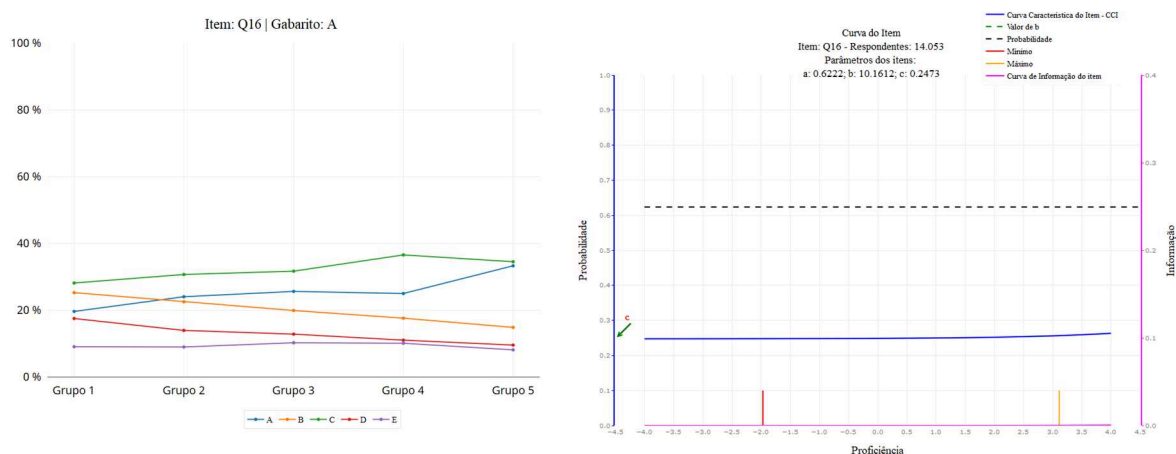
- II. No tecido pulmonar, a SGLT1 promove diminuição da concentração de glicose com conseqüente aumento do potencial hídrico intracelular.

A respeito dessas asserções, assinale a opção correta.

Fonte: Brasil (2017e)

Observa-se no gráfico da TCT (Figura 15) que os distratores C e E apresentam atratividade mesmo nos quintis superiores.

Figura 15: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q16 por quintis (abcissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita).



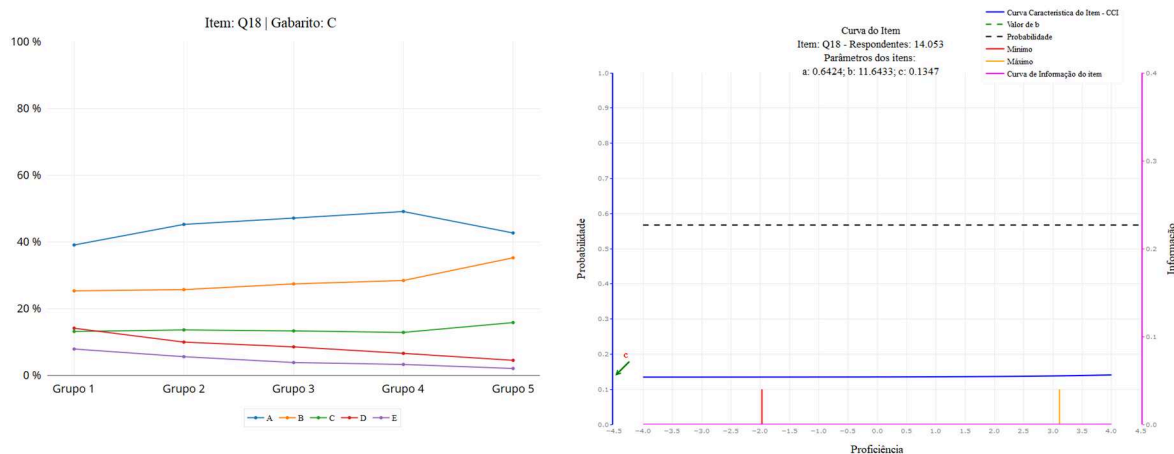
Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

No texto da razão (2ª sentença) há uma expressão incomum na fisiologia humana ao tratar de osmorregulação/osmolaridade, pois “potencial hídrico” é expressão comumente utilizada em fisiologia vegetal, também contém uma imprecisão (não esclarece se a redução na concentração de glicose é no meio intracelular ou extracelular) e ainda há uma imagem confusa e de baixa resolução no “texto base”, fatores que em conjunto geram dúvidas sobre a veracidade da sentença “razão”, mesmo entre os extratos com maior desempenho. Dado o exposto, e por não ser a única questão que trata dos conteúdos, será ratificada a recomendação para sua exclusão.

O item Q18, apresentado a seguir na Figura 16, foi indicado para eliminação pois sua dificuldade foi considerada desproporcional e o erro padrão das estimativas dos parâmetros “a” e “b” foi muito alto quando comparado com itens com estimativas de mesma magnitude. Trata-se de um item de asserção-razão que abrange conteúdos relacionados a biotecnologia, mas apontados como relativos aos conteúdos “Microbiologia, Imunologia e Parasitologia” (OC3) e Saúde-educação (OC11), logo esse item compartilha conteúdos com o Item Q21, podendo ser excluído. Conforme ilustrado na imagem a seguir, os distratores que apresentaram padrões diretamente proporcionais entre a elevação do escore nos quintis e a proporção de respostas foram as alternativas A e B que consideram a razão (2ª

sentença) verdadeira, quando o correto segundo o elaborador do item seria a alternativa C que considera a razão falsa.

Figura 16: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q18 por quintis (abscissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita).

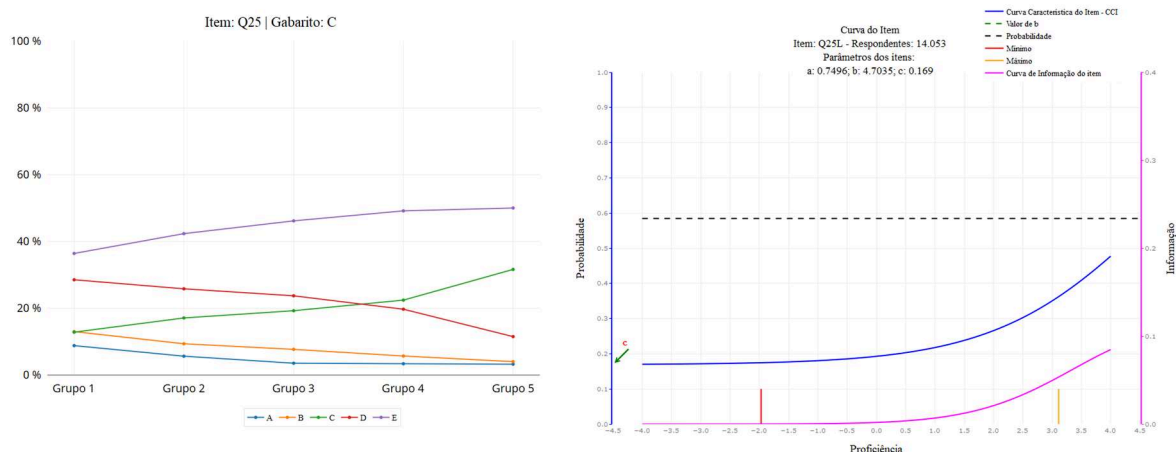


Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

O texto da razão (2ª sentença) avalia mais a compreensão textual do enunciado e contém uma indução ao erro (“pegadinha”) ao dar a entender que são necessários conhecimentos irrelevantes (nanotecnologia e eletrônica). Compreende-se que a razão (2ª sentença) está incorreta, pois não é o sensor, mas o dispositivo a ele acoplado, quem emite sinal sem fio, portanto o que tornaria o diagnóstico mais ágil é a possibilidade de conversão de um sinal bioquímico em um sinal eletrônico, e não o fato mencionado no texto base de que o dispositivo é capaz de emitir sinais sem fio.

Por fim, a questão Q25L trata de técnicas, procedimentos e metodologias de ensino, através do formato “múltipla escolha complexa”. Seus objetos de conhecimento e competências são compartilhados pelos itens Q26 e Q29. Conforme o gráfico de quintis exposto na Figura 17, apresentada a seguir, o distrator “E” teve atratividade crescente e superior ao gabarito.

Figura 17: Distribuição de proporções de escolha das alternativas (ordenadas) do item Q25L por quintis (abscissas) agrupando os participantes em níveis crescentes de escore parcial do CE (à esquerda) e Curva Característica do Item (CCI) pelo ML3P (à direita)



Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b)

O motivo de sua dificuldade está relacionado ao distrator presente na segunda sentença que afirma: “As tecnologias da informação e comunicação são ferramentas indispensáveis às práticas de ensino em Ciências e Biologia” e que tal sentença foi considerada falsa pelos elaboradores, mas plausível para grande proporção dos respondentes com boas performances. Conjectura-se que a cultura digital e os incentivos ao uso das TIC na educação por políticas públicas foram tão impactantes que os estudantes não reconhecem as TIC como dispensáveis para as práticas de ensino de Biologia. Recomenda-se para estudos futuros testar essa questão quanto a uma possível existência de DIF em comparações entre grupos de estudantes vinculados a IES com maior e menor disponibilidade de infraestrutura. Em síntese, a questão Q25L é apontada na seção 4.3.1 como uma das que se eliminadas resultaria em ampliação da consistência interna, possui um padrão de escolha de alternativas que se assemelha a um duplo gabarito, foi uma das questões eliminada pelo INEP por apresentar coeficiente ponto-bisserial fraco, e não trata de nenhum objeto de conhecimento com exclusividade; fatos que associados corroboram para a deliberação de sua eliminação do modelo final.

Sumarizando as decisões descritas e justificadas nessa seção, cinco questões foram eliminadas da construção do modelo final os itens: Q14, Q23, Q16,

Q18 e Q25 e mantidas por constituírem-se casos limítrofes e essenciais para a validade de conteúdo do teste os itens Q11, Q27L.

4.5.2 Ajuste do modelo

Os indicadores de ajuste do modelo inicial (com todos os 27 itens) e final, quando aplicada a exclusão de 5 itens conforme critérios descritos na seção 4.5.1.1, bem como dos modelos bi e tridimensionais conforme os valores apresentados no Quadro 5:

Quadro 5: Principais indicadores de informação restrita da adequação do modelo aos dados.

Modelo	Log Lik	M2	gl	p	Nº Parâm.	RMSEA	RMSEA (5%)	RMSEA (95%)
Inicial ML3P	-231,564,1	1622,1	297	0	81	0,01827	0,01740	0,01914
Final ML3P	-183,477,8	1007,0	187	0	66	0,01811	0,01702	0,01921
Inicial ML3P 2D	-220,017,4	690,1	271	0	107	0,01075	0,00976	0,01175
Inicial ML3P 3D	-219,913	483,3	246	0	132	0,00849	0,00737	0,00961
Modelo	SRMR	CFI	TLI		AICc	SABIC		
Inicial ML3P	0,01986	0,9258	0,9123		461,855,4	462,208,7		
Final ML3P	0,01812	0,9499	0,9382		367,031,6	367,316,3		
Inicial ML3P 2D	0,01202	0,9765	0,9696		440,250,6	440,711,4		
Inicial ML3P 3D	0,01019	0,9867	0,9810		440,092,6	440,660,6		

Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b). Nota: 2D=Modelo Bidimensional; 3D Modelo Tridimensional.

Kline (2005) recomenda sempre reportar o χ^2 ou uma estatística equivalente e apropriada para o tipo de modelo utilizado, que no caso da modelagem de traços latentes é a estatística M2, uma derivação do χ^2 . Mediante o teste do M2 evidencia-se que o modelo falhou ($p=0$) em atingir um ajuste exato com os dados empíricos, fato esperado dado o grande tamanho amostral e a alta sensibilidade desse teste a tamanhos de amostra grandes ($n>500$). Além disso, sabe-se que a matriz de avaliação engloba capacidades e saberes não correlacionados e que, portanto, interpretações mais precisas das proficiências de indivíduos e médias de cursos

demandariam o emprego de modelos multidimensionais; entretanto tal intenção desvia-se do escopo dessa pesquisa. Também não faz parte do escopo de investigação a intenção de realizar a modelagem exata do traço latente, mas tão somente avaliar se o instrumento de CE aplicado 2017 é válido, sugerindo-se a demanda de estudos futuros de avaliação da magnitude e identificação de prováveis causas para a não exatidão do ajuste do modelo aos dados empíricos.

A título de ilustração, faculta-se aos interessados em iniciar tal análise das causas e magnitudes do erro de especificação do modelo, uma consulta à matriz de resíduos no (Apêndice C), e os gráficos apresentando as CCI dos 22 itens calibrados pela TRI (Apêndice D), sobrepostas à curva empírica (formada pela sobreposição de 20 pontos empíricos que representam a localização estabelecida no cruzamento entre a proficiência e probabilidade de acerto ao item pelos indivíduos do grupo, sendo 20 grupos de indivíduos homogêneos quanto às proficiências).

4.6 A escala e sua interpretação

Antes da eliminação, os itens objetivos da prova de Conhecimentos Específicos apresentavam média de dificuldade $\bar{x}b_{inicial} = 2,2436$ (EP: 1,1175), após a exclusão dos itens com parâmetros inadequados o instrumento passou a ter uma dificuldade mais condizente com a população avaliada $\bar{x}b_{final} = 1,2931$ (EP: 0,1573) e também parâmetros com menor erro padrão, logo, com estimativas mais precisas. Apesar disso, pode-se considerar que a prova tem um nível difícil para a população, pois 94,33% da população (12616) estava abaixo do nível de dificuldade médio da prova. Posteriormente ao estabelecimento do modelo final, procedeu-se a triagem dos indivíduos que atendiam ao critério de pelo menos 3 respostas válidas (não nulas nem branco) dentre as respostas aos 22 itens disponíveis, resultando em uma amostra de 13.374 indivíduos.

Em uma escala composta por níveis dispostos em intervalos regulares com interstícios de 0,5 desvios padrões entre si, foi possível identificar um item âncora: Q24L em 3,5 dp, faixa de proficiência que não contém indivíduos na amostra de

2017, e um item quase âncora: Q17 ancorado em 1dp (com 15% de diferença de 0,5 dp, mas que não ancorou pois não cumpriu o terceiro critério).

Conforme Beaton e Allen (1992, p.193) o processo de ancoragem não apresenta uma garantia de resultado no fornecimento de descrições úteis dos pontos de ancoragem. Mesmo um teste bem desenhado para seu propósito pode não ter informação suficiente disponível para diferenciar os níveis selecionados na escala, algumas vezes devido à multidimensionalidade. Conforme os autores, em alguns casos pode ser necessário selecionar diferentes níveis na escala que sejam ancoráveis. Em outros casos, nenhum nível na escala terá número suficiente de itens ancorados para prover uma descrição adequada do que os estudantes naquele nível são capazes. Podem também ocorrer situações em que os itens em um nível âncora sejam tão diferentes uns dos outros que o nível torna-se não interpretável. Em todo caso, “os procedimentos de ancoragem fornecem uma maneira de selecionar e organizar os itens para facilitar a interpretação da escala e gerar teorias” (BEATON; ALLEN, 1992, p.193).

Devido à baixa granularidade da escala (apenas 1 item âncora com intervalos de 0,5dp), simulou-se o processo de ancoragem com interstícios regulares de 1 desvio padrão entre níveis, novamente nenhum item atendeu ao critério de Beaton e Allen (1992) para diferença de probabilidade entre níveis adjacentes ($\geq 30\%$), porém empregando-se o critério menos extremado proposto por Haberman, Sinharay e Lee (2011) com diferença de probabilidades de acertos entre níveis adjacentes superior a 15%, os itens Q10, Q21, Q24L e Q35L foram classificados como âncoras e outros doze itens como quase âncoras.

Dado o exposto, na Tabela 9, estão apresentados os parâmetros finais dos itens estimados pelo modelo logístico de três parâmetros, e o posicionamento de dezesseis itens (âncoras e quase âncoras) distribuídos entre os três níveis âncoras identificados na escala (0,1) de proficiência proposta.

Tabela 9: Parâmetros finais dos itens com respectivo erro padrão e posicionamento na escala (0,1).

Item	<i>a</i>	<i>EP(a)</i>	<i>b</i>	<i>EP(b)</i>	<i>C</i>	<i>EP (c)</i>	Nível
Q09	0,7731	0,0420	-0,4987	0,1292	0,1209	0,0391	0,5*
Q10	1,2354	0,0796	0,1046	0,0772	0,1285	0,0308	0,5**
Q11	0,5379	0,0464	1,1002	0,1597	0,1140	0,0345	
Q12	0,6123	0,0578	1,2284	0,1374	0,1194	0,0341	
Q13	1,1280	0,1436	1,8684	0,0807	0,2215	0,0159	2,5*
Q15	0,7293	0,0841	1,9496	0,0953	0,1091	0,0250	3,5*
Q17	1,4913	0,1246	0,6181	0,0536	0,1879	0,0221	1,5*
Q19	0,6766	0,0960	1,3577	0,1559	0,1845	0,0441	
Q20	0,7956	0,1316	2,3988	0,1374	0,1656	0,0237	3,5*
Q21	1,1881	0,0575	-0,0512	0,0613	0,0888	0,0247	0,5**
Q22	0,8295	0,2174	3,3708	0,4082	0,2282	0,0176	3,5*
Q24L	1,7152	0,4751	3,3681	0,4025	0,2114	0,0050	3,5**
Q26L	0,7183	0,1577	3,0199	0,2601	0,1672	0,0249	
Q27L	0,5793	0,1072	3,1212	0,2324	0,1237	0,0285	
Q28L	0,8082	0,0434	-0,4012	0,1186	0,1160	0,0371	0,5*
Q29L	0,7454	0,0494	1,0949	0,0793	0,0794	0,0217	2,5*
Q30L	0,9386	0,2353	3,4569	0,4158	0,1459	0,0132	
Q31L	0,7469	0,0440	0,6562	0,0839	0,0821	0,0237	1,5*
Q32L	0,7556	0,0467	0,6065	0,0908	0,0905	0,0261	1,5*
Q33L	0,7909	0,0398	0,0043	0,0883	0,0880	0,0268	1,5*
Q34L	0,8097	0,0804	1,1478	0,0972	0,1343	0,0312	2,5*
Q35L	0,9109	0,0397	-1,0725	0,0958	0,1069	0,0349	-0,5**

Fonte: O autor (2019). Legenda: a= parâmetro de discriminação; b=parâmetro de dificuldade; c= parâmetro de probabilidade de acertos ao acaso; ** Item âncora; * Item "quase âncora".

Conforme Beaton e Allen (1992, p.193), o processo de ancoragem na escala reflete diferenças entre itens que influenciam as respostas individuais aos itens, e lançam em hipótese que algumas vezes tais diferenças podem relacionar-se a variações no formato dos itens. Para verificar essa hipótese, apresenta-se na Tabela 10, a relação dos itens presentes no modelo final ordenados por nível de facilidade (b clássico) e classificados quanto ao formato de elaboração:

Tabela 10: Itens ordenados por nível crescente de dificuldade, e tipificados quanto ao formato dos itens, parâmetros, carga fatorial, comunalidade e seu posicionamento na escala de proficiência.

Item	% de acerto	a	b	c	Tipo de formato	Nível		
						CF	Com	
Q35L	72,1%	0,911	-1,073	0,107	Interpretação	0,41	0,17	-0,5**
Q09	63,0%	0,773	-0,499	0,121	Asserção-razão	0,59	0,35	0,5*
Q28L	61,4%	0,808	-0,401	0,116	Asserção-razão	0,30	0,09	0,5*
Q21	54,9%	1,188	-0,051	0,089	M.E. Complexa	0,34	0,11	0,5**
Q10	53,7%	1,235	0,105	0,129	Interpretação	0,55	0,31	0,5**
Q33L	53,4%	0,791	0,004	0,088	Interpretação	0,39	0,16	1,5*
Q17	46,3%	1,491	0,618	0,188	M.E. Complexa	0,66	0,43	1,5*
Q32L	45,0%	0,756	0,607	0,091	Interpretação	0,37	0,14	1,5*
Q31L	43,8%	0,747	0,656	0,082	M.E. Complexa	0,42	0,18	1,5*
Q11	43,3%	0,538	1,100	0,114	M.E. Complexa	0,57	0,33	
Q19	42,8%	0,677	1,358	0,185	Asserção-razão	0,44	0,19	
Q12	41,0%	0,612	1,228	0,119	Interpretação	0,71	0,50	
Q34L	39,5%	0,810	1,148	0,134	M.E. Complexa	0,39	0,15	2,5*
Q29L	37,5%	0,745	1,095	0,079	M.E. Complexa	0,32	0,10	2,5*
Q13	33,8%	1,128	1,868	0,222	M.E. Complexa	0,43	0,18	2,5*
Q15	30,0%	0,729	1,950	0,109	M.E. Complexa	0,40	0,16	3,5*
Q20	29,1%	0,796	2,399	0,166	M.E. Complexa	0,48	0,23	3,5*
Q22	28,5%	0,830	3,371	0,228	M.E. Complexa	0,40	0,16	3,5*
Q26L	26,5%	0,718	3,020	0,167	Interpretação	0,41	0,16	
Q27L	25,7%	0,579	3,121	0,124	M.E. Complexa	0,42	0,18	
Q24L	21,9%	1,715	3,368	0,211	Asserção-razão	0,43	0,18	3,5**
Q30L	19,1%	0,939	3,457	0,146	M.E. Complexa	0,47	0,22	

Fonte: O autor (2019). Legenda: a= parâmetro de discriminação; b= parâmetro de dificuldade; c= parâmetro de probabilidade de acertos ao acaso; CF= carga fatorial; Com= Comunalidade; ** Item âncora; * Item “quase âncora”.

Conforme observa-se na Tabela 10 não houve uma segmentação clara dos formatos dos itens entre os níveis, portanto, pelo processo de ancoragem suavizado com base no ML3P não há como afirmar indubitavelmente que as variações no tipo do item influenciam o processo de resposta.

Realizando-se o posicionamento de todos dos 22 itens mantidos no processo de calibração pelo ML3P, foi possível gerar interpretações pedagógicas sobre a escala de proficiência segmentada em 3 níveis plausíveis e 2 níveis extrapolados, pois conforme Beaton e Allen (1992) diferentemente da metodologia de ancoragem direta na qual não se deve estabelecer níveis âncoras em intervalos de escores com pequeno número de indivíduos (i.e. $n < 50$), mas na ancoragem por *smoothing* a CCI possibilita extrapolar os níveis âncoras mesmo em pontos da

escala de proficiência em que não haja indivíduos. Os descritores dos níveis (PLD) identificados na escala unidimensional estão expostos no Quadro 6:

Quadro 6: Interpretação da escala de proficiência da licenciatura construída sobre o mapeamento de itens âncoras calibrados pelo ML3P.

Grau	Descrição
Além do avançado (3,5 dp acima da média ou superior)	<ul style="list-style-type: none"> • Além das capacidades descritas para o nível anterior, caso existissem estudantes nesse nível eles seriam capazes de: • Compreender os conceitos de isolamento reprodutivo e variedade (táxon infra-específico em botânica) e é capaz de articulá-los para avaliar situações envolvendo organismos geneticamente modificados (transgênicos); • Saber que a membrana plasmática das células da raiz faz transporte ativo de íons e/ou compreender o conceito de pressão osmótica concebendo que o aumento de seus valores não pode ser gerado sem gasto de energia para o carregamento de íons para o interior da membrana semipermeável (raiz) e/ou identificar que os vasos do floema não estão envolvidos com o mecanismo de transporte de seiva bruta descrito no enunciado; • Realizar uma leitura literal do enunciado, valorizando a precisão vocabular mais que a dedução de fatos. É capaz de refletir que os princípios de gestão democrática não são a causa imediata da ampliação dos Conselhos escolares, outrossim, tal ampliação é mediada pela comunidade escolar, tanto que a rede estadual supera a federal em representação, a despeito do PNE ser elaborado em nível federal. Também compreende a tomada de decisão coletiva como necessária à gestão democrática; ○ Compreende os sentidos ocultos de frases e expressões relacionadas à leitura de mundo dos estudantes sobre sexualidade e as finalidades da educação formal. Ou ser capaz de buscar elementos intertextuais do contexto dos discursos apresentados. ○ Aplicar seu conhecimento sobre a importância da história da ciência para o ensino ao avaliar planejamentos de aula, reconhecendo que a atividade de ensino adequada para não deformar a descrição do trabalho científico deve conter elementos de história e epistemologia da ciência; ○ Compreender o currículo como arbítrio cultural e é capaz de julgar e problematizar demandas sociais contingentes, avalia o livro didático como dispositivo construído com finalidades que podem distinguir-se das demandas sociais e de seu uso na prática; ○ Compreender o que é uma espécie carismática e contra-intuitivamente reconhece o mico-leão-dourado como uma espécie que devido a sua pequena população, e posição na cadeia alimentar, apresenta pouca importância para o fornecimento de serviços ecossistêmicos à espécie humana. Reconhecer a função econômica das áreas protegidas; <p>São capazes de interpretar discursos elaborados aplicando raciocínio lógico construído a partir conhecimentos biológicos, inclusive em situações contra intuitivas.</p>

Continua na próxima página

Grau	Descrição
<p>Muito Avançado (2,5 dp acima da média ou superior)</p>	<p>Além das capacidades descritas para o nível anterior o sujeito classificado nesse nível:</p> <ul style="list-style-type: none"> • Avaliar e distinguir propostas pedagógicas capazes de ilustrar a aplicação de diretrizes e princípios presentes nas diretrizes curriculares do ensino médio (DCNEM), especialmente em situações em que haja a sobreposição de princípios como autonomia intelectual [do estudante] e integração com o mundo do trabalho. • Aplicar os pressupostos da teoria vygostskyana para avaliar afirmações abstratas; Através de descrições é capaz de diferenciar o aprendizado individual do aprendizado socialmente mediado. • Análise e Interpretação de representação esquemática de teias tróficas com interações ecológicas não tróficas, e dos efeitos indiretos de interações ecológicas (predação/competição) sob a dinâmica de comunidades. <ul style="list-style-type: none"> ○ Compreende a natureza bioquímica dos mecanismos de reconhecimento celular e resposta imune. ○ Conhecer a natureza e as propriedades da bicamada lipídica (polaridade molecular e natureza anfipática dos lipídios hidrofílico x hidrofóbico), e é capaz de aplicar esses conceitos em uma nova situação. ○ Lembra e compreende o conceito de homoplasia, e compreende os conceitos de variância, dispersão (sinônimo de especiação alopática), cosmopolitismo, <i>hotspot</i> biogeográfico. Lembra que ratitas não são cosmopolitas e que o grupo formado por dinossauros e crocodilomorpha não é monofilético. ○ Diferencia legalidade de moralidade e compreende que o campo da bioética está implicado com questões para além da legalidade.
<p>Avançado (1,5 dp acima da média ou superior) n=404 (3%)</p>	<p>Além das capacidades descritas para o nível anterior o estudante:</p> <ul style="list-style-type: none"> • Compreende a dinâmica das interações predador-presa e relações de proporcionalidade; e é capaz de aplicá-las em situações experimentais de cadeias tróficas com até três níveis, interpreta gráficos estatísticos e desenhos experimentais com amostra de controle local. • Compreende e aplica a política nacional de educação especial em um contexto de educação inclusiva para surdos; • Compreende os conceitos de práxis, método dialético, teoria histórico cultural e reconhecer a possibilidade de adoção desses conceitos em situações escolares de elaboração de conhecimento. Reconhece que a didática pode funcionar como atualizadora da prática docente através do relacionamento das questões sociais com o conhecimento historicamente construído. • Avalia a aplicação do conceito de gênero em práticas pedagógicas de modo convergente com as diretrizes e demonstra respeito às diferenças; Geralmente apresenta boa capacidade de interpretação e/ou compreensão textual.

Continua na próxima página

Continuação do Quadro 6

Grau	Descrição
Suficiente (0,5 dp acima da média ou superior) n=3.235 (24,2%)	Além das capacidades descritas para o nível anterior o sujeito classificado nesse nível: <ul style="list-style-type: none"> •Compreende os princípios de expressão gênica e as implicações da edição de genomas em células somáticas proposta em técnicas de terapia genética. •Compreende o sentido oculto e global do enunciado¹⁵, reconhecendo com válido o princípio de “conhecer para respeitar”. •Recorda a existência de mecanismos de diferentes naturezas na resposta imunológica e que alguns desses apresentam especificidade quanto aos antígenos dos tipos microbianos reconhecidos. •É capaz de interpretar representações esquemáticas de vias metabólicas, e aplicar relações de proporcionalidade a vias de sinalização.
Básico (0,5 dp abaixo da média ou superior) n=5.954 (44,5%)	<ul style="list-style-type: none"> •Identifica uma situação de aplicação da teoria construtivista em uma situação de sala de aula.
Abaixo do básico: (inferior a -0,5) n=3.774 (26%)	

Fonte: O autor (2019). ●= descritores estabelecidos com base em itens âncoras e quase âncoras; ○= descritores de itens não ancorados (apenas posicionados no ponto de ancoragem em que 65% dos indivíduos acertaram o item, mas o nível imediatamente inferior não teve proporção de acertos inferior a 50%)

Com base na metodologia de padronização e reescalamento descrita em Brasil (2018, ver equação 10 e tabela 1 daquele documento) foram calculados pontos de corte para a escala de proficiência (com valores de *theta*) equivalentes à nota padronizada do CE. É importante fazer a ressalva de que tal metodologia é empregada para a média do escore dos cursos (em uma lógica de 2 níveis), e leva em consideração todas as questões de ambos os componentes de prova, porém, apenas para ilustrar de outros valores para o posicionamento dos pontos de ancoragem, em equivalência aos cinco níveis do Conceito ENADE, simulou-se a ancoragem exposta na Tabela 11:

¹⁵ Q28L - asserção e razão são praticamente idênticas

Tabela 11: Simulação do estabelecimento de pontos de ancoragem através da metodologia de padronização e escalonamento do INEP.

Theta	Conceito ENADE	Número de itens âncoras / quase-âncoras	Número de sujeitos
[-1,9209;-1,0058)	Conceito 1		1.489
[-1,0058;-0,0374)	Conceito 2		5.030
[-0,0374; 0,9311)	Conceito 3	Quase âncoras = 1	5.099
[0,9311; 1,8995)	Conceito 4	Âncoras=1; quase-âncoras = 4	1.654
[1,8995; 2,9212]	Conceito 5	Quase-âncoras = 4	102

Fonte: O autor, baseado nos microdados 2017 (BRASIL, 2018b).

Percebe-se que caso fossem aplicados tais critérios de segmentação resultaria um número inferior de itens âncoras e quase âncoras, e que tais itens não estariam posicionados em níveis capazes de otimizar a quantidade de sujeitos a eles vinculados.

5 CONCLUSÕES

O objetivo geral desse trabalho foi avaliar a qualidade da prova de conhecimentos específicos da área de Licenciatura em Ciências Biológicas - do ENADE 2017, por meio de análise de conteúdo, da TCT, de Análise Fatorial e da TRI. Após o desenvolvimento da pesquisa exposto nas seções anteriores foi possível cumprir os objetivos específicos concluindo-se que:

Conforme a pesquisa documental e a análise de conteúdo foi possível constatar que as “diretrizes de avaliação”, que orientam a elaboração da matriz de referência, diferem das DNC e não são estáveis ao longo das edições, o mesmo pode ser dito do número de questões por subcomponente (pedagógicas e biológicas) e por tipo (discursivas e objetivas). O número de questões de cada subcomponente mudou de 2014 para 2017 sem prévia publicação aos interessados (*stakeholders*). A distribuição dos conteúdos e competências dentre os itens de prova aparenta guardar marcas circunstanciais (com conteúdos irrelevantes), preferências (superdimensionamento de áreas) e negligências (sub-representação de constructos) comuns aos processos democráticos de escolhas e amostragens curriculares e avaliativas.

Observando-se os padrões de resposta, conclui-se quanto aos formatos dos itens, que caso o INEP pretenda utilizar modelos dicotômicos da teoria de resposta ao item, seria fortemente recomendado a redução ou a eliminação de formatos de elaboração de itens que permitem acertos parciais e carregamento de cargas fatoriais em mais de um fator (i.e. asserção-razão e múltipla escolha complexos). De outra forma, dada a grande quantidade de dados disponíveis no atual formato de provas, sugere-se como perspectiva para pesquisas futuras o teste de modelos politômicos (nominal, ou créditos parciais) a cursos com grande tamanho amostral, pois apenas em tais cursos será possível utilizar esses modelos mantendo-se uma boa proporção entre o número de indivíduos e o número de parâmetros estimado. Conjectura-se que tais formatos de itens contribuem para fragilizar a validade das interpretações dos escores, pois tratados dicotomicamente não avaliam

adequadamente os sujeitos de proficiência abaixo da proficiência ótima para a resposta correta ao item, apesar dos microdados conterem tais informações.

A estrutura fatorial da prova revela uma relevante separação entre as dimensões pedagógica e disciplinar (saberes biológicos) do saber-fazer docente, pois a análise fatorial de informação restrita revela a existência de pelo menos duas dimensões relevantes e a de informação plena confirma a separação dos itens de conhecimentos biológicos carregando no fator predominante e os itens de conhecimentos pedagógicos (conjuntamente com o item Q15) carregando no segundo fator. Essa separação vem de encontro às Diretrizes Curriculares Nacionais para a Formação de Professores da Educação Básica (BRASIL, 2001a) que prescreve que saberes disciplinares e práticos (didáticos e pedagógicos) devem estar permeados ao longo de toda formação do professor. Por isso sugere-se que ao menos as questões pedagógicas exclusivas da matriz de avaliação de ciências biológicas demandem articulações mais explícitas entre os saberes disciplinares (biológicos) e pedagógicos em suas problematizações (textos bases e/ou enunciados). Mas cabe ponderar sobre a dimensionalidade de tais competências, pois, conforme alertam Primi, Hutz e Silva (2011), sua natureza é fundamentalmente multidimensional/relacional, tornando a operacionalização da avaliação por competências bastante difícil. Dado tal dificuldade de operacionalização de que nos falam os autores, conjectura-se se não seria adequado o fornecimento de subcores para melhor representar os dois subcomponentes das provas dos cursos de licenciatura.

Apesar do razoável ajuste dos padrões de respostas dados aos itens objetivos de Conhecimentos Específicos ao modelo unidimensional de três parâmetros, conclui-se que ainda há muito o que aprimorar na prova, tanto em sua extensão (quantidade de itens dado à quantidade de conteúdos e competências que se pretende avaliar) quanto na qualidade dos itens, haja vista que cinco, dos vinte e sete itens objetivos do CE, apresentaram parâmetros inferiores ao aceitável. Inicialmente, analisando-se toda a prova de conhecimentos específicos (todos os itens objetivos), a média de dificuldade da prova foi estimada pela TRI em 2,2436, na escala (0,1). Após a exclusão dos itens com parâmetros inadequados (dificuldade

extrema e baixa discriminação), o instrumento passou a ter uma dificuldade média mais condizente com a população avaliada $\bar{x}_b=1,2931$, bem como, com estimativas de dificuldade mais precisas, dado a redução do erro padrão (EP: 0,1573).

A atual interpretação referenciada na norma (padronização de escores pela TCT) apresenta limitações potencialmente geradora de injustiças, e exclusões (elitismo do conceito 5 no ENADE), por isso, assim como Silva (2001), compreende-se a importância do estabelecimento de referências baseadas em critérios, e que isso seria mais coerente com os propósitos de regulação do SINAES, além de potencialmente desencorajar o uso dos rankings entre instituições/cursos. A partir do procedimento de ancoragem na escala, conclui-se que o teste não apresentou informação suficiente para diferenciar os estudantes em cinco extratos de desempenho (conforme os cinco níveis do conceito ENADE), e que do modo como o exame está formatado (sem a pré-teste, e com itens concentrados nos extratos mais difíceis) foi possível diferenciar apenas 4 níveis, sendo que em apenas 3 havia itens âncoras e respondentes posicionados.

De maneira geral, é possível observar que a análise da ancoragem dos itens na escala, elaborada a partir do modelo logístico de três parâmetros permitiu o desenvolvimento de descrições interpretativas, distinguindo-se em cinco níveis as inferências sobre as competências e conhecimentos dominados pela população, isso tudo com uma heurística mais informativa e acurada do que possibilitam os escores e os parâmetros calculados pela TCT.

Outras perspectivas abertas a pesquisas futuras é a possibilidade de equalização entre as provas da licenciatura e do bacharelado, tendo em vista a existência de itens de ligação (de 12 a 17 dependendo da edição e da área). Outras demandas sociais que apontam para boas perspectivas de pesquisa relacionam-se à necessidade de acompanhamento da qualidade dos cursos ao longo do tempo, com a realização de pesquisas semelhantes à desenvolvida nessa dissertação e em Alves, Ferrazza e Borgatto (2019) torna-se possível a seleção dos itens com melhores parâmetros psicométricos, e a partir disso, a realização de estudos de equalização pela composição de testes de ligação entre as diversas edições do ENADE. Também a partir de resultados de pesquisas com essa, seria possível que o

INEP elaborasse rotinas para que a cada edição do ENADE fossem devolvidos “mapas de posicionamento dos itens nas escalas”, para que pesquisadores e as Comissões Assessoras de Áreas, com emprego de técnicas de estabelecimento de padrões de desempenho (ex: bookmark) estabelecessem o referencias em critérios para as escalas.

Entende-se que, apesar de no presente estudo ter sido analisado somente uma área e ano do ENADE, a abordagem metodológica utilizada pode ser testada nos demais cursos e anos dos ciclos avaliados.

REFERÊNCIAS

ALVES, Kauê Tortato; FERRAZZA, Adriana Cioato; BORGATTO, Adriano Ferreti. Construção de uma Escala de Proficiência do ENADE 2014 para Ciências Biológicas por meio da TRI. In: Anais da X Reunião da ABAVE, 2019, São Paulo. Anais do(a) Anais da X Reunião da ABAVE. Recife: Even3, 2019. Disponível em: <http://dx.doi.org/10.29327/15427.1-4> Acesso em: 11 out. 2019.

ANDRADE, Alline Nunes. **FLUXOGRAMA DA ELABORAÇÃO DA PROVA DO ENADE**. Brasília, 2017. 20 slides, color. Disponível em: http://download.inep.gov.br/educacao_superior/enade/seminarios/2017/seminario_enade_2017_apresentacao_fluxograma_Alline.pdf. Acesso em: 31 dez. 2019.

APA; AERA; NCME American Psychological Association; American Educational Research Association; National Council on Measurement in Education. **Standards for educational and psychological tests**. American Psychological Association, Washington, D.C.: 2014. pp. 11-109 Disponível em: <https://pt.scribd.com/document/344894060/AERA-APA-NCME-2014-Standars-for-Educational-and-Psychological-Testing> Acesso em: 19 dez. 2019.

BANGERT, Art. W. Two traditions for assessing student achievement. Retrieved December, 2006. *Teaching Learning Committee: Montana State University Bozeman*. 2006. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.517.3157&rep=rep1&type=pdf> Acesso em: 09 jun. 2019

BARBOSA, G.C; FREIRE, F.S.; CRISÓSTOMO, V.L. Análise dos indicadores de gestão das IFES e o desempenho discente no ENADE. *Avaliação*, Campinas, v. 16, n. 2, p. 317- 344, 2011. Disponível em: <http://www.scielo.br/pdf/aval/v16n2/a05v16n2.pdf> Acesso em 24 jan. 2018.

BEATON, Albert E.; ALLEN, Nancy L. Chapter 6: Interpreting Scales Through Scale Anchoring. *Journal of Educational Statistics*, v. 17, n. 2, p. 191-204, 1992. Disponível em: <https://www.jstor.org/stable/1165169> Acesso em: 31 dez. 2019

BIZZO, Nelio, et al. Ciências biológicas. In: BRASIL. Ministério da Educação. *Orientações curriculares nacionais do ensino médio*. Brasília, DF: MEC, 2004. Disponível em: <http://files.biopibid2011.webnode.com.br/200000018-e836be9301/Ci%C3%AAncias%20Biol%C3%B3gicas.pdf> Acesso em: 29 mai. 2019.

BLOOM, Benjamin S. Learning for Mastery. *Instruction and Curriculum*. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. Evaluation comment, v. 1, n. 2, p. n2, 1968. Disponível em: <https://files.eric.ed.gov/fulltext/ED053419.pdf> Acesso em: 01 out. 2019.

BOLLELA, Valdes Roberto; BORGES, Marcos de Carvalho; TRONCON, Luiz Ernesto de Almeida. Avaliação somativa de habilidades cognitivas: experiência envolvendo boas práticas para a elaboração de testes de múltipla escolha e a composição de exames. **Revista Brasileira de Educação Médica**, v. 42, n. 4, p. 74-85, 2018. Disponível em: <http://dx.doi.org/10.1590/1981-52712015v42n4rb20160065> Acesso em: 17 dez. 2019

BOLLELA, Valdes Roberto; BORGES, Marcos de Carvalho; TRONCON, Luiz Ernesto de Almeida. Avaliação somativa de habilidades cognitivas: experiência envolvendo boas práticas para a elaboração de testes de múltipla escolha e a composição de exames. **Revista Brasileira de Educação Médica**, v. 42, n. 4, p. 74-85, 2018. Disponível em: <http://www.scielo.br/pdf/rbem/v42n4/1981-5271-rbem-42-4-0074.pdf> Acesso em: 17 dez. 2019.

BORGATTO, Adriano Ferreti; DE ANDRADE, Dalton Francisco. Análise Clássica de Testes com diferentes graus de dificuldade. **Estudos em Avaliação Educacional**, v. 23, n. 52, p. 146-156, 2012. Disponível em: <http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/1733/1733.pdf> Acesso em: 01 set. 2018.

BRASIL. Lei N. 6.684 de 03 de setembro de 1979. Regulamenta a profissão do Biólogo. Disponível em: http://www.crbio04.gov.br/index.php?option=com_content&view=article&id=2245&Itemid=218

BRASIL. Lei nº 9.131, de 24 de novembro de 1995. Altera dispositivos da Lei nº 4.024, de 20 de dezembro de 1961 e dá outras providências. *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 24 nov. 1995. Edição Extra, p. 19257. Disponível em: http://www.planalto.gov.br/ccivil_03/Leis/L9131.htm.

BRASIL. Lei n.º 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L9394compilado.htm

BRASIL. Parecer n.º 776/CES/CNE, de 03 de dezembro de 1997. Orienta para as diretrizes curriculares dos cursos de graduação. Disponível em: http://portal.mec.gov.br/setec/arquivos/pdf_legislacao/superior/legisla_superior_parecer77697.pdf

BRASIL. Parecer n.º 009/2001CES, de 08 de maio de 2001. Diretrizes Curriculares Nacionais para a Formação de Professores da Educação Básica, em nível superior, curso de licenciatura, de graduação plena. 2002a Disponível em: http://portal.mec.gov.br/seesp/arquivos/pdf/res1_2.pdf

BRASIL. Parecer n.º 1301/2001CES, de 06 de novembro de 2001. Diretrizes Curriculares Nacionais para os Cursos de Ciências Biológicas. 2001b Disponível em: http://portal.mec.gov.br/cne/arquivos/pdf/2001/pces1301_01.pdf

BRASIL. Resolução n.º 1CNE/CES, de 18 de fevereiro de 2002. Estabelece as Diretrizes Curriculares para os cursos de Ciências Biológicas. 2002a Disponível em: http://portal.mec.gov.br/seesp/arquivos/pdf/res1_2.pdf

BRASIL. Resolução n.º 7CNE/CES, de 11 de março de 2002. Estabelece as Diretrizes Curriculares para os cursos de Ciências Biológicas. 2002b Disponível em: <http://portal.mec.gov.br/cne/arquivos/pdf/CES07-2002.pdf>

BRASIL. Lei n.º 10.861, de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior – SINAES e dá outras providências. *Leis*, Brasília, DF, 2004. Disponível em: <http://portal.mec.gov.br/arquivos/pdf/10861.pdf>

BRASIL. Ministério da Educação. Portaria normativa n.º 40, de 12 de dezembro de 2007. Institui o e-MEC, sistema eletrônico de fluxo de trabalho e gerenciamento de informações relativas aos processos de regulação da educação superior no sistema federal de educação. Disponível em: http://download.inep.gov.br/download/condicoes_ensino/2007/Portaria_n40.pdf

BRASIL. Webster Spiegel Cassiano. Coordenação Geral do Exame Nacional de Desempenho do Estudante. **Guia de Elaboração e Revisão de Itens**: Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2011. [Mimeo]

BRASIL. Relatório Síntese de Área: Ciências Biológicas, ENADE 2014 [Relatório]. Versão: 07/04/2016. 2016. Disponível em: http://download.inep.gov.br/educacao_superior/enade/relatorio_sintese/2014/2014_r_el_ciencias_biologicas.pdf

BRASIL. Ministério da Educação - MEC. Portaria normativa n.º 8, de 26 de abril de 2017. Estabelece o regulamento do Exame Nacional de Desempenho dos Estudantes para o ano de 2017. 2017a. Disponível em: http://download.inep.gov.br/educacao_superior/enade/legislacao/2017/portaria_normativa_n8_de_26042017.pdf

BRASIL. Ministério da Educação - MEC. Edital n.º 26 INEP/MEC, de 16 de junho de 2017. Dispõe sobre as diretrizes, os procedimentos e os prazos do Enade 2017. 2017b Disponível em: http://download.inep.gov.br/educacao_superior/enade/edital/2017/edital_n26_de_16_062017_enade_2017.pdf

BRASIL. Portaria INEP n.º 471 de 6 de junho de 2017. Dispõe que a prova do Enade 2017, no componente específico da área de Ciências Biológicas – Bacharelado. 2017c Disponível em: http://download.inep.gov.br/educacao_superior/enade/legislacao/2017/ciencias_biologicas_bacharelado_-_portaria_n_471_de_6_de_junho_de_2017.pdf

BRASIL. Portaria INEP nº 472 de 6 de junho de 2017. Dispõe que a prova do Enade 2017, no componente específico da área de Ciências Biológicas – Licenciatura. 2017d Disponível em: http://download.inep.gov.br/educacao_superior/enade/legislacao/2017/ciencias_biologicas_licenciatura_-_portaria_n_472_de_6_de_junho_de_2017.pdf

BRASIL. Portaria INEP n.º 510 de 6 de junho de 2017. Dispõe sobre Exame Nacional de Desempenho dos Estudantes (ENADE) da Área de Pedagogia. 2017e Disponível em: http://download.inep.gov.br/educacao_superior/enade/legislacao/2017/pedagogia_-_portaria_n_510_de_6_de_junho_de_2017.pdf

BRASIL. ENADE2017: Ciências Biológicas: Licenciatura [Prova]. 2017f. Disponível em: http://download.inep.gov.br/educacao_superior/enade/provas/2017/04_CIE_COM_LI_CENCIATURA_BAIXA.pdf

BRASIL. Relatório Síntese de Área: Ciências Biológicas (Bacharelado/Licenciatura), ENADE 2017 [Relatório]. 2018a. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>

BRASIL. Microdados ENADE 2017 [Banco de dados]. 2018b. Disponível em: <http://portal.inep.gov.br/web/guest/microdados>

BRASIL. Nota técnica n.º 16/2018/CGCQES/DAES de 14 de maio de 2018. Apresenta a metodologia utilizada no cálculo do Conceito ENADE referente ao ano de 2017. 2018c Disponível em: http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2017/nota_tecnica_n16_2018_calculo_conceito-enade.pdf

BRASIL; INEP, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Pedido de informação registrado sob o protocolo nº 23480-022380/2019-00 Sobre a Matriz de referência ENADE Ciências Biológicas 2011, 2014 e 2017.** [mensagem pessoal] Mensagem recebida por: <Kauê Tortato>. em: 01 out. 2019. (anexo 2)

BROWN, Timothy A. **Confirmatory factor analysis for applied research.** Guilford publications, 2015. Disponível em: <https://books.google.com.br/books?id=ITL2BQAAQBAJ&lpg=PP1&hl=pt-BR&pg=PP1#v=onepage&q&f=false> Acesso em: 24 fev. 2020.

CAI, Li; HANSEN, Mark. Limited-information goodness-of-fit testing of hierarchical item factor models. **British Journal of Mathematical and Statistical Psychology**, v. 66, n. 2, p. 245-276, 2013. Disponível em: DOI: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3760206/> Acesso em: 24 fev. 2020.

CAMARGO, Raphael Vinicius Weigert; Camargo, R. D. C. C. P.; de Andrade, D. F.; Bornia, A. C.. Desempenho dos alunos de Ciências Contábeis na prova Enade/2012: uma aplicação da Teoria da Resposta ao Item. **Revista de Educação e Pesquisa em Contabilidade**, v. 10, n. 3, p. 332-355, 2016. Disponível em: <http://dx.doi.org/10.17524/repec.v10i3.1401> Acesso em: 18 fev. 2019.

CAMPOS, Fernanda Cristina dos Santos. Elaboração da prova do ENADE no modelo do banco nacional de itens. Dissertação (mestrado profissional) - Universidade Federal de Juiz de Fora, Faculdade de Educação/CAEd. Programa de Pós-Graduação em Gestão e Avaliação da Educação Pública, 2013. 89 p. Disponível em: <http://www.mestrado.caedufff.net/elaboracao-da-prova-do-enade-no-modelo-do-banco-nacional-de-itens/> Acesso em: 08 set. 2018

CARROLL, John B.; SPEARRITT, Donald. A Study of a" Model of School Learning." Monograph Number 4. 1967. Disponível em: <https://files.eric.ed.gov/fulltext/ED045477.pdf> Acesso em: 22 out. 2019.

CHALMERS, R. Philip, et al. mirt: A multidimensional item response theory package for the R environment. **Journal of Statistical Software**, v. 48, n. 6, p. 1-29, 2012. Disponível em: <https://www.jstatsoft.org/htaccess.php?volume=48&type=i&issue=06&paper=true> Acesso em: 12 out. 2018

CIZEK, Gregory J.; BUNCH, Michael B.; KOONS, Heather. Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, v. 23, n. 4, p. 31-31, 2004. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.2004.tb00166.x> Acesso em: 29 dez. 2018

COELHO, Edy Célia. Teoria da resposta ao item: desafios e perspectivas em exames multidisciplinares. Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Métodos Numéricos em Engenharia. 2014. Disponível em: <http://hdl.handle.net/1884/36872> Acesso em: 02 mar. 2019

COELHO, Edy Célia; RIBEIRO JUNIOR, Paulo Justiniano; BONAT, Wagner Hugo. Exame nacional de desenvolvimento de estudantes de estatística-desafios e perspectivas pela TRI. **Revista da Estatística UFOP**, v. 3, p. 2, 2014a, Disponível em: https://www.researchgate.net/profile/Wagner_Bonat2/publication/281968326_EXAME_NACIONAL_DE_DESENVOLVIMENTO_DE_ESTUDANTES_DE_ESTADISTICA_DESAFIOS_E_PERSPECTIVAS_PELA_TRI/links/55ffe91e08aec948c4f9c384.pdf Acesso em: 19 jan. 2019

COELHO, Edy Célia; RIBEIRO JUNIOR, Paulo Justiniano; BONAT, Wagner Hugo. Teoria de Resposta ao Item: Desafios e perspectivas no ENADE 2009 de estatística. In: 4º SIMPÓSIO DE MÉTODOS NUMÉRICOS COMPUTACIONAIS, 2014b,

Curitiba. **Periódico do Simpósio de Métodos Numéricos Computacionais da Universidade Federal do Paraná**. Curitiba: Ufpr-ppgmne, 2014. p. 157 - 163. Disponível em: <http://www.smnc.ufpr.br/publicacoes/Periodico-2014.pdf>. Acesso em: 02 mar. 2019.

COHEN, R. J.; SWERDLIK, M. E.; STURMAN, E. D. **Testagem e avaliação psicológica**: introdução a testes e medidas (8a ed). Porto Alegre, RS: AMGH. 2014

CORTELAZZO, Angelo Luiz; RIBEIRO, Viktória Kövesdy. Enade 2005 e 2008: desempenho dos estudantes de biologia de instituições de Educação Superior estaduais e municipais de São Paulo. **Ciência & Educação (Bauru)**, Bauru, v. 19, n. 2, p. 409-424, 2013. Disponível em: <http://dx.doi.org/10.1590/S1516-73132013000200012> Acesso em: 18 jan. 2019.

DA MATA, Andreia Silva. Análise das habilidades e competências das provas do ENADE (2005 e 2008) dos cursos de engenharia civil. Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Educação, Campinas, SP. 2016. (334 p.). Disponível em: <http://www.repositorio.unicamp.br/handle/REPOSIP/321134> Acesso em: 21 fev. 2019

DAMÁSIO, Bruno Figueiredo. Uso da análise fatorial exploratória em psicologia. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, v. 11, n. 2, p. 213-228, 2012. Disponível em: <http://pepsic.bvsalud.org/pdf/avp/v11n2/v11n2a07.pdf> Acesso em: 17 nov. 2019

DE ANDRADE, Dalton Francisco; TAVARES, Heliton Ribeiro; DA CUNHA VALLE, Raquel. **Teoria da Resposta ao Item**: conceitos e aplicações. Associação Brasileira de Estatística (ABE), São Paulo, 2000. Disponível em: <http://www.ime.unicamp.br/~cnaber/LivroTRI.pdf> Acesso em: 13 mai. 2018

EISNER, Elliot. The Use of Qualitative Forms of Evaluation For Improving Educational Practice. *Educational Evaluation and Policy Analysis*. v.1, n.6. 1979. Doi: <https://doi.org/10.3102/01623737001006011>

CUMMING, Tammie; EWELL, Peter. Introduction: History and Conceptual Basis of Assessment in Higher Education. In: Cumming, T. & Miller, D. (Eds.), *Enhancing assessment in higher education: Putting psychometrics to work*. Sterling, VA: Stylus. 2017. Disponível em: https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1236&context=ny_pubs Acesso em: 10 jan. 2020.

FERREIRA, M. S.; SELLES, S. E. Entrelaçamentos históricos das Ciências Biológicas com a disciplina escolar Biologia: investigando a versão azul do BSCS. **Anais**: V Encontro Nacional de Pesquisa em Educação em Ciências, 2005.

FIGUEIREDO FILHO, Dalson Brito; SILVA JUNIOR, José Alexandre da. Visão além do alcance: uma introdução à análise fatorial. **Opinião pública**, Campinas, v. 16, n.

1, p. 160-185, Jun 2010 . Disponível em: <http://dx.doi.org/10.1590/S0104-62762010000100007>. Acesso em: 15 dez. 2018

ForGRAD, Fórum de Pró-Reitores de Graduação das Universidades Brasileiras. **Diretrizes Curriculares para os Cursos de Graduação**. 2000 Disponível em: <http://portal.mec.gov.br/sesu/arquivos/pdf/DocDiretoria.pdf> Acesso em: 07 ago. 2018

FRANCISCO, Reinaldo. Aplicação da teoria da resposta ao item (TRI) no exame nacional de cursos (ENC) da UNICENTRO. Dissertação (mestrado) 2005. Disponível em: https://www.acervodigital.ufpr.br/bitstream/handle/1884/2349/TRI_FINAL.pdf?sequen ce=1 Acesso em: 04 mar. 2019.

HABERMAN, Shelby J.; SINHARAY, Sandip; LEE, Yi-Hsuan. Statistical procedures to evaluate quality of scale anchoring. **ETS Research Report Series**, v. 2011, n. 1, p. i-20, 2011 Disponível em: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2333-8504.2011.tb02238.x> Acesso em: 30 dez. 2019

HAIR JUNIOR, Joseph F., et al. Exploratory Factor Analysis. In: HAIR JUNIOR, Joseph F., et al. **Multivariate Data Analysis**: Pearson New International Edition. 7. ed. Londres: Pearson, 2014. Cap. 3. p. 89-150. Disponível em: <https://is.muni.cz/el/1423/podzim2017/PSY028/um/ Hair - Multivariate data analysis 7th revised.pdf>. Acesso em: 10 jan. 2020.

KLEIN, Stephen; LIU, Ou Lydia; SCONING, James; et.al **Test Validity Study (TVS) Report**. 2009. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.173.3647&rep=rep1&type=pdf> Acesso em: 18 ago. 2019

KLINE, R. B. **Methodology in the social sciences**. Principles and practice of structural equation modeling 2^a ed. Guilford Press, 2005

LANDEIRA-FERNANDEZ, J.; PRIMI, Ricardo. Comparação do desempenho entre calouros e formandos no Provão de Psicologia 2000. *Psicologia: Reflexão e Crítica*, v. 15, n. 1, p. 219-234, 2002. Disponível em: <http://dx.doi.org/10.1590/S0102-79722002000100023> Acesso em 21 set. 2019

LAROS, Jacob A. O uso da análise fatorial: algumas diretrizes para pesquisadores. In: *Análise fatorial para pesquisadores*, v. 1, p. 145, 2012. Disponível em: <https://www.researchgate.net/publication/233735561> Acesso em: 16 jan. 2019

LEWIS, D. M.; MITZEL, H. C. An item response theory based standard setting procedure. In: **DR Green (Chair), Some uses of item response theory in standard setting**. Symposium conducted at the annual meeting of the California Educational Research Association, Lake Tahoe, NV. 1995.

LIMANA, A.; BRITO, M. R. F. O modelo de avaliação dinâmica e o desenvolvimento de competências: algumas considerações a respeito do Enade. **Avaliação: Revista**

da Avaliação da Educação Superior, Campinas, Sorocaba, São Paulo, v. 10, n. 2, p. 9-32, jun. 2005. Disponível em: <http://periodicos.uniso.br/ojs/index.php/avaliacao/article/view/1303> Acesso em: 24 set. 2019

LIRA, Sachiko Araki. **Análise de correlação: abordagem teórica e de construção dos coeficientes com aplicações**. 2004, 209f. 2004. Tese de Doutorado. Dissertação (Mestrado em Ciências). Programa de Pós-Graduação em Métodos Numéricos em Engenharia da Universidade Federal do Paraná. Paraná. Disponível em: http://www.ipardes.gov.br/biblioteca/docs/dissertacao_sachiko.pdf Acesso em: 20 ago. 2018

LOPES, F. L. **Equalização de provas de habilidades acadêmicas: Estudos a partir do ENADE via TRI**. Tese de Doutorado. Dissertação de Mestrado. Programa de Pós-Graduação Stricto Sensu em Psicologia, Universidade São Francisco, Itatiba. 2011. Disponível em: <https://www.usf.edu.br/galeria/getImage/385/514726631600918.pdf> Acesso em: 05 mar. 2019

LOPES, Fernanda Luzia; VENDRAMINI, Claudette Maria Medeiros. Equalização de provas acadêmicas via Teoria de Resposta ao Item. **Psico-USF**, v. 18, n. 1, p. 141-149, 2013. Disponível em: <http://www.scielo.br/pdf/psuf/v18n1/v18n1a15.pdf> Acesso em: 10 nov. 2018.

LOPES, Fernanda Luzia; VENDRAMINI, Claudette Maria Medeiros. Propriedades psicométricas das provas de pedagogia do ENADE via TRI. **Avaliação, Campinas; Sorocaba**, v. 20, n. 1, p. 27-47, 2015. Disponível em: Disponível em: <https://www.redalyc.org/pdf/4010/401036093015.pdf> Acesso em: 10 nov. 2018.

LORENZO-SEVA, Urbano; TIMMERMAN, Marieke E.; KIERS, Henk AL. The Hull method for selecting the number of common factors. **Multivariate behavioral research**, v. 46, n. 2, p. 340-364, 2011 Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00273171.2011.564527> Acesso em: 18 fev. 2020

MARANDINO, Martha; SELLES, Sandra Escovedo; FERREIRA, Marcia Serra. **Ensino de Biologia**: histórias e práticas em diferentes espaços educativos. São Paulo, SP: Cortez. 2009.

MARINHO-ARAUJO, Claisy Maria; RABELO, Mauro Luiz. Avaliação de perfil e de competências dos estudantes da educação superior no Brasil: a matriz de referência nas provas do Enade. **Psicologia, Educação e Cultura**, p. 9-26, 2016. Disponível em: <https://comum.rcaap.pt/bitstream/10400.26/28290/1/PEC%20Maio%202016.pdf#page=9> Acesso em: 06 out. 2019

MAROCO, João; GARCIA-MARQUES, Teresa. Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas?. **Laboratório de psicologia**, p. 65-90, 2006. Disponível em: <http://hdl.handle.net/10400.12/133> Acesso em: 26 ago. 2018

MOLCK, Adauto Marin; CALDERÓN, Adolfo Ignacio. Exame Nacional de Desempenho de Estudantes: mapeamento e tendências temáticas da produção científica brasileira (2004-2010). **Educação On-Line (PUC-Rio)**, v. 15, p. 57-77, 2014. Disponível em: http://www.academia.edu/download/36627373/ENADE_-_RJ.pdf Acesso em: 13 mai. 2018.

MOREIRA JUNIOR, Fernando de Jesus. Contribuições da Teoria da Resposta ao Item nas Avaliações Educacionais. **Ciência e Natura**, v. 36, p. 58-72, 2014. Disponível em: <https://www.redalyc.org/pdf/4675/467546183006.pdf> Acesso em: 01 jun. 2019.

NOGUEIRA, Samantha Oliveira. ENADE: Análise de Itens de Formação Geral e de Estatística pela TRI. 2008. 166 f. 2008. Dissertação (Mestrado em Psicologia)- Programa de Pós-Graduação Stricto Sensu em Psicologia, Universidade São Francisco, Itatiba. Disponível em: <http://livros01.livrosgratis.com.br/cp084008.pdf>. Acesso em: 17 jan. 2019.

NOJOSA, Ronald T. Modelos multidimensionais para a teoria de resposta ao item. Dissertação de Mestrado. Universidade Federal de Pernambuco, Recife. Pernambuco: UFPE, 2001. Disponível em: https://www.ime.unicamp.br/~cnaber/Ronald_dissertcao.pdf Acesso em: 22 dez. 2019.

NUSCHE, Deborah. **Assessment of Learning Outcomes in Higher Education: a comparative review of selected practices**. OECD Education Working Papers. OECD, 2008. n. 15. Disponível em: <https://doi.org/10.1787/244257272573>. Acesso em 29 jul. 2019

NYE, Christopher D.; DRASGOW, Fritz. Assessing goodness of fit: Simple rules of thumb simply do not work. **Organizational Research Methods**, v. 14, n. 3, p. 548-570, 2011. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/1094428110368562> Acesso em: 26 dez. 2019.

OCDE - Organização para a Cooperação e o Desenvolvimento Econômico. **Repensando a Garantia de Qualidade para o Ensino Superior no Brasil** [Rethinking Quality Assurance for Higher Education in Brazil: Reviews of National Policies for Education], 2018. Disponível em: http://download.inep.gov.br/acoes_internacionais/ocde/Repensando_a_Garantia_de_Qualidade_para_o_Ensino_Superior_no_Brasil_PT.pdf Acesso em: 31 out. 2019.

OLIVEIRA, K. S. **Avaliação do exame nacional de desempenho do estudante pela teoria de resposta ao item.** 2006. 96 f. 2006. Dissertação (Mestrado em Psicologia)-Programa de Pós-Graduação Stricto Sensu em Psicologia, Universidade São Francisco, Itatiba. Disponível em: <http://www.dominiopublico.gov.br/download/texto/cp019960.pdf> Acesso em: 30 set. 2018.

OLIVEIRA, André Luiz Santos de. Avaliação psicométrica da medida do componente de formação geral da prova do exame nacional de desempenho de estudantes (ENADE) de 2010, 2011 e 2012. Dissertação (mestrado) Programa de Pós-Graduação em Métodos e Gestão em Avaliação da Universidade Federal de Santa Catarina. 2017. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/182889/351614.pdf?sequence=1&isAllowed=y> Acesso em: 17 jan. 2019.

OTTER, S. Learning Outcomes in Higher Education, A Development Project Report, Unit for the Development of Adult Continuing Education (UDACE). p. 145. 1992. Disponível: <https://files.eric.ed.gov/fulltext/ED354397.pdf>

PASQUALI, et al. **Teoria e Métodos de Medida em ciências do comportamento.** Brasília: Laboratório de Pesquisa em Avaliação e Medida. Instituto de Psicologia, Universidade de Brasília. INEP. 432p. 1996. Disponível em: <http://www.dominiopublico.gov.br/download/texto/me002482.pdf>. Acesso em: 09 ago. 2018

PASQUALI, L. **Psicometria:** Teoria dos testes na Psicologia e na Educação. 5. ed. Petrópolis, RJ: Vozes, 2013.

PASQUALI, Luiz. **Teoria de resposta ao item:** teoria, procedimentos e aplicações. 1ªed. Appris: Curitiba, 2018. 277 p. ISBN: 978-8547315696

PASQUALI, Luiz; PRIMI, Ricardo. Fundamentos da teoria da resposta ao item: TRI. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, Porto Alegre ,v. 2, n. 2, p. 99-110, dez. 2003 . Disponível em http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712003000200002&lng=pt&nrm=iso Acesso em: 12 ago. 2018

PEDROSO, Carla Vargas; SELLES, Sandra Lucia Escovedo. Investigando a origem dos cursos de Ciências Biológicas no Brasil. **Anais:** X Encontro Nacional de Pesquisa em Educação em Ciências – X ENPEC Águas de Lindóia, SP – 24 a 27 de Novembro de 2015. Disponível em: <http://www.nutes.ufrj.br/abrapec/venpec/conteudo/artigos/3/doc/p892.doc> Acesso em: 30 mai. 2019

PENG, Chao-Ying Joanne. Biserial Correlation. In: LEWIS-BECK, Michael S., BRYMAN, Alan; LIAO, Tim Futing. **The SAGE Encyclopedia of Social Science**

Research Methods. Thousand Oaks, CA: Sage Publications, 2004. [Disponível em: http://dx.doi.org/10.4135/9781412950589.n70](http://dx.doi.org/10.4135/9781412950589.n70) Acesso em: 30 dez. 2019

PIKE, Gary R. Background, college experiences, and the ACT-COMP exam: Using construct validity to evaluate assessment instruments. **The Review of Higher Education**, v. 13, n. 1, p. 91-117, 1989. Disponível em: <https://doi.org/10.1353/rhe.1989.0005> Acesso em: 10 jan. 2020.

POLIDORI, Marlis Morosini. Políticas de avaliação da educação superior brasileira: Provão, SINAES, IDD, CPC, IGC e... outros índices. **Avaliação, Campinas; Sorocaba, SP**. v. 14, n. 2, p. 267-290, jul. 2009. Disponível em: <http://www.scielo.br/pdf/aval/v14n2/a09v14n2.pdf> Acesso em: 17 jan. 2018

PRIMI, Ricardo; DA SILVA, Marjorie Cristina Rocha; BARTHOLOMEU, Daniel. A validade do Enade para avaliação de cursos superiores. **Examen: Política, Gestão E Avaliação Da Educação**, v. 2, n. 2, p. 128-151, 2018. Disponível em: <https://examen.emnuvens.com.br/rev/article/download/89/47> Acesso em: 10 mai. 2019.

PRIMI, Ricardo; de CARVALHO, L. F.; MIGUEL, F. K.; da SILVA, M. C. R. Análise do funcionamento diferencial dos itens do Exame Nacional do Estudante (ENADE) de psicologia de 2006. **Psico-USF**, v. 15, n. 3, p. 379-393, 2010. Disponível em: <https://www.redalyc.org/pdf/4010/401036083011.pdf> Acesso em: 10 nov. 2018.

PRIMI, Ricardo; HUTZ, Cláudio S; SILVA, Marjorie Cristina Rocha da. A prova do ENADE de psicologia 2006: concepção, construção e análise psicométrica da prova. **Aval. psicol.**, Itatiba, v. 10, n. 3, p. 271-294, dez. 2011. Disponível em: <http://pepsic.bvsalud.org/pdf/avp/v10n3/v10n3a04.pdf> Acessos em 04 ago. 2018.

PRICE, Larry R.. Test Development. In: PRICE, Larry R.. **Psychometric methods: Theory into practice**. Nova York: Guilford Publications, 2017. p. 165-201. (Methodology in the social sciences).

R Core Team (2018). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, (Versão 3.6.1) [Software]. Vienna, Austria. Disponível em: <https://www.R-project.org>

RECKASE, Mark D. Unifactor latent trait models applied to multifactor tests: Results and implications. **Journal of educational statistics**, v. 4, n. 3, p. 207-230, 1979. Disponível em: <http://www.jstor.org/stable/1164671> Acesso em: 17 nov. 2019

REVELLE, William R. **psych: Procedures for personality and psychological research**. 2017.

REVELLE, William; REVELLE, Maintainer William. Package 'psych'. **The Comprehensive R Archive Network**, 2015. [Disponível em: http://www.test.personality-project.org/r/psych/psych-manual.pdf](http://www.test.personality-project.org/r/psych/psych-manual.pdf) Acesso em: 05 fev. 2019.

RISTOFF, Dilvo Ilvo; GIOLO, Jaime. O SINAES como sistema. RBPG. **Revista Brasileira de Pós-Graduação**, Brasília, v. 3, p. 193-213, 2006. Disponível em: <http://www2.capes.gov.br/rbpg/portal/conteudo/Est_Artigo2_n6.pdf> Acesso em: 17 jan. 2018

RStudio (2018). **RStudio**: Integrated development environment for R (Versão 1.1.456) [Software]. Boston, MA. Disponível em: <https://rstudio.com/products/rstudio/download/>

SANTO, Alexandre do Espírito. Técnicas empregadas na análise de itens. **Semina: Ciências Sociais e Humanas**, v. 1, n. 2, p. 37-42, 1978 Disponível em: <http://www.uel.br/seer/index.php/seminasoc/article/viewFile/7200/6362> Acesso em 14 ago. 2018.

SCHER, Vinícius Teodoro; Moreira Junior, F. D. J.; Correa, A. C.; Schuch Junior, V. F.; Andrade, D. F. D.; Bortolotti, S. L. V.. Uma aplicação da teoria da resposta ao item na avaliação do ENADE do Curso de Administração. 2014. Disponível em: <https://core.ac.uk/download/pdf/30408247.pdf> Acesso em: 07 jan. 2019.

SILVA, Marjorie CR. A validade do ENADE para avaliação da qualidade dos cursos: Modelo de rasch e multinível. Itatiba: Universidade São Francisco, 2011, 204 p. 2011. Tese de Doutorado. Tese (Doutorado)–Programa de Pós-graduação em Psicologia, Universidade São Francisco, Itatiba, São Paulo. Disponível em: <https://www.usf.edu.br/galeria/getImage/427/606693926076531.pdf> Acesso em 01 mai. 2019

SILVEIRA, Pablo Magno da. Criação de um índice de satisfação com a vida por meio da teoria da resposta ao item e fatores associados em trabalhadores brasileiros. Dissertação (mestrado) Programa de Pós-Graduação em Educação Física da Universidade Federal de Santa Catarina. 2015. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/135687/335077.pdf?sequence=1> Acesso em: 01 jan. 2020

SOARES, Tufi Machado. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. **Pesquisa Operacional**, v. 25, n. 1, p. 83-112, 2005. Disponível em: <http://dx.doi.org/10.1590/S0101-74382005000100006> Acesso em: 12 ago. 2018.

SOBRINHO, José Dias. Sobre a proposta do Sistema Nacional de Avaliação da Educação Superior. **Avaliação-Revista da Avaliação da Educação Superior**, v. 9, n. 1, 2004. Disponível em: <http://periodicos.uniso.br/ojs3/index.php/avaliacao/article/view/1257>

STERNBERG, Robert J., et al. **Beyond IQ: A triarchic theory of human intelligence**. CUP Archive, 1985.

STERNBERG, Robert J.; GRIGORENKO, Elena L. **Dynamic testing**: the nature and measurement of learning potential. Cambridge university press, 2002.

TABACHNICK, B.; FIDELL, L. **Using multivariate statistics**: Pearson new international edition. International edition of sixth edition. ed: United Kingdom: Pearson Education, 2014.

TREVISAN, Ligia Maria Vettorato Barbetta, P. A., de Andrade, D. F., Rocha, G. T., & de Macedo Azevedo, T. C. A. Dimensionalidade e escala de proficiência em uma prova interdisciplinar. **Estudos em Avaliação Educacional**, v. 30, n. 74, p. 392-419, 2019. Disponível em: <http://dx.doi.org/10.18222/eae.v30i74.5324> Acesso em: 26 dez. 2019

UBRIACO, Fabiana E. C. A. Interpretação de Escalas de Medida de Competência Matemática. Dissertação de Mestrado, Faculdade de Educação, UFMG 2009. Disponível em: <http://hdl.handle.net/1843/FAEC-84RNKN> Acesso em: 05 jan. 2020

VAN DER LINDEN, Wim J. **Handbook of Item Response Theory**, Volume Three: Applications. Chapman and Hall/CRC, 2017. Disponível em: <https://doi.org/10.1201/9781315117430> Acesso em: 05 jan. 2020

VENDRAMINI, Claudette Maria Medeiros. (2009). Análise Fatorial baseada na TRI do componente de formação geral do ENADE. **Revista de Educação ANEC**, v.38, n.150, p. 153-163, 2009.

VENDRAMINI, Claudette Maria Medeiros; DIAS, Anelise Silva. Teoria de Resposta ao Item na análise de uma prova de estatística em universitários. **Psico-USF**, v. 10, n. 2, p. 201-210, 2005. Disponível em: <http://www.scielo.br/pdf/psuf/v10n2/v10n2a12> Acesso em: 22 dez. 2019.

VERHINE, Robert Evan; DANTAS, Lys Maria Vinhaes; SOARES, José Francisco. From the National Course Exam (Provão) to ENADE: a comparative analysis of national exams used in Brazilian High School. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 14, n. 52, p. 291-310, set. 2006. Disponível em: <http://dx.doi.org/10.1590/S0104-40362006000300002> Acesso em: 24 jan. 2018.

VILARINHO, Ana Paula Lima. Uma proposta de análise de desempenho dos estudantes e de valorização da primeira fase da OBMEP. Dissertação Departamento de Matemática da Universidade de Brasília Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT 2015. Disponível em: https://repositorio.unb.br/bitstream/10482/19335/3/2015_AnaPaulaLimaVilarinho.pdf Acesso em: 07 jan. 2020.

WARREN, Jonathan. Learning as an indicator of educational quality. **Studies in Higher Education**, v. 17, n. 3, p. 337-348, 1992. Disponível em: <http://dx.doi.org/10.1080/03075079212331382587> Acesso em: 20 jun. 2019

YORK, T. T.; GIBSON, C.; RANKIN, S. Defining and measuring academic success. **Practical Assessment, Research & Evaluation**, v. 20, n.5, p.2, 2015. Disponível em: <https://scholarworks.umass.edu/pare/vol20/iss1/5> Acesso em: 07 jun. 2019

ZLATKIN-TROITSCHANSKAIA, Olga; SHAVELSON, Richard J.; KUHN, Christiane. The international state of research on measurement of competency in higher education. **Studies in Higher Education**, v. 40, n. 3, p. 393-411, 2015. <https://doi.org/10.1080/03075079.2015.1004241> Acesso em: 01 jun. 2019

APÊNDICE A – Reconstituição da Matriz de Referência inferida a partir do “Anexo IX Concepção e Elaboração das Provas do ENADE” do Relatório de Área (BRASIL, 2018a)

Competência (C#)	C1	C2	C4	C6	C7	C8	C9	C10
Caract. de Perfil (CP#)								
CP1		Q10[Oc1,Oc2] Q11[Oc6,Oc7,Oc13] Q12[Oc2,Oc4,Oc11] Q13[Oc7,Oc8,Oc9] Q14[Oc6,Oc7,Oc13] Q16[Oc1,Oc4,Oc11] Q21[Oc3,Oc5,Oc11]	Q17[Oc9,Oc12]		Q09[Oc2,Oc5]			
CP3	Q19[Oc14] Q22[Oc9,Oc10]					Q15[Oc11,Oc14]	QL27[Oc17,Oc18]	
CP4				Q23[Oc1,Oc6,Oc7]			QL26[Oc16,Oc17] QL29[Oc16,Oc17]	
CP5				QL24[Oc12,Oc15]				
CP6		Q20[Oc2,Oc8,Oc11]			Q18[Oc3,Oc11]		QL25[Oc16,Oc17]	
CP7				QL28[Oc14,Oc15,Oc18]				QL30[Oc15,Oc17,Oc18]

Continua na próxima página

Competência (C#) Caract. de Perfil (CP#)	C4(P)	C3(P)	C6(P)2	C5(P)
CP4(P)	QL31[OcP5,OcP7,OcP14] QL34[OcP4]			
CP3(P)		QL32[OcP11,OcP19]	QL33[OcP11]	
CP2(P)				QL35[OcP5,OcP14,OcP15]

Legenda: CP1. observador, crítico[a] e integrador ao interpretar e avaliar os padrões e processos biológicos e suas interfaces com outras áreas do saber; CP3. ético, com responsabilidade social e ambiental, comprometido com a sua contínua atualização profissional e com a divulgação científica; CP4. consciente de sua responsabilidade como educador frente à comunidade, nos vários contextos de atuação profissional[e], compreendendo a ciência como uma atividade social com potencialidades e limitações; CP5. empático, propositivo e colaborativo nas relações interpessoais que envolvem o mundo do trabalho; CP6 criativo e empreendedor na concepção de ideias inovadoras para o desenvolvimento humano e de sociedades sustentáveis; CP7. sensível às questões ligadas aos direitos humanos, à diversidade sociocultural e ambiental e à identidade de gênero no contexto escolar. C1. elaborar, executar e gerenciar programas e projetos nas áreas de Biodiversidade e Ambiente, Saúde e Biotecnologia; C2. identificar e problematizar questões inerentes às Ciências Biológicas, incluindo seus aspectos científicos, históricos e filosóficos; C4. executar técnicas básicas e aplicadas, em laboratório e em campo, no contexto educacional; C6. instrumentalizar a comunidade escolar para a prática social crítica, emancipatória e transformadora; C7. comunicar e divulgar o conhecimento científico; C8. desenvolver produtos e processos de inovação científica, tecnológica e educacional; C9. elaborar, aplicar e avaliar estratégias pedagógicas e materiais didáticos; C10. desenvolver e aplicar estratégias de avaliação na prática docente. C4(P) compreender as abordagens do conhecimento pedagógico que fundamentam o processo educativo na Educação Infantil e anos iniciais do Ensino Fundamental; C3(P) planejar, desenvolver e avaliar situações de aprendizagem de crianças, jovens e adultos, observando as dimensões física, cognitiva, afetiva, estética, cultural, lúdica, artística, ética e biossocial; C6(P) planejar, implementar e avaliar projetos educativos contemplando e articulando a diversidade e as múltiplas relações das esferas social, cultural, ética, estética, científica e tecnológica; C5(P) planejar, desenvolver e avaliar situações de ensino-aprendizagem propondo metodologias específicas para as diferentes áreas, considerando as múltiplas dimensões da formação humana;

I. elaborar, executar e gerenciar programas e projetos nas áreas de <i>Biodiversidade e Ambiente, Saúde e Biotecnologia</i> .	II. Identificar e problematizar questões inerentes às Ciências Biológicas, incluindo seus aspectos científicos, históricos e filosóficos;	III. realizar consultorias, <i>educacionais e outras atividades afins</i> definidas na legislação e em políticas públicas.	IV. executar técnicas básicas e aplicadas, em laboratório e em campo, <i>no contexto educacional</i> ;	V. realizar atividades de gestão educacional nos diversos espaços institucionais;	VI. instrumentalizar a comunidade escolar para a prática social crítica, emancipatória e transformadora;	VII. comunicar e divulgar o conhecimento científico;	VIII. desenvolver produtos e processos de inovação científica, tecnológica e educacional;	IX. elaborar, aplicar e avaliar estratégias pedagógicas e materiais didáticos;	X. desenvolver e aplicar estratégias de avaliação na prática docente.	IV(P) compreender as abordagens do conhecimento pedagógico que fundamentam o processo educativo na Educação Infantil e anos iniciais do Ensino Fundamental ;	III(P) planejar, desenvolver e avaliar situações de aprendizagem de crianças, jovens e adultos, observando as dimensões física, cognitiva, afetiva, estética, cultural, lúdica, artística, ética e biossocial;	VI(P) planejar, implementar e avaliar projetos educativos contemplando e articulando a diversidade e as múltiplas relações das esferas social, cultural, ética, estética, científica e tecnológica;	VI(P) planejar, desenvolver e avaliar fluxos de ensino-aprendizagem propondo metodologias específicas para as diferentes áreas, considerando as múltiplas dimensões da formação humana.	Total de itens	% Itens	Total de objetos de conhecimento
I. observador, crítico[a] e integrador ao interpretar e avaliar os padrões e processos biológicos e suas interfaces com outras áreas do saber;	Q10[C1,C2], Q11[C6,C7]		Q17[C9,C12]			Q09[C2,C5]								9	33%	24
II. comprometido com a produção de conhecimento , a transformação social e a educação emancipatória ;														0	0%	0
III. ético, com responsabilidade social e ambiental, comprometido com a sua contínua atualização profissional e com a divulgação científica ;	Q19[C14], Q22[C9,C10]						Q15[C11], Q127[C17,C18]							4	15%	7
IV. consciente de sua responsabilidade como educador frente à comunidade, nos vários contextos de atuação profissional[e], compreendendo a ciência como uma atividade social com potencialidades e limitações;				Q23[C1,C6,C7]		Q26[C16,C17], Q129[C16,C17]								3	11%	7
V. empático, propositivo e colaborativo nas relações interpessoais que envolvem o mundo do trabalho;				Q24[C12,C15]										1	4%	2
VI. criativo e empreendedor na concepção de ideias inovadoras para o desenvolvimento humano e de sociedades sustentáveis;	Q20[C2,C8,C11]			Q18[C3,C11]		Q25[C16,C17]								3	11%	7
VII. sensível às questões ligadas aos direitos humanos, à diversidade sociocultural e ambiental e à identidade de gênero no contexto escolar.				Q128[C14,C15,C18]		Q130[C15,C17,C18]								2	7%	6
IV(P) Comprometido com a fundamentação epistemológica implicada nos estudos teórico-práticos, na investigação e na reflexão crítica no campo da educação, tendo em vista os conhecimentos filosófico, histórico, antropológico, ambiental-ecológico, psicológico, linguístico, sociológico, político, econômico e cultural;									Q131[CP5,CP7,CP14], Q134[CP4]					2	7%	4
III(P) Envolvido com a construção de uma sociedade incluyente, equânime, justa e solidária, sensível às diferenças de natureza ambiental-ecológica, étnico-racial, de gêneros, faixas geracionais, classes sociais, religiões, necessidades especiais, orientação sexual, entre outras										Q132[CP1], Q133[CP11]				2	7%	3
II(P) Responsável no exercício do planejamento, da organização, da avaliação e da gestão, em contextos escolares e não escolares;						Q135[CP5]								1	4%	3
Total de itens	2	8	0	1	0	3	2	1	4	1	2	1	1	27		
% Itens	7%	30%	0%	4%	0%	11%	7%	4%	15%	4%	7%	4%	4%	4%		
Total de conhecimentos	3	23	0	2	0	8	4	2	8	3	4	2	1	3		63

APÊNDICE B – Análise fatorial de informação completa com três fatores/dimensões para subsidiar a análise da validade discriminante do componente de Conhecimentos Específicos em contraste com o componente de Formação Geral aplicada à área de Licenciatura em Ciências Biológicas no ENADE 2017.

Item	F1	F2	F3	h2
Q01	0,06	-0,06	0,84	0,65
Q02	-0,22	0,05	-0,04	0,04
Q03	-0,21	0,03	0,34	0,26
Q04	-0,33	0,03	0,30	0,33
Q05	-0,15	0,05	0,32	0,19
Q06	-0,51	0,00	0,24	0,47
Q07	-0,21	-0,01	0,28	0,20
Q08	-0,37	0,11	0,22	0,29
Q09	-0,10	-0,18	0,28	0,16
Q10	-0,11	-0,06	0,50	0,34
Q11	-0,01	-0,16	0,21	0,08
Q12	-0,09	-0,11	0,20	0,08
Q13	-0,01	0,00	0,58	0,35
Q14	-0,17	0,10	0,20	0,12
Q15	-0,06	-0,36	0,22	0,20
Q16	-0,11	-0,40	-0,49	0,36
Q17	-0,12	-0,11	0,64	0,53
Q18	0,18	0,35	-0,27	0,30
Q19	-0,18	0,02	0,17	0,10
Q20	-0,14	-0,01	0,39	0,25
Q21	-0,26	-0,19	0,27	0,28
Q22	-0,28	0,24	0,35	0,37
Q23	-0,04	0,00	0,18	0,04
Q24L	-0,32	0,26	0,44	0,53
Q25L	-0,02	0,16	0,14	0,05
Q26L	-0,19	-0,09	0,32	0,23
Q27L	-0,01	-0,58	0,04	0,34
Q28L	-0,30	-0,31	0,02	0,22
Q29L	-0,30	-0,15	0,03	0,14
Q30L	0,07	-0,58	0,08	0,33
Q31L	-0,24	-0,30	0,05	0,19
Q32L	-0,41	-0,02	0,00	0,17
Q33L	-0,52	-0,03	-0,09	0,22
Q34L	-0,30	-0,26	0,08	0,21
Q35L	-0,54	-0,05	-0,06	0,26
“Oblimin” $\Sigma(Cf^2)$	2,19	1,69	3,52	

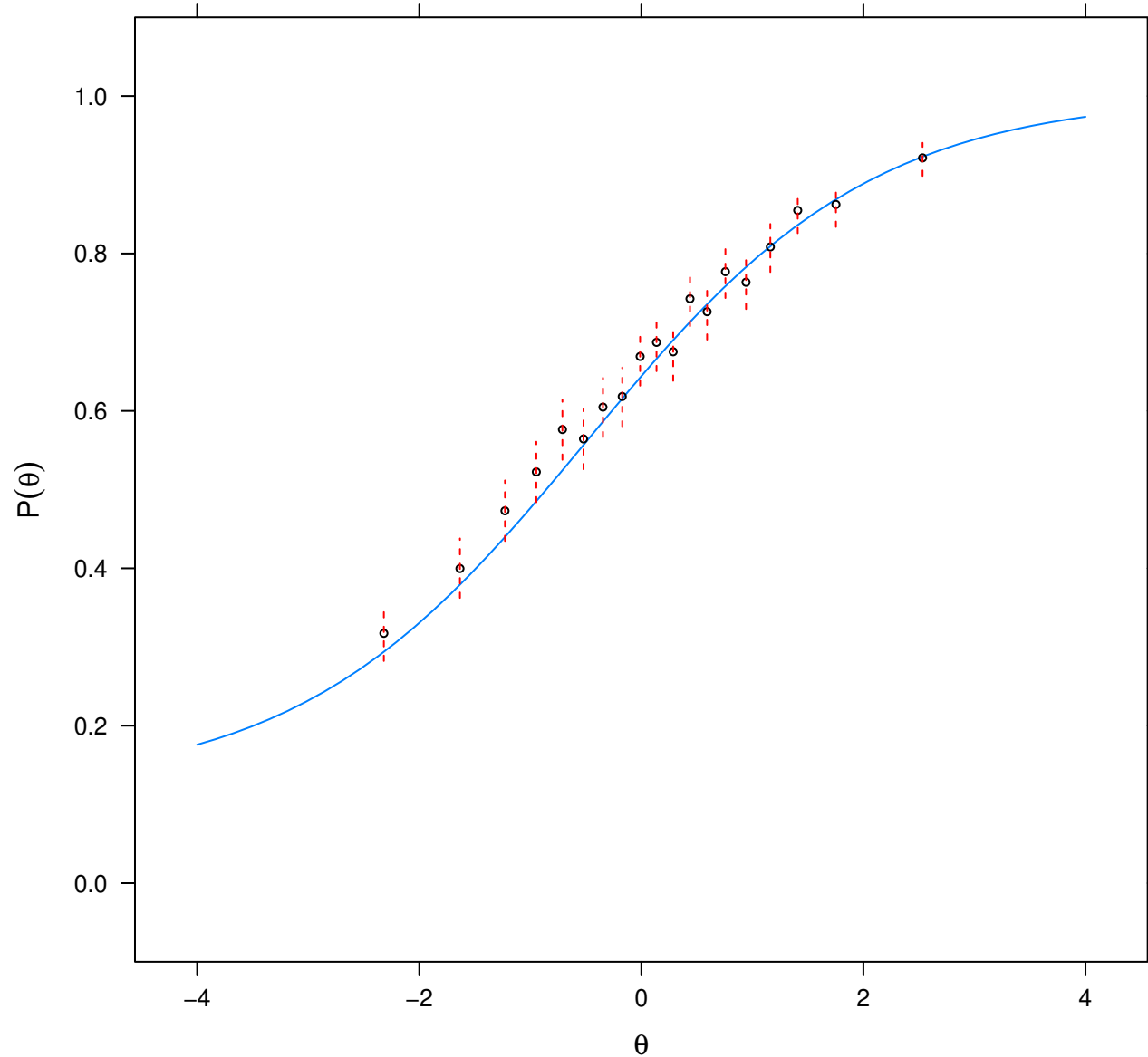
APÊNDICE C – Matriz de Resíduos do Modelo Logístico de Três Parâmetros Unidimensional gerado a partir dos 22 itens selecionados para manutenção na estimação das proficiências e construção da escala.

	Q09	Q10	Q11	Q12	Q13	Q15	Q17	Q19	Q20	Q21	Q22	Q24L	Q26L	Q27L	Q28L	Q29L	Q30L	Q31L	Q32L	Q33L	Q34L		
Q10		15,5																					
Q11		5,3	3,9																				
Q12		1,6	3,3	-0,8																			
Q13		0,0	10,4	-7,9	0,9																		
Q15		11,1	-0,9	0,1	-0,4	-5,3																	
Q17		-0,1	15,1	5,3	1,5	22,3	-1,3																
Q19		-0,6	5,7	-0,2	-3,8	0,1	-3,7	2,8															
Q20		-3,5	1,9	-1,9	-0,2	1,1	-8,2	10,8	0,5														
Q21		-0,4	1,0	-0,6	-0,5	0,0	0,0	-0,5	2,2	6,3													
Q22		-1,4	4,2	-4,1	0,0	2,7	-6,7	5,5	-0,2	5,5	2,0												
Q24L		-0,4	-0,7	-0,1	0,0	1,6	0,2	1,5	-1,2	2,3	0,7	3,3											
Q26L		0,0	0,0	-0,1	-0,9	1,3	0,0	1,3	-0,2	-2,0	0,1	-0,1	0,0										
Q27L		0,0	-22,5	3,4	1,9	-8,1	38,3	-7,2	-7,4	-3,6	0,1	-14,8	-3,6	-0,5									
Q28L		1,7	-16,1	0,1	0,0	-3,9	-0,4	-1,8	-4,8	-0,1	0,0	-1,0	-10,6	0,0	32,6								
Q29L		-0,1	-3,6	0,4	-1,6	-5,5	0,0	-1,1	-1,4	-3,0	1,0	-1,0	-3,6	-5,9	1,2	2,0							
Q30L		4,1	-6,4	0,1	3,9	-7,4	21,6	-7,0	-9,0	-6,9	1,3	-14,6	0,0	-0,6	96,6	16,0	-0,3						
Q31L		1,0	-16,8	4,0	-0,1	-10,3	13,0	-13,9	-1,8	-6,7	4,0	-7,5	-4,9	0,1	34,4	23,6	0,7	0,6					
Q32L		-1,4	0,0	-3,4	-0,7	-0,3	0,0	-0,9	0,3	-2,9	0,1	-0,6	0,0	0,8	-2,3	-1,6	0,0	-1,4	-0,4				
Q33L		-4,8	-3,0	-4,8	0,4	0,0	-2,3	-4,4	-0,6	-0,2	-2,9	1,5	-0,5	0,7	-5,8	1,8	1,7	-0,6	-0,2	22,8			
Q34L		-3,1	-0,1	0,4	-0,6	-14,4	6,0	-2,4	0,0	0,2	-0,9	-4,2	1,9	-4,5	6,0	0,0	5,0	6,3	12,6	-0,7	2,2		
Q35L		-3,7	-1,0	-1,6	0,0	-4,0	-3,2	-2,4	2,1	-0,7	0,0	-0,2	-0,3	-0,3	-3,3	2,9	6,3	-8,9	0,4	14,5	38,9	-0,3	

APÊNDICE D – Comparação entre as Curvas Empíricas e Curvas Características dos Itens Geradas a partir do Modelo Logístico de Três Parâmetros Unidimensional Calibrado a partir dos 22 itens Mantidos para Construção da Escala.

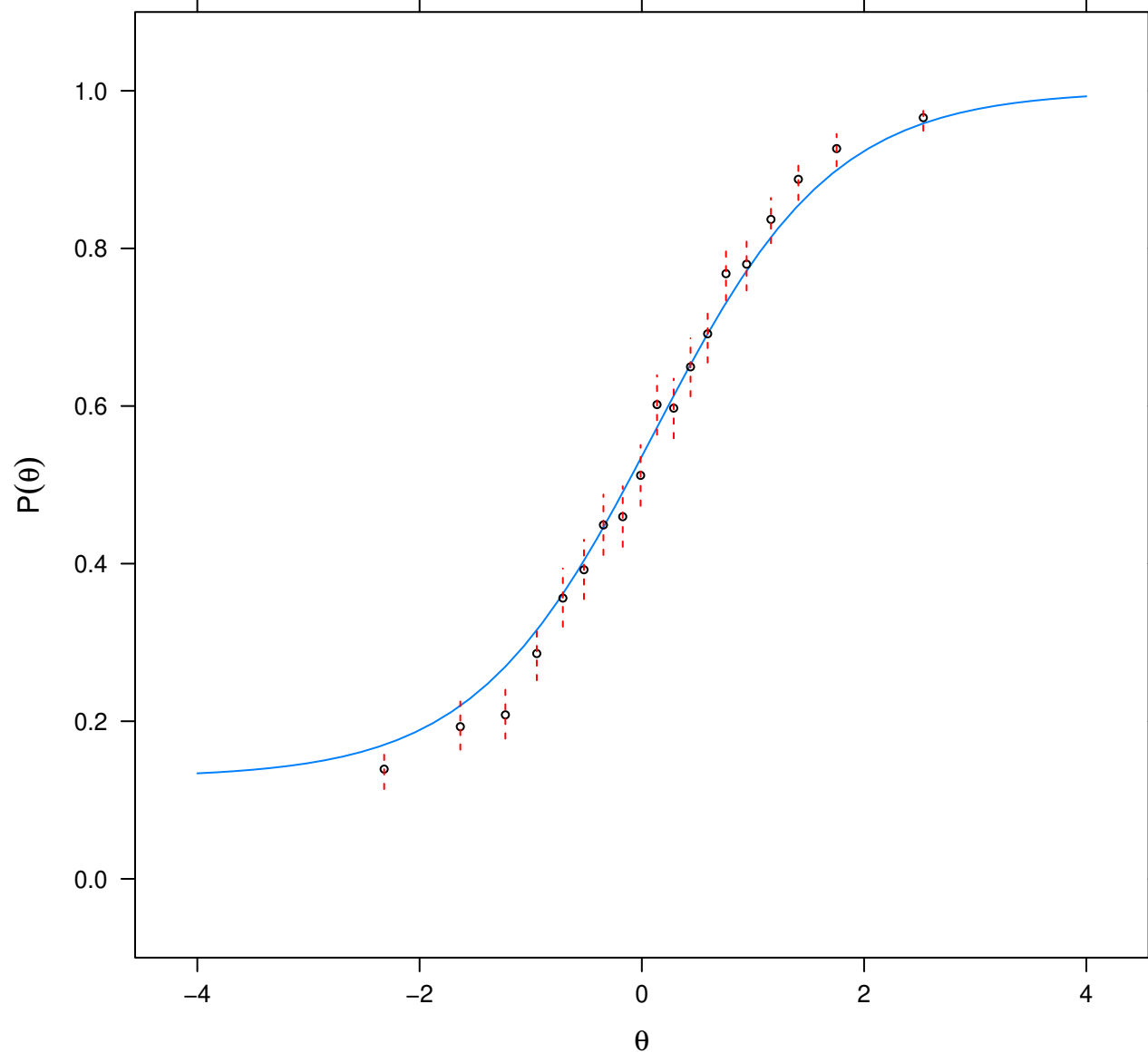
Curva Empírica do Item

Item: Q09



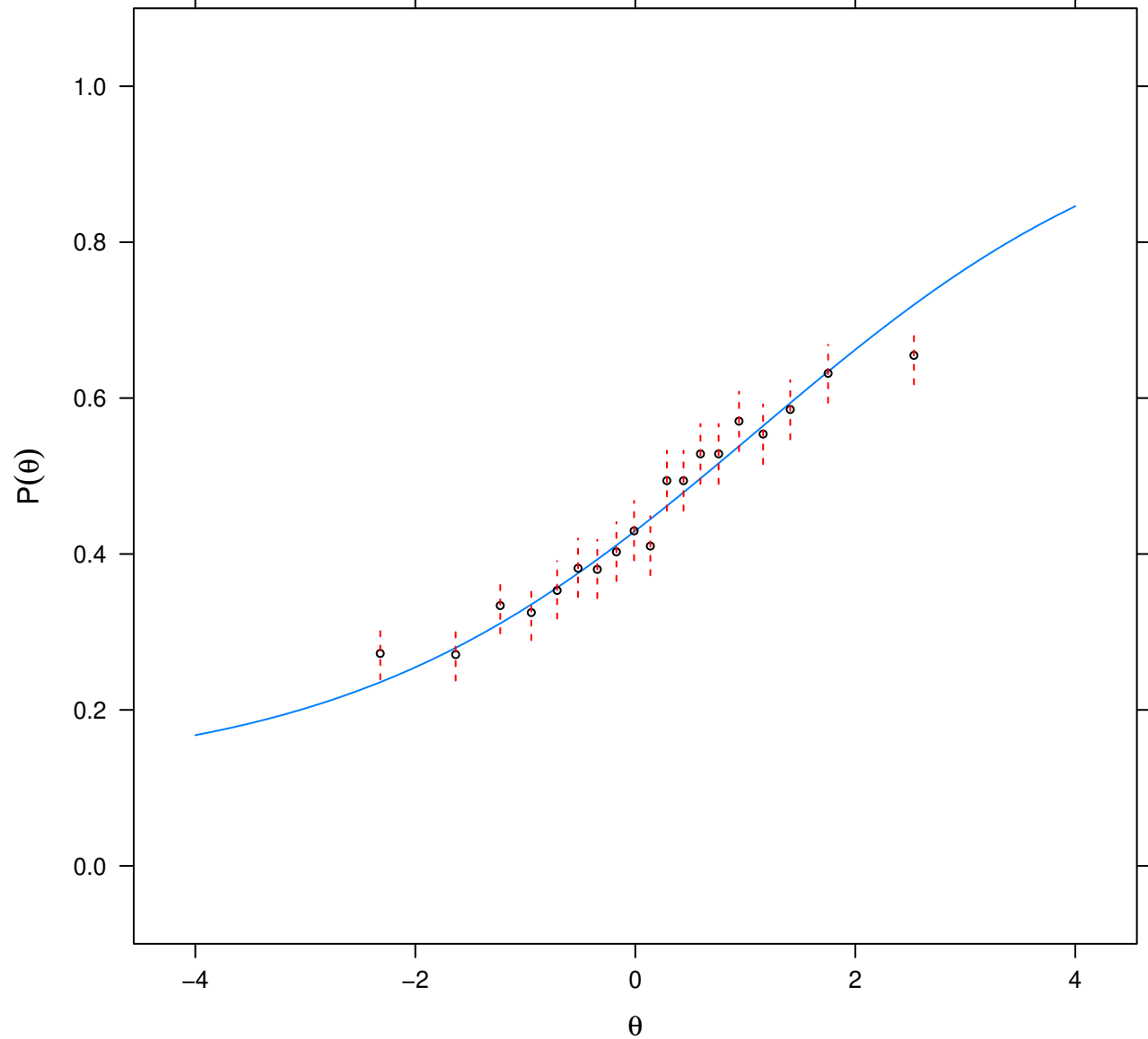
Curva Empírica do Item

Item: Q10



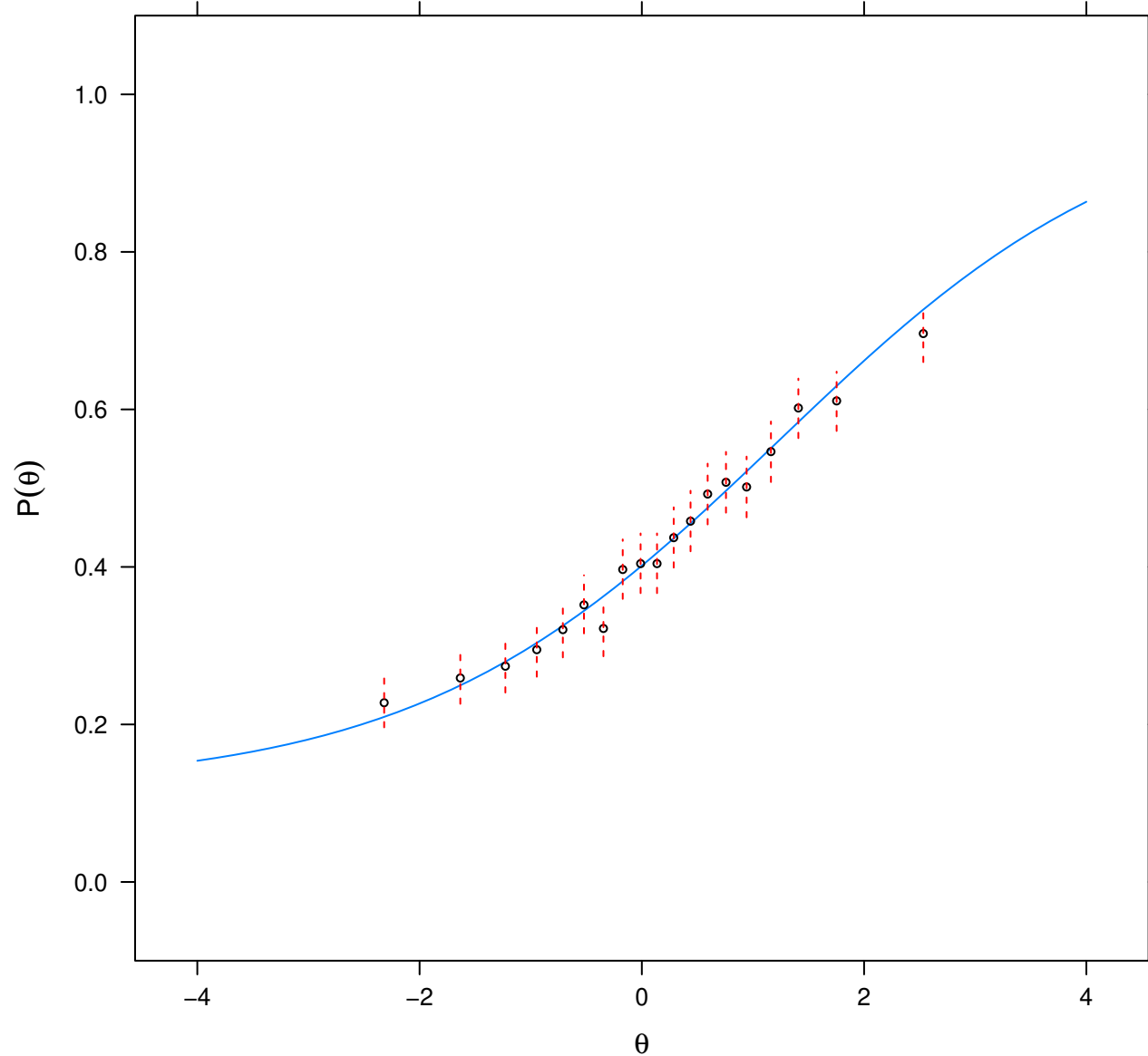
Curva Empírica do Item

Item: Q11



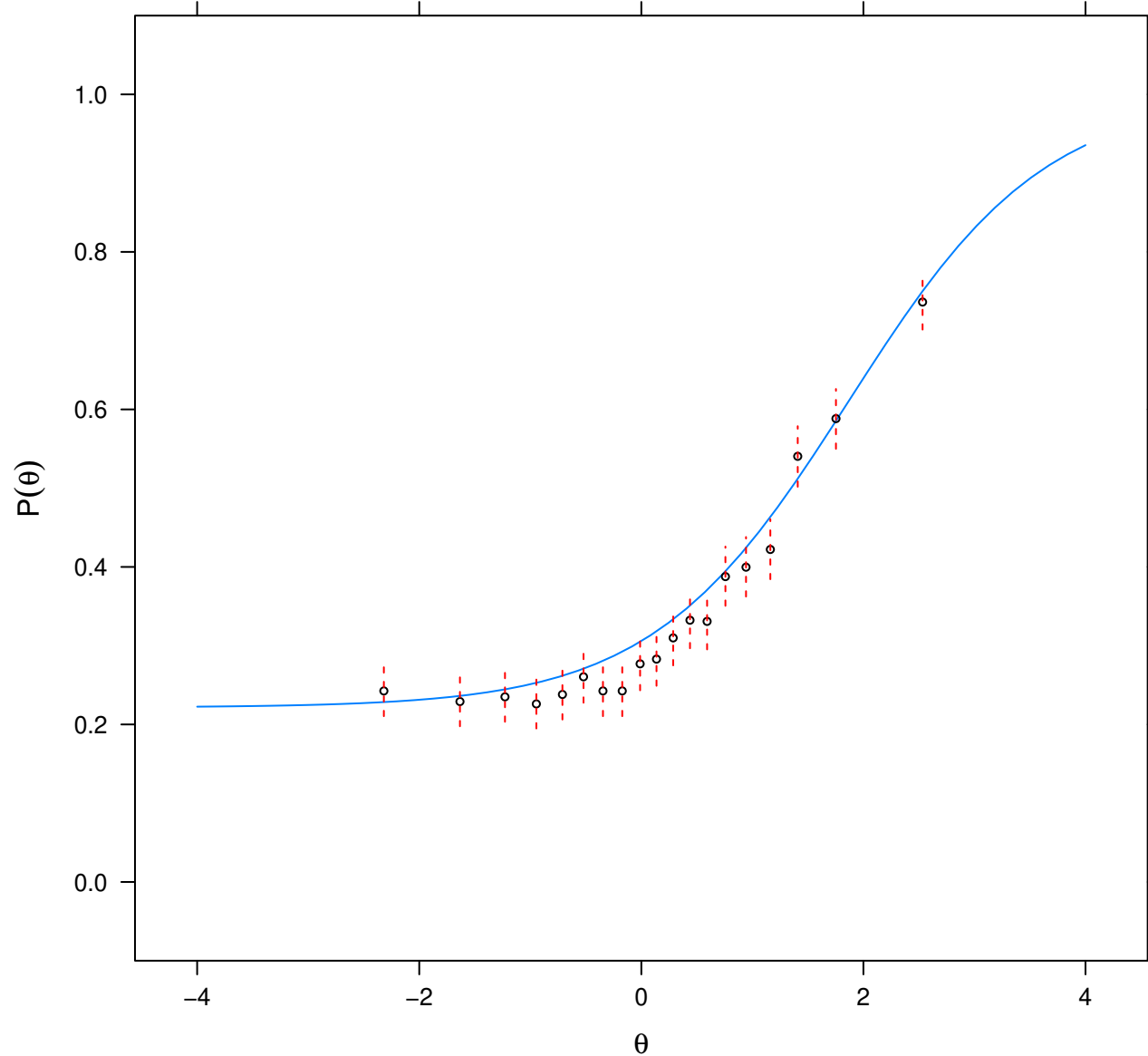
Curva Empírica do Item

Item: Q12



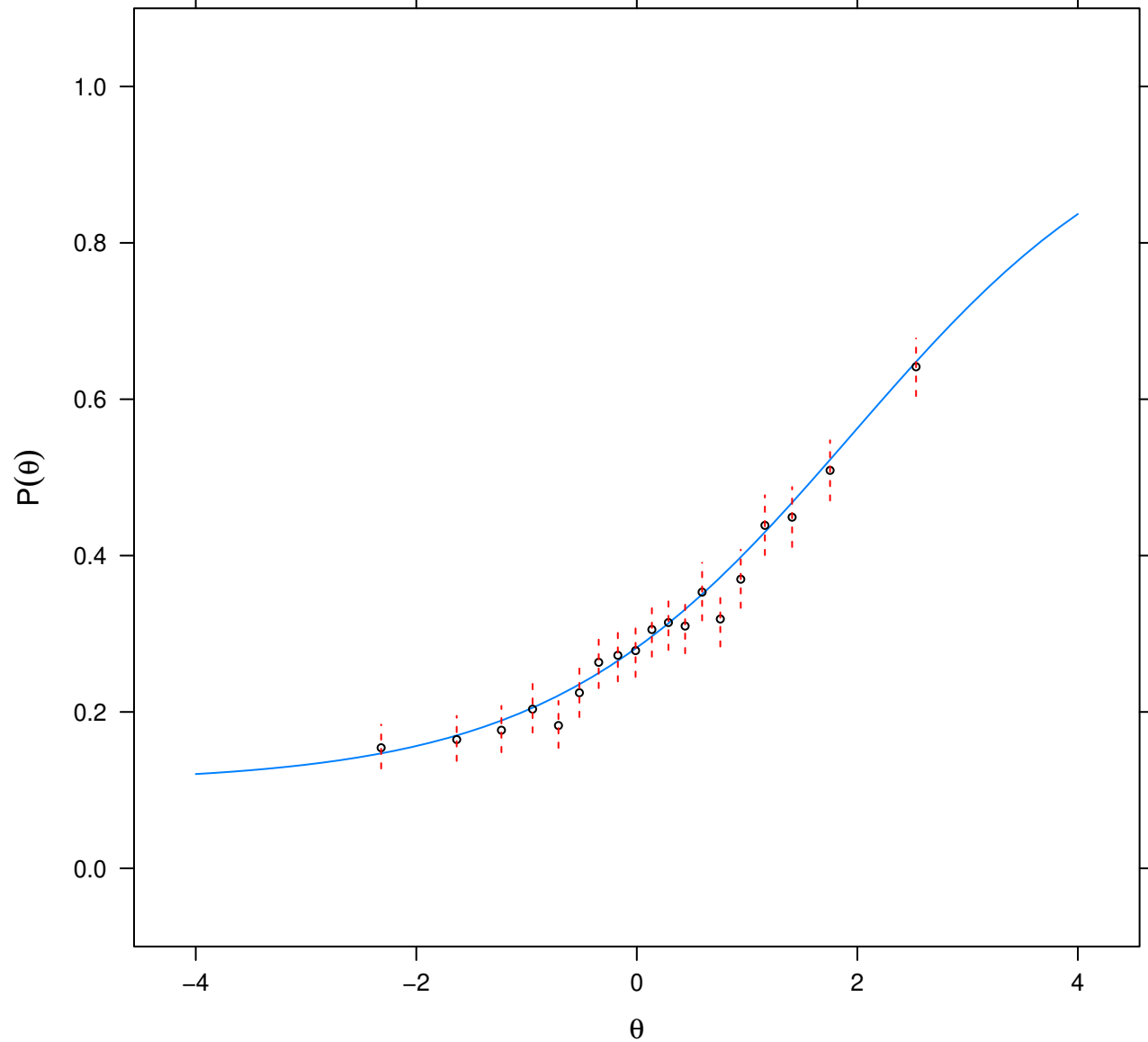
Curva Empírica do Item

Item: Q13



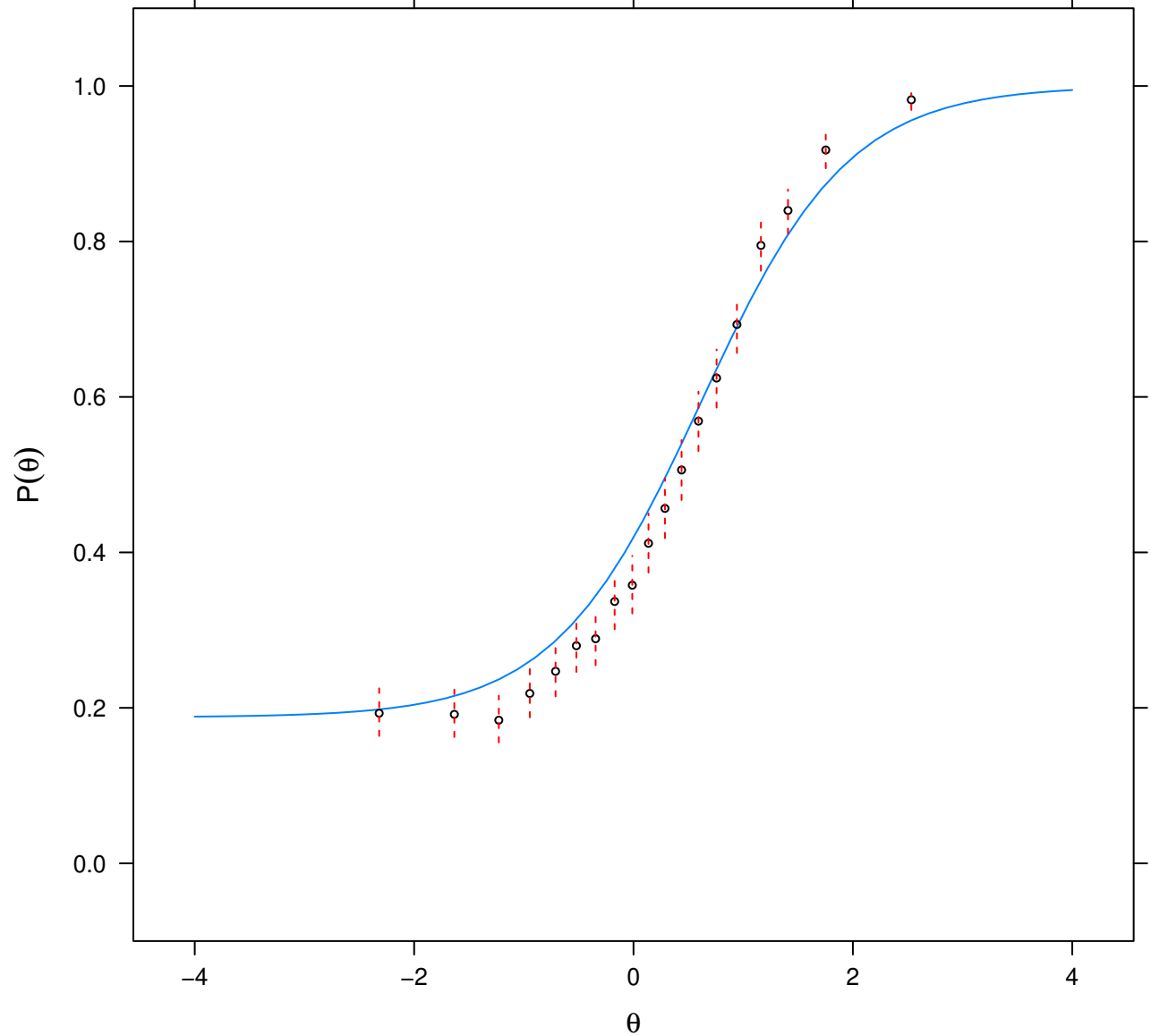
Curva Empírica do Item

Item: Q15



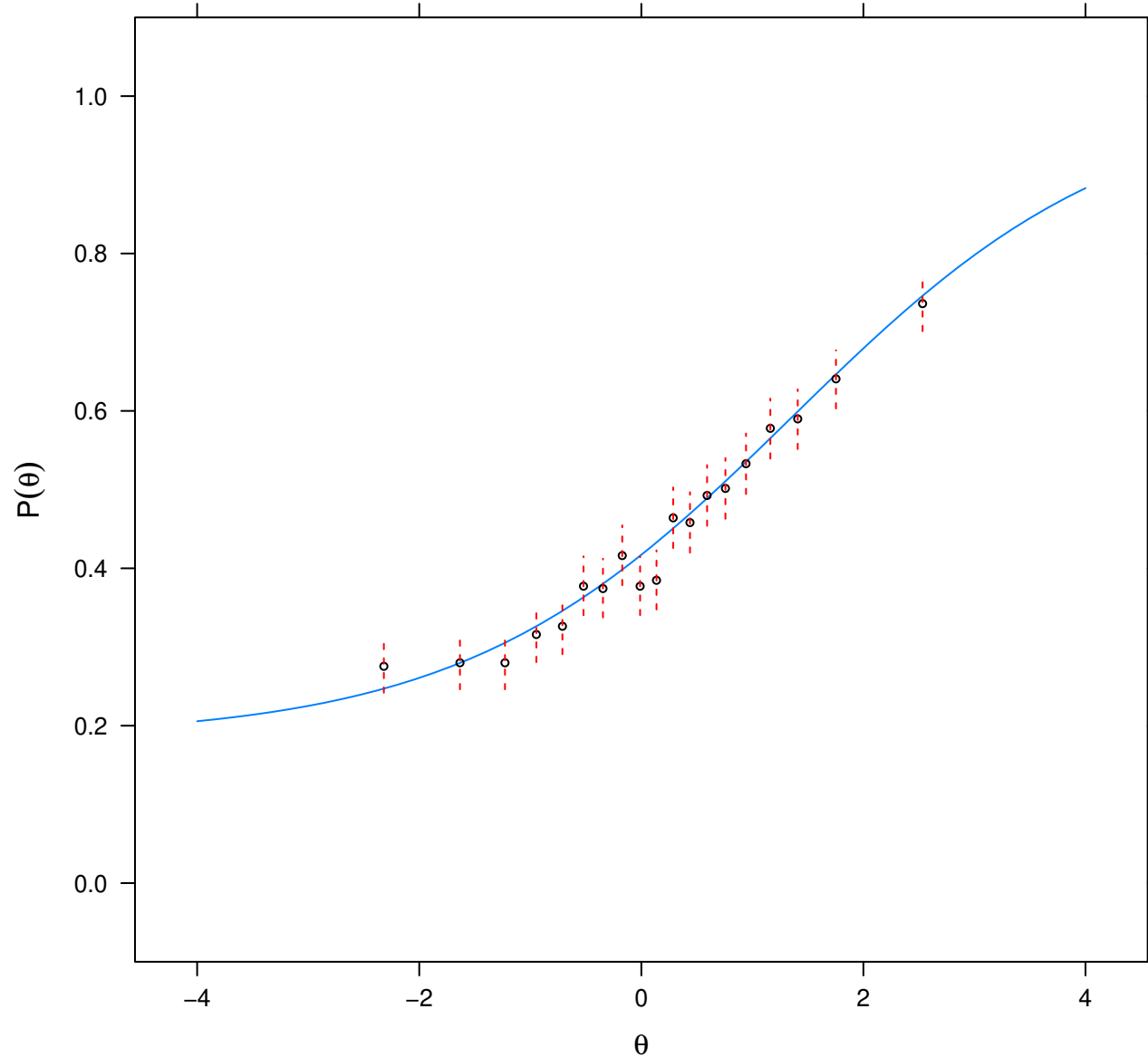
Curva Empírica do Item

Item: Q17



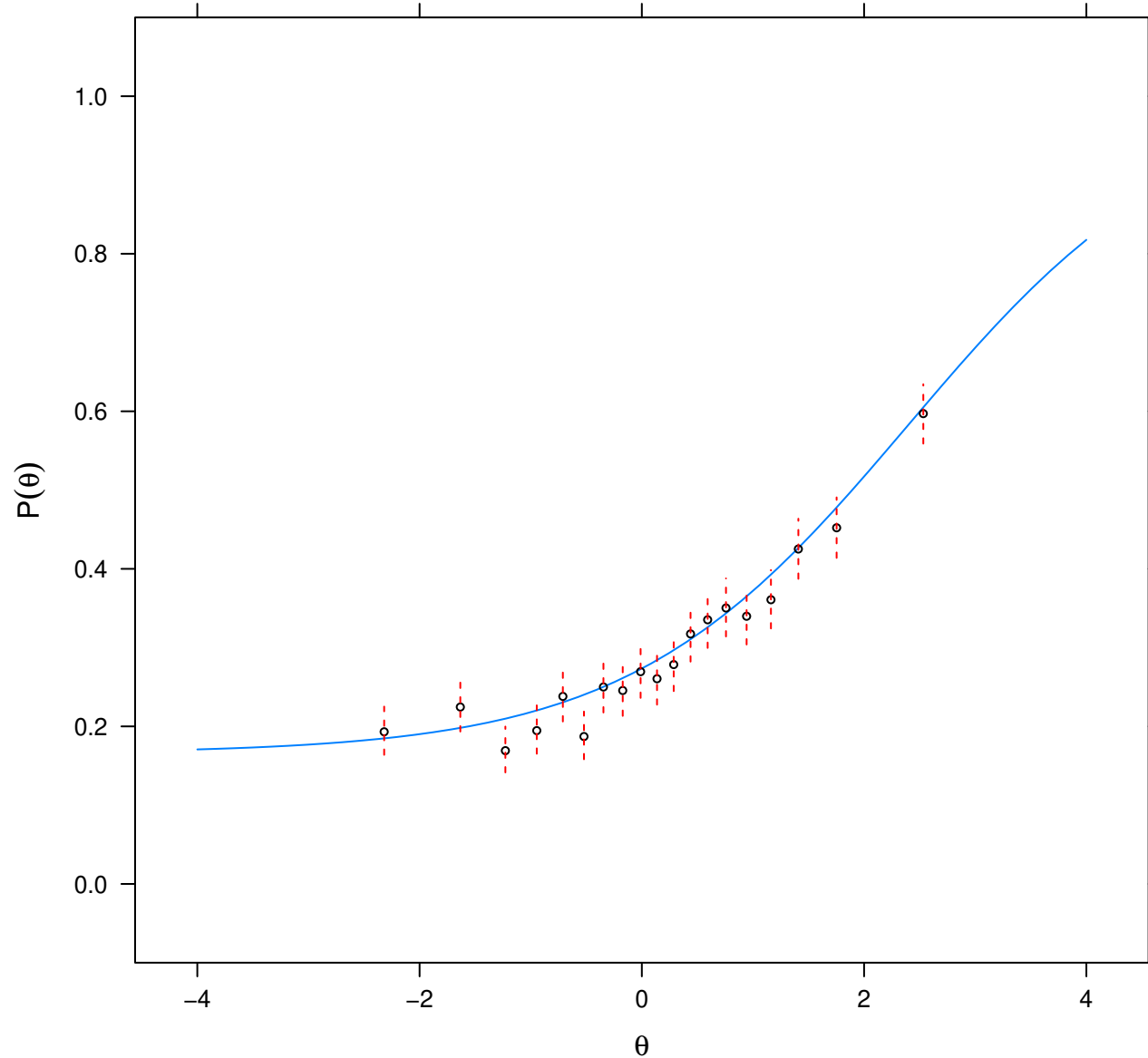
Curva Empírica do Item

Item: Q19



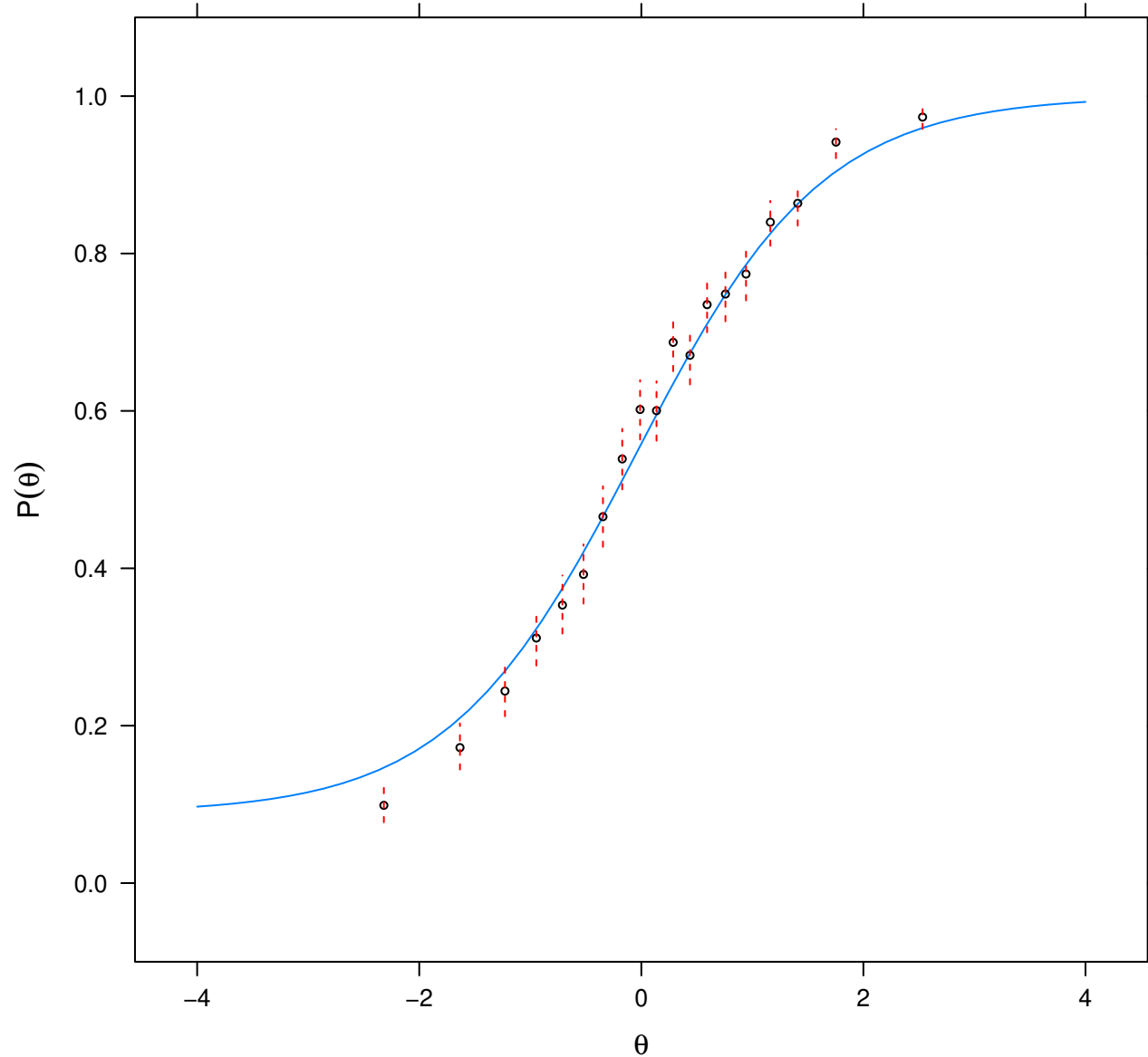
Curva Empírica do Item

Item: Q20



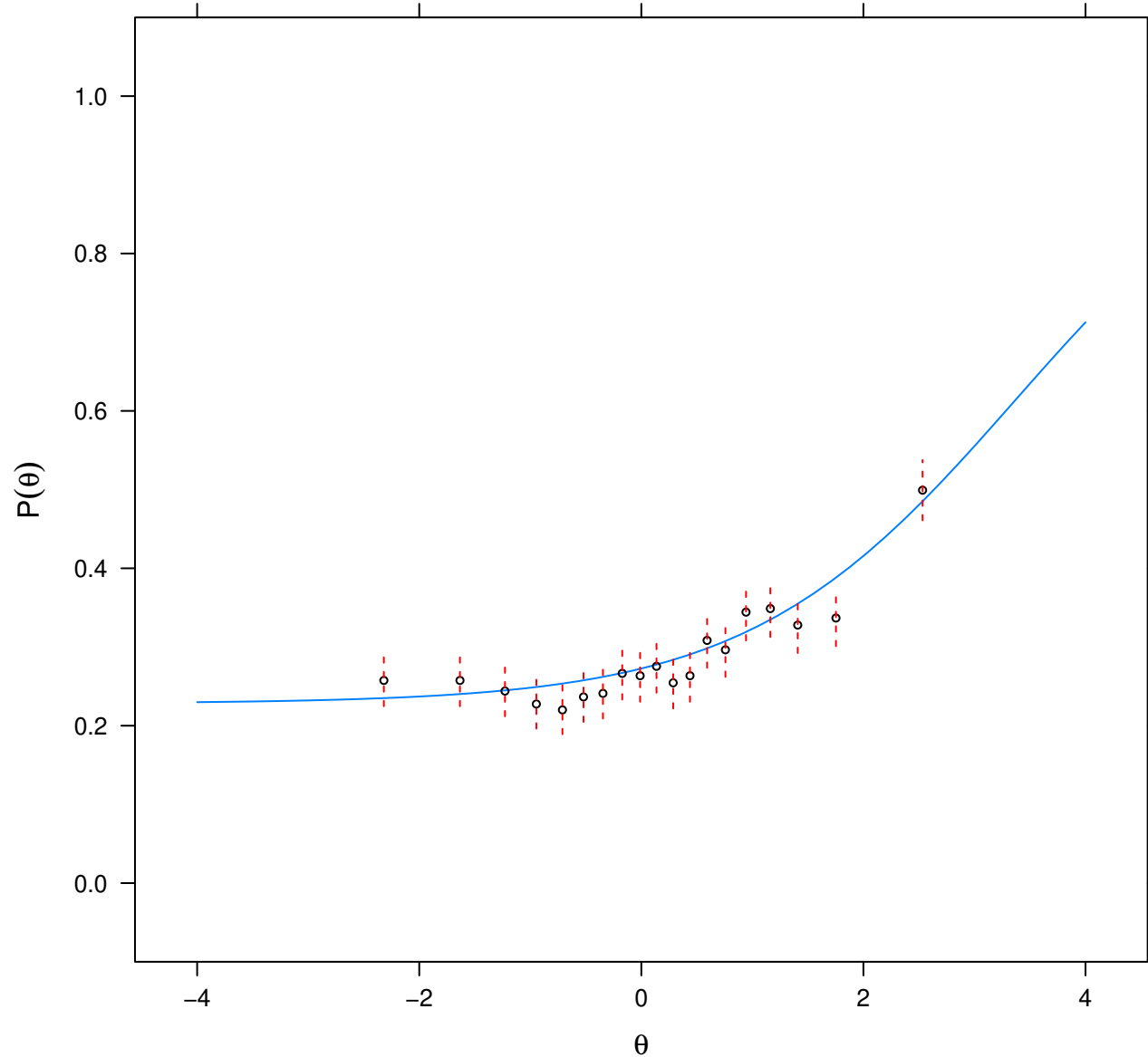
Curva Empírica do Item

Item: Q21



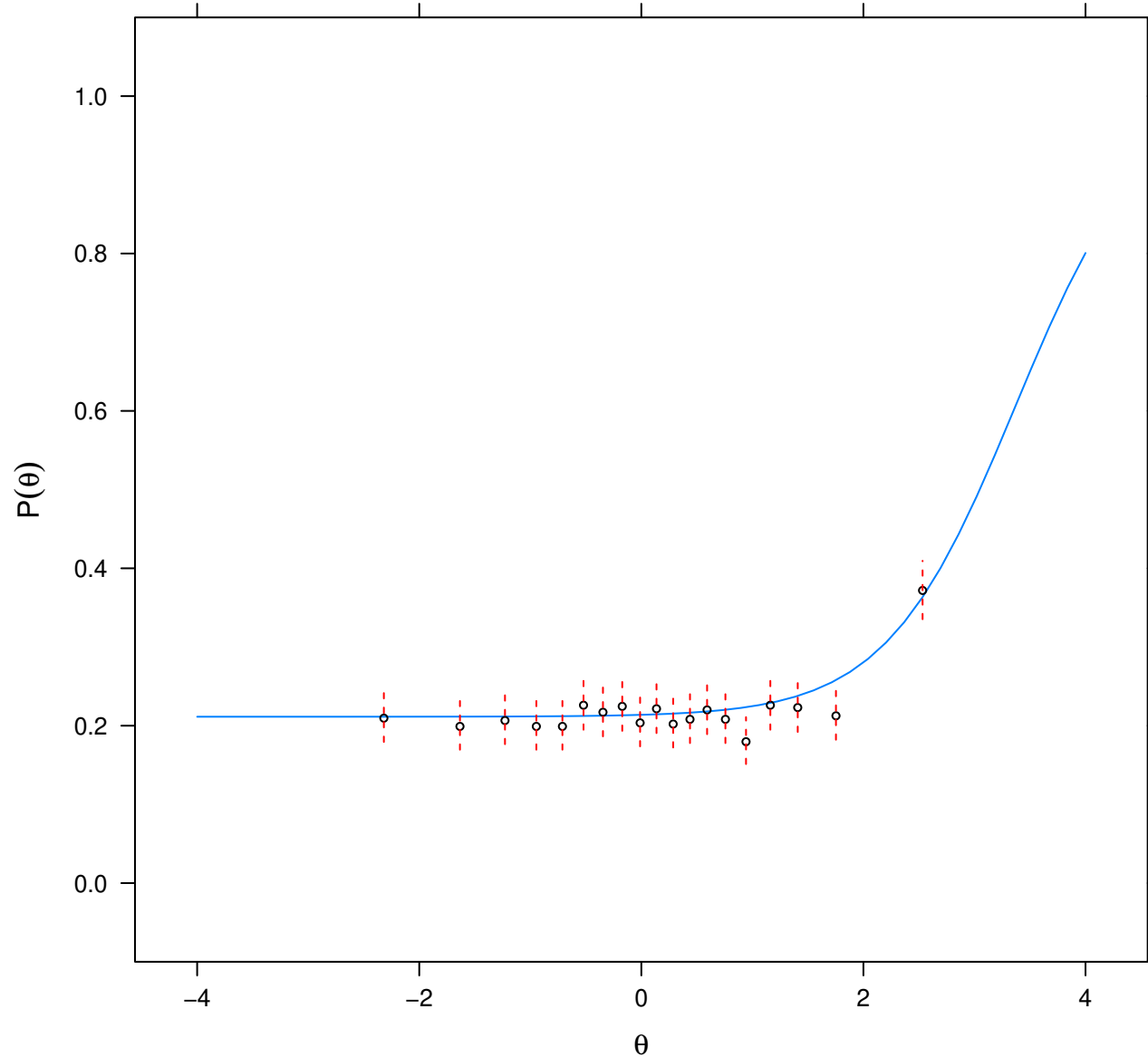
Curva Empírica do Item

Item: Q22



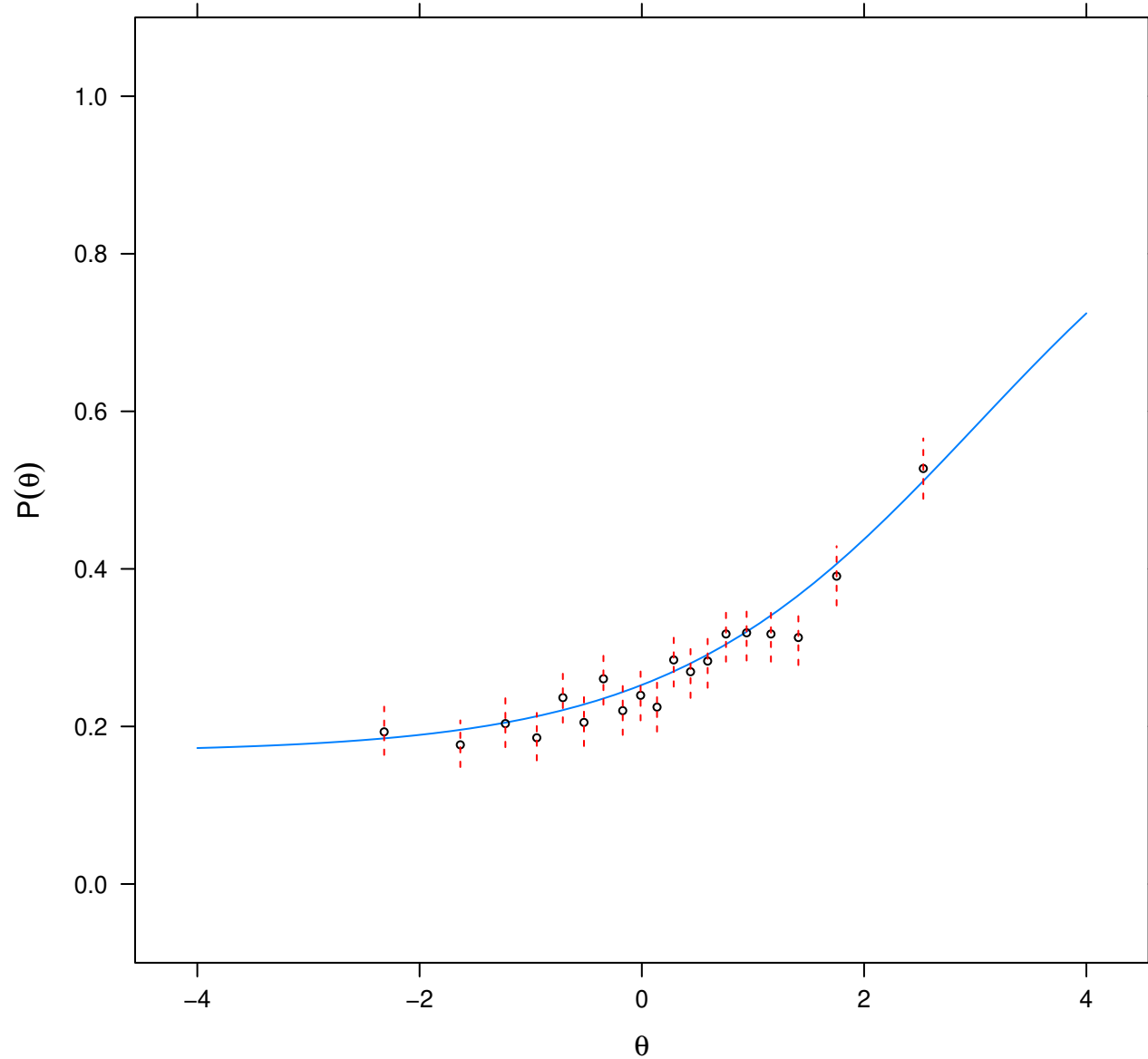
Curva Empírica do Item

Item: Q24L



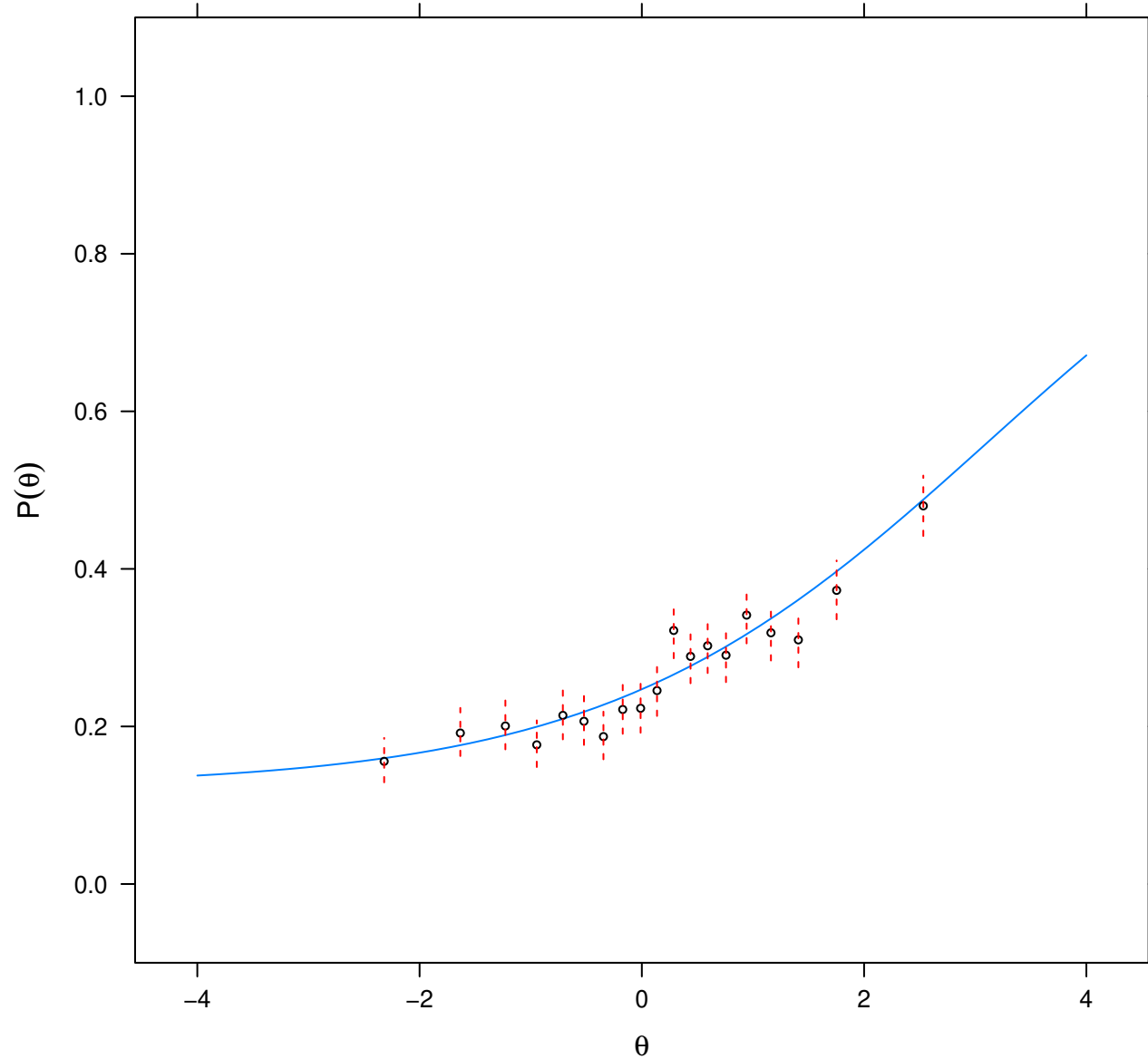
Curva Empírica do Item

Item: Q26L



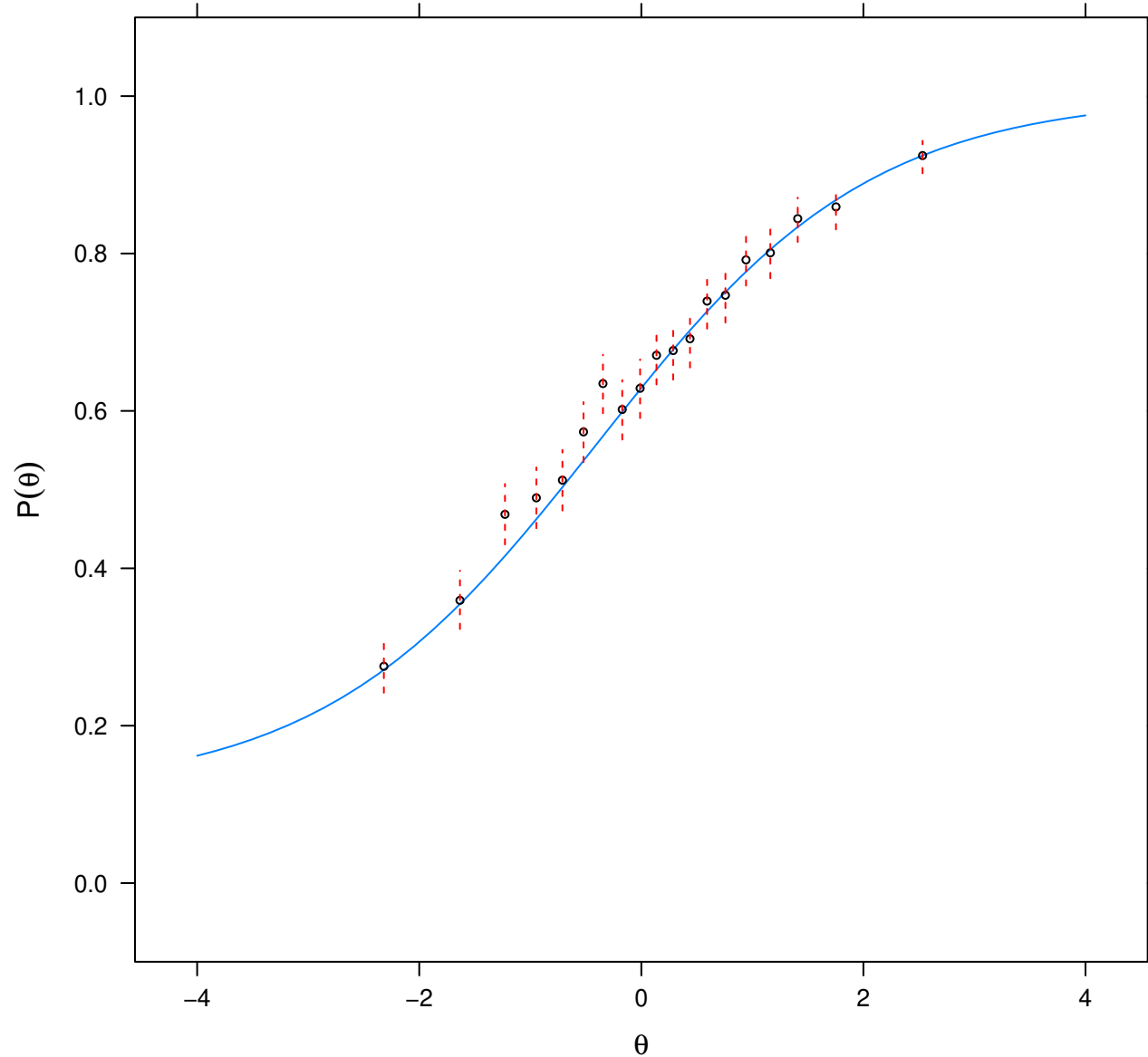
Curva Empírica do Item

Item: Q27L



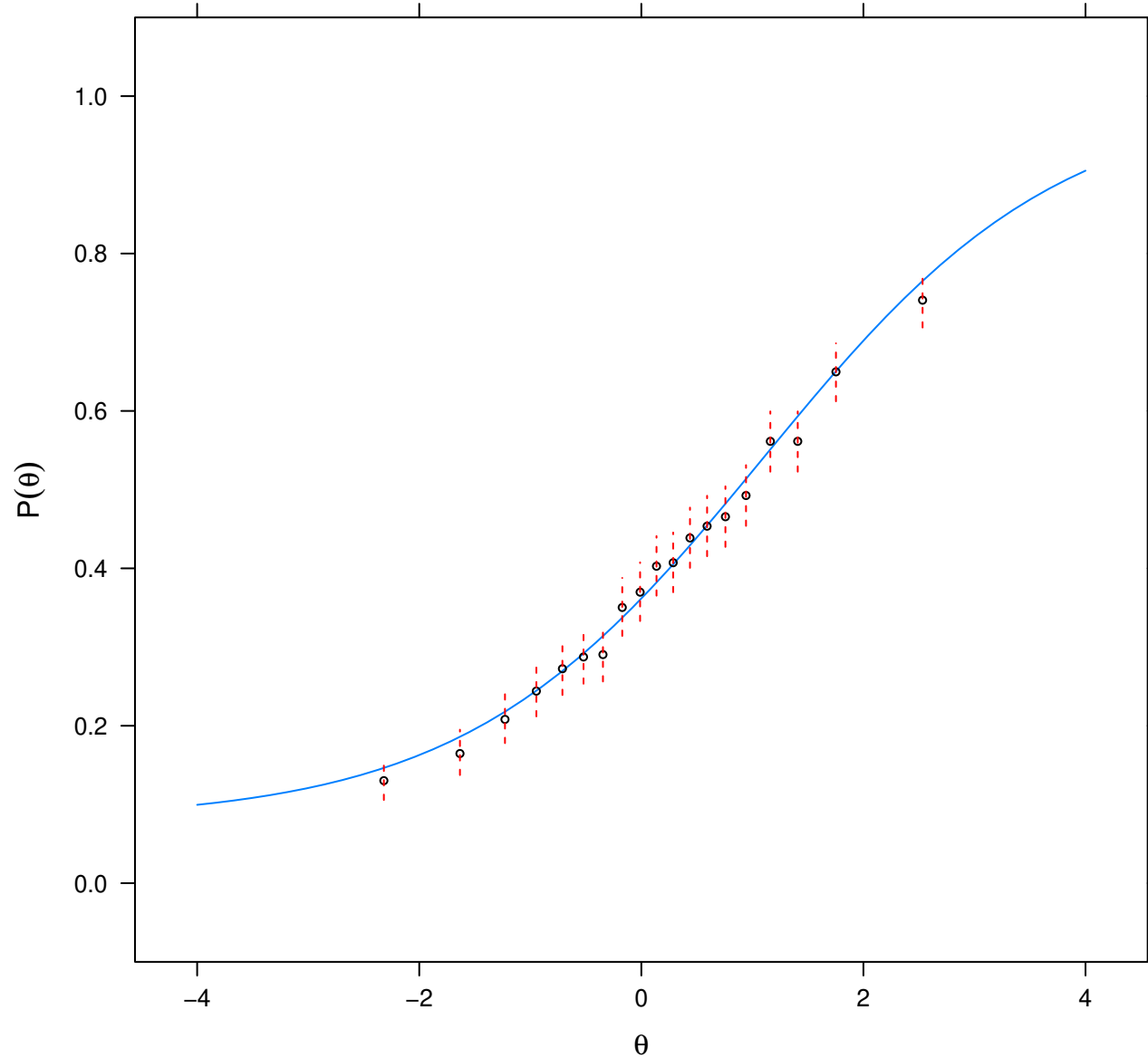
Curva Empírica do Item

Item: Q28L



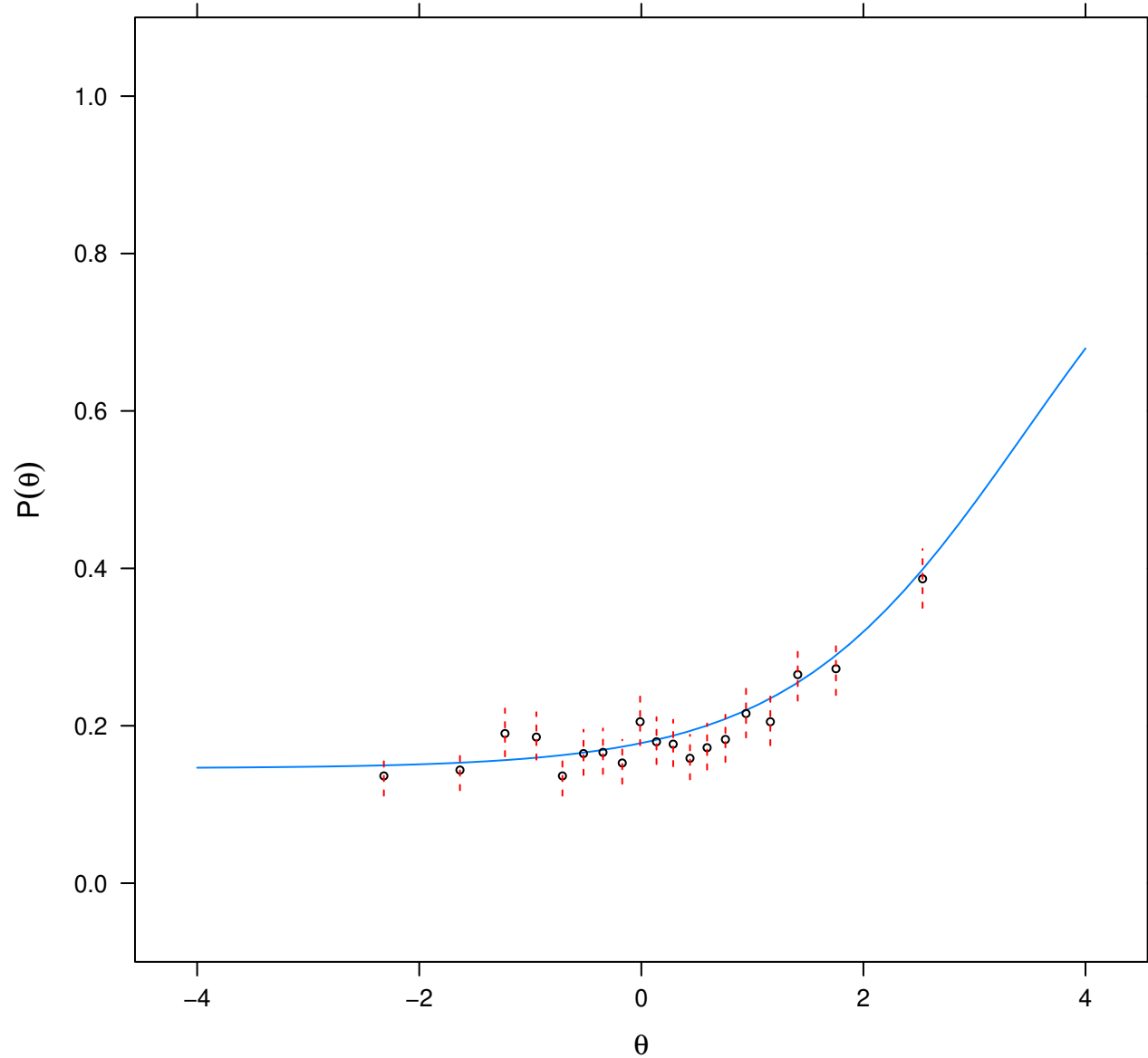
Curva Empírica do Item

Item: Q29L



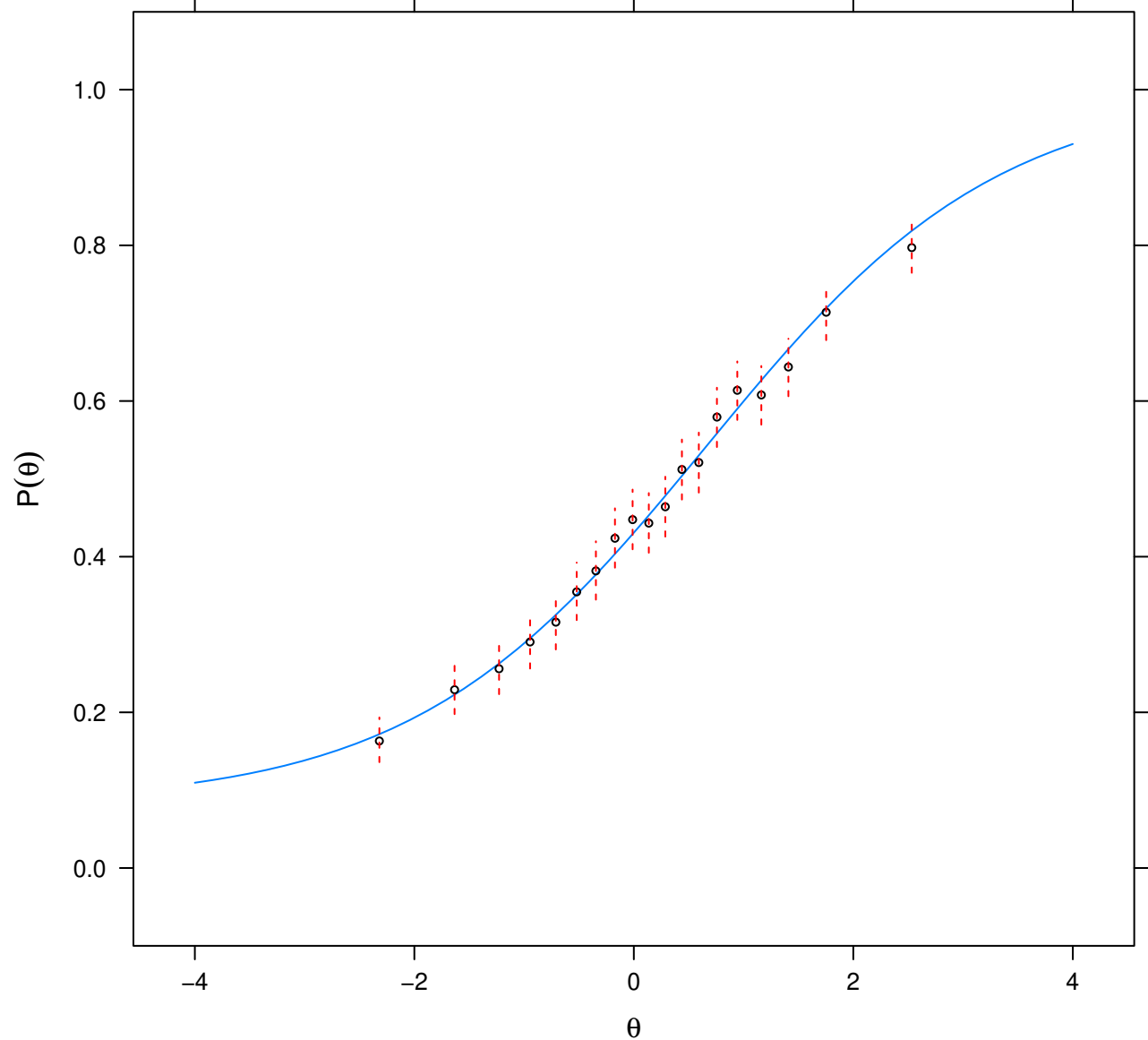
Curva Empírica do Item

Item: Q30L



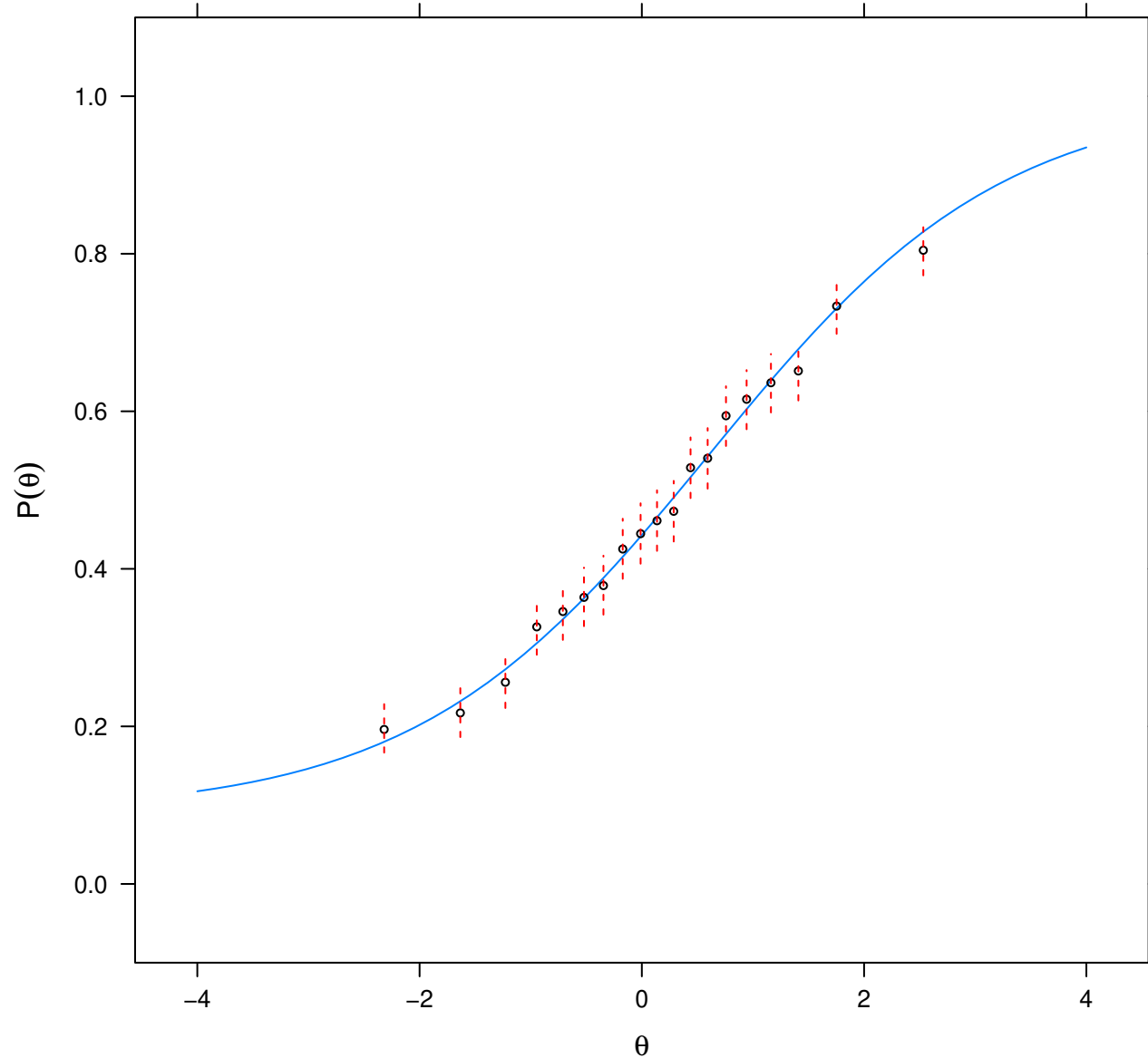
Curva Empírica do Item

Item: Q31L



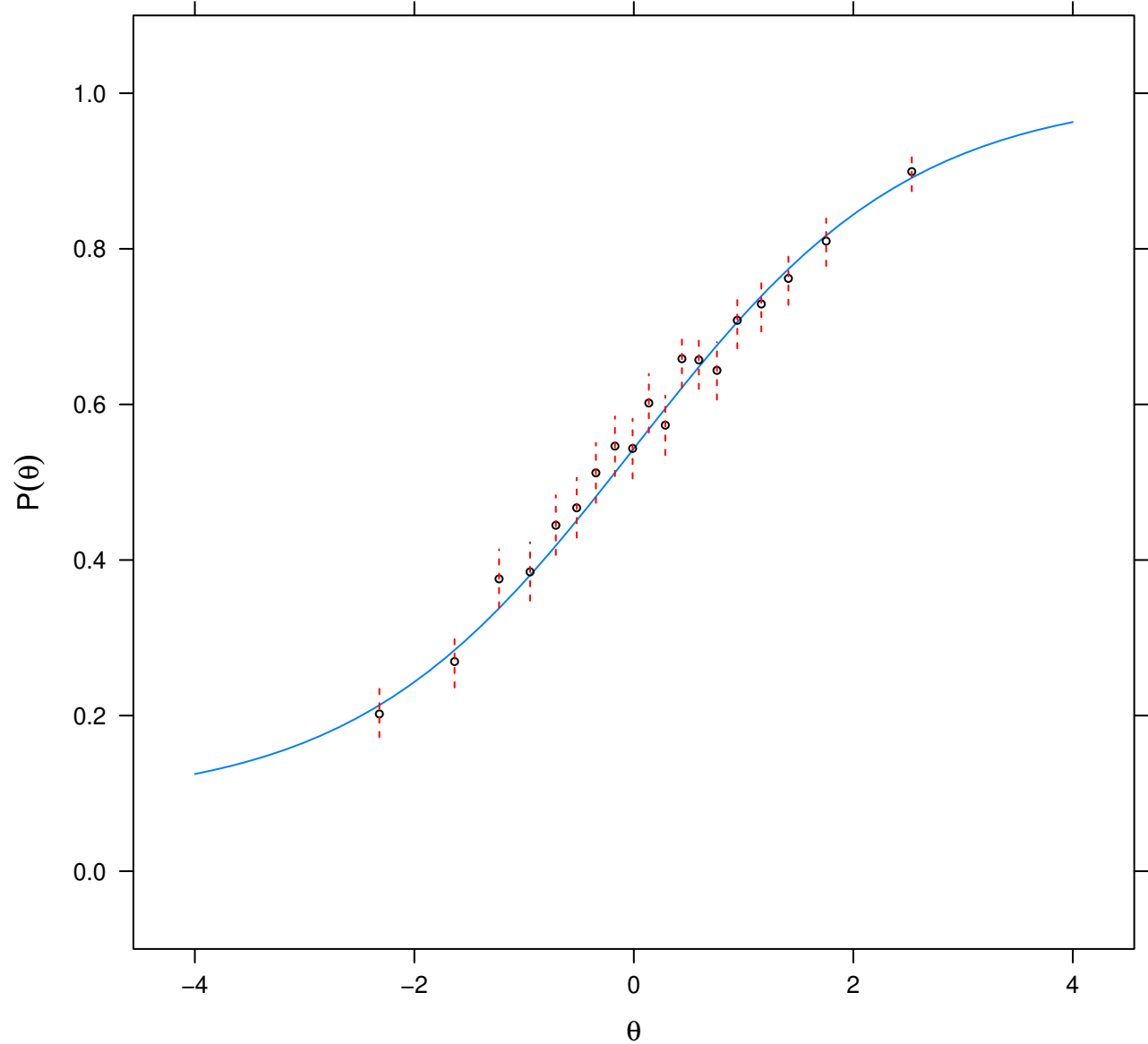
Curva Empírica do Item

Item: Q32L



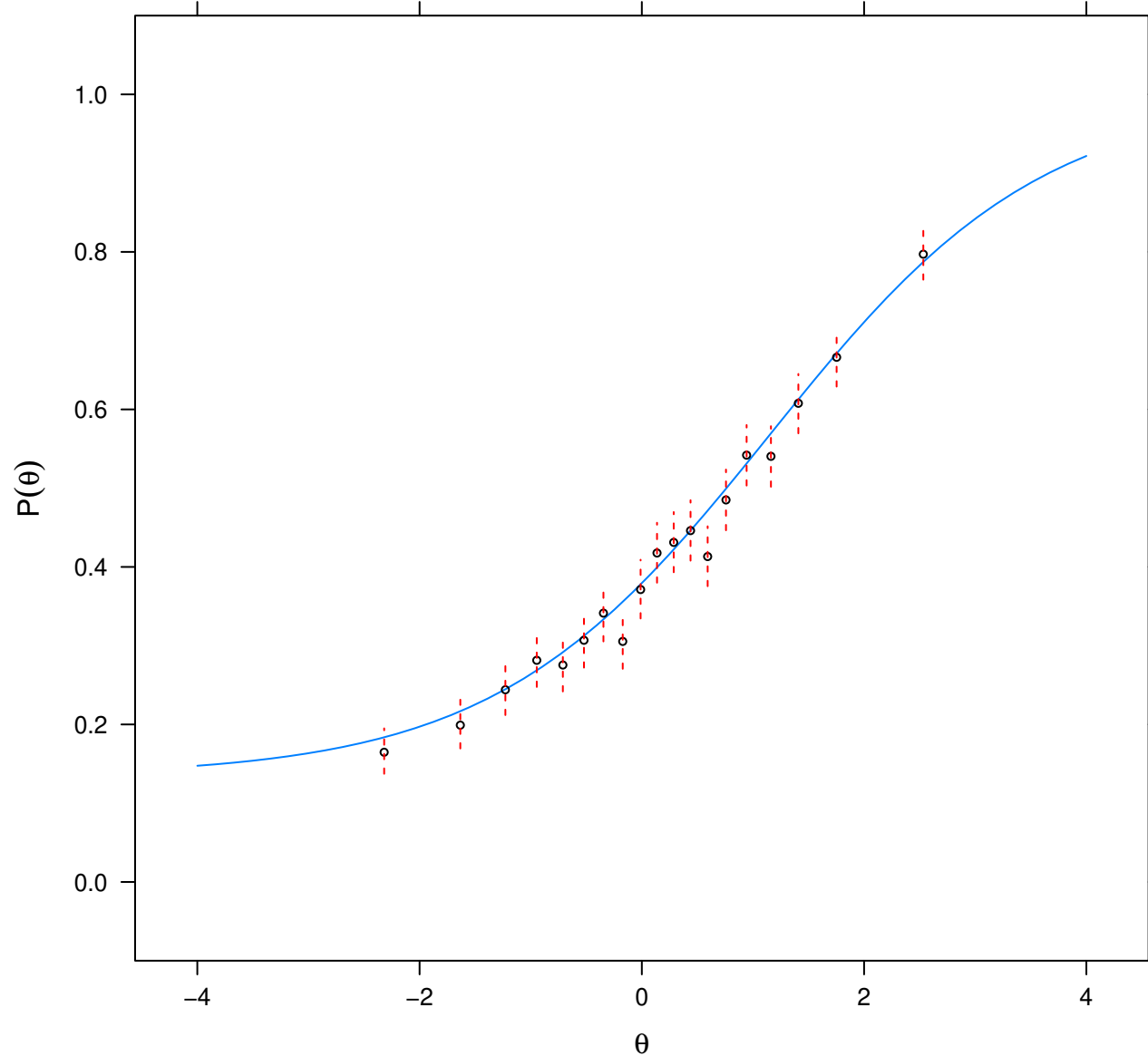
Curva Empírica do Item

Item: Q33L



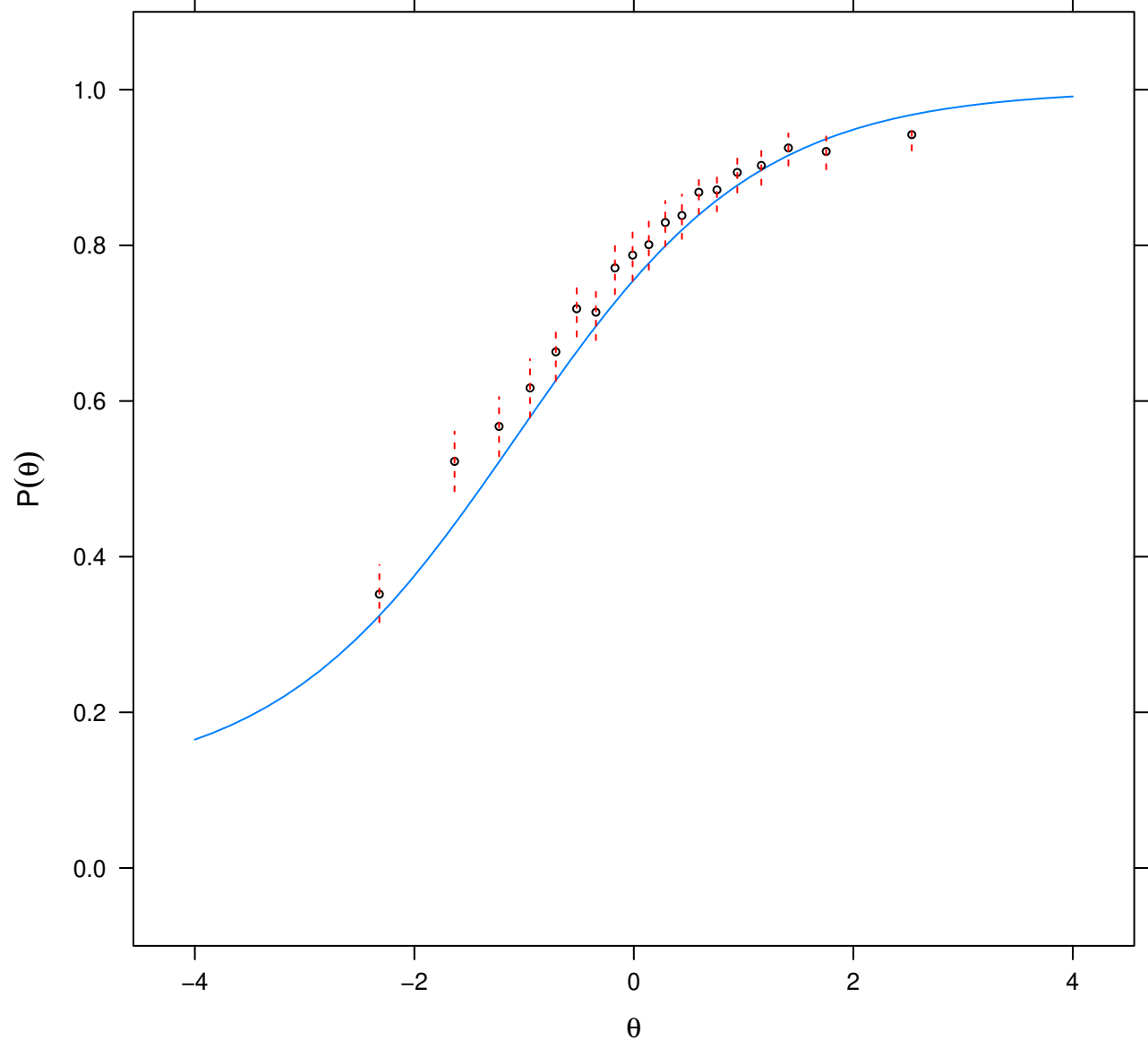
Curva Empírica do Item

Item: Q34L



Curva Empírica do Item

Item: Q35L



**ANEXO A – Prova do ENADE 2017 Aplicada aos Concluintes dos Cursos de
Licenciatura em Ciências Biológicas**



enade2017

CIÊNCIAS BIOLÓGICAS LICENCIATURA

06

Novembro/17

LEIA COM ATENÇÃO AS INSTRUÇÕES ABAIXO.

06

1. Verifique se, além deste Caderno, você recebeu o **CARTÃO-RESPOSTA**, destinado à transcrição das respostas das questões de múltipla escolha, das questões discursivas (D) e das questões de percepção da prova.
2. Confira se este Caderno contém as questões discursivas e as objetivas de múltipla escolha, de formação geral e de componente específico da área, e as relativas à sua percepção da prova. As questões estão assim distribuídas:

Partes	Número das questões	Peso das questões no componente	Peso dos componentes no cálculo da nota
Formação Geral: Discursivas	D1 e D2	40%	25%
Formação Geral: Objetivas	1 a 8	60%	
Componente Específico: Discursivas	D3 a D5	15%	75%
Componente Específico: Objetivas	9 a 35	85%	
Questionário de Percepção da Prova	1 a 9	-	-

3. Verifique se a prova está completa e se o seu nome está correto no **CARTÃO-RESPOSTA**. Caso contrário, avise imediatamente ao Chefe de Sala.
4. Assine o **CARTÃO-RESPOSTA** no local apropriado, **com caneta esferográfica de tinta preta, fabricada em material transparente**.
5. As respostas da prova objetiva, da prova discursiva e do questionário de percepção da prova deverão ser transcritas, com caneta esferográfica de tinta preta, fabricada em material transparente, para o **CARTÃO-RESPOSTA** que deverá ser entregue ao Chefe de Sala ao término da prova.
6. Responda cada questão discursiva em, no máximo, 15 linhas. Qualquer texto que ultrapasse o espaço destinado à resposta será desconsiderado.
7. Você terá quatro horas para responder as questões de múltipla escolha, as questões discursivas e o questionário de percepção da prova.
8. Ao terminar a prova, levante a mão e aguarde o Chefe de Sala em sua carteira para proceder a sua identificação, recolher o seu material de prova e coletar a sua assinatura na Lista de Presença.
9. Atenção! Você deverá permanecer na sala de aplicação, no mínimo, por uma hora a partir do início da prova e só poderá levar este Caderno de Prova quando faltarem 30 minutos para o término do Exame.



FORMAÇÃO GERAL

QUESTÃO DISCURSIVA 01

TEXTO 1

Em 2001, a incidência da sífilis congênita — transmitida da mulher para o feto durante a gravidez — era de um caso a cada mil bebês nascidos vivos. Havia uma meta da Organização Pan-Americana de Saúde e da Unicef de essa ocorrência diminuir no Brasil, chegando, em 2015, a 5 casos de sífilis congênita por 10 mil nascidos vivos. O país não atingiu esse objetivo, tendo se distanciado ainda mais dele, embora o tratamento para sífilis seja relativamente simples, à base de antibióticos. Trata-se de uma doença para a qual a medicina já encontrou a solução, mas a sociedade ainda não.

Disponível em: <<http://www1.folha.uol.com.br>>. Acesso em: 23 jul. 2017 (adaptado).

TEXTO 2

O Ministério da Saúde anunciou que há uma epidemia de sífilis no Brasil. Nos últimos cinco anos, foram 230 mil novos casos, um aumento de 32% somente entre 2014 e 2015. Por que isso aconteceu?

Primeiro, ampliou-se o diagnóstico com o teste rápido para sífilis realizado na unidade básica de saúde e cujo resultado sai em 30 minutos. Aí vem o segundo ponto, um dos mais negativos, que foi o desabastecimento, no país, da matéria-prima para a penicilina. O Ministério da Saúde importou essa penicilina, mas, por um bom tempo, não esteve disponível, e isso fez com que mais pessoas se infectassem. O terceiro ponto é a prevenção. Houve, nos últimos dez anos, uma redução do uso do preservativo, o que aumentou, e muito, a transmissão.

A incidência de casos de sífilis, que, em 2010, era maior entre homens, hoje recai sobre as mulheres. Por que a vulnerabilidade neste grupo está aumentando?

As mulheres ainda são as mais vulneráveis a doenças sexualmente transmissíveis (DST), de uma forma geral. Elas têm dificuldade de negociar o preservativo com o parceiro, por exemplo. Mas o acesso da mulher ao diagnóstico também é maior, por isso, é mais fácil contabilizar essa população. Quando um homem faz exame para a sífilis? Somente quando tem sintoma aparente ou outra doença. E a sífilis pode ser uma doença silenciosa. A mulher, por outro lado, vai fazer o pré-natal e, automaticamente, faz o teste para a sífilis. No Brasil, estima-se que apenas 12% dos parceiros sexuais recebam tratamento para sífilis.

Entrevista com Ana Gabriela Travassos, presidente da regional baiana da Sociedade Brasileira de Doenças Sexualmente Transmissíveis. Disponível em: <<http://www.agenciapatriciagalvao.org.br>>. Acesso em: 25 jul. 2017 (adaptado).

TEXTO 3

Vários estudos constatam que os homens, em geral, padecem mais de condições severas e crônicas de saúde que as mulheres e morrem mais que elas em razão de doenças que levam a óbito. Entretanto, apesar de as taxas de morbimortalidade masculinas assumirem um peso significativo, observa-se que a presença de homens nos serviços de atenção primária à saúde é muito menor que a de mulheres.

GOMES, R.; NASCIMENTO, E.; ARAUJO, F. Por que os homens buscam menos os serviços de saúde do que as mulheres? As explicações de homens com baixa escolaridade e homens com ensino superior. **Cad. Saúde Pública** [online], v. 23, n. 3, 2007 (adaptado).





A partir das informações apresentadas, redija um texto acerca do tema:

Epidemia de sífilis congênita no Brasil e relações de gênero

Em seu texto, aborde os seguintes aspectos:

- a vulnerabilidade das mulheres às DSTs e o papel social do homem em relação à prevenção dessas doenças;
- duas ações especificamente voltadas para o público masculino, a serem adotadas no âmbito das políticas públicas de saúde ou de educação, para reduzir o problema.

(valor: 10,0 pontos)

RASCUNHO	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

Área livre



QUESTÃO DISCURSIVA 02

A pessoa *trans* precisa que alguém ateste, confirme e comprove que ela pode ser reconhecida pelo nome que ela escolheu. Não aceitam que ela se autodeclare mulher ou homem. Exigem que um profissional de saúde diga quem ela é. Sua declaração é o que menos conta na hora de solicitar, judicialmente, a mudança dos documentos.

Disponível em: <<http://www.ebc.com.br>>. Acesso em: 31 ago. 2017 (adaptado).

No chão, a travesti morre
Ninguém jamais saberá seu nome
Nos jornais, fala-se de outra morte
De tal homem que ninguém conheceu

Disponível em: <<http://www.aminoapps.com>>. Acesso em: 31 ago. 2017 (adaptado).

Usava meu nome oficial, feminino, no currículo porque diziam que eu estava cometendo um crime, que era falsidade ideológica se eu usasse outro nome. Depois fui pesquisar e descobri que não é assim. Infelizmente, ainda existe muita desinformação sobre os direitos das pessoas *trans*.

Disponível em: <<https://www.brasil.elpais.com>>. Acesso em: 31 ago. 2017 (adaptado).

Uma vez o segurança da balada achou que eu tinha, por engano, mostrado o RG do meu namorado. Isso quando insistem em não colocar meu nome social na minha ficha de consumação.

Disponível em: <<https://www.brasil.elpais.com>>. Acesso em: 31 ago. 2017 (adaptado).

Com base nessas falas, discorra sobre a importância do nome para as pessoas transgêneras e, nesse contexto, proponha uma medida, no âmbito das políticas públicas, que tenha como objetivo facilitar o acesso dessas pessoas à cidadania. (valor: 10,0 pontos)

RASCUNHO	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	



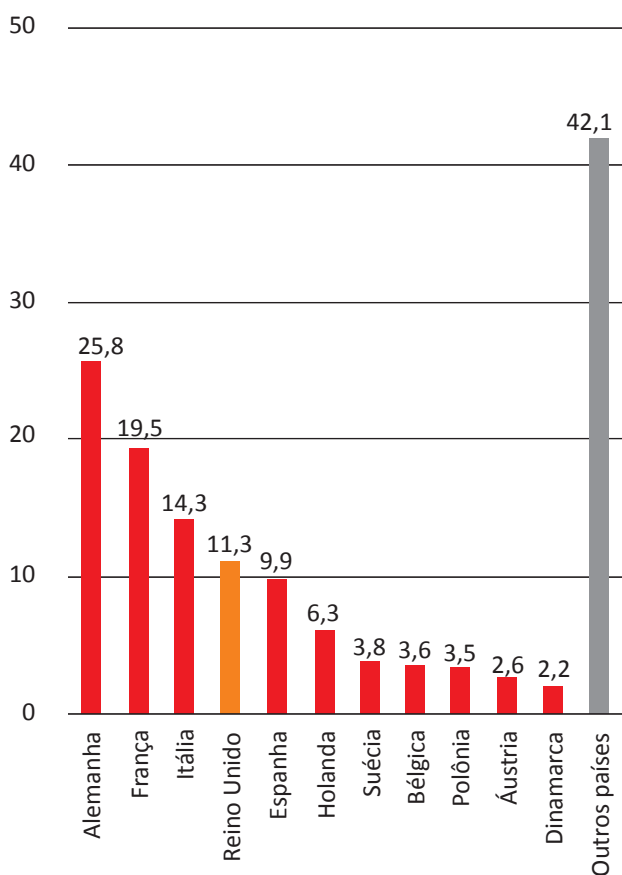


QUESTÃO 01

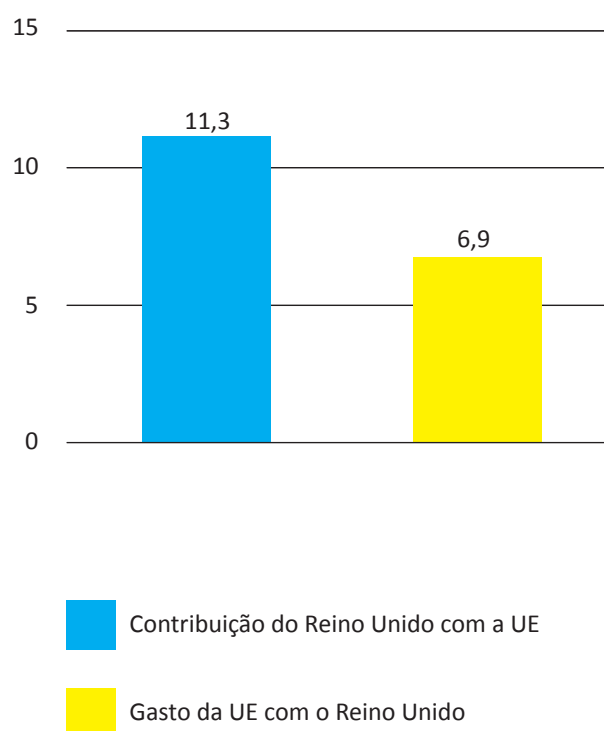
Os britânicos decidiram sair da União Europeia (UE). A decisão do referendo abalou os mercados financeiros em meio às incertezas sobre os possíveis impactos dessa saída.

Os gráficos a seguir apresentam, respectivamente, as contribuições dos países integrantes do bloco para a UE, em 2014, que somam € 144,9 bilhões de euros, e a comparação entre a contribuição do Reino Unido para a UE e a contrapartida dos gastos da UE com o Reino Unido.

Contribuições para a UE
 Dados de 2014, em € bilhões



Reino Unido e UE
 Dados de 2014, em € bilhões



Disponível em: <<http://www.g1.globo.com>>. Acesso em: 6 set. 2017 (adaptado).

Considerando o texto e as informações apresentadas nos gráficos acima, assinale a opção correta.

- A** A contribuição dos quatro maiores países do bloco somou 41,13%.
- B** O grupo “Outros países” contribuiu para esse bloco econômico com 42,1%.
- C** A diferença da contribuição do Reino Unido em relação ao recebido do bloco econômico foi 38,94%.
- D** A soma das participações dos três países com maior contribuição para o bloco econômico supera 50%.
- E** O percentual de participação do Reino Unido com o bloco econômico em 2014 foi de 17,8%, o que o colocou entre os quatro maiores participantes.



QUESTÃO 02

Segundo o relatório da Organização das Nações Unidas para a Alimentação e a Agricultura de 2014, a agricultura familiar produz cerca de 80% dos alimentos no mundo e é guardiã de aproximadamente 75% de todos os recursos agrícolas do planeta. Nesse sentido, a agricultura familiar é fundamental para a melhoria da sustentabilidade ecológica.

Disponível em: <<http://www.fao.org>>. Acesso em: 29 ago. 2017 (adaptado).

Considerando as informações apresentadas no texto, avalie as afirmações a seguir.

- I. Os principais desafios da agricultura familiar estão relacionados à segurança alimentar, à sustentabilidade ambiental e à capacidade produtiva.
- II. As políticas públicas para o desenvolvimento da agricultura familiar devem fomentar a inovação, respeitando o tamanho das propriedades, as tecnologias utilizadas, a integração de mercados e as configurações ecológicas.
- III. A maioria das propriedades agrícolas no mundo tem caráter familiar, entretanto o trabalho realizado nessas propriedades é majoritariamente resultante da contratação de mão de obra assalariada.

É correto o que se afirma em

- A I, apenas.
- B III, apenas.
- C I e II, apenas.
- D II e III, apenas.
- E I, II e III.

Área livre





QUESTÃO 03

O sistema de tarifação de energia elétrica funciona com base em três bandeiras. Na bandeira verde, as condições de geração de energia são favoráveis e a tarifa não sofre acréscimo. Na bandeira amarela, a tarifa sofre acréscimo de R\$ 0,020 para cada kWh consumido, e na bandeira vermelha, condição de maior custo de geração de energia, a tarifa sofre acréscimo de R\$ 0,035 para cada kWh consumido. Assim, para saber o quanto se gasta com o consumo de energia de cada aparelho, basta multiplicar o consumo em kWh do aparelho pela tarifa em questão.

Disponível em: <<http://www.aneel.gov.br>>. Acesso em: 17 jul. 2017 (adaptado).

Na tabela a seguir, são apresentadas a potência e o tempo de uso diário de alguns aparelhos eletroeletrônicos usuais em residências.

Aparelho	Potência (kW)	Tempo de uso diário (h)	kWh
Carregador de celular	0,010	24	0,240
Chuveiro 3 500 W	3,500	0,5	1,750
Chuveiro 5 500 W	5,500	0,5	2,250
Lâmpada de LED	0,008	5	0,040
Lâmpada fluorescente	0,015	5	0,075
Lâmpada incandescente	0,060	5	0,300
Modem de internet em <i>stand-by</i>	0,005	24	0,120
Modem de internet em uso	0,012	8	0,096

Disponível em: <<https://www.educandoseubolso.blog.br>>. Acesso em: 17 jul. 2017 (adaptado).

Considerando as informações do texto, os dados apresentados na tabela, uma tarifa de R\$ 0,50 por kWh em bandeira verde e um mês de 30 dias, avalie as afirmações a seguir.

- I. Em bandeira amarela, o valor mensal da tarifa de energia elétrica para um chuveiro de 3 500 W seria de R\$ 1,05, e de R\$ 1,65, para um chuveiro de 5 500 W.
- II. Deixar um carregador de celular e um *modem* de internet em *stand-by* conectados na rede de energia durante 24 horas representa um gasto mensal de R\$ 5,40 na tarifa de energia elétrica em bandeira verde, e de R\$ 5,78, em bandeira amarela.
- III. Em bandeira verde, o consumidor gastaria mensalmente R\$ 3,90 a mais na tarifa de energia elétrica em relação a cada lâmpada incandescente usada no lugar de uma lâmpada LED.

É correto o que se afirma em

- A** II, apenas.
- B** III, apenas.
- C** I e II, apenas.
- D** I e III, apenas.
- E** I, II e III.



QUESTÃO 04

Sobre a televisão, considere a tirinha e o texto a seguir.

TEXTO 1



Disponível em: <<https://www.coletivando.files.wordpress.com>>. Acesso em: 25 jul. 2015.

TEXTO 2

A televisão é este contínuo de imagens, em que o telejornal se confunde com o anúncio de pasta de dentes, que é semelhante à novela, que se mistura com a transmissão de futebol. Os programas mal se distinguem uns dos outros. O espetáculo consiste na própria sequência, cada vez mais vertiginosa, de imagens.

PEIXOTO, N. B. As imagens de TV têm tempo? In: NOVAES, A. **Rede imaginária**: televisão e democracia. São Paulo: Companhia das Letras, 1991 (adaptado).

Com base nos textos 1 e 2, é correto afirmar que o tempo de recepção típico da televisão como veículo de comunicação estimula a

- A** contemplação das imagens animadas como meio de reflexão acerca do estado de coisas no mundo contemporâneo, traduzido em forma de espetáculo.
- B** fragmentação e o excesso de informação, que evidenciam a opacidade do mundo contemporâneo, cada vez mais impregnado de imagens e informações superficiais.
- C** especialização do conhecimento, com vistas a promover uma difusão de valores e princípios amplos, com espaço garantido para a diferença cultural como capital simbólico valorizado.
- D** atenção concentrada do telespectador em determinado assunto, uma vez que os recursos expressivos próprios do meio garantem a motivação necessária para o foco em determinado assunto.
- E** reflexão crítica do telespectador, uma vez que permite o acesso a uma sequência de assuntos de interesse público que são apresentados de forma justaposta, o que permite o estabelecimento de comparações.

Área livre

QUESTÃO 05

Hidrogéis são materiais poliméricos em forma de pó, grão ou fragmentos semelhantes a pedaços de plástico maleável. Surgiram nos anos 1950, nos Estados Unidos da América e, desde então, têm sido usados na agricultura. Os hidrogéis ou polímeros hidrorretentores podem ser criados a partir de polímeros naturais ou sintetizados em laboratório. Os estudos com polímeros naturais mostram que eles são viáveis ecologicamente, mas ainda não comercialmente.

No infográfico abaixo, explica-se como os polímeros naturais superabsorventes, quando misturados ao solo, podem viabilizar culturas agrícolas em regiões áridas.

Por dentro dos hidrogéis

Saiba como funcionam os polímeros superabsorventes que ajudam a reter no solo, por mais tempo, a água da chuva ou da irrigação.



Disponível em: <<http://www.revistapesquisa.fapesp.br>>. Acesso em: 18 jul. 2017 (adaptado).

A partir das informações apresentadas, assinale a opção correta.

- A** O uso do hidrogel, em caso de estiagem, propicia a mortalidade dos pés de café.
- B** O hidrogel criado a partir de polímeros naturais deve ter seu uso restrito a solos áridos.
- C** Os hidrogéis são usados em culturas agrícolas e florestais e em diferentes tipos de solos.
- D** O uso de hidrogéis naturais é economicamente viável em lavouras tradicionais de larga escala.
- E** O uso dos hidrogéis permite que as plantas sobrevivam sem a água da irrigação ou das chuvas.

Área livre



QUESTÃO 06

A imigração haitiana para o Brasil passou a ter grande repercussão na imprensa a partir de 2010. Devido ao pior terremoto do país, muitos haitianos redescobriram o Brasil como rota alternativa para migração. O país já havia sido uma alternativa para os haitianos desde 2004, e isso se deve à reorientação da política externa nacional para alcançar liderança regional nos assuntos humanitários.

A descoberta e a preferência pelo Brasil também sofreram influência da presença do exército brasileiro no Haiti, que intensificou a relação de proximidade entre brasileiros e haitianos. Em meio a esse clima amistoso, os haitianos presumiram que seriam bem acolhidos em uma possível migração ao país que passara a liderar a missão da ONU.

No entanto, os imigrantes haitianos têm sofrido ataques xenofóbicos por parte da população brasileira. Recentemente, uma das grandes cidades brasileiras serviu como palco para uma marcha anti-imigração, com demonstrações de um crescente discurso de ódio em relação a povos imigrantes marginalizados.

Observa-se, na maneira como esses discursos se conformam, que a reação de uma parcela dos brasileiros aos imigrantes se dá em termos bem específicos: os que sofrem com a violência dos atos de xenofobia, em geral, são negros e têm origem em países mais pobres.

SILVA, C. A. S.; MORAES, M. T. A política migratória brasileira para refugiados e a imigração haitiana. *Revista do Direito*. Santa Cruz do Sul, v. 3, n. 50, p. 98-117, set./dez. 2016 (adaptado).

A partir das informações do texto, conclui-se que

- A** o processo de acolhimento dos imigrantes haitianos tem sido pautado por características fortemente associadas ao povo brasileiro: a solidariedade e o respeito às diferenças.
- B** as reações xenofobas estão relacionadas ao fato de que os imigrantes são concorrentes diretos para os postos de trabalho de maior prestígio na sociedade, aumentando a disputa por boas vagas de emprego.
- C** o acolhimento promovido pelos brasileiros aos imigrantes oriundos de países do leste europeu tende a ser semelhante ao oferecido aos imigrantes haitianos, pois no Brasil vigora a ideia de democracia racial e do respeito às etnias.
- D** o nacionalismo exacerbado de classes sociais mais favorecidas, no Brasil, motiva a rejeição aos imigrantes haitianos e a perseguição contra os brasileiros que pretendem morar fora do seu país em busca de melhores condições de vida.
- E** a crescente onda de xenofobia que vem se destacando no Brasil evidencia que o preconceito e a rejeição por parte dos brasileiros em relação aos imigrantes haitianos é pautada pela discriminação social e pelo racismo.

Área livre



QUESTÃO 07

A produção artesanal de panela de barro é uma das maiores expressões da cultura popular do Espírito Santo. A técnica de produção pouco mudou em mais de 400 anos, desde quando a panela de barro era produzida em comunidades indígenas. Atualmente, apresenta-se com modelagem própria e original, adaptada às necessidades funcionais da culinária típica da região. As artesãs, vinculadas à Associação das Paneleiras de Goiabeiras, do município de Vitória-ES, trabalham em um galpão com cabines individuais preparadas para a realização de todas as etapas de produção. Para fazer as panelas, as artesãs retiram a argila do Vale do Mulembá e do manguezal que margeia a região e coletam a casca da *Rhizophora mangle*, popularmente chamada de mangue vermelho. Da casca dessa planta as artesãs retiram a tintura impermeabilizante com a qual açoitam as panelas ainda quentes. Por tradição, as autênticas moqueca e torta capixabas, dois pratos típicos regionais, devem ser servidas nas panelas de barro assim produzidas. Essa fusão entre as panelas de barro e os pratos preparados com frutos do mar, principalmente a moqueca, pelo menos no estado do Espírito Santo, faz parte das tradições deixadas pelas comunidades indígenas.

Disponível em: <<http://www.vitoria.es.gov.br>>. Acesso em: 14 jul. 2017 (adaptado).

Como principal elemento cultural na elaboração de pratos típicos da cultura capixaba, a panela de barro de Goiabeiras foi tombada, em 2002, tornando-se a primeira indicação geográfica brasileira na área do artesanato, considerada bem imaterial, registrado e protegido no Instituto do Patrimônio Histórico e Artístico Nacional (Iphan), no Livro de Registro dos Saberes e declarada patrimônio cultural do Brasil.

SILVA, A. Comunidade tradicional, práticas coletivas e reconhecimento: narrativas contemporâneas do patrimônio cultural. **40º Encontro Anual da Anpocs**. Caxambu, 2016 (adaptado).

Atualmente, o trabalho foi profissionalizado e a concorrência para atender ao mercado ficou mais acirrada, a produção que se desenvolve no galpão ganhou um ritmo mais empresarial com maior visibilidade publicitária, enquanto as paneleiras de fundo de quintal se queixam de ficarem ofuscadas comercialmente depois que o galpão ganhou notoriedade.

MERLO, P. Repensando a tradição: a moqueca capixaba e a construção da identidade local. **Interseções**. Rio de Janeiro. v. 13, n. 1, 2011 (adaptado).

Com base nas informações apresentadas, assinale a alternativa correta.

- A** A produção das panelas de barro abrange interrelações com a natureza local, de onde se extrai a matéria-prima indispensável à confecção das peças ceramistas.
- B** A relação entre as tradições das panelas de barro e o prato típico da culinária indígena permanece inalterada, o que viabiliza a manutenção da identidade cultural capixaba.
- C** A demanda por bens culturais produzidos por comunidades tradicionais insere o ofício das paneleiras no mercado comercial, com retornos positivos para toda a comunidade.
- D** A inserção das panelas de barro no mercado turístico reduz a dimensão histórica, cultural e estética do ofício das paneleiras à dimensão econômica da comercialização de produtos artesanais.
- E** O ofício das paneleiras representa uma forma de resistência sociocultural da comunidade tradicional na medida em que o estado do Espírito Santo mantém-se alheio aos modos de produção, divulgação e comercialização dos produtos.



QUESTÃO 08

Os Objetivos de Desenvolvimento Sustentável (ODS) compõem uma agenda mundial adotada durante a Cúpula das Nações Unidas sobre o Desenvolvimento Sustentável, em setembro de 2015. Nessa agenda, representada na figura a seguir, são previstas ações em diversas áreas para o estabelecimento de parcerias, grupos e redes que favoreçam o cumprimento desses objetivos.



Disponível em: <<http://www.stockholmresilience.org>>. Acesso em: 26 set. 2017 (adaptado).

Considerando que os ODS devem ser implementados por meio de ações que integrem a economia, a sociedade e a biosfera, avalie as afirmações a seguir.

- I. O capital humano deve ser capacitado para atender às demandas por pesquisa e inovação em áreas estratégicas para o desenvolvimento sustentável.
- II. A padronização cultural dinamiza a difusão do conhecimento científico e tecnológico entre as nações para a promoção do desenvolvimento sustentável.
- III. Os países devem incentivar políticas de desenvolvimento do empreendedorismo e de atividades produtivas com geração de empregos que garantam a dignidade da pessoa humana.

É correto o que se afirma em

- A** II, apenas.
- B** III, apenas.
- C** I e II, apenas.
- D** I e III, apenas.
- E** I, II e III.



COMPONENTE ESPECÍFICO

QUESTÃO DISCURSIVA 03

As mitocôndrias são organelas que surgiram da integração endossimbiótica de uma protobactéria com uma célula hospedeira (protoeucarionte). Durante a evolução eucariótica, o genoma e o proteoma dessas mitocôndrias foram significativamente modificados. Embora as mitocôndrias sejam relevantes para os processos de transformação de energia, alguns eucariotos encontrados em intestinos de mamíferos podem apresentar mitocôndrias reduzidas em tamanho e, em alguns casos, até ausentes.

Estudos realizados por meio do sequenciamento completo do genoma e de dois RNAs de um flagelado do gênero *Monocercomonoides* revelaram que este não apresenta a organela e nem genes a ela relacionados, como os da cadeia transportadora de elétrons e do ciclo dos ácidos tricarbóxicos.

KARNKOWSKA, A. et al. A Eukaryote without a mitochondrial organelle. *Current Biology*, n. 26, p. 1274–1284, 2016 (adaptado).

Considerando o tema tratado no texto, faça o que se pede nos itens a seguir.

- Descreva uma estratégia utilizada por seres desprovidos de mitocôndria para obtenção de energia, indicando o local onde ocorre. (valor: 5,0 pontos)
- Explique de que forma os sequenciamentos de DNA e RNA podem ser utilizados como indicadores da atividade mitocondrial. (valor: 5,0 pontos)

RASCUNHO	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

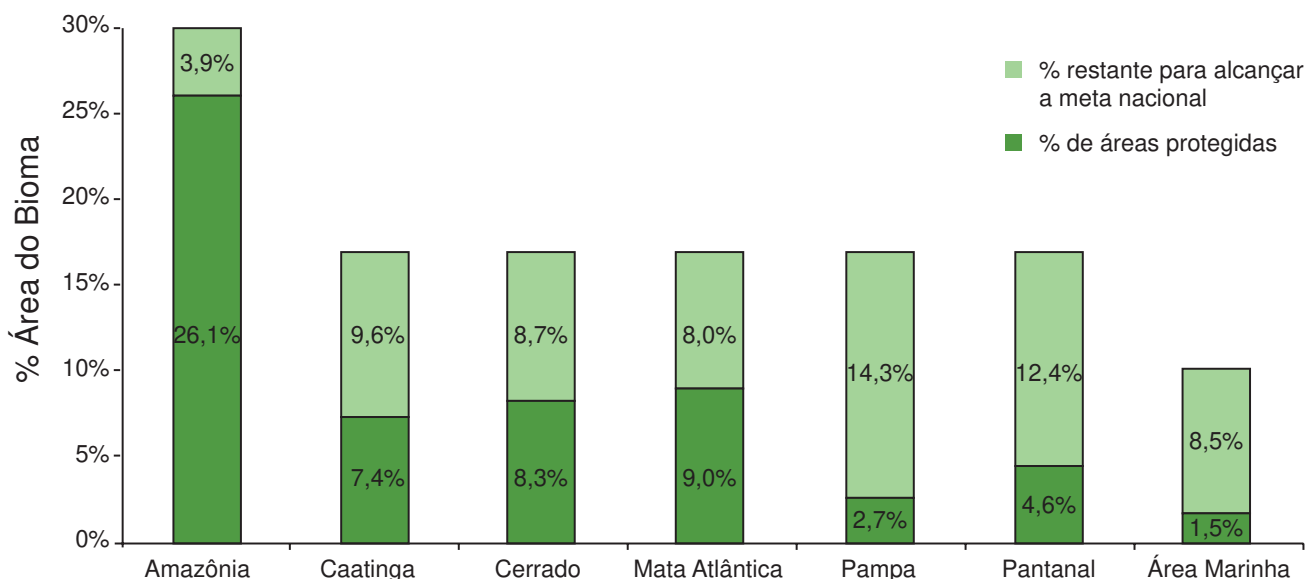
Área livre



QUESTÃO DISCURSIVA 04

Em 2013, a Comissão Nacional de Biodiversidade (CONABIO), cumprindo com suas obrigações legais e com o compromisso internacional com a Convenção sobre Diversidade Biológica (CDB), aprovou as Metas Nacionais de Biodiversidade para o período de 2011 a 2020. Do total de 20 metas, destacam-se as que preveem a conservação de 20% das áreas terrestres e águas continentais e 17% das áreas marinhas e costeiras, por meio de sistemas de unidades de conservação e outras categorias de áreas oficialmente protegidas. A figura a seguir representa a contribuição das áreas protegidas no âmbito do Sistema Nacional de Unidades de Conservação (SNUC) para o alcance da meta nacional.

Unidades de Conservação nos Biomas Brasileiros



BRASIL. Ministério do Meio Ambiente. Secretaria de Biodiversidade e Florestas. 5º relatório nacional para a Convenção Sobre Diversidade Biológica. Brasília, 2016 (adaptado).

No que se refere à biodiversidade em políticas públicas de ciência, tecnologia e inovação, dadas as características do Brasil, com elevada diversidade biológica e cultural, e elevado grau de ameaça em todos os biomas brasileiros, é recomendável que, nos editais e chamadas públicas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sejam incentivadas abordagens da biodiversidade integradas a um contexto sociocultural conservacionista, reconhecendo-se e valorizando-se a diversidade de saberes e de possibilidades de engajamento para a conservação da biodiversidade.

OLIVEIRA, D. **Biodiversidade em políticas públicas de ciência, tecnologia e inovação**: caracterização e perspectivas na integração do fomento à divulgação e educação em ciências. 230f. Tese (doutorado) – Universidade Federal do Rio Grande – FURG, Faculdade de Educação, Programa de Pós-graduação em Educação em Ciências: Química da Vida e Saúde, 2016 (adaptado).

Considerando a temática apresentada nos textos, faça o que se pede nos itens a seguir.

- Descreva dois possíveis efeitos da ampliação das unidades de conservação sobre a biodiversidade. (valor: 4,0 pontos)
- Apresente três exemplos de ações que integrem o reconhecimento e a valorização do contexto sociocultural com a conservação da biodiversidade. (valor: 6,0 pontos)



RASCUNHO	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

Área livre



QUESTÃO DISCURSIVA 05

Educação e Saúde são setores com afinidades no campo das políticas públicas, por se basearem na universalização de direitos fundamentais preconizados na Constituição Federal, e as ações de saúde realizadas no contexto escolar alcançam diretamente a comunidade, pois possibilitam a atuação mais próxima aos fatores de risco e determinantes sociais.

NAKAGAWA, C. *Promoção da saúde nas ações de controle e combate à dengue nas escolas de Ceilândia*.
Brasília: FCE/UnB, 2013 (adaptado).

Entre as políticas públicas educacionais brasileiras, destaca-se o Programa Saúde na Escola (PSE), cujo objetivo é contribuir para a formação integral dos estudantes por meio de ações de promoção, prevenção e atenção à saúde. O tema “Comunidade Escolar Mobilizada Contra o *Aedes aegypti*!” foi escolhido pelo PSE para a 6ª Semana Saúde na Escola 2017.

Considerando o tema proposto pelo Programa Saúde na Escola para 2017, faça o que se pede nos itens a seguir.

- Utilize dois argumentos que justifiquem o papel do professor de Ciências/Biologia como agente promotor de saúde na comunidade escolar. (valor: 4,0 pontos).
- Apresente um objetivo, os recursos e o público alvo de um projeto para ações educativas sobre o tema “Comunidade Escolar Mobilizada Contra o *Aedes aegypti*!”. (valor: 6,0 pontos).

RASCUNHO	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

Área livre



QUESTÃO 09

A técnica de edição genética conhecida como CRISPR (repetições palindrômicas curtas agrupadas e regularmente interespaçadas) tem sido observada com entusiasmo pela comunidade científica. Nessa técnica, o genoma de uma célula pode ser modificado em pontos específicos com grande precisão. Uma enzima associada a uma molécula de RNA-guia identifica um ponto alvo no genoma, clivando-o para a remoção ou a inserção de uma sequência de nucleotídeos.

SUZUKI, K., et al. *In vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. **Nature**. n. 540, p. 144-149, 2016 (adaptado).

Com relação às informações apresentadas, avalie as seguintes asserções e a relação proposta entre elas.

- I. A edição de genomas em células somáticas pode representar uma alternativa viável em terapias gênicas, entretanto deve ser utilizada com cautela.

PORQUE

- II. A edição de sequências genômicas pode provocar o surgimento de efeitos não previstos (*off target*), como a alteração da expressão de genes relacionados a outras funções.

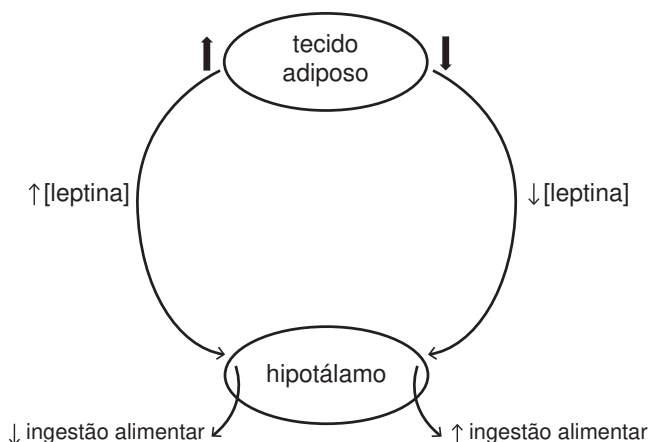
A respeito dessas asserções, assinale a opção correta.

- A** As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- B** As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- C** A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- D** A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- E** As asserções I e II são proposições falsas.

Área livre

QUESTÃO 10

O mecanismo de ação da leptina, hormônio produzido no tecido adiposo, está representado no esquema a seguir.



Com base na análise da figura, assinale a opção correta.

- A** O aumento da síntese de leptina inibe a sua produção pelo tecido adiposo, o que caracteriza um processo de retroalimentação positiva.
- B** O aumento da concentração plasmática de leptina sinaliza a saciedade, o que implica a diminuição do consumo de alimentos.
- C** O aumento da massa corporal gera maior deposição de tecido adiposo, o que diminui a síntese e a liberação de leptina.
- D** A diminuição da liberação de leptina estimula maior atividade do hipotálamo, diminuindo o consumo de alimentos.
- E** A leptina tem ação no hipotálamo e no tecido conjuntivo que a produz, regulando o gasto energético.

Área livre



QUESTÃO 11

Razanandrongobe sakalavae é um grande Archosauria predador do Jurássico médio da Ilha de Madagascar. Em estudo recente, pesquisadores analisaram novos fragmentos da mesma região e da mesma espécie e identificaram a espécie *R. sakalavae* como pertencente ao clado Crocodylomorpha. As profundas e gigantescas mandíbulas do crocodilo pré-histórico eram armadas com enormes dentes serrados e muito semelhantes em tamanho e forma às dos dinossauros terópodes e, especialmente, às do carnívoro *Tiranosaurus rex*. Essas mandíbulas são uma notável evidência de que esses répteis crocodilianos se alimentaram de tecidos orgânicos duros, como ossos e tendões.

DAL SASSO, C. et al. *Razanandrongobe sakalavae*, a gigantic mesoeucrocodylian from the Middle Jurassic of Madagascar, is the oldest known notosuchian. *PeerJ* 5:e3481, 2017 (adaptado).

A evolução das aves ratitas tem sido atribuída à especiação vicariante, impulsionada pela dissolução do supercontinente Gondwana no período Cretáceo. O isolamento inicial da África e de Madagascar é indicativo de que o avestruz e os pássaros-elefante-de-madagascar – esse último extinto – devem ser as linhagens de ratitas mais antigas. O sequenciamento do genoma mitocondrial de dois pássaros-elefante e análises filogenéticas revelaram que essas aves são os parentes mais próximos do kiwi da Nova Zelândia e estão distantes da linhagem basal de avestruzes. Essas evidências sugerem que a convergência para o gigantismo e a ausência de voo foi facilitada no Terciário, após a extinção dos dinossauros, com a exploração do nicho de herbivoria diurna.

MITCHELL, K. J. et al. Ancient DNA reveals elephant bird and kiwi are sister taxa and clarifies ratite bird Evolution. *Science*, v. 344, n. 898, 2014 (adaptado).

Considerando os textos apresentados, avalie as afirmações a seguir.

- I. A semelhança de dentições entre os dinossauros terópodes e o crocodilo fóssil exemplifica uma característica homoplásica.
- II. A divergência a respeito da evolução das linhagens de aves ratitas constitui exemplo de hipóteses de vicariância e de dispersão.
- III. A distribuição biogeográfica das espécies de aves ratitas é considerada cosmopolita.
- IV. A ilha de Madagascar representa um *hotspot* biogeográfico dada, entre outras características, a presença de endemismo de espécies.

É correto apenas o que se afirma em

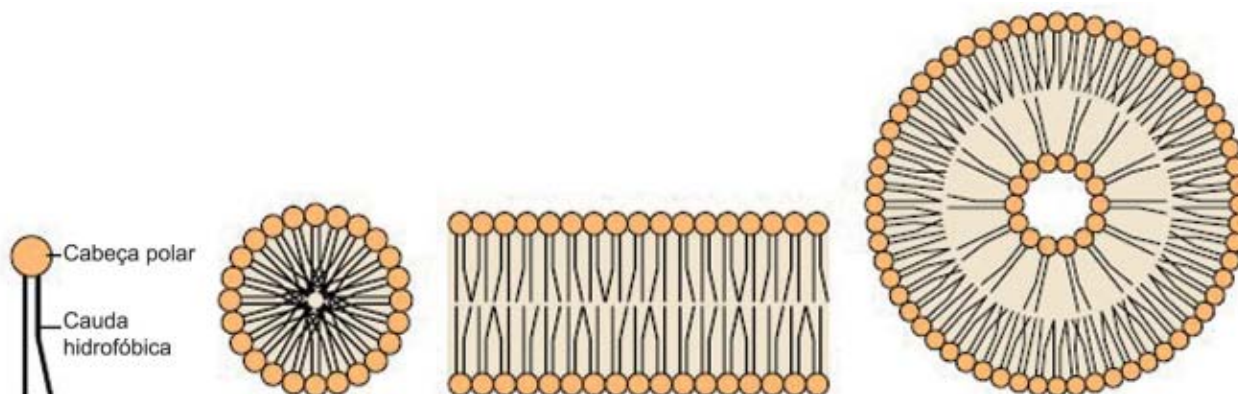
- A** I e III.
- B** II e III.
- C** II e IV.
- D** I, II e IV.
- E** I, III e IV.

Área livre



QUESTÃO 12

A figura a seguir representa a estrutura de um fosfolípideo e suas diferentes interações em meio aquoso.



Interações de fosfolípeos em um meio aquoso

DEVLIN, T. M. **Manual de bioquímica e correlações clínicas**. São Paulo: Blucher, 2011 (adaptado).

Os lipossomos, formados a partir de interações fosfolípídicas, têm sido utilizados na administração de uma variedade de substâncias terapêuticas para alcançar órgãos e tecidos específicos. A incorporação de drogas de rápida metabolização em lipossomos permite liberação dessas substâncias por um período mais longo, o que aumenta sua eficácia.

A respeito das características biofísicas e bioquímicas dos lipossomos, assinale a opção correta.

- A** A aplicação dos lipossomos é restrita a agentes antibióticos, antimaláricos e anti-inflamatórios.
- B** A captação de lipossomos pode ser realizada por fusão com a membrana plasmática ou por exocitose.
- C** Os lipossomos preparados a partir de fosfolípeos purificados são reconhecidos pelo sistema imune, o que constitui uma vantagem de sua utilização.
- D** Os lipossomos aniônicos interagem com a carga da superfície da membrana celular e com carboidratos específicos para se ligarem a receptores celulares.
- E** Tanto os compostos hidrofílicos quanto os hidrofóbicos podem ser carreados, sendo os primeiros no interior aquoso e os últimos no interior da bicamada lipossômica.

Área livre

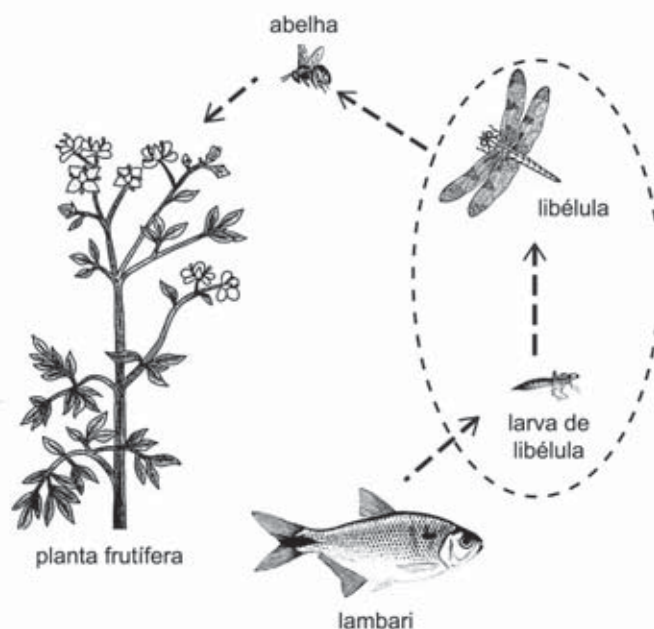


QUESTÃO 13

No sistema Alto Rio Paraná predominam espécies de peixes de pequeno porte, que habitam restritamente riachos e cabeceiras. Entre as espécies do local, cerca de 25% não são nativas. As principais causas dessa ocorrência são a construção de barragens/reservatórios, que permite a dispersão de muitas espécies alóctones, e o escape de várias espécies exóticas de pisciculturas, tal como *Coptodon rendalli* (tilápia). Essas espécies introduzidas podem, eventualmente, apresentar dominância, o que resulta em um decréscimo no número de espécies nativas como bagrinhos, cascudos e lambaris. Desse modo, pode-se dizer que as espécies alóctones são responsáveis pelo desequilíbrio desse ecossistema aquático e da biota do entorno.

LANGANI, F. et al. Diversidade da ictiofauna do Alto Rio Paraná: composição atual e perspectivas futuras. *Biota Neotropica*, v. 7, n. 3, p. 181-197, 2007 (adaptado).

A figura a seguir ilustra a interação ecológica de uma espécie nativa de lambari de riacho e outras espécies de comunidades adjacentes. Nela, as setas indicam efeitos diretos entre populações de algumas espécies.



KNIGHT, T. M. et al. Trophic cascades across ecosystems. *Nature*, v. 437, n. 7060, p. 880-883, 2005 (adaptado).

Considerando as interações ecológicas apresentadas, avalie as afirmações a seguir.

- I. A introdução de peixes exóticos competidores tróficos do lambari terá efeito indireto negativo na abundância de abelhas, o que acarretará efeito indireto negativo para as plantas frutíferas do entorno do riacho.
- II. A introdução de peixes exóticos predadores do lambari terá efeito direto positivo na abundância das larvas de libélulas, o que acarretará efeito indireto positivo na densidade de abelhas.
- III. A extinção do lambari, devido à introdução de predadores exóticos, terá efeito indireto negativo sobre o sucesso reprodutivo das plantas frutíferas do entorno do riacho.

É correto o que se afirma em

- A I, apenas.
- B III, apenas.
- C I e II, apenas.
- D II e III, apenas.
- E I, II e III.

QUESTÃO 14

Acerca dos tentilhões de Galápagos, Darwin faz as colocações transcritas a seguir.

Observando essa transição gradual, essa diversidade, em um grupo de aves pequeno e intimamente aparentado, pode-se postular que, a partir de uma escassez original de aves, uma espécie tenha sido tomada e modificada para fins distintos. Infelizmente, a maior parte dos espécimes da tribo dos tentilhões mesclaram-se; mas tenho motivos fortes para suspeitar que algumas das espécies do subgrupo *Geospiza* se encontram confinadas a ilhas distintas. Se cada uma das ilhas tem representantes do *Geospiza*, isso pode ajudar a explicar o grande, e singular, número de espécies desse subgrupo neste pequeno arquipélago e, como consequência do seu número, a série graduada perfeita no tamanho dos seus bicos.

DARWIN, C. *Journal of Researches into the Natural History and Geology of the countries visited during the voyage round the world of H.M.S. Beagle*, Revised Edition, London: Henry Colburn, 1845, p. 403-420 (adaptado).

Estudos posteriores mostraram que as modificações na forma do bico dos tentilhões de Galápagos também alteraram a frequência do canto de atração sexual.

A partir dessas informações, avalie as asserções a seguir e a relação proposta entre elas.

- I. Todas as espécies de tentilhões de Galápagos são derivadas de um ancestral comum.

PORQUE

- II. Pressões seletivas que resultaram na modificação da forma dos bicos dos tentilhões, somadas ao isolamento geográfico, levaram ao isolamento reprodutivo pré-zigótico etológico.

A respeito dessas asserções, assinale a opção correta.

- A** As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- B** As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- C** A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- D** A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- E** As asserções I e II são proposições falsas.

QUESTÃO 15

O cultivo de plantas transgênicas, assim como o consumo de seus derivados, reveste-se de interesses, impactos e conflitos múltiplos. As principais críticas ao uso de transgênicos referem-se aos riscos ambientais e de saúde pública. Nesse sentido, estudos de impacto ambiental fornecerão elementos que poderão favorecer a preservação dos ecossistemas no que diz respeito à sustentabilidade para a produção dos transgênicos e, com relação à segurança alimentar, é necessário investir em pesquisas mais aprofundadas para identificar os potenciais riscos socioambientais e à saúde.

NODARI, R. O.; GUERRA, M. P. Plantas transgênicas e seus produtos: impactos, riscos e segurança alimentar (Biossegurança de plantas transgênicas). *Revista de Nutrição*, v. 16, n. 1, p. 105-116, 2003 (adaptado).

Sobre o uso dos transgênicos, avalie as afirmações a seguir.

- I. A biossegurança e a adoção do preceito internacional da precaução implica responsabilidade pelas atuais e futuras gerações e pelo ambiente, estando a percepção pública do risco e segurança no uso de produtos transgênicos evidenciada pela rotulagem obrigatória.
- II. Um dos possíveis perigos desses produtos à segurança ambiental consiste no escape gênico, uma vez que os transgênicos não são isolados reprodutivamente de variedades não transgênicas.
- III. Os riscos relacionados à segurança alimentar referem-se às reações adversas dos alimentos derivados de transgênicos para a saúde humana, que podem acarretar alergias, toxicidades e intolerâncias.

É correto o que se afirma em

- A** I, apenas.
- B** II, apenas.
- C** I e III, apenas.
- D** II e III, apenas.
- E** I, II e III.

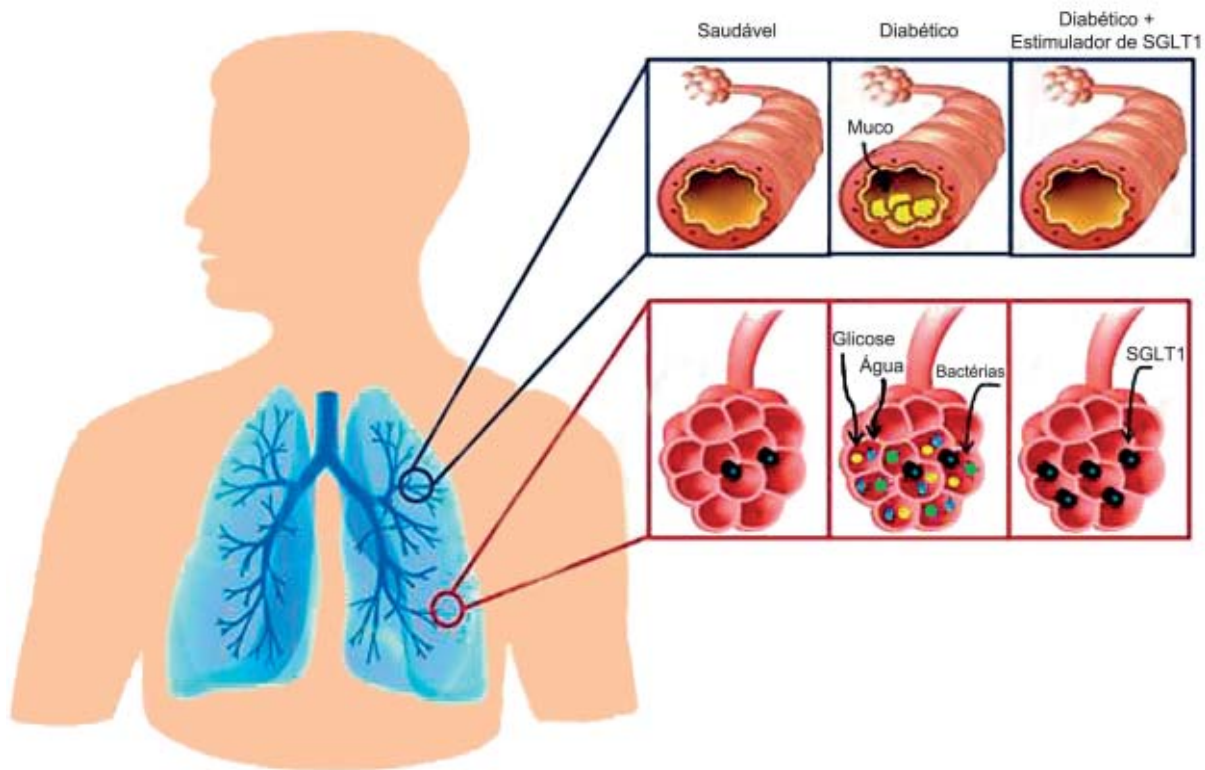
Área livre



QUESTÃO 16

A proteína transportadora de glicose (SGLT1) acoplada ao íon sódio está naturalmente presente em diversos órgãos, sendo responsável pela reabsorção de glicose e pelo transporte contra o gradiente de concentração de glicose e a favor do gradiente de concentração do sódio.

Pesquisas na área de biotecnologia têm demonstrado que estimuladores dessa proteína podem auxiliar no combate a infecções bacterianas em pacientes diabéticos, conforme ilustrado na figura abaixo.



Disponível em: <<http://www.comunica.ufu.br>>. Acesso em: 18 jul. 2017 (adaptado).

A respeito da atuação da SGLT1 nos alvéolos pulmonares de diabéticos, avalie as asserções a seguir e a relação proposta entre elas.

- I. O uso terapêutico de fármacos estimuladores da SGLT1 em pacientes diabéticos promove a redução do volume de muco e, conseqüentemente, da proliferação bacteriana.

PORQUE

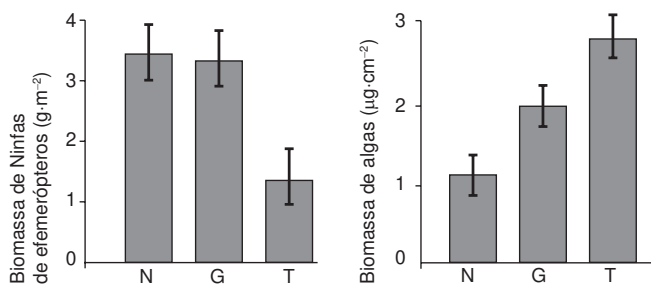
- II. No tecido pulmonar, a SGLT1 promove diminuição da concentração de glicose com conseqüente aumento do potencial hídrico intracelular.

A respeito dessas asserções, assinale a opção correta.

- A** As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- B** As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- C** A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- D** A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- E** As asserções I e II são proposições falsas.

QUESTÃO 17

Em experimento realizado em um riacho, foram instaladas gaiolas para avaliar a biomassa de algas e de ninfas de efemerópteros em relação à ocorrência de duas espécies de peixes predadores da ninfa (G e T). Permitiu-se a colonização das algas durante 12 dias em blocos casualizados em um trecho do riacho. Os tratamentos foram 3: um com as ninfas e sem os peixes (experimento N), outro com a presença do predador G em densidade natural e outro com presença do predador T em densidade natural. Após 12 dias, as algas e os invertebrados foram amostrados e os principais resultados estão apresentados nos gráficos a seguir.



FLECKER, A. S.; TOWNSEND C. R. Community wide consequences of trout introduction in New Zealand streams. *Ecological Applications*, v.4, p. 798-807, 1994 (adaptado).

Com base nos resultados obtidos, avalie as afirmações a seguir.

- I. A presença do predador G não alterou significativamente a biomassa de ninfas.
- II. No que se refere ao controle de ninfas, a espécie predadora T é mais eficiente que a espécie predadora G.
- III. A biomassa de ninfas é diretamente proporcional à biomassa de algas no tratamento T.
- IV. A presença do predador G pode influenciar o comportamento de herbivoria das ninfas.

É correto apenas o que se afirma em

- A I e III.
- B I e IV.
- C II e III.
- D I, II e IV.
- E II, III e IV.

QUESTÃO 18

As técnicas de detecção e identificação de microrganismos causadores de infecções humanas são frequentemente baseadas na utilização de protocolos de cultivo em meios de cultura e posterior contagem de unidades formadoras de colônias (UFC). Como esse processo é dependente da multiplicação das células microbianas, o diagnóstico se torna um processo lento. Recentemente, foi desenvolvido um nanossensor para a detecção, em tempo real, de bactérias em dentes e equipamentos hospitalares. O dispositivo é fabricado em grafeno recoberto com eletrodos de ouro que contém proteínas capazes de se ligar a microrganismos específicos e que emite sinais por conexão sem fio. O sistema é capaz de detectar diferentes tipos de bactérias e até mesmo uma única célula bacteriana em um ambiente.

MANNOOR, M. S. et al. Graphene-based wireless bacteria detection on tooth enamel. *Nature Communications*. v. 3, n. 763, 2012 (adaptado).

Com relação às informações apresentadas, avalie as asserções a seguir e a relação proposta entre elas.

- I. O uso de nanossensores poderia facilitar o diagnóstico de determinadas doenças bacterianas, personalizar a prescrição de antibióticos e ainda ser utilizada para detecção e prospecção de microrganismos em amostras ambientais.

PORQUE

- II. A emissão de laudos e pareceres diagnósticos poderia ser substancialmente agilizada, uma vez que os nanossensores emitem sinais por conexão sem fio.

A respeito dessas asserções, assinale a opção correta.

- A As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- B As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- C A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- D A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- E As asserções I e II são proposições falsas.



* R 0 6 2 0 1 7 2 4 *

QUESTÃO 19

A bioética pode ser entendida como a aplicação prática da ética, tratando de problemas persistentes e emergentes. Os problemas persistentes dizem respeito a situações tradicionalmente discutidas em diversas áreas, como o aborto e a eutanásia. Por outro lado, as situações emergentes tratam de questões referentes à reprodução assistida, à saúde pública e coletiva, à engenharia genética, ao ambiente, entre outras.

GARRAFA, V.; COSTA, S. I. F. *A Bioética no século XXI*. Brasília: Editora Universidade de Brasília, 2000 (adaptado).

Em relação à bioética, avalie as asserções a seguir e a relação proposta entre elas.

- I. A preocupação da ética ambiental está centrada no pressuposto “cuidar”, com base no qual o indivíduo atua com responsabilidade social na preservação e conservação socioambiental.

PORQUE

- II. Atualmente, a bioética foca nas questões individuais, avaliadas sob a perspectiva da legalidade.

A respeito dessas asserções, assinale a opção correta.

- A** As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- B** As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- C** A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- D** A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- E** As asserções I e II são proposições falsas.

Área livre

QUESTÃO 20

Sob o ponto de vista socioeconômico, a hidroponia é um sistema de produção agrícola considerado ecologicamente viável, devido à grande possibilidade de ser aplicado com êxito em condições e usos distintos. Por meio dessa técnica, plantas de interesse econômico, como verduras, frutas, flores, ervas aromáticas e plantas ornamentais, são cultivadas sem terra e, dessa forma, suas raízes ficam suspensas. Ao sistema é adicionado o líquido hidropônico, uma solução nutritiva contendo íons solubilizados, que é necessária para o desenvolvimento normal da planta.

SARDARE, M. D.; ADMANE, S. V. A Review on plant without soil – hydroponics. *International Journal of Research in Engineering and Technology*, v. 2, n. 3, p. 299-304, 2013 (adaptado).

Considerando o texto, avalie as afirmações a seguir.

- I. A absorção de água pela raiz de uma planta hidropônica é consequência do transporte passivo de íons para o interior do floema desse órgão, que provoca aumento da pressão osmótica da raiz em relação ao líquido hidropônico.
- II. A concentração da solução nutritiva está relacionada com a produtividade agrícola das plantas cultivadas em um sistema hidropônico até o limite da capacidade de absorção de nutrientes pelas raízes.
- III. A condução da água por capilaridade das raízes em direção a outras partes da planta cultivada em um sistema hidropônico correlaciona os processos de absorção e transpiração.

É correto o que se afirma em

- A** I, apenas.
- B** II, apenas.
- C** I e III, apenas.
- D** II e III, apenas.
- E** I, II e III.

QUESTÃO 21

A exposição a microrganismos e parasitas na infância beneficia o desenvolvimento do sistema imunológico humano. Por outro lado, essa exposição pode desencadear uma resposta imune aumentada, o que resulta em doenças alérgicas e autoimunes. Em alguns países europeus com cultura de extrema limpeza, a incidência de alergia e doenças autoimunes em crianças de até dois anos de idade pode ser até 5 ou 6 vezes maior que naqueles em que há uma menor cultura higienista. Análises metagenômicas indicam que a presença de microrganismos – encontrados em maior frequência em crianças de países com cultura de extrema limpeza – pode produzir substâncias que inibem as respostas imunológicas, enquanto a presença da bactéria *Escherichia coli* – mais comum em crianças de países com menor cultura higienista – consegue tornar o sistema imune mais eficiente, o que reduz a prevalência de alergias e doenças.

VATANEN, T. et al. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell*. n. 165, p. 842-853, 2016 (adaptado).

A respeito da temática tratada no texto, avalie as afirmações a seguir.

- I. Microrganismos de diferentes espécies desencadeiam os mesmos eventos imunológicos, o que favorece o desenvolvimento de respostas imunes mais eficazes.
- II. Uma maior exposição a determinados grupos microbianos e de parasitas aumenta a capacidade de resposta do sistema imune.
- III. O processo imunológico é multifatorial por estar associado à composição genética dos indivíduos e a fatores ambientais.
- IV. Fatores epigenéticos podem resultar em modificações do padrão de resposta imunológica.

É correto apenas o que se afirma em

- A** I e III.
- B** I e IV.
- C** II e IV.
- D** I, II e III.
- E** II, III e IV.

QUESTÃO 22

O *Leontopithecus rosalia* (mico-leão-dourado) é uma espécie endêmica, com distribuição restrita a remanescentes florestais severamente antropizados da Mata Atlântica no estado do Rio de Janeiro. Embora a população dessa espécie esteja aumentando como resultado de esforços conservacionistas e da criação de áreas protegidas, ela ainda sofre com a fragmentação e seus efeitos secundários e permanece ameaçada de extinção.

Nesse contexto, as áreas protegidas podem ser estabelecidas com vistas à conservação de espécies únicas, como a “megafauna carismática”, que cativa o público, a exemplo do *L. rosalia*. Essas áreas de conservação geram benefícios diretos para toda a sociedade, no que diz respeito à proteção de mananciais de água, à regulação do clima, à contenção de erosões, ao oferecimento de oportunidades de lazer com apreciação de paisagens únicas, à manutenção das riquezas culturais e às possibilidades econômicas sustentáveis de desenvolvimento.

PRIMACK, R. B.; RODRIGUES, E. **Biologia da conservação**. Londrina: E. Rodrigues, 2001 (adaptado).

Em relação ao tema tratado no texto, avalie as afirmações a seguir.

- I. As espécies carismáticas ao público geral, como o mico-leão-dourado, são importantes para a conservação de outras espécies, que, embora menos carismáticas, são tão importantes ecologicamente quanto as primeiras.
- II. O valor de existência do mico-leão-dourado decorre de sua importância ecológica na recuperação de ambientes antropizados, além de oferecer serviços ecossistêmicos para a espécie humana.
- III. As áreas protegidas perfazem uma estratégia de sustentabilidade financeira ao fornecer benefícios econômicos diretos e indiretos à sociedade humana.

É correto o que se afirma em

- A** I, apenas.
- B** II, apenas.
- C** I e III, apenas.
- D** II e III, apenas.
- E** I, II e III.

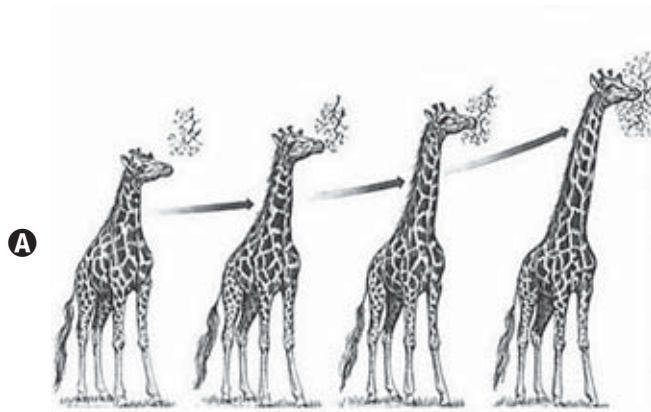


QUESTÃO 23

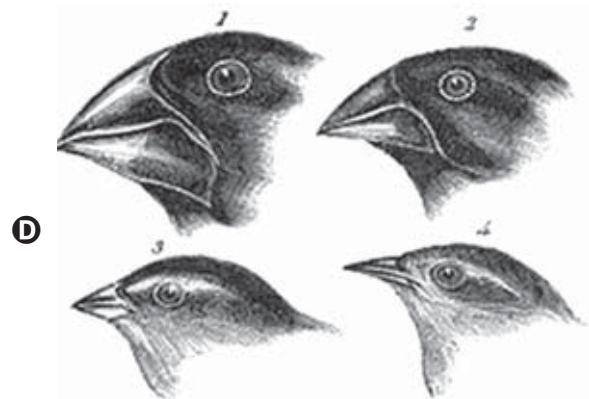
Os organismos produzem descendência em número muito maior em relação aos que serão capazes de sobreviver, o que resulta em uma luta e/ou competição pela sobrevivência. A seleção natural age sobre quaisquer entidades capazes de se reproduzir, de apresentar herança de suas características de uma geração para a outra e de variar em aptidão (isto é, número relativo de descendentes diretos que elas produzirem) de acordo com as características que possuem.

RIDLEY, M. **Evolução**. Porto Alegre: Artmed, 2006 (adaptado).

Qual das imagens a seguir melhor representa o processo de seleção natural no contexto da divulgação científica?



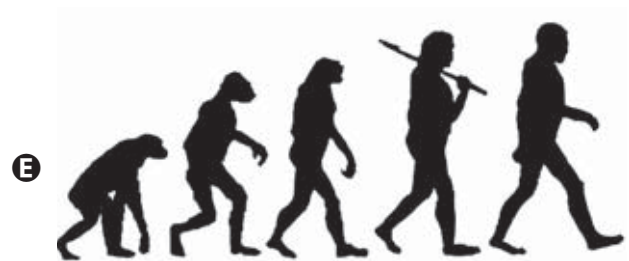
Disponível em: <<https://biologiafacil.wordpress.com>>.
Acesso em: 17 jul. 2017 (adaptado).



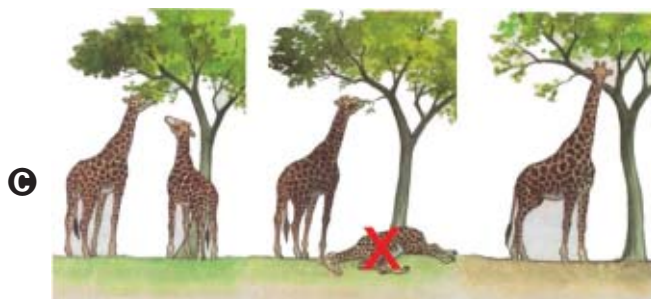
Disponível em: <<http://educador.brasilecola.uol.com.br>>. Acesso em: 17 jul. 2017 (adaptado).



Disponível em: <<http://www.ib.usp.br>>.
Acesso em: 17 jul. 2017 (adaptado).



Disponível em: <<http://www.ich.pucminas.br>>.
Acesso em: 17 jul. 2017 (adaptado).



Disponível em: <<https://biologiacosmos.wordpress.com>>.
Acesso em: 17 jul. 2017 (adaptado).





QUESTÃO 24

A participação social é um direito humano e sua efetivação no cotidiano escolar é fundamental para a formação cidadã de crianças, adolescentes, jovens e adultos, conforme previsto na Constituição brasileira e na Lei de Diretrizes e Base. O artigo 9º e a Meta 19 do Plano Nacional de Educação (PNE) reafirmam esse princípio e vão além ao estabelecerem com maior clareza as políticas de gestão democrática, sendo importantes estratégias a constituição e o fortalecimento de conselhos escolares (CE) e conselhos municipais de educação.

A tabela a seguir apresenta dados sobre a participação de diferentes grupos no conselho escolar por redes de ensino nos anos de 2013 e 2015.

Grupos representados no conselho escolar	Federal		Estaduais		Municipais	
	2013	2015	2013	2015	2013	2015
Professores, funcionários, alunos e pais/responsáveis	24.0%	25.4%	77.0%	80.1%	53.0%	55.6%
Professores, funcionários e pais/responsáveis	4.4%	7.5%	10.2%	10.5%	25,6%	29.1%
Professores, alunos e pais/responsáveis	2.7%	0.0%	2.3%	3.0%	1.9%	2.2%
Professores, funcionários e alunos	16.9%	13.4%	1.1%	1.0%	0.7%	0.7%
Professores e pais/responsáveis	2.7%	1.5%	1.0%	1.3%	2.0%	2.1%
Outros	8.7%	14.9%	0.8%	0.7%	0.8%	0.9%
CE não existe	37.2%	37.3%	5.9%	2.9%	13.0%	8.9%
Total	183	67	20 176	19 280	34 387	34 502

Fonte: MEC/Inep - Prova Brasil - elaboração Cenpec

OBSERVATÓRIO DO PLANO NACIONAL DE EDUCAÇÃO — OPNE. 3 anos do Plano Nacional de Educação.

Disponível em: <<http://www.observatoriodopne.org.br>>. Acesso em: 15 jul. 2017 (adaptado).

A partir da análise dos dados apresentados na tabela, avalie as asserções a seguir e a relação proposta entre elas.

- I. De 2013 a 2015, houve a diminuição de escolas sem CE nas redes estaduais e municipais e a ampliação da representação dos grupos nos conselhos escolares, sendo esse indicador maior nas redes estaduais de ensino.

PORQUE

- II. Para a efetivação da gestão democrática, é necessária a tomada coletiva de decisões, com a participação efetiva do maior número de grupos que compõem a comunidade escolar.

A respeito dessas asserções, assinale a opção correta.

- A** As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- B** As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- C** A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- D** A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- E** As asserções I e II são proposições falsas.



* R 0 6 2 0 1 7 2 8 *

QUESTÃO 25

A escolha da modalidade didática adequada depende do conteúdo e dos objetivos selecionados, da classe a que se destina, do tempo e dos recursos didáticos e infraestrutura disponíveis, assim como dos valores e convicções do professor. Qualquer conteúdo deve incluir uma diversidade de modalidades didáticas, pois cada situação exige uma solução própria; além disso, a variação das atividades pode atrair e interessar os alunos, atendendo às diferenças individuais inerentes de uma sala de aula.

KRASILCHIK, M. *Prática de ensino de biologia*. São Paulo: Edusp, 2016 (adaptado).

Considerando o texto e as diferentes situações de aplicação de modalidades didáticas, avalie as afirmações a seguir.

- I. A aula expositiva é uma modalidade didática comumente utilizada para o trabalho com conteúdos extensos, em turmas com grande número de alunos.
- II. As tecnologias da informação e comunicação são ferramentas indispensáveis às práticas de ensino em Ciências e Biologia.
- III. Demonstrações práticas podem ser realizadas mesmo quando não há material suficiente para que todos possam manusear.

É correto o que se afirma em

- A** I, apenas.
- B** II, apenas.
- C** I e III, apenas.
- D** II e III, apenas.
- E** I, II e III.

Área livre

QUESTÃO 26

A literatura da área de Ensino de Ciências frequentemente mostra a concepção deformada do trabalho científico, trazendo consequências negativas ao processo de ensino-aprendizagem dos conhecimentos biológicos. É comum, por exemplo, que docentes destaquem o papel “neutro” da observação e da experimentação, deixando de lado tanto o papel essencial das hipóteses como norte da investigação científica, quanto a relevância dos conhecimentos (teorias), que orientam todo o processo investigativo. Um dos resultados desse tipo de concepção da ciência é a execução de aulas transmissivas, nas quais os conhecimentos são apresentados como acabados, sem apresentar as dificuldades encontradas ao longo do processo. Apesar do discurso que coloca a experimentação em evidência, de uma forma geral, o ensino de Ciências ainda é essencialmente livresco.

PÉREZ, G. et al. Para uma imagem não deformada do trabalho científico. *Ciência & Educação* v. 7, n. 2, p. 125-153, 2001 (adaptado).

Uma estratégia para o planejamento de uma atividade de ensino na sala de aula que promova uma imagem não deformada do trabalho científico deve envolver

- A** apresentação de uma problematização; levantamento das hipóteses discentes; explicação de um conceito inserido em um episódio histórico; avaliação da argumentação.
- B** apresentação de um fenômeno; investigação das concepções da turma; explicação de um conceito; avaliação para a comparação dos conhecimentos anteriores e posteriores.
- C** apresentação de um experimento; discussão dos resultados; explicação dos conceitos relacionados com o fenômeno em questão; solicitação de um relato de atividades.
- D** apresentação de uma situação cotidiana; investigação das hipóteses discentes; explicação dos conceitos relacionados à situação; avaliação dos conceitos assimilados.
- E** apresentação de um conceito; exposição do método científico; execução de um experimento; solicitação da explicação do fenômeno por meio de um relatório.



QUESTÃO 27

Os textos a seguir apresentam trechos de relatos de pesquisas a respeito de livros didáticos de Ciências e/ou Biologia.

TEXTO 1

O uso racional de medicamentos deve ser considerado um tópico importante da promoção da saúde, que atinge todos os setores da sociedade, devendo ser enfatizado na educação básica.

CORRÊA, A. D. et al. Uma abordagem sobre o uso de medicamentos nos livros didáticos de biologia como estratégia de promoção de saúde. *Ciência e Saúde Coletiva*, v. 18, p. 3071-3081, 2013 (adaptado).

TEXTO 2

Essas obras ainda apresentam erros e insuficiências, o que exige dos professores uma análise mais crítica e detalhada das obras que serão escolhidas para suas aulas de Ciências.

FERREIRA, A. M.; SOARES, C. A. Aracnídeos peçonhentos: análise das informações nos livros didáticos de ciências. *Ciência & Educação (Bauru)*, v.14, n.2, p. 307-314, 2008 (adaptado).

TEXTO 3

Alguns fatores limitam a utilização do livro didático como instrumento colaborativo para as ações desencadeadas no ambiente escolar com vistas ao controle da dengue.

ASSIS, S. S.; PIMENTA, D. N.; SCHAL, V. A dengue nos livros didáticos de ciências e biologia indicados pelo Programa Nacional do Livro Didático. *Ciência & Educação (Bauru)*, v.19, n.3, p. 633-656, 2013 (adaptado).

A respeito dos fragmentos de texto apresentados, avalie as afirmações a seguir.

- I. A defesa para que determinado assunto seja enfatizado na educação escolar, tal como se identifica no texto 1, evidencia o caráter contingente e questionável de um currículo.
- II. O alerta para deficiências encontradas em livros didáticos, tal como identificado no texto 2, convoca docentes a mobilizarem conhecimentos que se esperam construídos desde a formação inicial.
- III. A denúncia da limitação dos livros didáticos quanto a ações cujas finalidades excedem o pedagógico, tal como se identifica no texto 3, ilustra a expectativa de que a escola resolva problemas sociais.

É correto o que se afirma em

- A** I, apenas.
- B** II, apenas.
- C** I e III, apenas.
- D** II e III, apenas.
- E** I, II e III.

Área livre



QUESTÃO 28

Os componentes curriculares e as áreas de conhecimento devem articular a seus conteúdos a abordagem de temas abrangentes e contemporâneos que afetam a vida humana em escala global, regional e local, bem como na esfera individual. Temas como saúde, sexualidade e gênero, vida familiar e social, assim como os direitos das crianças e adolescentes, preservação do ambiente, educação para o consumo, educação fiscal, trabalho, ciência e tecnologia, diversidade cultural, devem permear o desenvolvimento dos conteúdos do currículo.

BRASIL. Ministério da Educação. Secretaria de Educação Básica. **Diretrizes Curriculares Nacionais Gerais da Educação Básica**. Brasília, 2013 (adaptado).

Com base nas informações apresentadas, avalie as asserções a seguir e a relação proposta entre elas.

- I. Conhecimentos socioeducativos, artísticos, culturais e/ou ambientais, fundamentados no pressuposto do respeito e da valorização das diferenças, podem ser trabalhados transversalmente no currículo de Biologia e, assim, contribuir para o combate às discriminações, racismos e preconceitos, favorecendo a reflexão sobre comportamentos responsáveis e solidários em relação ao que é diferente de nós e ao ambiente.

PORQUE

- II. O conhecimento sobre valores, crenças e modos de vida de grupos que estiveram por muito tempo excluídos dos currículos pode contribuir para a superação de preconceitos e discriminações, que têm, historicamente, alimentado os processos de exclusão das pessoas com deficiência e outras, bem como as desigualdades sociais, étnico-raciais, de gênero e diversidade sexual.

A respeito dessas asserções, assinale a opção correta.

- A** As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.
- B** As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.
- C** A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.
- D** A asserção I é uma proposição falsa, e a II é uma proposição verdadeira.
- E** As asserções I e II são proposições falsas.

Área livre



QUESTÃO 29

O Ensino Médio tem como objetivo situar o educando como sujeito produtor de conhecimento e participante do mundo do trabalho, e considerar o desenvolvimento da pessoa, como “sujeito em situação” — cidadão. Esse nível de ensino integra, em uma mesma e única modalidade, as seguintes finalidades:

- a formação da pessoa, de maneira a desenvolver valores e competências necessárias à integração de seu projeto individual ao projeto da sociedade em que se situa;
- o aprimoramento do educando como pessoa humana, incluídos a formação ética e o desenvolvimento da **autonomia intelectual** e do pensamento crítico;
- a preparação e a orientação básica para a sua integração ao mundo do trabalho, com as competências que garantam seu aprimoramento profissional e permitam acompanhar as mudanças que caracterizam a produção no nosso tempo;
- o desenvolvimento das competências para continuar aprendendo, de **forma autônoma** e crítica, em níveis mais complexos de estudos.

BRASIL. Ministério da Educação. **Parâmetros Curriculares Nacionais (Ensino Médio)**.
Brasília, 2000 (adaptado).

Considerando as finalidades do Ensino Médio apresentadas no texto, avalie as propostas pedagógicas a seguir.

- I. Apresentação de plano de trabalho para resolver o problema de reciclagem do lixo na escola, no qual estudantes devem **executar passo a passo as ideias previamente elaboradas por docentes** da escola.
- II. Aula expositiva em que se apresenta historicamente uma situação problema que envolva o conceito de raça, a fim de fornecer argumentos biológicos que contraponham ideias de segregação racial.
- III. Simulação no formato de dramatização (*role play*) que envolva discentes na discussão a respeito do aborto, distribuindo entre os estudantes da classe personagens com diferentes pontos de vista sobre a situação, a fim de que possam formar juízo de valor em relação ao assunto a partir da coleta de informações pertinentes para a análise do problema.
- IV. Demonstração, pela/o docente, do processo de fermentação, seguida de discussão e **estudo dirigido** a fim de que a classe compreenda o papel dos fungos na fabricação de alimentos, tais como pães e pizzas.

É correto apenas o que se propõe em

- A** I e III.
- B** I e IV.
- C** II e III.
- D** I, II e IV.
- E** II, III e IV.

Área livre



QUESTÃO 30

Os textos a seguir constituem relatos de docentes e estudantes em que se identificam elementos de avaliação da prática pedagógica para docentes que se interessam pelo acolhimento das diferenças de corpos, gêneros e sexualidades em suas aulas.

TEXTO 1

“Antes eu tinha uma matéria de Programa de Saúde. Programa de Saúde que era só sobre essa questão de prevenção às DSTs. Era mais a questão de prevenção de DST, não era muito abordada a sexualidade em si. Era mais a prevenção de DST.”

REIS, N.; PINHO, R. Ensino de Biologia e gênero: o que dizem as narrativas não-binárias?
Revista de Ensino de Biologia da Associação Brasileira de Ensino de Biologia (SBEnBio), v. 9, p. 3716-3727, 2016 (adaptado).

TEXTO 2

“Há inúmeros pontos que poderia discutir dessa minha gratificante experiência em sala de aula, mas quero finalizar esse texto com o que consegui concluir ao final da dinâmica que realizei. Ainda há muita associação entre gênero atrelado à genitália externa, entre sexo atrelado à heteronormatividade e a (re)produção de estereótipos de gêneros”.

CARVALHO FILHO, E. G. Ser ou não ser: eis as questões (de gênero).
Revista de Ensino de Biologia da Associação Brasileira de Ensino de Biologia (SBEnBio), v. 9, p. 1412-1423, 2016 (adaptado).

TEXTO 3

“Nesta questão foram obtidas quinze respostas e, dentre elas, sete mencionavam que não haveria outro lugar melhor do que a escola para o trabalho com o assunto sexualidade.”

CANDIDO, C.; FERNANDES, J. A. B. A abordagem de sexualidade na escola atual segundo a concepção de alunos do ensino médio.
Revista de Ensino de Biologia da Associação Brasileira de Ensino de Biologia (SBEnBio), v. 7, p. 1043-1054, 2014 (adaptado).

Considerando os relatos apresentados, avalie as afirmações a seguir.

- I. No trecho “Era mais a questão de prevenção de DST, não era muito abordada a sexualidade em si” (texto 1), evidencia-se um modo biológico-higienista de se trabalhar os aspectos de saúde relacionados às sexualidades.
- II. No trecho “associação entre gênero atrelado à genitália externa” (texto 2), é explícita a percepção segundo a qual o corpo seria a justificativa biológica para as desigualdades de gênero baseadas no sexo.
- III. No trecho “não haveria outro lugar melhor do que a escola para o trabalho com o assunto sexualidade” (texto 3), é focalizada a abordagem pedagógica de um fenômeno humano que diz respeito à forma como prazeres e desejos são vividos.

É correto o que se afirma em

- A** I, apenas.
- B** I e II, apenas.
- C** I e III, apenas.
- D** II e III, apenas.
- E** I, II e III.

Área livre





QUESTÃO 31

A didática escolar cumpre funções de caráter político, educativo e científico a um só tempo. A integralização dessas funções pela didática escolar torna essa disciplina acadêmica algo mais complexo que a simples procura e implementação de procedimentos de ensino. Por meio desse processo, a unidade dialética da teoria e da prática assume as características de uma verdadeira investigação científica da realidade cotidiana da prática pedagógica.

RAYS, O. A. A relação teoria-prática na didática escolar crítica. In: VEIGA, I. P. A. (Org.). **Didática: o ensino e suas relações**. 7. ed. Campinas: Papirus, 2003 (adaptado).

A partir das informações apresentadas, avalie as afirmações a seguir.

- I. A práxis pedagógica envolve a adoção do método dialético no processo de elaboração do conhecimento em articulação com a teoria histórico-cultural.
- II. A apropriação crítica e histórica do conhecimento é um instrumento de compreensão da realidade social e de atuação crítica para a transformação da sociedade.
- III. A Didática é uma área do conhecimento que utiliza os elementos do cotidiano escolar e das questões sociais para atualizar a prática docente.

É correto o que se afirma em

- A I, apenas.
- B III, apenas.
- C I e II, apenas.
- D II e III, apenas.
- E I, II e III.

Área livre



QUESTÃO 32

Um aluno da rede pública de ensino, com 11 anos de idade, está matriculado no 5º ano do Ensino Fundamental e tem surdez profunda bilateral. Ele é bem humorado, brincalhão e bastante sociável. É fluente na língua brasileira de sinais (Libras), mas apresenta dificuldades de leitura e escrita da língua portuguesa. Tem potencial cognitivo elevado, embora necessite de constante interferência e auxílio da professora para realizar suas atividades.

Disponível em: <<http://www.cepae.faced.ufu.br>>.
Acesso em: 7 jul. 2017 (adaptado).

Considerando a situação apresentada e o que estabelece a Política Nacional da Educação Especial na perspectiva da Educação Inclusiva, deve-se assegurar a esse aluno

- A** escolarização que atenda à proposta educacional bilíngue, considerando-se a língua de sinais como primeira língua.
- B** atendimento educacional especializado, priorizando-se o ensino da língua portuguesa, de modo a garantir a educação bilíngue.
- C** processo avaliativo que priorize o uso da língua portuguesa na modalidade escrita, dada a importância da manutenção do registro da aprendizagem.
- D** ensino da língua brasileira de sinais (Libras) após a aquisição da língua portuguesa na modalidade escrita, em processo análogo ao da alfabetização de aluno ouvinte.
- E** educação inclusiva, apesar de a surdez não se enquadrar no campo da deficiência física ou das limitações de mobilidade.

Área livre

QUESTÃO 33

As escolas brasileiras não têm um único jeito de ensinar sobre gênero e sexualidade; pesquisas evidenciam currículos e práticas pedagógicas e de gestão marcadas pela discriminação. Distinções sexistas nas aulas, na chamada, nas filas de meninos e de meninas, nos uniformes, no tratamento e nas expectativas sobre alunos ou alunas, tolerância da violência verbal e até física entre os meninos, representações de homens e mulheres nos materiais didáticos, abordagem quase exclusivamente biológica da sexualidade no livro didático, estigmatização referente à manifestação da sexualidade das adolescentes, perseguição sofrida por homossexuais, travestis e transexuais, evidenciam o quanto a escola (já) ensina, em diferentes momentos e espaços, sobre masculinidade, feminilidade, sexo, afeto, conjugalidade, família.

Disponível em: <<http://www.spm.gov.br>>.
Acesso em: 11 jul. 2017 (adaptado).

Nesse contexto, para construir uma prática pedagógica que promova transformações no sentido da igualdade de gênero a partir do respeito às diferenças, espera-se que a escola

- A** incorpore o conceito de gênero nos diferentes componentes do currículo de maneira transversal.
- B** realize atividades em seu cotidiano que definam para as crianças o que é masculino e o que é feminino.
- C** se valha das diferenças sexuais naturais entre meninos e meninas para conduzir a classe e manter a disciplina.
- D** se refira à questão de gênero de forma tangencial, suficiente para promover vivência menos intransigente e mais equânime entre homens e mulheres.
- E** reforce modelos de comportamentos socialmente atribuídos a homens e mulheres que formam um conjunto de representações sobre masculinidade e feminilidade.



QUESTÃO 34

Lev Semenovitch Vygotsky, psicólogo russo, elaborou sua teoria tendo por base o desenvolvimento do indivíduo como resultado de um processo sócio-histórico, enfatizando o papel da linguagem e da aprendizagem nesse desenvolvimento. Esse pressuposto teórico, conhecido como Teoria Histórico-Cultural, apresenta como questão central a apropriação de conhecimentos pela interação do sujeito com o contexto social.

Considerando os pressupostos da teoria vygotskyana, avalie as afirmações a seguir.

- I. O desenvolvimento cognitivo é produzido no processo de internalização da interação social com a cultura.
- II. Ao acessar a língua escrita, o indivíduo se apropria das técnicas inerentes a este instrumento cultural, modificando suas funções mentais superiores.
- III. A apropriação da linguagem específica do meio sociocultural transforma os rumos do desenvolvimento individual.
- IV. O desenvolvimento das funções psíquicas superiores decorre de funções existentes no indivíduo.
- V. A educação sistemática e organizada pode contribuir com o processo de aquisição dos sistemas de conceitos científicos, o que modifica a estrutura do pensamento do indivíduo.

É correto apenas o que se afirma em

- A** I e IV.
- B** I e V.
- C** II, III e IV.
- D** I, II, III e V.
- E** II, III, IV e V.

Área livre

QUESTÃO 35

A professora de uma escola pública tem sua prática pedagógica fundamentada na teoria de Jean Piaget. Essa professora irá desenvolver com uma turma do 5º ano do Ensino Fundamental uma aula de Ciências sobre o tema força e movimento, utilizando a abordagem construtivista.

Nesse contexto, qual deverá ser a proposta de trabalho elaborada pela professora?

- A** Demonstrar aos estudantes, em laboratório, experimentos relacionados ao tema e realizar avaliação do conteúdo trabalhado.
- B** Utilizar livro didático e figuras previamente selecionadas para sintetizar conceitos e informações relacionados ao conteúdo trabalhado.
- C** Aplicar exercícios de fixação em níveis crescentes de complexidade para a internalização dos conteúdos pelos estudantes.
- D** Partir do saber do cotidiano do estudante sobre a relação entre força e movimento para provocar o surgimento de hipóteses, criar conflitos cognitivos para desenvolvimento do conceito desejado.
- E** Realizar leituras informativas sobre o conteúdo e, a partir da apresentação de *slides* ilustrativos, descrever o conceito de força e de movimento, apresentando exemplos.

Área livre

ANEXO B –Pedido de informação registrado sob o protocolo nº 23480-022380/2019-00 Sobre a Matriz de referência ENADE Ciências Biológicas 2011, 2014 e 2017.



Close

Matriz de referência ENADE Ciências Biológicas 2011, 2014 e 2017 - Pedido 2348002238..

01/10/2019 - Acesso Concedido

Dados do Pedido


Órgão Destinatário: INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
Classificação de Resposta: Resposta solicitada inserida no e-SIC
Categoria do Pedido: Educação
Subcategoria do Pedido: Educação superior
Link Detalhes: http://www.consultaesic.cgu.gov.br/busca/_layouts/15/DetallePedido/DetallePedido.aspx?n..

Pergunta	Sou mestrando no Programa de Pós-graduação em Métodos e Gestão da Avaliação (PPMGGA/UFSC) e esto desenvolvendo uma dissertação que pretende, pela aplicação da TRI, avaliar a qualidade da prova de conhecimento específicos de ciências biológicas. Conversando com os meus professores Adriano Borgatto e Dalton Andrade, surgiram dúvidas para as quais não encontrei a documentação no portal do INEP/ENADE. Foi elaborada alguma matriz de referência (no estilo da imagem anexa) previamente à elaboração das provas de 2011, 2014 e 2017 de ciência biológicas (não falo da descrição presente no relatório, mas de algo elaborado antes da encomenda do BNI)? Gostaria de ter acesso a tais documentos (incluindo se possível os materiais de treinamento dos elaboradores do BNI)?
----------	---

Resposta
21/10/2019

Prezado(a) Senhor(a), Em atendimento ao pedido de informação registrado sob o protocolo nº 23480-022380/2019-00, segue resposta elaborada pela unidade responsável: "Em resposta à solicitação, informamos que todos os documentos e relatórios públicos atinentes ao Enade estão disponibilizados no portal do INEP, sendo que os relatórios anuais possuem todas as informações, inclusive com os relatórios sínteses de área. Nos relatórios constam as informações atinentes à encomenda de itens, com os cruzamentos de características de perfil, competências e conteúdos que correspondem a cada um dos itens da prova, o que pode ser observado no final do relatório da área de Ciências Biológicas (Bacharelado) do Enade 2017. Especificamente no tocante ao ano de 2017, a tabela está em outro formato, porém com as informações solicitadas atinentes à encomenda de itens. Encontram-se ainda disponibilizados no Portal INEP/ENADE os relatórios e demais informações relativas às informações do curso de Biologia em 2011 e Ciências Biológicas no ano de 2014. O que mais existir e não for visualizado no portal do INEP, ou na biblioteca local do INEP, faz parte do conteúdo de documento sigiloso. Para que se tenha vista dos mesmos é necessário que a solicitação atenda a procedimento próprio, mediante solicitação ao serviço de atendimento ao pesquisador. Contato: sap@inep.gov.br O Serviço de Atendimento ao Pesquisador (SAP) foi criado no âmbito da Diretoria de Estudos Educacionais (Direde), a fim de atender às solicitações de acesso controlado à base de dados restritos do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), consoante ao Decreto 6.317, de 20 de dezembro de 2007, e à Portaria nº 467, de 19 de setembro de 2014. Consultores, professores e pesquisadores podem solicitar acesso às informações. Para garantir transparência e segurança ao processo, exclusivamente para fins de pesquisa e de estudo, foi criado um ambiente seguro para a consulta. O espaço, na sede do Inep, em Brasília, permite aos pesquisadores o acesso a bancos de dados para pesquisas sobre censos e avaliações. O pedido deve ser protocolado em formulário específico, com justificativa para o estudo. O objetivo desse procedimento é qualificar a permissão, respeitada a individualidade e a transparência dos dados. Quem pode solicitar o acesso aos dados? • Servidores do Inep; • Servidores públicos externos ao Inep que estejam trabalhando na produção de pesquisas de interesse do Estado; • Bolsistas, consultores e colaboradores do Inep, trabalhando na produção de pesquisas de interesse da Instituição; • Pessoas físicas ou jurídicas. Bases disponíveis Caso se faça necessário, é permitido o acesso simultâneo a mais de uma base de dados, desde que as condições de restrição e de controle de acesso de cada uma das bases sejam respeitadas Caso queira solicitar mais informações, é necessário registrar nova demanda no e-SIC, para que corram os prazos de atendimento previstos pela Lei de Acesso à Informação. Quando for negado o pedido de acesso à informação, o Decreto nº 7.724, de 16 de maio de 2012, estabelece que se resguarda ao interessado a possibilidade de apresentar recurso, no prazo de 10 (dez) dias." Caso queira solicitar mais informações, é necessário registrar uma nova demanda no e-SIC, para que corram os prazos de atendimento previstos pela Lei de Acesso à Informação. Quando for negado o pedido de acesso à informação, o Decreto nº 7.724, de 16 de maio de 2012, estabelece que se resguarda ao interessado a possibilidade de apresentação de recurso, no prazo de 10 (dez) dias. Nesse caso, o recurso será direcionado ao dirigente da Diretoria de Avaliação da Educação Superior. Atenciosamente, Serviço de Informação ao Cidadão – SIC Ouvidoria do Inep Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira Edifício Villa Lobos – Sede do Inep, térreo Setor de Indústrias Gráficas, quadra 04, lote 327 CEP: 70610-908 – Brasília/DF e-SIC:
<http://www.acessoainformacao.gov.br/sistema/>

Anexos:

 PEDIDO_Matriz.png (http://www.consultaesic.cgu.gov.br/busca/dados/Lists/Pedido/Attachments/807688/PEDIDO_Matriz.png)
(http://www.consultaesic.cgu.gov.br/busca/dados/Lists/Pedido/Attachments/807688/PEDIDO_Matriz.png)