

Paulo Augusto da Costa Silva

EQUALIZAÇÃO DA PROVA BELÉM NA ESCALA SAEB

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

Orientador: Prof. Dr. Adriano Borgatto
Coorientador: Prof. Dr. Héilton Tavares

Florianópolis
Março/2017

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

Silva, Paulo Augusto da Costa.

Equalização da Prova Belém na escala SAEB / Paulo Augusto da Costa Silva; Orientador, Adriano Borgatto; coorientador, Héilton Tavares. - Florianópolis, SC, 2017. 128 p.

Dissertação (mestrado profissional) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Métodos e Gestão em Avaliação.

Inclui referências

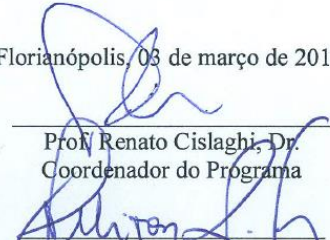
1. Métodos e Gestão em Avaliação. 2. Equalização de Testes Educacionais. 3. Teoria de Resposta ao Item. 4. Avaliação de Larga Escala. I. Borgatto, Adriano. II. Tavares, Héilton. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Métodos e Gestão em Avaliação. IV. Título.

Paulo Augusto da Costa Silva

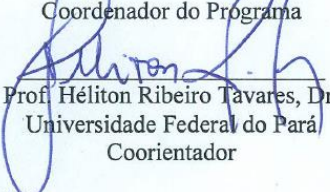
TÍTULO: Equalização da Prova Belém na Escala SAEB

Esta dissertação foi julgada adequada para obtenção do título de mestre e aprovada em sua forma final pelo Programa de Pós-Graduação em Métodos e Gestão em Avaliação.

Florianópolis, 03 de março de 2017.

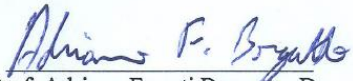


Prof. Renato Cislighi, Dr.
Coordenador do Programa

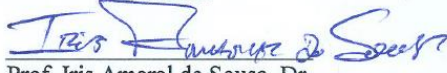


Prof. Héilton Ribeiro Tavares, Dr.
Universidade Federal do Pará
Coorientador

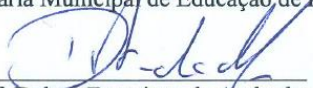
Banca Examinadora:




Prof. Adriano Ferreti Borgatto, Dr.
Universidade Federal de Santa Catarina
Orientador



Prof. Iris Amaral de Sousa, Dr.
Secretaria Municipal de Educação de Belém



Prof. Dalton Francisco de Andrade, Dr.
Universidade Federal de Santa Catarina



Prof. Silvana Ligia Vincenzi Bortolotti, Dra.
Instituto Federal do Paraná

Este trabalho é dedicado a minha esposa HELENA MARGARETH, por todo apoio, incentivo, compreensão e afeto transmitidos. E aos meus queridos pais (*in memoriam*).

AGRADECIMENTOS

À Universidade Federal de Santa Catarina - UFSC;

Ao Programa de Pós-graduação em Métodos e Gestão da Avaliação - PPMGA;

Ao Núcleo de Informática Educativa, da secretaria Municipal de Educação de Belém - NIED/SEMEC;

Ao prof. Dr. Adriano Borgatto, meu orientador, que com dedicação e paciência, pode tornar este trabalho uma realidade;

Ao prof. Dr. Héilton Tavares, meu coorientador, pelo incentivo a participar do processo seletivo do mestrado e apoio durante o mesmo;

Aos colegas do curso, pelo estímulo e trocas realizadas ao longo dos estudos;

A todos os colegas professores do Núcleo de Informática Educativa que contribuíram para esse projeto;

A todos os amigos que, de uma forma ou de outra, me ajudaram neste trabalho.

RESUMO

Este estudo objetivou equalizar a Prova Belém, aplicada aos estudantes da rede pública municipal nos moldes da Prova Brasil, com a escala SAEB via Teoria de Resposta ao Item. Utilizou-se um banco de dados simulado contendo informações de 100.000 alunos para o grupo de referência e o banco de dados contendo informações de 11.271 estudantes da rede municipal de Belém do Ciclo II 1º e 2º anos (equivalentes ao 4º e 5º anos do ensino fundamental/EF de 09 anos) que responderam à Prova Belém em dezembro/2014 de Língua Portuguesa e Matemática. Os dados revelam que as respostas do banco de dados simulado produziram resultados bastante satisfatórios para equalizar os itens da prova Belém na escala SAEB, pois as médias das proficiências estimadas para os alunos do 4º e 5º anos do ensino fundamental por meio da Prova Belém, Língua Portuguesa (139.3/156.5) e Matemática (145.6/166.5), ficaram próximas dos resultados divulgados pelo Sistema Paraense de Avaliação Educacional - SisPAE em 2014 para os alunos dos mesmos anos das escolas municipais de Belém, Língua Portuguesa (148.8/167.6) e Matemática (150.7/166.5), que também utiliza a escala SAEB como referência. Os resultados permitem avançar na interpretação pedagógica dos itens posicionando-os na Escala SAEB, tornando mais consistentes a orientação da formação continuada dos professores do Ciclo II e do planejamento da escola.

Palavras-chave: Equalização de testes. Teoria de Resposta ao Item. Testes Educacionais. Avaliação em larga escala.

ABSTRACT

This study aimed to equalize the Belem Test, applied to public school students hall along the lines of the Brazil Test, with the SAEB scale via Item Response Theory. A simulated database containing information from 100,000 students was used for the reference group and the database containing information about 11,271 students from the municipal network of Belem from Cycle II 1° and 2° Years (equivalent to the 4th and 5th years of elementary school) who answered the Belem Test in December / 2014 of Portuguese Language and Mathematics. The results showed that the answers of the simulated database produced quite satisfactory results to equalize the items of the Belem test in the SAEB scale, since the averages of the proficiencies estimated for the students of the 4th and 5th year of elementary school through the Belem Portuguese Language Test (139.3 / 156.5) and Mathematics (145.6 / 166.5), were close to the results published by the Paraense System of Educational Evaluation - SisPAE in 2014 for students of the same years of the municipal schools of Belem, Portuguese Language (148.8 / 167.6) and Mathematics (150.7 / 166.5), which also uses The SAEB scale as a reference.

Keywords: Equated Scores. Item response theory. Educational Testing. Large Scale Testing.

LISTA DE GRÁFICOS

Gráfico 01 - Situações quanto ao nº de grupos e de tipos de provas	42
Gráfico 02 - Número de Itens utilizados em Língua Portuguesa	51
Gráfico 03 - Número de Itens utilizados em Matemática.....	52
Gráfico 04 - Curva Característica do Item (CCI)	53
Gráfico 05 - Grupos utilizados para equalização	57
Gráfico 06 - Relação dos parâmetros na amostra 100mil (LP)	64
Gráfico 07 - Distribuição das habilidades em Língua Portuguesa	67
Gráfico 08 - Boxplot da proficiência em Língua Portuguesa.....	68
Gráfico 09 - CCI e CII do item LP54	70
Gráfico 10 - CCI e CII do item LP98 de maior informação.....	71
Gráfico 11 - LP Curva de Informação Total (CIT) do Teste.....	72
Gráfico 12 - Alunos (%) nos níveis por categorias de desempenho.....	73
Gráfico 13 - Relação dos parâmetros a, b, c na amostra 100mil - MT..	82
Gráfico 14 - Distribuição das habilidades em Matemática.....	84
Gráfico 15 - Boxplot da proficiência em Matemática	85
Gráfico 16 - CCI e CII do item mais difícil (MT40).....	87
Gráfico 17 - CCI e CII do item MT75 de maior informação	88
Gráfico 18 - MT Curva de Informação Total (CIT) do teste.....	89
Gráfico 19 - Alunos por nível e categoria de desempenho - MT	92

LISTA DE TABELAS

Tabela 2.1 - Critérios para o índice de Facilidade (IF).....	33
Tabela 2.2 - Design dos blocos de itens na Prova Brasil 2007.....	40
Tabela 3.1 - Distribuição dos blocos na Prova Belém 2014.....	49
Tabela 3.2 - Blocos de itens na Prova Belém 2014.....	49
Tabela 3.3 - Blocos da Prova Belém (Ciclo II 2º ano - LP)	50
Tabela 3.4 - Blocos da Prova Belém (Ciclo II 1º ano - LP)	50
Tabela 3.5 - Blocos da Prova Belém (Ciclo II 2º ano - Matemática)	51
Tabela 3.6 - Blocos do Ciclo II 1º ano - Matemática	52
Tabela 3.7 - Ilustração de dados simulados.....	56
Tabela 3.8 - Número de estudantes na Prova Belém em dez/2014	56
Tabela 4.01 - LP Comparação dos dados simulados - parâmetro “a” ...	63
Tabela 4.02 - LP comparação dos dados simulados - parâmetro “b”	63
Tabela 4.03 - LP comparação dos dados simulados - parâmetro “c”	65
Tabela 4.04 - LP amostra de 100mil (Escala -1,1) grupo de 5º ano	66
Tabela 4.05 - Categorias de desempenho em Língua Portuguesa	66
Tabela 4.06 - Estatísticas da Proficiência em Língua Portuguesa.....	69
Tabela 4.07 - Itens que chamaram a atenção em Língua Portuguesa ...	71
Tabela 4.08 - Língua Portuguesa: níveis de desempenho 4º e 5º ano ...	73
Tabela 4.09 - Posicionamento dos itens de LP - 5º ano	74
Tabela 4.10 - Posicionamento dos itens de LP - 4º ano	76
Tabela 4.11 - Percentual de acertos do item LP32.....	78
Tabela 4.12 - Percentual de acertos do item LP34.....	79
Tabela 4.13 - MT comparação dos dados simulados- parâmetro “a”	80
Tabela 4.14 - MT comparação dos dados simulados - parâmetro “b”... ..	81
Tabela 4.15 - MT Comparação dos dados simulados - parâmetro “c” ..	81
Tabela 4.16 - MT amostra de 100mil (Escala -1,1) grupo de 5º ano.....	83
Tabela 4.17 - Categorias de desempenho em Matemática	83
Tabela 4.18 - Estatísticas da Proficiência em Matemática	85
Tabela 4.19 - Itens que chamaram a atenção em Matemática.	88
Tabela 4.20 - Matemática: escala de níveis 4º e 5º anos	89
Tabela 4.21 - Posicionamento dos itens de Matemática - 4º ano	90
Tabela 4.22 - Posicionamento dos itens de Matemática - 5º ano	93
Tabela 4.23 - Percentual de respostas às alternativas do item MT02....	94
Tabela 4.24 - Percentual de respostas às alternativas do item MT04....	95

LISTA DE ABREVIATURAS E SIGLAS

ALFAMAT - Programa de Alfabetização, Matemática, Leitura e Escrita.
BIB - Blocos Incompletos Balanceados.
CINBESA - Companhia de Informática de Belém
CCI - Curva Característica do Item
CII - Curva de Informação do Item
CIT - Curva de Informação Total
EF - Ensino Fundamental
IDEB - Índice de desenvolvimento da Educação Básica
ID - Índice de Discriminação
IF - Índice de Facilidade
JCSEE - Joint Committee on Standards for Educational Evaluation
LDB - Lei de Diretrizes e Bases
LP - Língua Portuguesa
INEP - Instituto Nacional de Estudos e Pesquisa Educacional "Anísio Teixeira"
MEC - Ministério da Educação
ML3 - Modelo Logístico de três parâmetros
MRA - Matriz de Referência da Avaliação
MT - Matemática
NAEP - National Assessment of Educational Progress
NIED - Núcleo de Informática Educativa da SEMEC/Belém
PDE - Plano de Desenvolvimento da Educação
PDET - Plano Decenal de Educação para Todos
PNUD - Programa das Nações Unidas para o Desenvolvimento
PPMGA - Programa de Pós-graduação em Métodos e Gestão da Avaliação
RME - Rede Municipal de Ensino
SAEB - Sistema de Avaliação da Educação Básica
SEMEC - Secretaria Municipal de Educação de Belém
SisPAE - Sistema Paraense de Avaliação Educacional
TCT - Teoria Clássica dos Testes
TRI - Teoria de Resposta ao Item
UFSC - Universidade Federal de Santa Catarina
UNESCO - Organização das Nações Unidas para a Educação, a Ciência e a Cultura
UNICEF - Fundo das Nações Unidas para a Infância

SUMÁRIO

INTRODUÇÃO	21
1.1 CONTEXTUALIZAÇÃO	21
1.1.1 Apuração e divulgação dos testes	23
1.2 DEFINIÇÃO DO PROBLEMA	25
1.3 OBJETIVOS	26
1.3.1 Objetivo geral	26
1.3.2 Objetivos específicos	26
1.4 ESTRUTURA DO TRABALHO	26
2 REFERENCIAL TEÓRICO	29
2.1 AVALIAÇÕES EDUCACIONAIS NO BRASIL	29
2.2 TEORIA CLÁSSICA E TEORIA DE RESPOSTA AO ITEM	33
2.3 ESCALA SAEB	37
2.4 EQUALIZAÇÃO	41
3 METODOLOGIA	47
3.1 CLASSIFICAÇÃO DA PESQUISA	47
3.2 PROVA APLICADA AOS ALUNOS DO CICLO II	47
3.2.1 Prova Belém de Língua Portuguesa	49
3.2.2 Prova Belém de Matemática	51
3.3 MODELO DA TEORIA DE RESPOSTA AO ITEM	52
3.3.1 Geração das respostas do grupo de referência	55
3.3.2 Método de Equalização utilizado	56
3.4 METODOLOGIA DE POSICIONAMENTO DOS ITENS	58
4 RESULTADOS	61
4.1 BASE CLONE DE LÍNGUA PORTUGUESA	62
4.2 ESCALA DE LÍNGUA PORTUGUESA	66
4.3 INTERPRETAÇÃO DOS ITENS DE LÍNGUA PORTUGUESA .	72
4.4 BASE CLONE DE MATEMÁTICA	80
4.5 ESCALA DE MATEMÁTICA	83
4.6 INTERPRETAÇÃO DOS ITENS DE MATEMÁTICA.....	89
CONSIDERAÇÕES FINAIS	97
REFERÊNCIAS	99
APÊNDICE A - ITENS NOS BLOCOS DE LÍNGUA PORTUGUESA	105
APÊNDICE B - ITENS NOS BLOCOS DE MATEMÁTICA	109

APÊNDICE C - BISSERIAIS EM LÍNGUA PORTUGUESA.....	115
APÊNDICE D - ESTIMATIVAS DOS PARÂMETROS - LP	117
APÊNDICE E - VALORES BISSERIAIS EM MATEMÁTICA	119
APÊNDICE F - ESTIMATIVAS DOS PARÂMETROS - MT	123
ANEXO I - MATRIZES DE LÍNGUA PORTUGUESA- MRA	125
ANEXO II - MATRIZES DE MATEMÁTICA - MRA	127

INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Os testes aplicados no programa de formação continuada de professores do Ciclo II, alfabetização, matemática, leitura e escrita (ALFAMAT), surgiram a partir das avaliações em larga escala, desenvolvidas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP/MEC) - Prova Brasil - realizada a cada dois anos no Brasil. O Ministério da Educação (MEC) preconiza que as avaliações em larga escala, destinadas ao Ensino Fundamental e Ensino Médio, devem contribuir para a melhoria da qualidade do ensino, redução das desigualdades e democratização da gestão do ensino público. Além disso, espera-se que induzam ao desenvolvimento de uma cultura avaliativa que estimule o controle social sobre os processos e resultados de ensino.

O processo de realização das oficinas e testes do ALFAMAT teve início, em 2009, após a realização pela Secretaria Municipal de Educação de Belém - SEMEC de um evento de apresentação dos dados do IDEB de Belém a todos os professores da rede municipal de ensino. Na ocasião foi divulgada a metodologia e o desempenho obtido pelos alunos das escolas municipais na Prova Brasil. Dentre os encaminhamentos definidos pela SEMEC, o Núcleo de Informática Educativa (NIED) ficou responsável pelo processo de formação dos professores cujos estudantes participariam da Prova Brasil.

O Ensino Fundamental (EF) de 09 anos, previsto para crianças de 06 aos 14 anos, está organizado na Rede Municipal de Ensino (RME) de Belém em ciclos de formação da seguinte forma:

- Ciclo I (três anos, equivalentes aos 1º, 2º e 3º anos do Ensino Fundamental de nove (09) anos, ou Alfabetização, 1ª e 2ª séries Ensino Fundamental de oito (08) anos);
- Ciclo II (dois anos, equivalente aos 4º e 5º anos do Ensino Fundamental de nove anos, ou 3ª e 4ª séries Ensino Fundamental de oito anos);
- Ciclo III (dois anos, equivalente aos 6º e 7º anos do Ensino Fundamental de nove anos, ou 5ª e 6ª séries Ensino Fundamental de oito anos);
- Ciclo IV (dois anos, equivalente aos 8º e 9º anos do Ensino Fundamental de nove anos, ou 7ª e 8ª séries Ensino Fundamental de oito anos);

Para alcançar sua nova missão, o NIED optou por realizar oficinas metodológicas, seguidas por aplicação dos testes, inicialmente bimestrais aos alunos, para levantar os seus desempenhos e dificuldades em Língua Portuguesa e Matemática, assim como familiarizá-los com a marcação de folha-resposta similar ao proposto pela Prova Brasil. Além disso, os testes procuravam fornecer indicadores para replanejar tanto as oficinas propostas pelo NIED, como as aulas realizadas pelos professores nas turmas do Ciclo II.

Assim, a Secretaria Municipal de Educação de Belém/SEMEC, por meio do NIED iniciou a aplicação de avaliações em larga escala aos alunos do Ciclo II, dentro do programa de formação continuada de professores - ALFAMAT, com os seguintes objetivos: (a) qualificar o professor em sua prática pedagógica nas disciplinas de Língua Portuguesa e Matemática; (b) propiciar a aprendizagem dos alunos do Ciclo II, especialmente em Língua Portuguesa e Matemática, a partir do uso das diversas ferramentas disponíveis no espaço escolar; (c) contribuir para a melhoria do desempenho dos alunos do Ciclo II 2º ano nas áreas da Língua Portuguesa e Matemática, na perspectiva de refletir positivamente nos resultados da Prova Brasil; (d) integrar as ações educativas dos professores de sala de aula e de informática da RME com base nos descritores, de Língua Portuguesa e Matemática, avaliados na Prova Brasil; (e) elaborar e orientar a aplicação de testes, a fim de que os alunos se familiarizem com as questões de múltipla escolha e com a folha de respostas que caracterizam a Prova Brasil; (f) diagnosticar, a partir dos resultados dos testes aplicados aos alunos, os avanços e dificuldades com vista a rever o planejamento pedagógico adequando-o aos desafios apontados.

No início do programa ALFAMAT os professores participaram de oficinas elaboradas de Língua Portuguesa e Matemática, a partir dos descritores utilizados na Prova Brasil. Após uma série de oficinas coordenadas pelos formadores do NIED, os professores da RME foram orientados a vivenciar a oficina com os alunos. Em seguida aplicou-se teste de múltipla escolha no mesmo formato da Prova Brasil aos alunos do Ciclo II das escolas para verificar o desempenho escolar. Atualmente, o programa ALFAMAT continua com as mesmas diretrizes e mesmas sistemáticas de oficinas articuladas com testes.

Os resultados quantitativos dos testes focaram-se, inicialmente, em apenas dois pontos: o desempenho dos alunos no teste/disciplina e o desempenho por questão (cada questão estava relacionada a um dos descritores propostos pela Prova Brasil). Esses resultados foram apresentados pelo NIED e discutidos com diretores, coordenadores

pedagógicos e professores das escolas da RME de Belém com a finalidade de buscar estratégias para superar as dificuldades evidenciadas; e também, divulgados online para as escolas através do Sistema Integrado de Gestão Acadêmica (SIGA), da Companhia de Informática de Belém (CINBESA), constituindo um banco de dados do desempenho acadêmico. Atualmente, além das reuniões, o NIED solicita às escolas um Plano de Intervenção Pedagógica para os alunos com baixo desempenho.

1.1.1 Apuração e divulgação dos testes

O processo de aplicação e divulgação dos testes sofreu evolução ao longo dos anos. Inicialmente, em 2009, a apuração e divulgação eram realizadas de forma manual; e em 2011 passou à apuração digitalizada e os resultados divulgados online através do Sistema Integrado de Gestão Acadêmica (SIGA). De 2013 até 2016 organizaram-se as provas e os itens num banco de dados *online* e a estrutura de montagem da prova através de blocos, buscando se aproximar da metodologia da Prova Brasil.

Nos dois primeiros anos (2009-2010) no processo de apuração dos resultados dos testes aplicados no Programa ALFAMAT os próprios professores realizavam o preenchimento de mapas com base nas alternativas marcadas pelos alunos na folha-resposta. E os procedimentos de aplicação desses testes iniciavam com a elaboração das questões a partir dos descritores e modelos de itens propostos pela Prova Brasil. A reprodução dos testes foi realizada via fotocópia, e o envelopamento por turma/escola. Após a aplicação dos testes cada escola informava por e-mail ao NIED as alternativas marcadas pelos alunos numa planilha e ficavam com os cadernos de testes para comentar e corrigir com os alunos.

Esses procedimentos davam margens a vários erros, como por exemplo, de digitação e transcrição, além de envolver várias pessoas até a compatibilização final ocasionando atrasos no resultado. Para reduzir os erros humanos e buscar mais agilidade e rapidez no resultado, a SEMEC com o assessoramento da Companhia de Informática de Belém (CINBESA), discutiu uma forma de digitalizar e automatizar a correção dos testes, chegando à proposta de uma solução composta do seguinte kit: scanner para digitalizar as folhas respostas; Software específico para elaboração das folhas respostas, leitura, verificação e exportação das alternativas marcadas e computadores para operacionalizar o processo de apuração.

A partir de 2011, após aquisição e treinamento do *software* específico de leitura digital, o processo de apuração ganhou novo formato com cartões personalizados dos alunos, reuniões para orientação de preenchimento, e a ampliação do Sistema Integrado de Gestão Acadêmica (SIGA), desenvolvida pela CINBESA, com o módulo “programas pedagógicos” de acompanhamento do desempenho dos alunos resultantes da aplicação dos testes e para organização da frequência dos professores nas oficinas que acabou sendo descontinuado.

No começo os relatórios apresentavam apenas a média de acertos dos alunos no teste. Porém, logo se percebeu que esta medida fornecia poucos subsídios para orientar a análise dos resultados dos testes, por isso, inseriu uma medida de dispersão (desvio padrão) para verificar como os acertos dos alunos nos testes se distanciavam da média.

Por meio da média e do desvio padrão foi feita uma classificação que se propõe a auxiliar os professores e a escola no mapeamento da turma a partir do resultado dos testes, auxiliando-os na organização do planejamento didático, como também a mobilizar a gestão da escola na busca de apoio para a superação dos limites e das dificuldades dos alunos no processo de ensino e aprendizagem. Porém, a utilização apenas da média e do desvio padrão ainda apresentava muitos limites na utilização pedagógica dos resultados.

Em 2013, assumiu uma nova gestão municipal, que deu continuidade à proposta. A experiência estava iniciando e vários desafios precisavam ser superados na busca da melhoria da qualidade do processo, entre eles, utilizar metodologias que não dependessem do grupo dos respondentes e permitissem a comparabilidade dos resultados ao longo do tempo. Cientes de que as avaliações, por si só, não solucionam os problemas detectados, mas oferecem possibilidades de reflexões sobre os métodos e os processos de ensino e aprendizagem na perspectiva de soluções.

Os dados revelados a partir dos testes estavam longe de apresentar um quadro avaliativo completo que abrangesse os aspectos cognitivos do aluno, como desejaria os educadores. Vianna (2003) lembra que muitas vezes importantes avaliações têm o seu impacto bastante restrito, ou até mesmo inexistente, em relação ao sistema e a suas escolas, por várias razões. Entre elas, destaca-se a dificuldade dos relatórios chegarem às mãos dos professores para fins de análise, discussão e estabelecimento de linhas de ação; linguagem demasiadamente técnica dificultando o entendimento dos relatórios pelo professor; resultados apresentados em termos globais de rede e escola

dificultando ações pontuais do professor em relação ao aluno ou turma; certa resistência por parte de professores e alunos com certa dose de incredulidade em relação à generalização das conclusões; falta de motivação dos estudantes para os testes, geralmente, monótonos, cansativos, levando-os a dar respostas mecânicas que levem a conclusões equivocadas.

As causas apontadas por Vianna (2003) instigaram a busca de estratégias para intensificar a utilização dos resultados disponível no SIGA pela gestão das escolas, pelos professores, seja nas reuniões de conselho de ciclo e nas reuniões de pais; seja no planejamento dos recursos que chegam à escola: PDE, Mais Educação etc. Além disso, procurar correlações entre as questões e o desempenho dos alunos que ajudassem o professor a entender os problemas de aprendizagens.

O primeiro passo se deu no estudo da Teoria Clássica dos Testes (TCT) e da metodologia da Teoria de Resposta ao Item (TRI) para dar mais consistência teórica na elaboração das questões e na composição dos testes. Principalmente, com vista à constituição de um banco de questões que sejam reutilizadas e validadas durante a aplicação dos testes.

O próximo passo está no sentido de articular a experiência municipal de Belém com a política de avaliação promovida pelo governo federal, através do Sistema de Avaliação da Educação Básica (SAEB), isso passa pela possibilidade de comparabilidade entre a Prova Brasil e a Prova Belém, que se torna possível por meio dessa proposta de equalização da Prova Belém na escala SAEB.

As avaliações de larga escala cresceram nos últimos anos no Brasil, e segundo Freitas et al. (2014), as avaliações de rede de ensino seriam mais eficazes se planejadas e conduzidas no nível municipal, tornando viável o envolvimento de professores, especialistas das secretarias e das escolas, na montagem do sistema de avaliação.

1.2 DEFINIÇÃO DO PROBLEMA

O presente estudo está inserido na linha de pesquisa Métodos Quantitativos em Avaliação Educacional do Programa de Pós-graduação em Métodos e Gestão da Avaliação (PPGMGA) da Universidade Federal de Santa Catarina (UFSC), focado nos métodos estatísticos, especialmente na avaliação educacional de grande escala.

O problema pesquisado é o fato da Prova Belém não se encontrar numa escala interpretada, nem numa métrica que permita comparações entre resultados dos vários anos da prova, nem com outras avaliações

nacionais e estaduais. Espera-se que a equalização da Prova Belém na escala SAEB possibilite a solução dos problemas de comparabilidade, e forneça uma escala interpretada para leitura dos resultados, tornando-os mais consistentes.

O problema nesse estudo pode ser apresentado por meio da seguinte pergunta: É possível que os resultados da Prova Belém sejam comparáveis aos resultados da Prova Brasil/Escala SAEB?

1.3 OBJETIVOS

A seguir apresenta-se o objetivo geral do presente trabalho e os objetivos específicos.

1.3.1 Objetivo geral

- Realizar estudo sobre o processo de equalização da Prova Belém com a Prova Brasil/Escala SAEB.

1.3.2 Objetivos específicos

- Simular respostas que reproduzam os parâmetros dos itens da prova Brasil;
- Analisar os itens das provas pelos seus índices psicométricos com vistas a organizá-los num banco de itens;
- Calibrar os itens da prova Belém de 2014 do 4º e 5º anos EF na escala SAEB;
- Posicionar os itens da prova Belém do 4º e 5º anos na escala SAEB;
- Gerar a proficiência dos alunos da prova Belém de 2014 do 4º e 5º anos E. F. na escala SAEB.

1.4 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em cinco capítulos, levando-se em consideração as atividades necessárias para alcançar os objetivos propostos na seção 1.3. configura-se como uma pesquisa aplicada numa abordagem quantitativa, desenvolvida e redigida de acordo com as normas acadêmicas. Sendo que, no primeiro capítulo, realizou-se introdução ao assunto, contextualização do tema, problemática de pesquisa, e expõem-se os objetivos.

No capítulo dois apresenta-se a fundamentação teórica sobre os temas mais relevantes que fornecem suporte ao trabalho: Avaliações Educacionais no Brasil, Teoria da Resposta ao Item (TRI) e metodologias de equalização de testes educacionais.

Detalha-se, no capítulo três, a prova utilizada como instrumento de medida e os procedimentos metodológicos utilizados, os aspectos considerados e a organização dos itens.

No capítulo quatro procede-se a apresentação e análise dos resultados da aplicação da TRI no conjunto de itens, a interpretação pedagógica da escala e a discussão sobre os resultados.

No quinto capítulo apresentam-se as conclusões e recomendações para trabalhos futuros, ao que seguem as referências.

2 REFERENCIAL TEÓRICO

Nesse capítulo, após uma panorâmica das avaliações externas nos estados e municípios brasileiros, apresentam-se os conceitos básicos da Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI) que embasam o estudo da medida psicométrica e a sua aplicação no processo de equalização de testes educacionais para compor o referencial teórico-prático do trabalho realizado de equalização na escala SAEB.

2.1 AVALIAÇÕES EDUCACIONAIS NO BRASIL

Nas últimas décadas, o Brasil experimentou uma crescente expansão de dados sobre a educação brasileira por meio dos sistemas de avaliação de desempenho educacional. Essa expansão começa a se organizar na esfera federal em 1990 a partir do compromisso internacional assumido na Conferência Mundial sobre Educação para Todos, realizada na Tailândia. Em seguida, em 1993 o Brasil - convocado por entidades internacionais como PNUD, UNICEF, UNESCO, Banco Mundial - elaborou o Plano Decenal de Educação para Todos (PDET), que entre outros pontos ressaltava a importância de uma avaliação nacional como forma de melhorar a qualidade do ensino, o que em 1996 foi ratificado na Lei de Diretrizes e Bases da educação nacional - LDB. Ribeiro (2002), Coelho (2008), Horta (2007).

Estudo de Brooke et al. (2011) apresenta 03 elementos que contribuíram para a expansão da avaliação externa dos sistemas educacionais: (a) a disseminação de um modelo calcado na avaliação educacional e voltado para a melhoria da qualidade; (b) a incorporação pelos governos de metodologias de gestão baseadas em critérios de eficiência e de planejamento estratégico e voltadas para a melhoria dos resultados dos serviços públicos; e (c) o incentivo criado pela inauguração de um indicador nacional de desenvolvimento educacional e pela liderança do MEC ao fixar o foco nos resultados e no uso da avaliação educacional entre as diretrizes principais do Plano de Metas Compromisso Todos pela Educação.

Atualmente vários estados brasileiros desenvolvem seus próprios métodos de avaliação do ensino, o levantamento realizado por Brooke et al. (2011) apontou a existência de 22 estados com sistemas de avaliação próprios implantados, descontinuados ou em fase de implantação. Atualmente, dentre os estados podem-se citar Minas Gerais, Rio Grande do Sul, Bahia, Ceará, Paraná, Pernambuco, São Paulo, e Pará que

realizam avaliações censitárias de suas escolas, visando, principalmente, ao direcionamento de políticas públicas, no sentido de melhorar a qualidade de suas redes de ensino e a melhoria da prática docente. Esses estados deram origem aos seguintes sistemas de avaliação no âmbito das Secretarias de Estado de Educação: (Gatti, 2009; Sousa & Oliveira, 2010; Oliveira, 2011; Lopes, 2007).

- SPAECE - Sistema Permanente de Avaliação da Educação Básica do Ceará. Criado em 1992 e ampliado em 1996;
- SARESP - Sistema de Avaliação de rendimento escolar do estado de São Paulo, em 1998;
- SABE - Sistema de avaliação baiano de educação; Projeto de avaliação externa do ensino fundamental no estado da Bahia, em 1999; é aperfeiçoado em 2007;
- SIMAVE - Sistema mineiro de Avaliação da Educação pública, em 2000;
- SAEPE - Sistema de avaliação educacional de Pernambuco, em 2000;
- SAERS - Sistema de avaliação do rendimento escolar do Rio Grande do Sul, em 2007;
- SAEP - Sistema de Avaliação da Educação Básica do Paraná, em 2012.
- SisPAE - Sistema Paraense de Avaliação Educacional, em 2013.

Essa expansão alcançou também os municípios, Bauer et al. (2015), realizaram pesquisa que atingiu 4.309 municípios (77,4% do total dos municípios brasileiros) em que mapeou e caracterizou iniciativas relacionadas às avaliações em larga escala em desenvolvimento nos municípios brasileiros. Essa pesquisa apontou as concepções e usos dos resultados que as gestões desses entes federados fazem das diversas avaliações e evidenciou a consolidação das diversas avaliações como um instrumento de gestão educacional nas municipalidades. Assim Bauer et al. (2015) constata que:

Desde os anos 2000, estados e municípios vem pouco a pouco assimilando o modelo de gestão da educação pública implantado pelo governo federal a partir da década de 1990, com o crescente uso dos resultados das avaliações em larga escala como principal indicador de qualidade (p. 332).

Em 2009, Belém tendo em vista a preocupação com a qualidade da educação municipal e a aprendizagem dos seus estudantes, sintetizada no Índice de Desenvolvimento da Educação Básica (IDEB), iniciou através da secretaria municipal de educação de Belém o

Programa de Alfabetização Matemática, Leitura e Escrita (ALFAMAT), articulando formação continuada de professores e provas elaboradas de acordo com a matriz de referência da Prova Brasil para os alunos do Ciclo II, chamadas a partir de 2011 de Prova Belém.

A Prova Belém se organizou a partir do Sistema de Avaliação da Educação Básica (SAEB) instituído em 1990 pelo governo federal, e reestruturado em 2005 com a componente voltada para o ensino fundamental chamada de Avaliação Nacional do Rendimento Escolar - ANRESC (Prova Brasil). O objetivo do SAEB é subsidiar, a partir dos resultados das avaliações, ações com vista à melhoria do ensino, produzindo informações para a formulação, reformulação e o monitoramento das políticas na área educacional. Essa iniciativa do governo federal tem motivado cada vez mais os governos estaduais e municipais a criar seus próprios sistemas de avaliação. Segundo Araújo & Luzio (2005):

O Sistema Nacional de Avaliação da Educação Básica (SAEB) é uma das primeiras ações brasileiras para conhecer os resultados de aprendizagem dos alunos. Constitui-se, hoje, o mais amplo instrumento de avaliação externa da qualidade do desenvolvimento de habilidades e competências dos estudantes do país e um dos mais sofisticados e amplos sistemas de avaliação em larga escala da América Latina (p. 13).

Para que as avaliações de larga escala efetivamente sirvam de subsídios para a regulação nacional, dos sistemas ou redes, e a microrregulação local, da escola e sala de aula, é necessário o envolvimento e definição de estratégias pelas três esferas do poder público - federal, estadual, municipal - “com foco no aperfeiçoamento institucional e no processo de ensino-aprendizagem nos diferentes graus e modalidades de ensino” (Oliveira, 2011).

Uma característica importante das avaliações estaduais é a preocupação de comparabilidade dos resultados do estado com os resultados do país. Para tanto, a parceria com o INEP, por meio de disponibilização de itens e bases de dados do SAEB foram imprescindíveis.

Com a criação da Prova Brasil, o INEP começou em 2005 a avaliar de forma censitária os alunos do ensino fundamental da rede pública de ensino, com o objetivo de fornecer maiores subsídios para os estados, municípios e escolas. No entanto, o que se nota é que alguns estados continuam com seus sistemas de avaliação, pois estes possuem

algumas características diferentes da Prova Brasil, e, a princípio, não querem abandonar a metodologia empregada ao longo dos anos. Por exemplo, Minas Gerais, que possui uma série histórica de avaliação nos anos de 2000, 2002, 2003, 2006, 2007 e 2008, aplica testes, em dias distintos, de Língua Portuguesa e Matemática, em todas as escolas do Estado, independentemente do número de alunos; diferente da Prova Brasil que aplica Língua Portuguesa e Matemática no mesmo caderno e só em escolas com mais de 15 alunos; Ou Rio Grande do Sul, que avalia a 5ª série/6º ano do ensino fundamental e 1º ano do ensino médio, o que difere das séries avaliadas pela Prova Brasil; ou Pará que através do SisPAE, em 2014, além do 4º, 5º, 8º e 9º anos do ensino fundamental, avaliou todo o ensino médio, fornecendo os resultados por escola, município, Unidade Regional de Educação e Região de Integração; ou Ceará que forneceu os resultados por aluno avaliado, contrapondo-se à Prova Brasil cuja menor unidade de avaliação é a escola (Silva & Soares, 2010).

As diferenças são justificativas, por parte das redes, para a elaboração de um sistema próprio de avaliação em larga escala, em vez de utilizar as avaliações nacionais já disponíveis. Gimenes (2013), em seu estudo em diferentes redes de ensino agrupou as justificativas em quatro grupos: (a) o caráter censitário e anual das avaliações externas promovidas localmente, enquanto a avaliação nacional acontece bianualmente; (b) possibilidade de detalhar e trabalhar os dados nas escolas com maior rapidez e agilidade, pois os dados das avaliações nacionais levam muito tempo para serem liberados; (c) a amplitude dos/as anos/séries avaliadas pelo modelo de avaliação adotado pelas redes de ensino, pois a avaliação nacional avalia apenas o 5º e 9º ano do ensino fundamental e o 3º ano do ensino médio; (d) a possibilidade de identificação nominal dos resultados de cada aluno, os resultados nacionais apresentam apenas os dados das escolas.

A Prova Belém procurou reduzir o tempo entre as avaliações e fornecer resultados mais rápidos para o monitoramento das políticas na área educacional do município, no entanto, ficou apenas voltada para a aprendizagem dos alunos e, ainda não possui medida relacionada aos fatores de contexto correlacionados com o desempenho escolar, procurando utilizar as fornecidas pela Prova Brasil, pois ambas contribuem para a melhoria da qualidade, equidade e eficiência do ensino (INEP, 2011).

2.2 TEORIA CLÁSSICA E TEORIA DE RESPOSTA AO ITEM

A Prova Belém, inicialmente, forneceu apenas resultados quantitativos focados em apenas dois pontos: o desempenho dos alunos no teste/disciplina e o desempenho por questão (cada questão relacionava-se a um dos descritores propostos pela Matriz de Referência da Avaliação - MRA da Prova Brasil). Esses pontos estão ligados à Teoria Clássica dos Testes (TCT), que se preocupa em explicar o resultado por meio da soma das respostas atribuídas a uma série de itens, conhecida como escore total (Pasquali, 2013), desta forma o desempenho numa questão está relacionado ao teste como um todo.

Para a adequação dos itens à população testada, a Teoria Clássica utiliza dois critérios principais para analisar os itens: índice de dificuldade e o índice de discriminação. O índice de dificuldade, também conhecido como Índice de Facilidade (IF), representa a proporção de alunos que acertou o item e pretende medir as diferenças individuais no que diz respeito ao rendimento alcançado no teste em que existam respostas do tipo certo ou errado. Quanto mais alunos acertam a um determinado item, mais fácil é esse item. Ele varia de 0 (zero) a 1 (um), zero significa que não houve acerto no item e um significa que todos os alunos acertaram o item.

Na tabela 2.1, apresenta-se a elaboração de Cerdá (1972), apud Erthal (2009), onde estabelece o modo de apresentação dos itens em um teste heterogêneo e sua interpretação.

Tabela 2.1 - Critérios para o índice de Facilidade (IF)

Classificação	Número de itens	Limite aproximado do (IF)
Muito fáceis	10%	0,75 a 0,95
Fáceis	20%	0,55 a 0,74
Normais	40%	0,45 a 0,54
Difíceis	20%	0,25 a 0,44
Muito difíceis	10%	0,05 a 0,24

Fonte: Cerdá E. (1972) in Erthal (2009).

Para análise dos itens, a Prova Belém utilizou como critério a sugestão reduzida de Condé (2002) para interpretação do IF, em que o índice médio de dificuldade dos itens fica em torno de 0.5, pois segundo Pasquali (2013), dificuldades do teste em torno desse valor produzem distribuições de escores no teste com maior variação:

- Item fácil: $IF > 0.70$;
- Item de dificuldade média: $0.30 < IF \leq 0.70$;
- Item difícil: $IF \leq 0.30$.

O Índice de Discriminação (ID) estabelece relação entre escores totais altos ou baixos no teste e as respostas corretas ou incorretas dadas a um item. Esse índice é o indicador de consistência interna. Para um teste discriminar bem o grupo de indivíduos na variável medida, é necessário que apresente um bom grau de consistência interna entre os itens e a totalidade do teste. Espera-se de uma resposta a um item discriminativo que os alunos com melhor desempenho na prova, de modo geral, acertem-no com mais frequência, e por sua vez, aqueles com menor desempenho, acertam-no com menos frequência. Pode-se calcular o ID pela correlação bisserial por pontos, pela correlação bisserial simples e pelo método dos 27%.

O SAEB utiliza a correlação bisserial, uma medida de associação entre o desempenho no item e o desempenho na prova que estima a correlação entre a variável de desempenho no teste e uma variável não observável com distribuição normal que, por hipótese, representa a habilidade que determina o acerto ou erro do item (Borgatto & Andrade, 2012). A correlação bisserial é menos influenciada pela dificuldade do item e tende a apresentar menos variação de uma situação de testagem para outra (Rodrigues, 2006).

A Prova Belém utilizou o método dos 27% para verificar o poder de discriminação. Esse método estabelece a determinação do grupo dos 27% dos respondentes que tiveram os mais baixos acertos (P_{inf}) e o grupo dos 27% dos respondentes que tiveram os mais altos acertos (P_{sup}), a diferença entre a porcentagem de acerto do grupo superior e a porcentagem de acerto do grupo inferior, é o índice de discriminação que varia de -1 a 1 (Ferreira, 2009).

Espera-se que, para cada item, os indivíduos do Grupo Superior apresentem uma proporção de acerto maior que a do Grupo Inferior. Quanto maior a diferença ($P_{sup} - P_{inf}$), maior será o potencial de discriminação do item. Ebel (1965) propõe a seguinte escala para a discriminação:

- Muito bom: $ID \geq 0.40$
- Satisfatório: $0.30 \leq ID < 0.40$
- Aceitável, mas convém revisar: $0.20 \leq ID < 0.30$.
- O item deve ser eliminado ou modificado: $ID < 0.20$

Quanto maior a discriminação, maior a capacidade do item de discriminar grupos de estudantes que construíram determinada competência e habilidade, daqueles que não as construíram. Os itens com coeficiente baixo não diferenciam o indivíduo que construiu, daquele que não construiu determinada competência e habilidade. Esse

índice é uma medida estatística capaz de identificar itens com formulação inadequada ou com erro de gabarito.

A capacidade de um item discriminar entre indivíduos é, em parte, função do seu nível de facilidade/dificuldade. Assim, ao construir um teste selecionam-se itens de acordo com o seu IF e seu ID, para Erthal (2009), na elaboração de um teste, o que se procura são itens com pequeno ou moderado grau de inter-relacionamento entre si, mas com boa correlação com o escore total, essas informações são obtidas pelo índice de discriminação (ID).

O INEP, responsável pela aplicação da Prova Brasil e a construção da escala SAEB, a partir de 1995 utiliza, além da TCT, a TRI, na elaboração e interpretação da Prova Brasil. Isso permitiu a possibilidade de comparabilidade entre os resultados das avaliações ao longo do tempo, o que levou a TRI a ser amplamente difundida como um sistema de medição que se faz pertinente em diversas áreas e saberes e, principalmente, na avaliação educacional (Andrade, Tavares & Valle, 2000; Pasquali, 2013).

A TRI se desenvolveu tendo como um dos objetivos suprir deficiências da TCT. Embora a TRI não entre em contradição com os princípios da TCT, ela traz uma nova proposta de análise centrada nos itens que supera as principais limitações da TCT, além de apresentar novos recursos tecnológicos para a avaliação. Portanto, a TRI não veio para substituir toda a análise via TCT, mas apenas parte dela, particularmente na análise dos itens e no tema da fidedignidade da medida.

A TRI compreende um conjunto de modelos matemáticos onde a probabilidade de resposta a um item é modelada como função da proficiência (habilidade) do aluno (variável latente, não observável) e de parâmetros que expressam certas propriedades dos itens, a partir da suposição que quanto maior a proficiência do aluno, maior a probabilidade de ele acertar o item (processo de resposta acumulativo). Existem modelos da TRI para processos cujas respostas não sejam acumulativas, são os chamados modelos de desdobramento, porém eles tem tido pouco ou nenhum uso em avaliações educacionais (Araujo et al., 2009).

Por se tratar de um modelo estatístico, a TRI permite calcular a probabilidade de um sujeito acertar um determinado item estimando a sua habilidade latente necessária para tal acerto (Andrade & Valle, 1998; Pasquali & Primi, 2003; Lopes & Vendramini, 2013), assim diferente da TCT, o cálculo do nível de habilidade do sujeito independe da amostra de itens utilizados, ou seja, a habilidade do sujeito é

independente do teste. Isso permite que na Prova Belém as habilidades calculadas dos alunos do Ciclo II 2º ano através dos seus escores sejam comparáveis com dos alunos do Ciclo II 1º ano através de um processo de equalização.

A TRI mudou o foco de análise da prova como um todo para a análise de cada item. O cálculo dos parâmetros dos itens (dificuldade e discriminação), que na TCT dependiam muito dos sujeitos amostrados, na TRI esse cálculo independe da amostra de sujeitos utilizada, ou seja, os parâmetros dos itens são independentes dos sujeitos. Além disso, também as proficiências dos indivíduos são invariantes, exceto pela escolha de origem e escala. Essa propriedade associada a outros procedimentos estatísticos permite comparar alunos, estimar a distribuição de proficiências e assim monitorar os progressos de um sistema educacional ao longo do tempo. A TRI constitui um modelo, em que a proficiência de cada sujeito está associado a um erro de medida, diferente da TCT que supõe erros de medida iguais para todos os respondentes, o que aparentemente é improvável (Pasquali & Primi, 2003).

Na TCT as análises e interpretações estão sempre associadas à prova como um todo e ao grupo de indivíduos respondentes, porque os resultados dependem do particular conjunto de questões e de indivíduos envolvidos na prova. Assim, a comparação entre indivíduos ou grupos de indivíduos somente é possível quando eles são submetidos às mesmas provas ou, pelo menos, ao que se denomina de provas paralelas. Desta maneira, fica muito difícil fazer comparações quando diferentes indivíduos fazem provas diferentes, como por exemplo, avaliações ao longo dos anos (Klein, 2003). Tais comparações entre as várias avaliações nas esferas governamentais podem ser potencializadas se utilizarem métodos psicométricos que se alinhem, permitindo assim uma integração dos sistemas avaliativos.

A metodologia fornecida pelas novas abordagens à construção de testes pela TRI tem complementando de forma eficaz - e em alguns casos substituído - os métodos tradicionais de construção e delineamento de testes (Urbina, 2007). A TRI muda o foco de análise da prova como um todo para a análise de cada item (Klein, 2003), levando as técnicas de análise de itens a uma transição gradual que vem alterando fundamentalmente a natureza dos testes. Esta transição se deve, em parte, à facilidade e à eficácia com que os dados de teste podem ser coletados, armazenados, analisados, recuperados e disseminados com o uso de computadores.

A TRI trouxe uma nova abordagem à questão de comparabilidade de diferentes testes ou populações, utilizando itens comuns, que será apresentada nas seções seguintes desse capítulo.

2.3 ESCALA SAEB

A associação de profissionais nos EUA/Canadá ligados a qualidade da avaliação (*Joint Committee on Standards for Educational Evaluation - JCSEE*) define avaliação como “a determinação sistemática do valor ou mérito de um objeto” (Stufflebeam, 1994). Para uma melhor compreensão da avaliação Luckesi (2000) considera que o ato de avaliar implica em dois processos articulados e indissociáveis: diagnosticar e decidir; onde o processo de diagnosticar se constitui de uma constatação e de uma qualificação do objeto da avaliação para subsidiar a tomada de decisão em função de um objetivo que se tem a alcançar.

Normalmente, o diagnóstico está apoiado numa medida que serve para estabelecer o nível de qualidade do objeto, ou sua capacidade para algo, com objetos do mesmo tipo e na mesma situação.

Desse ponto de vista, o SAEB é um sistema de avaliação que gera medidas para verificar a efetividade dos sistemas de ensino em atingir patamares aceitáveis de desempenho, refletindo melhor o desenvolvimento cognitivo dos alunos (Araújo & Luzio, 2005). Para Luckesi (2000), esses patamares de comparação da aprendizagem do aluno, devem ser estabelecidos no planejamento de ensino e sustentados em uma teoria pedagógica.

De uma maneira geral, a medida consiste na atribuição de números às propriedades das coisas segundo certas regras, de forma a garantir que os axiomas dos números (a identidade, a ordem e a aditividade), ou a maioria deles, sejam salvos. Pasquali (2013) ressalta que não se mede objetos, pessoas, ou coletividades enquanto tais; se mede os atributos (variáveis) observáveis delas.

A Escala SAEB procura medir a proficiência dos alunos da educação básica em Matemática e Língua Portuguesa, por meio de duas escalas uma para cada disciplina.

Assim, a escala de medição do SAEB é o conjunto dos possíveis valores da proficiência dos alunos da educação básica colocados num contínuo de valores ordenados correlativamente, que admite um ponto inicial e outro final. Stevens (1946) distinguiu quatro escalas ou níveis de medição: nominal, ordinal, intervalar e racional.

A escala SAEB não é nominal ou ordinal, pois na escala nominal os números atribuídos à medida são usados apenas para identificá-los, e

na escala ordinal, os números atribuídos à medida representam apenas uma ordem. Nesses níveis as escalas são chamadas de categóricas, comumente usadas para variáveis qualitativas. Nos níveis intervalar e racional são chamadas de escalas numéricas sendo mais adequadas para a medição de variáveis quantitativas (Padilla, 2007).

A escala SAEB, portanto, é uma escala numérica intervalar em que os números atribuídos à medida podem começar em qualquer posição (inclusive zero), numa ordem crescente e distante entre si igualmente, ou seja, o importante são as relações de ordem existentes entre seus pontos equidistantes e não necessariamente sua magnitude.

Na escala razão os números fornecem informações acerca da relação ou proporção entre as características medidas, sendo, por isso, a escala de mais alto nível. Esse nível de medida não é usado em psicologia ou educação, pois até o momento não se conseguiu demonstrar a existência de um zero absoluto nos fenômenos psicológicos (total ausência de uma característica ou atributo).

No SAEB, a proficiência medida através da TRI pode assumir teoricamente qualquer valor real entre $-\infty$ e $+\infty$, não possuindo um zero absoluto. Entretanto, é necessário estabelecer uma origem e uma unidade de variação para a definição da escala. Esses valores foram arbitrados de modo a representar, respectivamente, o valor médio (250) e o desvio-padrão (50) para a distribuição das proficiências dos alunos do grupo de referência, ficando o SAEB numa escala (250;50), sendo que geralmente os resultados de desempenho dos alunos variam, na prática, de 100 a 400 (Klein, 2003). Normalmente devido a facilidades computacionais, utiliza-se a escala com média igual a 0 e desvio-padrão igual a 1, que é representada por escala (0;1) para a estimativa dos parâmetros e das proficiências dos alunos, realizando, quando necessário, a transformação para a escala (250;50).

A escala SAEB, para resolver problemas de indeterminação de escala, definiu o grupo dos alunos da 8ª série de 1997, como sendo a população referência, ou seja, com desempenho médio 250, então todas demais populações anteriores e futuras são posicionadas com relação a essa referência.

Como o número em uma escala está sempre associado a uma interpretação, na escala utilizada pelo SAEB cada intervalo numérico representa um nível de desempenho de grupos de alunos e vem acompanhado de uma interpretação das competências e habilidades que eles já construíram no seu processo de desenvolvimento. A criação de uma interpretação possibilita a verificação dos conteúdos que os alunos das séries avaliadas dominam, dos conteúdos onde há falhas e

acompanhar a “evolução do conhecimento” de uma série para outra etc. (Santos, 2008; Andrade, Tavares & Valle, 2000).

A escala é caracterizada pelos itens, logo a qualidade da escala depende da qualidade dos itens. Basicamente, será uma adequada elaboração dos itens, associado ao um bom planejamento e uma boa execução das Provas que definirão a qualidade da escala. Os itens são elaborados a partir de unidades denominadas “descritores”, partes menores dos conteúdos da matriz de referência de disciplina, que especificam o que será medido. No SAEB a matriz de referência da Avaliação (anexo I e II) foram montadas para as disciplinas de Língua Portuguesa, com 15 descritores organizados em 06 tópicos; e Matemática, com 28 descritores organizados em 04 temas, para o 5º ano do ensino fundamental. Segundo o INEP (2011):

A Matriz de Referência é o referencial curricular do que será avaliado em cada disciplina e série, informando as competências e habilidades esperadas dos alunos (p.17).

A partir das respostas dadas aos itens nos testes, verifica-se o que os alunos sabem e conseguem fazer com os conhecimentos adquiridos (Rodrigues, 2006).

Os itens são aplicados a um grupo de alunos com as mesmas características e após a análise dos seus índices psicométricos e pedagógicos são escolhidos para compor uma prova que assegure uma ampla e representativa cobertura do conteúdo da avaliação expressa na MRA. Contudo, nem sempre é viável ou desejável que todos os itens escolhidos para representar todos os descritores da MRA no teste sejam administrados a todos os respondentes. Assim, para permitir a aplicação de uma grande quantidade de itens, o SAEB adotou, a partir de 1995, uma metodologia baseada na amostragem matricial de itens, que utiliza o esquema de montagem e aplicação de testes por Blocos Incompletos Balanceados (BIB).

O BIB é delineamento usado pelo sistema de avaliação norte-americano - *National Assessment of Educational Progress (NAEP)* - onde são montados, primeiramente, blocos de itens, e a partir da combinação desses blocos de itens são montados cadernos de testes. Em essência, isso significa que um conjunto completo de itens é dividido em um número menor de blocos. Os blocos são, então, designados para os cadernos, de modo que cada bloco seja emparelhado com outro bloco para formar um caderno.

Essa distribuição assegura que sejam superados fatores intervenientes na resposta aos itens, como o cansaço do aluno ou a falta

de tempo para responder às questões que aparecem no fim da prova. Além disso, permite a aplicação de um maior número de itens, pois seria inviável a aplicação de todos os itens a um único aluno. Normalmente, se utiliza, ainda, itens comuns entre os anos e as séries investigadas para viabilizar o processo de equalização.

Com o esquema amostral e o delineamento BIB adotado no SAEB, consegue-se uma aplicação em que os alunos que respondem a um determinado caderno apresentem, proporcionalmente, características semelhantes aos grupos que responderam aos outros cadernos, visto que a alocação dos cadernos aos alunos é aleatória. Em outras palavras, os grupos de alunos que respondem a cada um dos cadernos de teste do SAEB são equivalentes (Conde & Laros, 2007).

Na Tabela 2.2 é apresentada a distribuição do número de itens, blocos e cadernos para a edição de 2007 da Prova Brasil para a 4ª série e 8ª série.

Tabela 2.2 - Design dos blocos de itens na Prova Brasil 2007

Série	Nº de bloco	Itens / bloco	Itens / caderno	Nº de caderno	Total de item
4 EF	7 LP e 7 MT	11	44	21	77 LP e 77 MT
8 EF	7 LP e 7 MT	13	52	21	91 LP e 91 MT

Fonte: Silva & Soares (2010).

O SAEB utiliza como instrumento de coleta de dados, a prova ou teste procurando avaliar as habilidades dos alunos em Matemática e Língua Portuguesa, os chamados testes de habilidade, onde as respostas são avaliadas como certas ou erradas e cujos respondentes recebem escores de acerto (1) ou erro (0). Para Luckesi (2000), são necessários cuidados para que os instrumentos sejam adequados:

- a) ao tipo de conduta e de habilidade a ser avaliado;
- b) aos conteúdos essenciais planejados e, de fato, realizados no processo de ensino;
- c) na linguagem, na clareza e na precisão da comunicação;
- d) ao processo de aprendizagem do aluno.

Testes e avaliações não são sinônimos. Os testes são uma das ferramentas usadas no processo de coleta de informações da avaliação. A testagem é uma técnica que produz resultados mais eficientes, principalmente acerca do domínio cognitivo, apesar de ser também usado no domínio afetivo e psicomotor. Além da testagem, existem outras duas técnicas de coleta de informações, de acordo com Mediano (1976): observação e inquirição. No SAEB utiliza-se questionário para

levantar os dados para os fatores associados e a testagem para estimar a proficiência dos alunos.

Os testes do SAEB são do tipo padronizado, aqueles instrumentos construídos por especialistas que se utilizam itens previamente experimentados, e a aplicação deve ser realizada de modo rigoroso para que não haja interferência de variáveis no processo, com normas de aplicação, de correção e de interpretação. Portanto, as provas do SAEB têm características diferentes dos testes não padronizados, aqueles instrumentos de que o professor se utiliza para medir a aprendizagem dos seus alunos.

Segundo Valle (2001), uma boa escala de habilidade é fruto da aplicação de um número razoável de itens de boa qualidade (com altos níveis de discriminação, diferentes níveis de dificuldades, etc.), em um número suficiente de indivíduos com os mais variados níveis de habilidade caracterizados por itens que possibilitam a interpretação dos pontos da escala. Essas características por estarem presentes na escala SAEB tornaram-na referência para os sistemas de avaliação do Brasil.

2.4 EQUALIZAÇÃO

No dicionário Aurélio XXI, equalizar significa uniformizar, igualar; no contexto de testes de larga escala, o conceito de equalização está relacionado com equiparar, tornar comparável, colocar os parâmetros dos itens provenientes de testes diferentes e traços latentes de respondentes de diferentes grupos na mesma métrica, tornando os itens e os respondentes comparáveis (Andrade, Tavares & Valle, 2000).

Os processos de comparabilidade de resultados entre avaliações (*linking*) em função da precisão da comparação que se deseja obter podem ser subdividida em quatro tipos em ordem decrescente de robustez: Equalização, Calibração, Projeção e Moderação (Silva & Soares, 2010). Nesse trabalho equalizar significa colocar parâmetros de itens vindos de provas distintas ou habilidades de respondentes de diferentes grupos, na mesma métrica, isto é, numa escala comum, tornando os itens e/ou as habilidades comparáveis.

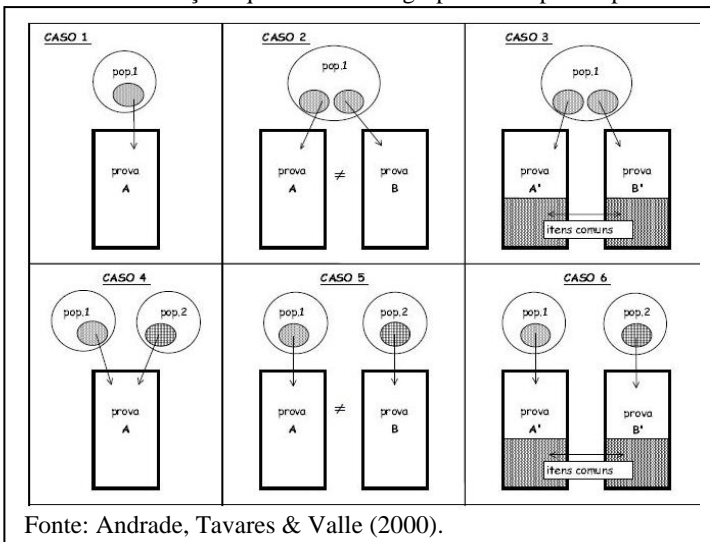
Assim, os métodos de equalização de testes são utilizados para colocar as estimativas dos parâmetros dos itens e as habilidades dos indivíduos, obtidos pela TRI para um teste, comparáveis às obtidas para outro teste, por meio de uma escala única de habilidades para todos os examinados de diversas séries que tenham respondido a testes distintos e permitindo acompanhar a evolução da aprendizagem dos alunos ao longo dos anos.

Existem dois tipos de equalização: (a) via população, quando um único grupo de respondentes é submetido a testes diferentes, ou seja, sem itens comuns; (b) via itens, quando grupos diferentes respondem testes diferentes com itens comuns entre os testes. O segundo tipo de equalização pode ser realizado de dois modos: *a posteriori* e simultaneamente, através da utilização de modelos de grupos múltiplos (Araújo, Andrade & Bortolotti, 2009).

A equalização é uma importante ferramenta quando o objetivo é comparar o desempenho de indivíduos pertencentes ou não à mesma população ao longo do tempo, pois permite a comparação entre itens e/ou níveis de habilidades, possibilitando que os parâmetros de itens e/ou níveis de habilidades dos respondentes, de grupos diferentes, possam ser analisados e compreendidos em uma mesma métrica, ou seja, numa escala comum. Andrade, Tavares & Valle (2000), lista os seis casos possíveis, quanto ao número de grupos e de tipos de prova envolvidos. Esses casos estão esquematizados no gráfico 01.

Um método natural de equalização é estimar as habilidades e parâmetros dos itens simultaneamente através de algum modelo de probabilidade da TRI, dado que se têm as respostas dos alunos nos diferentes testes com itens comuns. Esse procedimento tem sido denominado de calibração simultânea, e é, provavelmente, o método mais extensivamente aplicado (Vital & Soares, 2002).

Gráfico 01 - Situações quanto ao nº de grupos e de tipos de provas



De acordo com Kim & Cohen (1998), as principais limitações do método de calibração simultânea estão associadas ao fato de serem observados problemas quando a localização e a variabilidade das distribuições de habilidade dos grupos apresentam grandes diferenças. Além disso, em muitas aplicações, uma destas limitações está associada ao fato do número de itens para os grupos exceder a capacidade de softwares disponíveis.

Outro método de equalização comum, conforme Kolen & Brennan (1995), também chamado de equalização *a posteriori* utiliza estatísticas (média, média geométrica, desvio padrão, e outras medidas) para estabelecer uma transformação linear entre as escalas criadas separadamente.

O processo de equalização do SAEB considera que grupos são submetidos a várias provas diferentes, mas que têm alguns itens comuns ligando os grupos, sendo neste caso, a equalização realizada via itens comuns através de calibração simultânea, pois os parâmetros dos itens são estimados conjuntamente para todos os itens (que são todos de múltipla escolha) de todas as séries de uma mesma disciplina, permitindo a comparação entre anos para a mesma disciplina e a obtenção dos parâmetros dos itens numa mesma escala. Assim, a equalização entre séries no SAEB utiliza a calibração conjunta dos itens incluindo-se na estimativa as respostas individuais de todos os alunos da amostra e o grupo de referência (alunos da 8ª série de 1997, equivalente atualmente ao 9º ano EF) com o uso da extensão da TRI a grupos múltiplos não equivalentes, onde as diferentes séries seriam os grupos. Dessa maneira, tem-se uma estimativa única dos parâmetros, não importando a série e evita-se a propagação de erros na equalização quando se estima os parâmetros série a série (Klein, 2003).

A equalização dos resultados da avaliação do rendimento da escola pública do Estado do Rio Grande do Norte com o SAEB-95, realizado pela Fundação Carlos Chagas em 1997, é um exemplo de equalização *a posteriori*, pois a calibração dos itens foi realizada em separado para cada uma das populações, depois utilizando o princípio de invariância dos itens, fez-se a equalização dos valores dos parâmetros dos itens comuns por meio de transformações lineares (Andrade & Valle, 1998).

Vital & Soares (2002) compararam o desempenho de alguns métodos de equalização de testes educacionais com base nos resultados dos testes de História do Programa de Avaliação da Rede Pública de Educação Básica (PROEB) de Minas Gerais, aplicados aos alunos da 4ª e 8ª séries do ensino fundamental e 3ª série do ensino médio, em 2001 e

dos resultados dos testes de História do Sistema Nacional de Avaliação da Educação Básica (SAEB/97), concluindo que o critério de equalização *a posteriori* utilizando a média aritmética e a média geométrica que vem sendo utilizado para produzir os resultados do SIMAVE/PROEB parece produzir bons resultados, não sendo muito discrepante dos resultados produzidos pelos outros métodos. No entanto, para Andrade & Valle (1998), espera-se que a equalização *a posteriori* seja menos precisa do que a equalização simultânea pela TRI.

A equalização das habilidades de indivíduos de diferentes populações exige que indivíduos sejam submetidos a provas com itens comuns, pois o uso de itens comuns entre provas distintas aplicadas a populações distintas permite que todos os parâmetros e habilidades possam ser colocados na mesma escala ao final dos processos de estimação, possibilitando comparações e a construção de “escalas do conhecimento” interpretáveis, que são de grande importância na área educacional.

Andrade & Valle (1998) enfatizaram que uma das grandes vantagens na utilização da TRI, trata-se da possibilidade de equalizar as habilidades dos sujeitos, sejam eles pertencentes à mesma população ou não, submetidos ao mesmo item ou conjunto de itens, cujos resultados permitem a comparação de seus desempenhos na tarefa solicitada.

Na literatura especializada tem-se verificado estudos comparativos entre os diversos métodos de equalização de testes. A maioria desses estudos utiliza testes simulados como, por exemplo, os de Kim & Cohen (1998), e os de Andrade (2001). Em Kim & Cohen (1998), os autores comparam o método de calibração simultânea com o Método da Curva Característica/funcional, e no estudo conduzido por Andrade (2001), o autor analisa o desempenho do método *mean-sigma* comparativamente ao do Método dos Grupos Múltiplos. Estudos com dados reais eram mais raros, porém, atualmente vem se ampliando a partir do acesso aos dados das avaliações da larga escala realizada no Brasil e no mundo.

A partir da equalização da Prova Brasil e Prova Belém é possível acompanhar o desempenho dos alunos da secretaria municipal numa escala que permita comparar os seus desempenhos com os dos alunos brasileiros do respectivo ciclo ou ano equivalente. Essa equalização de provas, uma das ferramentas da TRI, permite a criação de uma escala de referência em que os desempenhos dos alunos possam ser passíveis de comparação. Com isso, além da comparação entre os desempenhos, também é possível a investigação do progresso desses alunos durante os diferentes anos de ensino, não somente demonstrando sua importância

para a questão da qualidade oferecida, mas também servindo de instrumento para diagnósticos que possam servir de base para futuras tomadas de decisões em relação à formação continuada dos professores a partir de “escalas do conhecimento” interpretáveis.

3 METODOLOGIA

Neste capítulo apresenta-se a classificação da pesquisa, o instrumento utilizado na coleta de dados, o modelo utilizado na análise dos dados e o processo de equalização na escala SAEB, procurando assim, descrever a metodologia aplicada na análise dos dados.

3.1 CLASSIFICAÇÃO DA PESQUISA

O presente trabalho trata-se de uma pesquisa aplicada, onde se utiliza as técnicas de equalização de escalas para tornar a Prova Belém comparável, entre as suas várias edições, com a Prova Brasil. Esse processo de equalização é baseado na TRI, e se desenvolveu com a aplicação de um teste, construído a partir da Matriz de Referência de Avaliação (MRA) do SAEB, nas disciplinas de Língua Portuguesa e Matemática (anexo I e II).

A pesquisa configura-se como uma abordagem quantitativa, pois está relacionada à quantificação de dados obtidos mediante prévia aplicação do teste e em técnica estatística para o tratamento dos dados. Em relação aos objetivos, esta pesquisa configura-se como descritiva, pois sua característica mais significativa está na utilização de técnicas padronizadas de coleta e análise de dados.

Os métodos específicos mais adotados nas ciências sociais, segundo Gil (2008), são: o experimental, o observacional, o comparativo, o estatístico, o clínico e o monográfico. Alguns autores ampliam consideravelmente o elenco desses métodos, incluindo o método do questionário, da entrevista, dos testes e muitos outros. Neste estudo, valeu-se de procedimentos observacionais, estatísticos e monográficos.

3.2 PROVA APLICADA AOS ALUNOS DO CICLO II

Em 1964, com o objetivo de fortalecer o ensino de nível primário e a promoção da cultura, foi criada a Fundação Educacional do Município de Belém. Quatro anos depois, a fundação se transformou em Secretaria Municipal de Educação e Cultura (SEMEC). Atualmente, a cultura é gerida por um órgão específico da Prefeitura de Belém, ficando a SEMEC responsável pela educação na esfera municipal. Em 2014, a SEMEC mantinha na capital, incluindo a região insular, cerca de 72 mil alunos de educação infantil ao 9º ano do Ensino Fundamental.

Desde 2009, a Secretaria Municipal de Educação de Belém/SEMEC, por meio do Núcleo de Informática Educativa/NIED, vem aplicando testes aos alunos do Ciclo II (equivalente aos 4º e 5º anos do Ensino Fundamental) nos moldes da Prova Brasil, com vistas a melhorar seus indicadores de qualidade de ensino expressos pelo IDEB (Índice de Desenvolvimento da Educação Básica), articulado com um programa de formação continuada de professores. Esse trabalho irá utilizar os dados de Língua Portuguesa e Matemática de um dos testes de 2014 para equalizar as habilidades dos alunos na escala SAEB.

No sentido de aprimorar as avaliações já existentes, esse trabalho buscou aplicar, no teste de dezembro de 2014, critérios metodológicos que garantisse sua comparabilidade com os resultados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), autarquia do MEC responsável por Estatísticas Educacionais. Tais critérios são similares aos da Prova Brasil/SAEB e abrangeu de forma censitária, a avaliação da aprendizagem de alunos matriculados em turmas do Ciclo II 1º e 2º anos na RME de Belém, nos conhecimentos em Língua Portuguesa e Matemática.

Os itens do teste em estudo, chamado de Prova Belém, foram elaborados a partir dos descritores da Matriz de Referência da Avaliação (MRA) de Língua Portuguesa e Matemática para o 5º ano do Ensino Fundamental proposto pelo SAEB (anexo I e II), que especifica, no contexto das avaliações em larga escala, as habilidades dos alunos a serem avaliadas. O descritor é uma associação entre conteúdos curriculares e operações mentais desenvolvidas pelo aluno, que traduzem certas competências e habilidades indicadas na MRA, e constituem a referência para seleção e elaboração dos itens que devem compor uma prova de avaliação (INEP, 2011).

A edição da Prova Belém foi organizada em blocos de itens por disciplina e por ano/série, que segundo uma combinação apropriada, resultou em 16 cadernos para os dois anos (4º e 5º anos EF). Cada aluno respondeu a um total de 44 itens em cada caderno, organizados em blocos de 11 itens, conforme distribuição na tabela 3.1.

Observa-se na tabela 3.1 que cada caderno de prova teve sua composição formada por quatro blocos de itens (1, 2, 3, 4) em quatro posições (I, II, III e IV), contendo dois blocos de itens de Língua Portuguesa e dois de Matemática. Na distribuição, cada bloco de item (1, 2, 3, 4) apareceu uma vez em cada posição (I, II, III, IV), sendo que os cadernos de 1 a 4 e 9 a 12 começaram com os blocos de Língua Portuguesa e os cadernos de 5 a 8 e 13 a 16 por matemática.

Tabela 3.1- Distribuição dos blocos na Prova Belém 2014

Caderno 5º ano EF	Português		Matemática		Caderno 4º ano EF	Português		Matemática	
	I	II	III	IV		I	II	III	IV
1	1	2	1	2	9	1	2	1	2
2	2	3	2	3	10	2	3	2	3
3	3	4	3	4	11	3	4	3	4
4	4	1	4	1	12	4	1	4	1
	Matemática		Português			Matemática		Português	
	I	II	III	IV		I	II	III	IV
5	1	2	1	2	13	1	2	1	2
6	2	3	2	3	14	2	3	2	3
7	3	4	3	4	15	3	4	3	4
8	4	1	4	1	16	4	1	4	1

Fonte: Elaboração do autor.

Para cada disciplina e ano/série, a prova Belém utilizou itens aplicados em testes anteriores e itens já calibrados na escala SAEB. A forma de construção da prova fez com que houvesse blocos comuns entre os cadernos e itens comuns entre os anos, desejável para a calibração dos parâmetros dos itens e estimação das habilidades numa mesma escala. Os quantitativos de blocos de itens, cadernos de prova e itens utilizados estão sintetizados na tabela 3.2 e serão detalhados nas próximas seções desse capítulo.

Tabela 3.2 - Blocos de itens na Prova Belém 2014

Ciclo / Ano	Nº de blocos	Nº de cadernos	Total de itens
Ciclo II 1º ano	4 LP e 4 MT	08	44 LP e 44 MT
Ciclo II 2º ano	4 LP e 4 MT	08	38 LP e 38 MT

Fonte: Elaboração do autor, LP=Português e MT=Matemática.

3.2.1 Prova Belém de Língua Portuguesa

Em Língua Portuguesa, para os alunos do Ciclo II 2º ano (5º ano do ensino fundamental), foram elaborados 4 blocos de itens cada um com 11 itens, entre os quais 3 itens comuns já calibrados na escala SAEB. Os quatro blocos de itens foram identificados através de uma numeração sequencial de 1 a 4, e cada caderno foi montado usando um bloco de item par e outro ímpar conforme a Tabela 3.1, sendo os 3 itens na escala SAEB de cada bloco repetidos da seguinte forma: 3 itens SAEB nos blocos pares e outros 3 itens SAEB nos blocos ímpares,

utilizando um total de 6 itens do SAEB e fazendo com que todos os alunos respondessem a eles.

Para os alunos do Ciclo II 1º ano (4º ano do ensino fundamental), também foram elaborados outros 4 blocos de 11 itens cada um. Do total de itens desse ano, 14 itens foram comuns com os blocos do Ciclo II 2º ano. Assim, na tabela 3.3 e 3.4, são apresentados as quantidades de itens dos blocos de Língua Portuguesa com um total de 68 itens (apêndice A), sendo 6 itens já calibrados pela escala SAEB, 32 itens elaborados para o Ciclo II 2º ano (total de 38 itens no Ciclo II 2º ano) e 30 outros itens exclusivos para o Ciclo II 1º ano.

Tabela 3.3 - Blocos da Prova Belém (Ciclo II 2º ano - LP)

Bloco 1	Bloco 2	Bloco 3	Bloco 4	Total
8 itens não calibrados	8 itens não calibrados	8 itens não calibrados	8 itens não calibrados	32 itens não calibrados
3 itens SAEB (Ímpar)	3 itens SAEB (Par)	3 itens SAEB (Ímpar)	3 itens SAEB (Par)	6 itens SAEB
11 itens	11 itens	11 itens	11 itens	38 itens

Fonte: Elaboração do autor.

Os itens calibrados do SAEB ficaram posicionados no meio do bloco e os itens por calibrar foram escolhidos procurando equilibrar as dificuldades e discriminação (IF e ID) a partir dos parâmetros da TCT obtidos em aplicações anteriores. Cada item estava associado a um descritor da matriz de referência do SAEB, e para distribuir todos os descritores de Língua Portuguesa foram colocados 8 (oito) descritores dos tópicos I, II e III nos blocos de itens ímpares, e os 7 (sete) demais descritores dos tópicos IV, V e VI, nos blocos de itens pares (anexo I).

Tabela 3.4 - Blocos da Prova Belém (Ciclo II 1º ano - LP)

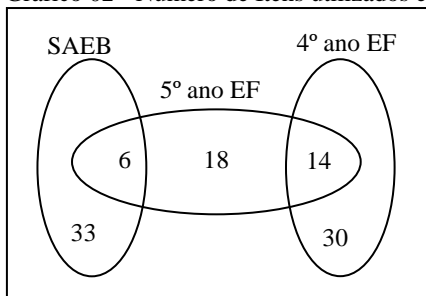
Bloco 1	Bloco 2	Bloco 3	Bloco 4	Total
6 não calibrados	6 não calibrados	10 não calibrados	8 não calibrados	30 itens não calibrados
5 itens comuns	5 itens comuns	1 item comum	3 itens comuns	14 itens comuns ao CII 2º ano
11 itens	11 itens	11 itens	11 itens	44 itens

Fonte: Elaboração do autor.

No gráfico 02, representa-se a quantidade de itens diferentes utilizados para elaboração dos quatro blocos de itens de Língua Portuguesa (nos diferentes grupos utilizados para efetuar a equalização do teste na escala SAEB), procurando mostrar de uma forma mais integrada os dados apresentados nas Tabela 3.3 e Tabela 3.4.

Não houve troca de gabarito entre itens comuns, nem item duplicado ou anulado, não sendo detectado qualquer outro problema de formulação.

Gráfico 02 - Número de Itens utilizados em Língua Portuguesa



Fonte: Elaboração do autor.

3.2.2 Prova Belém de Matemática

Em Matemática, para os alunos do Ciclo II 2º ano (5º ano do ensino fundamental) foram elaborados 4 blocos, cada um com 11 itens. De maneira semelhante ao efetuado com Língua Portuguesa, utilizou-se 6 itens já calibrados na escala SAEB nos cadernos.

Para os alunos do Ciclo II 1º ano foram criados outros 4 blocos, de 11 itens cada, sendo que desde total de itens 12 eram comuns com os itens do Ciclo II 2º ano. Portanto, nos 8 blocos de Matemática para o Ciclo II 1º e 2º anos, foram utilizados 70 itens (apêndice B) no total, sendo 6 itens já calibrados do SAEB, 32 itens elaborados para o Ciclo II 2º ano e 32 itens exclusivos para o Ciclo II 1º ano.

Tabela 3.5 - Blocos da Prova Belém (Ciclo II 2º ano - Matemática)

Bloco 1	Bloco 2	Bloco 3	Bloco 4	Total
8 não calibrados	8 não calibrados	8 não calibrados	8 não calibrados	32 não calibrados
3 itens SAEB (Ímpar)	3 itens SAEB (Par)	3 itens SAEB (Ímpar)	3 itens SAEB (Par)	6 itens SAEB
11 itens	11 itens	11 itens	11 itens	38 itens

Fonte: Elaboração do autor.

Os blocos de Matemática foram organizados de forma similar aos de Língua Portuguesa, ou seja, os itens calibrados ficaram estrategicamente posicionados no meio do bloco, de forma que os 28 descritores da matriz de referência da avaliação do SAEB (anexo II), foram distribuídos da seguinte maneira: todos os 14 (quatorze)

descritores dos Temas I, II e IV ficaram nos blocos ímpares e todos os 14 (quatorze) descritores do Tema III nos blocos pares.

Representa-se no gráfico 03 a quantidade de itens diferentes utilizados para elaboração dos quatro blocos de itens de Matemática, nos diferentes grupos utilizados para efetuar a equalização do teste na escala SAEB, procurando mostrar de uma forma mais integrada os dados apresentados nas Tabela 3.5 e 3.6.

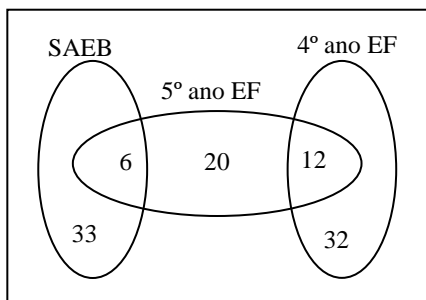
Tabela 3.6 - Blocos do Ciclo II 1º ano - Matemática

Bloco 1	Bloco 2	Bloco 3	Bloco 4	Total
7 não calibrados	6 não calibrados	11 não calibrados	8 não calibrados	32 itens não calibrados
4 itens comuns	5 itens comuns	0comuns	3 itens comuns	12 itens comuns ao CII 2º ano
11 itens	11 itens	11 itens	11 itens	44 itens

Fonte: Elaboração do autor.

Na correção dos cadernos de Matemática, verificou-se diferença no gabarito em dois itens comuns do Ciclo II 2º ano com o Ciclo II 1º ano (MT46 e MT66) para o qual se efetuou a recodificação das respostas. Além disso, o item MT55 ficou duplicado no bloco 2 do Ciclo II 1º ano.

Gráfico 03 - Número de Itens utilizados em Matemática



Fonte: Elaboração do autor.

3.3 MODELO DA TEORIA DE RESPOSTA AO ITEM

Para a calibração dos parâmetros dos itens e estimativas das proficiências dos alunos na Prova Belém foi utilizado o modelo da ogiva normal conforme feito pelo SAEB, sendo a probabilidade de acerto do aluno com uma habilidade θ , pode ser aproximada pela equação abaixo:

$$P(U_{ij} = 1 | \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-1,7a_i(\theta_j - b_i)}}, \quad (1)$$

Com $i = 1, 2, \dots, I$; e $j = 1, 2, 3, 4, 5, \dots, n$; sendo I igual ao número do item e n ao número de alunos.

Onde:

$$U_{ij} = \begin{cases} 1, & \text{quando o aluno } j \text{ responde corretamente o item } i, \\ 0, & \text{quando o aluno } j \text{ não responde correto ao item } i. \end{cases}$$

θ_j representa a habilidade (traço latente θ) do j – ésimos aluno.

$P(U_{ij} = 1 | \theta_j)$ é a probabilidade de um aluno j com habilidade θ_j responder corretamente ao item i .

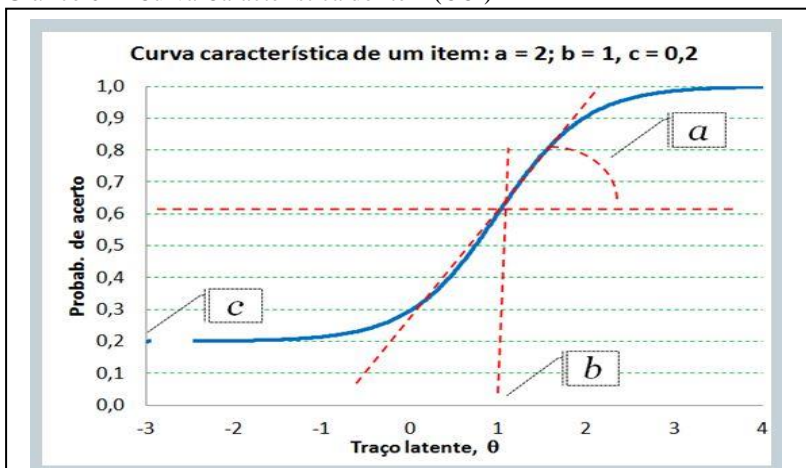
b_i é o parâmetro de dificuldade (ou de posição) do item i .

a_i é parâmetro de discriminação (ou de inclinação) do item i .

c_i é parâmetro do item que representa a probabilidade de alunos com baixa habilidade responderem corretamente ao item i .

Esse modelo fornece para cada item uma curva exemplificada pelo Gráfico 04, denominado de Curva Característica do Item (CCI), onde se destaca a representação gráfica dos parâmetros a , b , c .

Gráfico 04 - Curva Característica do Item (CCI)



Fonte: Barbetta et al., 2014.

a : discriminação ou inclinação do item;

b : dificuldade ou posição (medido na mesma escala do traço latente θ).

c : probabilidade de acerto para alunos com baixa proficiência.

No Gráfico 04, o eixo do traço latente está representado de -3 a 4, em uma escala (0;1), no qual a média do grupo de referência é arbitrada em zero, e o desvio padrão dos traços latente em um. Essa métrica é usual, principalmente nos resultados fornecidos pelos programas computacionais que trabalham com a TRI. Nessa seção serão feitas

algumas considerações sobre esses parâmetros considerando essa escala (0;1) na equalização da prova Belém com a prova do SAEB, posteriormente os resultados são facilmente colocados na escala SAEB (250;50) por meio da transformação linear:

Língua Portuguesa ==>NOTA250P = 249.98499+55.0933628*Prof01;

Matemática ==> NOTA250M = 249.964381+55.8923279*Prof01.

O parâmetro de dificuldade b_i do item i corresponde ao ponto na escala de habilidade θ onde a probabilidade de resposta é $\frac{(1+c)}{2}$, assim, quanto maior o valor do parâmetro b , mais difícil é o item. Os valores padronizados do parâmetro b tipicamente se situam entre -3 para itens fáceis, e +3 para itens difíceis, quando os valores de b estão na escala de média 0 e desvio padrão 1. Como o parâmetro b é medido na mesma unidade da habilidade θ , a dificuldade do item e a habilidade dos alunos se tornam comparáveis.

O parâmetro a_i no modelo, indica o quão o item está relacionado ao traço latente que se pretende medir, quanto maior seu valor mais esse item consegue medir a habilidade que se pretende. O parâmetro a_i pode variar de 0 a ∞ , mas tipicamente itens que estão bem associados ao traço latente apresentam $a > 0.7$ para o modelo logístico ($a > 0.4$ para a ogiva normal), e é representada pela inclinação da curva (ângulo) no ponto de inflexão, onde a probabilidade de resposta correta é $\frac{(1+c)}{2}$. O item com valor de a_i abaixo de 0.7 tem pouco poder de discriminação, pois os alunos com habilidades em torno de b_i têm aproximadamente a mesma probabilidade de responder corretamente ao item, por outro lado, valores mais altos indicam itens com CCI “íngremes”, que discriminam os alunos basicamente em dois grupos: os que possuem habilidades abaixo do valor do parâmetro b_i e os que possuem habilidades acima do valor do parâmetro b_i (Andrade, Tavares & Valle, 2000).

O parâmetro c_i é expresso pela assíntota inferior da curva CCI, e não depende da escala, pois representa a probabilidade dos alunos com baixa habilidade responderem corretamente ao item, e como tal, assume sempre valores entre 0 e 1. Em testes que não são de múltipla escolha, $c_i = 0$ e b_i representa o ponto na escala da habilidade onde a probabilidade de acertar o item é 0,5.

Além da CCI, a TRI utiliza a Curva de Informação do Item (CII), que indica a precisão com que se pode estimar θ através da resposta de um item; e a Curva de Informação Total (CIT) do teste que é a somatória da informação de todos os itens e representa a qualidade do instrumento de uma maneira geral, servindo para verificar a partir de

qual ponto da escala o teste é mais informativo. Exemplos dessas curvas serão apresentados no próximo capítulo ao mostrar os resultados do estudo.

3.3.1 Geração das respostas do grupo de referência

Para o processo de equalização são necessárias as respostas do grupo de referência, que na escala SAEB ficou definido como os respondentes da prova de 1997, da 8ª série do ensino fundamental. Além dos itens calibrados na escala SAEB é necessária uma amostra dos dados das respostas do grupo de referência, porém como só estavam disponíveis os itens com os seus parâmetros (a, b, c) calibrados na escala SAEB, foi necessário simular esses dados de respostas.

O modelo unidimensional da TRI para itens dicotômicos baseia-se na suposição de que quanto maior o traço latente do indivíduo na área avaliada, maior é a probabilidade dele responder corretamente ao item, portanto as habilidades a serem estimadas se aproximam de uma distribuição normal, onde se arbitra média igual a 0, e desvio padrão igual a 1.

Para realizar as simulações das respostas do grupo de referência, geraram-se valores aleatórios para as habilidades θ de cada indivíduo j , a partir de uma distribuição normal (0;1), e posteriormente obteve-se as respostas de cada indivíduo a cada um dos itens a partir de uma distribuição de Bernoulli de proporção $P(U_{ij} = 1 | \theta_j)$, calculado pela equação apresentada na equação (1) que aproxima a ogiva normal do modelo logístico de três parâmetros (ML3). Portanto, a partir das probabilidades de um indivíduo j com uma habilidade θ , acertar o item i com parâmetros conhecidos, foram simuladas respostas com valores 0 ou 1 (U_{ij} = variável dicotômica) para o número de indivíduos da amostra.

Na tabela 3.7, n representa o número de itens, no caso desse estudo $n=39$; e N o número de indivíduos que responderam aos n itens, no caso desse estudo foram feitas simulação com 10mil, 30mil, 50mil, 80mil e 100mil indivíduos até encontrar o melhor ajuste. A qualidade da base simulada será verificada estimando novamente os parâmetros dos itens por meio da base simulada e os parâmetros dos itens foram comparados com os valores reais fornecidos pelo INEP. Os resultados dessa comparação serão apresentados no capítulo 4.

Para avaliar as amostras plotou-se graficamente, por meio de planilha eletrônica, a relação entre os valores dos parâmetros: reais e

estimados. Em seguida, traçou-se a linha de tendência ou regressão linear expressa na equação $y = \alpha x + \beta$, e obteve-se coeficiente de determinação (R^2) entre os valores estimados e os valores reais. Utilizou-se como critério de escolha, a amostra em que os valores α e β se aproximassem de 1 e 0, respectivamente, além disso, o R^2 fosse o mais próximo do valor 1 (100%), por melhor explicar os valores observados.

Tabela 3.7 - Ilustração de dados simulados

Indivíduos	1	2	...	N
1	u_{11}	u_{12}	...	u_{1n}
2	u_{21}	u_{22}	...	u_{2n}
.
.
.
N	u_{N1}	u_{N2}	...	u_{Nn}

Fonte: Elaboração do autor, adaptado de Tavares (2015).

3.3.2 Método de Equalização utilizado

No processo de equalização da Prova Belém na escala SAEB, procurou-se não alterar os quatro fatores que influenciam a confiabilidade dos resultados obtidos nos processos de equalização, segundo Silva e Soares (2010): (a) conteúdo do teste, (b) formato do teste, (c) usos e consequências, e (d) erro do método estatístico de equalização. Assim, elaborou-se a prova com o mesmo conteúdo e no mesmo formato, deram-se os mesmos usos e utilizou-se o mesmo método estatístico, alterando apenas parte da população alvo, enquanto que no SAEB as três populações alvo são 4ª série e 8ª série Ensino Fundamental e 3ª série do Ensino Médio, na Prova Belém a população alvo foi Ciclo II 1º e 2º anos. A quantidade de respondente utilizada em cada um dos grupos está na tabela 3.8 e no gráfico 05.

Tabela 3.8 - Número de estudantes na Prova Belém em dez/2014

Grupo	Ciclo /ano	Nº alunos
01	8ª série de 1997 (simulado) – Referência	100.000
02	Ciclo II 2º ano (5º ano do ensino fundamental-EF)	5.966
03	Ciclo II 1º ano (4º ano do ensino fundamental-EF)	5.305

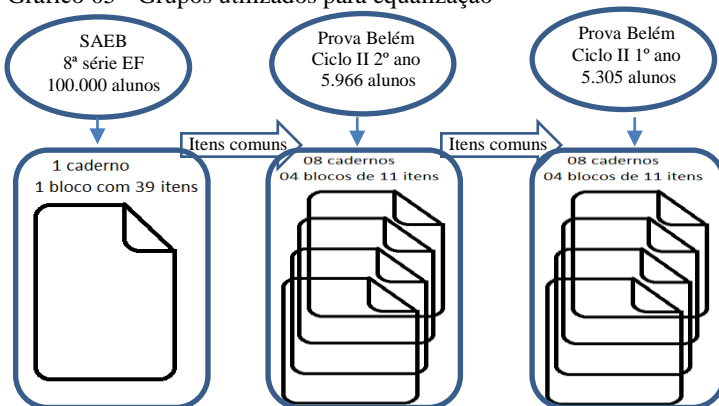
Fonte: Elaboração do autor.

A equalização da Prova Belém se enquadra no caso em que dois ou mais grupos são submetidos a duas ou mais provas diferentes, mas

que têm alguns itens comuns, sendo um exemplo prático de equalização via itens comuns feita de forma simultânea.

A equalização entre o Ciclo II 1º e 2º anos com a escala do SAEB foi feita por meio da calibração conjunta dos itens, incluindo-se na estimativa as respostas individuais dos alunos respondentes da Prova Belém e as respostas simuladas do grupo de referência (alunos da 8ª série de 1997). Dessa maneira, tem-se uma estimativa única dos parâmetros, não importando o ano e evitando a propagação de erros na equalização quando se estima os parâmetros em cada ano (Klein, 2003).

Gráfico 05 - Grupos utilizados para equalização



Fonte: Elaboração do autor.

Desse modo, na equalização do Ciclo II 1º e 2º anos com a escala SAEB consideraram-se além das provas descritas na secção anterior, uma terceira prova composta de 39 itens de Língua Portuguesa e 39 itens de Matemática com todos os itens já calibrados na escala SAEB, como mostra o gráfico 06. A população simulada de respondentes da terceira prova se constituiu em grupo referência e teve o único objetivo de possibilitar a equalização, não havendo interesse em estudar o seu desempenho.

A equalização da prova de Língua Portuguesa foi desenvolvida a partir das respostas a 68 itens (apêndice A) distintos apresentados aos alunos do Ciclo II 1º e 2º anos, dos quais 06 eram itens calibrados do SAEB; mais as respostas simuladas aos 39 itens do SAEB, dos quais 33 não apresentados aos referidos alunos. Já a equalização da prova de Matemática foi desenvolvida a partir de 70 itens distintos (apêndice B) apresentados aos alunos e 33 itens calibrados não apresentados (Gráfico 02 e 03).

No caso de itens dicotômicos, o software utilizado pelo INEP na análise do SAEB é o Bilog-MG, *software* que pode ser considerado uma extensão do Bilog, permitindo a calibração simultânea com grupos diferentes. Os parâmetros dos itens são estimados através do método de estimação de máxima verossimilhança marginal (MMLE) e as habilidades dos alunos pelo método do valor esperado *a posteriori* (EAP) (Vital & Soares, 2002).

O uso de itens comuns entre provas distintas aplicadas a populações distintas permite que todos os parâmetros estejam na mesma escala ao final dos processos de estimação, possibilitando comparações e a construção de “escalas do conhecimento” interpretáveis, que são de grande importância na área educacional (Santos, 2008).

3.4 METODOLOGIA DE POSICIONAMENTO DOS ITENS

As escalas obtidas no processo de equalização ordenam os desempenhos dos alunos (do nível mais baixo ao mais alto) e os parâmetros dos itens em um *continuum* e precisam ser interpretadas para ganharem sentido prático. As escalas são interpretadas através da escolha de pontos, chamados de níveis âncoras, que permitam descrever as habilidades que os alunos demonstraram possuir quando situados em torno desses pontos. Portanto, níveis âncora são pontos selecionados na escala da habilidade para serem interpretados pedagogicamente. Geralmente, escolhe-se a média como um ponto para nível âncora, e os demais pontos em intervalos de um desvio padrão ou meio desvio padrão.

Cada nível âncora definido na escala é caracterizado por conjuntos de itens, denominados itens âncoras, que servem para explicar ou interpretar o nível efetuando uma ligação dos valores da escala com os conteúdos verificados nos itens.

Um Item é definido como âncora de um nível Z, quando atendem as três condições abaixo satisfeitas simultaneamente, onde Y e Z são níveis âncoras consecutivos com $Y < Z$ (Andrade, Tavares & Valle, 2000):

1. $P(U = 1|\theta = Z) \geq 0,65$;
2. $P(U = 1|\theta = Y) < 0,50$;
3. $P(U = 1|\theta = Z) - P(U = 1|\theta = Y) \geq 0,30$

Em outras palavras, para um item ser âncora em um determinado nível âncora da escala, ele precisa ser respondido corretamente por uma grande proporção de alunos (pelo menos 65%) com este nível de habilidade e por uma pequena proporção de alunos (no máximo 50%)

com o nível de habilidade imediatamente anterior. Além disso, a diferença entre a proporção de alunos com esses níveis de habilidade deve ser de pelo menos 30%. Assim, um item âncora é típico daquele nível, ou seja, bastante acertado por alunos com aquele nível de habilidade e pouco acertado por alunos com um nível de habilidade imediatamente inferior (Valle, 2001).

A metodologia para interpretação das escalas inclui dois procedimentos principais: identificação de itens âncoras pela análise estatística e a apresentação desses itens a especialistas para descrição do nível âncora a partir do estudo do conteúdo abordado no conjunto de itens que definem cada nível âncora, lembrando que na prova Belém elaborou-se cada item associado a apenas um descritor da matriz de referência.

Esse critério de seleção dos itens âncoras foi utilizado nos SAEB 95 e 97, porém foram observados vários problemas (Klein, 2003): (a) para ter muitos itens âncora, os níveis selecionados têm que ser bem espaçados; (b) para ter mais itens para auxiliar a interpretação é necessário ampliar o conceito de nível âncora para quase âncora; (c) dificuldade dos membros dos painéis de especialistas de utilizarem o conceito de discriminação entre níveis, só usando praticamente itens considerados dominados pelos alunos em torno do nível; (d) alguns descritores cobertos por alguns itens não aparecerem na descrição, pois mesmo sendo bons itens, não foram classificados como itens âncora.

A decisão tomada pelo SAEB para enfrentar os problemas encontrados foi selecionar os pontos da escala em intervalos de meio desvio padrão (25), incluindo o ponto da média arbitrada da distribuição de proficiências da 8^a série/9^o ano EF nas disciplinas de Língua Portuguesa e Matemática (250). E para cada item foi estimado o percentual de acerto dos alunos em cada nível, calculando-se o percentual de acerto dos alunos com proficiência no intervalo que contém o nível.

Nesta nova abordagem, um item é dito "âncora" em um nível se: (a) O número de alunos no nível que respondeu ao item é maior que 50; (b) O percentual de acerto do item nos níveis anteriores é menor que 65%; (c) O percentual de acerto do item no nível considerado e nos níveis acima é maior ou igual a 65%; (d) O ajuste da curva é bom, com parâmetro "a" elevado (Klein, 2003).

Nesse estudo os níveis âncoras foram escolhidos de 25 em 25 pontos na escala (250;50), iniciando a partir de 125 pontos, de acordo como foram definidos na escala SAEB. Os itens foram posicionados nos níveis, levando

em consideração a proficiência necessária para que o aluno tivesse uma probabilidade de 0,65 de acertar o item.

4 RESULTADOS

Esse capítulo apresenta os resultados encontrados na pesquisa, iniciando por Língua Portuguesa e na sequência Matemática. Os parâmetros dos itens e as proficiências foram estimados na escala SAEB (0;1), ou seja, com média 0 e desvio padrão 1 e posteriormente transformadas para a escala SAEB (250;50), através dos valores de transformação utilizados pelo INEP, informados na seção 3.3. Também são apresentados os resultados encontrados nas gerações dos dados das bases clones para equalização de Língua Portuguesa e de Matemática.

Os níveis definidos pelo SAEB para o 5º ano do ensino fundamental e as categorias de desempenhos propostos pelo Sistema Paraense de Avaliação Educação (SisPAE) - programa de avaliação permanente, aplicado anualmente nas escolas paraenses, para orientar o planejamento das ações das Secretarias de Educação foram utilizados neste trabalho. Em 2014, participaram do SisPAE estudantes dos 4º, 5º, 8º e 9º anos do Ensino Fundamental e das 1ª, 2ª e 3ª série do Ensino Médio, com os seus resultados equalizados na escala do SAEB em Língua Portuguesa e Matemática.

Para a equalização dos itens da prova Belém, os parâmetros dos itens e as proficiências foram estimados supondo-se que os alunos pertenciam a um dos seguintes grupos: Grupo1, alunos do 9º ano - Prova Brasil 1997, considerado como referência, usando um banco de respostas simuladas a 39 itens calibrados na escala SAEB; Grupo 2, alunos do Ciclo II 2º ano (5º ano EF); e Grupo 3, alunos do Ciclo II 1º ano (4º ano EF). Desta forma, as estimativas dos parâmetros dos itens e proficiências foram automaticamente obtidas na mesma escala estabelecida para o grupo de referência.

Para as estimativas dos parâmetros e das proficiências dos alunos utilizou-se o software Bilog-MG, para os 3 grupos, como definido acima, fixando os parâmetros conhecidos do Grupo 1 (referência). Na fase 1 do programa foram calculados, entre outros, os valores das correlações bisseriais, utilizados para detectar problema no funcionamento do item. Utilizou-se o critério proposto pelo INEP (Andrade, Laros & Gouveia, 2010), que estabelece que o item deva ser revisto quando o coeficiente bisserial: (a) do gabarito for menor 0,15; (b) de dois distratores forem maiores que 0,10; ou (c) de um distrator for maior que da alternativa correta.

Na segunda rodada, os valores dos parâmetros foram estimados na mesma escala do grupo de referência (escala SAEB), em seguida foram analisados os valores, bem como a curva característica do item

(CCI) e na última rodada as proficiências dos alunos foram geradas. Os resultados desses passos foram apresentados nesse capítulo utilizando-se de tabelas e gráficos.

Em relação aos dados, consideraram-se todas as respostas dos alunos presentes, mesmo daqueles que deixaram o cartão em branco ou fizeram várias marcações no mesmo item, como também dos alunos com necessidades especiais e de turmas multisseriadas da região das ilhas. Portanto, não foi aplicado nenhum filtro aos dados brutos coletados.

Por fim, tratou-se da interpretação da escala de desempenho, que significa escolher alguns níveis na escala e descrever os conhecimentos e habilidades que os alunos demonstram possuir quando situados em torno desses níveis. Os níveis foram definidos variando de $\frac{1}{2}$ desvio padrão, iniciando no 5º ano do Ensino Fundamental, com nível 0 para os alunos com proficiências menores que 125, tanto para Língua Portuguesa como para Matemática. Conclui-se exemplificando a interpretação pedagógica de alguns itens aplicados na avaliação e posicioná-los nos níveis âncoras da escala.

4.1 BASE CLONE DE LÍNGUA PORTUGUESA

O INEP disponibilizou somente os parâmetros (a, b, c) dos itens calibrados na escala SAEB. Portanto, para equalização dos itens novos da prova Belém foi necessário simular uma base de dados com respostas que gerassem os mesmos parâmetros fornecidos pelo INEP. Realizou-se simulações com amostras de 10mil, 30mil, 50mil, 80mil e 100 mil indivíduos utilizando os procedimentos descritos na seção 3.3.1, e compararam-se os valores dos parâmetros (a, b, c) estimados pelo Bilog-MG para a base de dados simulada com os valores reais, a fim de escolher qual amostra utilizar na equalização.

Para avaliar as amostras, traçou-se graficamente, com o auxílio de uma planilha eletrônica, a relação entre os valores estimados dos parâmetros para a amostra e os valores reais dos parâmetros. Procurando a linha de tendência expressa na equação linear ($y = \alpha x + \beta$) em que os valores α e β tendesse a 1 e 0, respectivamente. Além disso, procurou-se o gráfico em que o coeficiente de determinação (R^2) fosse o mais próximo do valor 1, pois o R^2 representa uma medida de ajustamento do modelo de regressão linear, indicando, em porcentagem, o quanto o modelo consegue explicar os valores observados. Essa análise gráfica foi realizada em cada amostra e para cada um dos parâmetros.

Nos resultados a seguir pode-se observar exemplos dos gráficos para o parâmetro “a” com amostra de 10 mil indivíduos, para o parâmetro “b” com amostra de 50 mil, e para o parâmetro “c” para amostra de 100 mil respondentes. Para auxiliar a comparação das equações das linhas de tendências e do R^2 organizou-se em tabelas os valores encontrados.

Na tabela 4.01, observa-se que o parâmetro “a”, de acordo com valor de R^2 , se ajusta melhor aos dados para 80mil respondentes, já em relação aos coeficientes da equação os dados de 50mil responderam melhor, pois o coeficiente linear mais se aproxima de zero e o coeficiente angular de 1.

Tabela 4.01 - LP Comparação dos dados simulados - parâmetro “a”

Parâmetro “a”	Equação ($\alpha x + \beta$)	α	β	R^2
10mil	0,9764x+0,0518	0,98	0,05	0,96
30mil	0,9561x+0,0513	0,96	0,05	0,98
50mil	1,0051x+0,018	1,01	0,02	0,99
80mil	0,9690x+0,0433	0,97	0,04	1,00
100mil	0,9802x+0,0323	0,98	0,03	0,99

Fonte: Elaboração do autor.

Na tabela 4.02, observa-se que o parâmetro “b”, melhor se ajustou aos dados da amostra de 80mil respondentes, de acordo com valor de R^2 e em relação aos coeficientes da equação.

Tabela 4.02 - LP comparação dos dados simulados - parâmetro “b”

Parâmetro “b”	Equação ($\alpha x + \beta$)	α	B	R^2
10mil	0,8839x-0,0394	0,88	0,04	0,96
30mil	0,9346x-0,0294	0,93	0,03	0,98
50mil	0,9402-0,0142	0,94	0,01	0,99
80mil	0,9532x-0,0142	0,95	0,01	0,99
100mil	0,9260x-0,058	0,93	-0,06	0,99

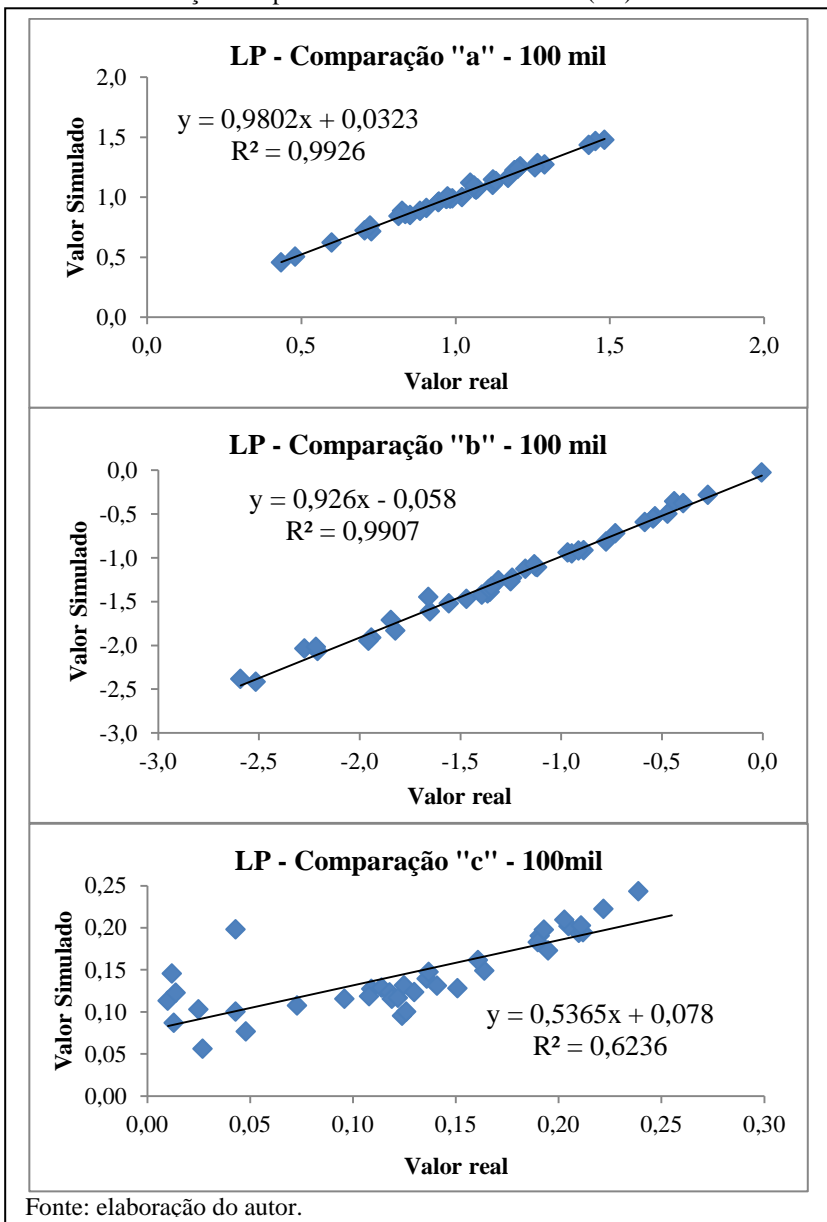
Fonte: Elaboração do autor.

Para o parâmetro “c”, os dados para 80mil foi o que melhor se ajustou, seja de acordo com valor de R^2 , seja em relação aos coeficientes da equação.

Observa-se na tabela 4.03 que o parâmetro c tem problemas de estimação, pois os valores α e R^2 não estão próximos de 1 e o valor de β também está distante de zero. Entretanto, para fixar os parâmetros na escala isso não é nenhum problema, pois o programa Bilog-MG consegue fixar os valores do parâmetro “c” mesmo que a simulação não consiga reproduzir os resultados. O problema seria se esse comportamento ocorresse com os parâmetros “a” e “b”, pois o programa

não conseguiria fixar esses valores, com a base não reproduzindo o valor original.

Gráfico 06 - Relação dos parâmetros na amostra 100mil (LP)



No Gráfico 06 observam-se as linhas de tendência para os valores dos parâmetros a , b , c para a amostra de 100mil indivíduos.

Tabela 4.03 - LP comparação dos dados simulados - parâmetro “c”

Parâmetro “c”	Equação ($\alpha x + \beta$)	α	B	R^2
10mil	0,2715x+0,1364	0,27	0,14	0,16
30mil	0,4402x+0,1064	0,44	0,11	0,39
50mil	0,5850x+0,0812	0,59	0,08	0,64
80mil	0,6422x+0,0644	0,64	0,06	0,79
100mil	0,5365x+0,078	0,54	0,08	0,62

Fonte: Elaboração do autor.

Salienta-se que a base clone de LP não conseguiu reproduzir o parâmetro “c” original com tanta precisão, possivelmente devido as respostas terem sido simuladas para uma população da 8a série (9º ano Ensino Fundamental) e utilizados itens da 4a série (5º ano Ensino Fundamental). Uma solução a essa situação seria efetuar uma mudança de escala dos parâmetros dos itens do grupo de referência do 9º ano para uma escala mais adequada para itens do grupo do 5º ano, por meio de uma transformação linear.

Conforme relatório do SisPAE 2014, em Língua Portuguesa a média de proficiência dos alunos da SEMEC ficou no nível básico, sendo que a diferença entre o desempenho dos alunos do 5º ano EF e do 9º ano EF nesse nível é de 50 pontos (5º ano EF = 150 e 9º ano EF = 200) numa escala (250;50), ou seja, um (01) desvio padrão. Portanto, é razoável considerar uma mudança da escala (0;1) em que se encontram os parâmetros do grupo de referência do 9º ano; para uma escala (-1;1) para um grupo de referência do 5º ano EF.

Para efetuar a transformação linear, utiliza-se a propriedade de invariância que permite considerar que os parâmetros a e b de um item apresentado a dois grupos diferentes (5º ano e 9º ano, no caso em estudo) devem satisfazer, a menos de flutuações amostrais, as seguintes relações lineares: $b_{9^\circ ano} = \alpha b_{5^\circ ano} + \beta$ e $a_{9^\circ ano} = \frac{1}{\alpha} a_{5^\circ ano}$; onde $b_{5^\circ ano}$ e $b_{9^\circ ano}$ são os valores do parâmetro de dificuldade e $a_{5^\circ ano}$ e $a_{9^\circ ano}$ são os valores do parâmetro de discriminação nos grupos 5º ano e 9º ano, respectivamente. Um dos métodos que determina os coeficientes α e β , é o denominado média-desvio (*mean-sigma*) em que $\alpha = \frac{\sigma_{9^\circ ano}}{\sigma_{5^\circ ano}}$, e $\beta = M_{9^\circ ano} - \alpha M_{5^\circ ano}$ (Andrade, Tavares & Valle, 2000).

Aplicando o método média-desvio, e sabendo que a Média $_{9^\circ ano}$ =0; Desvio $_{9^\circ ano}$ =1; Média $_{5^\circ ano}$ = -1; Desvio $_{5^\circ ano}$ = 1 temos que: $\alpha=1$; $\beta=1$;

$b_{9^{\circ}ano} = b_{5^{\circ}ano} + 1$; e $a_{9^{\circ}ano} = a_{5^{\circ}ano}$. Efetuando essas transformações a base clone conseguiu reproduzir o parâmetro c com melhor precisão e também os demais parâmetros melhoraram ou se mantiveram, conforme mostrado na Tabela 4.04.

Tabela 4.04 - LP amostra de 100mil (Escala -1,1) grupo de 5º ano

Parâmetro	Equação ($\alpha x + \beta$)	α	β	R^2
a	$0,9785x+0,0398$	0,98	0,04	0,9954
b	$0,9549x+0,0399$	0,95	0,04	0,9949
c	$0,7428x+0,0486$	0,74	0,05	0,9109

Fonte: Elaboração do autor.

Inicialmente, comparando as várias amostras, a simulação com 80mil indivíduos em Língua Portuguesa apresenta melhor ajuste e poderia ter sido a escolhida. Porém, como a amostra de 100mil teve resultado similar, optou-se por efetuar o mesmo procedimento para Matemática e escolher uma amostra que se ajustasse bem nos parâmetros e apresentasse também a mesma quantidade de dados no processo de calibração em ambas as disciplinas.

4.2 ESCALA DE LÍNGUA PORTUGUESA

A escala SAEB permite distribuir os alunos em categorias de desempenhos qualitativos, em relação ao desempenho adequado esperado para os alunos. Conforme relatório do INEP 2002 (INEP/MEC/Brasil, 2002), a partir de 200 ou mais pontos os alunos do 5º ano do ensino fundamental se encontram com desempenho adequado ao seu ano de escolarização, mesmo critério utilizado pelo SisPAE que também estabelece que os alunos do 4º ano a partir de 185 pontos atingem a aprendizagem adequada. Assim, a classificação dos alunos em Língua Portuguesa no 4º e 5º ano do Ensino fundamental em 4 categorias de desempenho pode ser expressa na tabela 4.05.

Tabela 4.05 - Categorias de desempenho em Língua Portuguesa

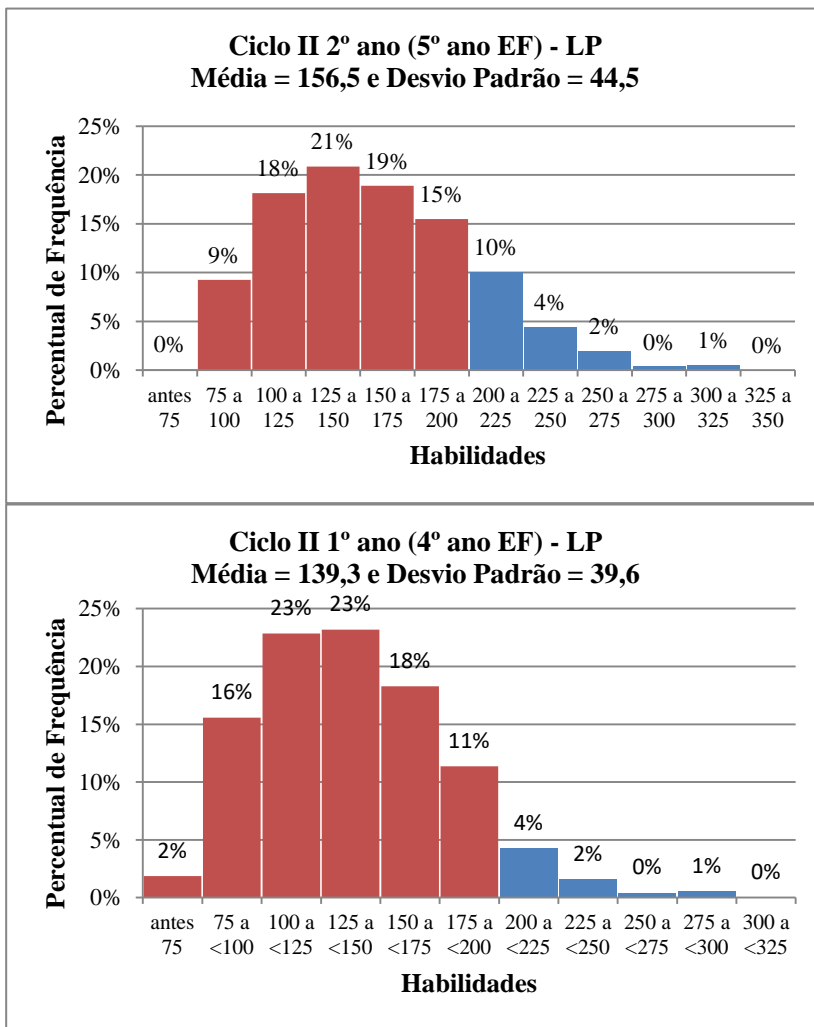
Categoria	4º ano E.F.	5º ano E.F.
Abaixo do Básico	menor que 135	menor que 150
Básico	135 a <185	150 a <200
Adequado	185 a <235	200 a <250
Avançado	Igual ou maior a 235	Igual ou maior a 250

Fonte: SisPAE 2014 (<https://sispae.vunesp.com.br/Padroes.aspx>, 16/12/2016).

O gráfico 07 mostra as distribuições das habilidades dos alunos do Ciclo II 1º ano e do Ciclo II 2º ano em 2014 na Prova Belém 2014, na escala (250; 50). Pode-se observar que a porcentagem de alunos do

Ciclo II 2º ano com habilidade maior ou igual a 200 (17%) - aprendizagem adequada no Ciclo II 5º ano EF - foi maior que a porcentagem de alunos do Ciclo II 1º ano com mesma faixa de habilidade (7%), isto mostra uma diferença positiva (10%) entre a porcentagem de alunos com habilidades iguais ou maiores que 200 de um ano para outro.

Gráfico 07 - Distribuição das habilidades em Língua Portuguesa

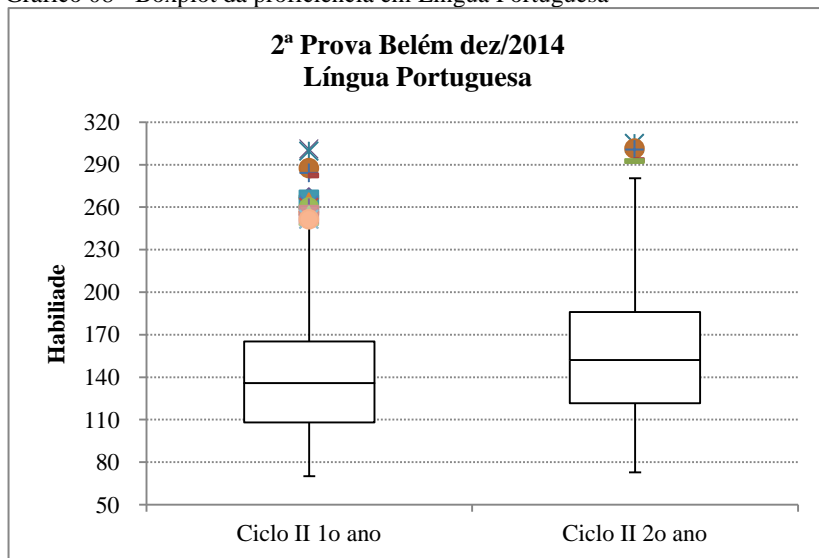


Fonte: Elaboração do autor.

No gráfico 08, observam-se as habilidades do grupo do Ciclo II 1º e 2º anos em Língua Portuguesa, em que metade dos alunos do Ciclo II 1º ano ficou abaixo de 136 pontos e metade dos alunos do Ciclo II 2º ano ficou abaixo de 153 pontos. Os pontos com grande afastamento dos demais alunos da série, *outliers*, no gráfico *boxplot* do Ciclo II 2º ano, representam 48 alunos, concentrados de 292,61 pontos a 304,91 pontos; já no Ciclo II 1º ano, representam 53 alunos, distribuídos de 251,31 pontos a 301,14 pontos da escala. Salienta-se que esses *outliers* são alunos com desempenho bem superior aos demais alunos de sua série.

Na Tabela 4.06 são apresentadas as principais estatísticas para os dados de Língua Portuguesa na 2ª Prova Belém 2014, onde o *alfa de Cronbach* (α), que mede a consistência interna dos formulários aplicados, ficou em 0,830 para o Ciclo II 2º ano e 0,815 para o Ciclo II 1º ano indicando boa consistência interna ($> 0,8$), segundo Pestana & Gajero (2014). Como já se esperava os alunos do Ciclo II 1º ano tem proficiência média (139,3) menor que os alunos do Ciclo II 2º ano (156,5), além disso, as médias são maiores que as medianas indicando uma assimetria positiva para a curva dos dados.

Gráfico 08 - Boxplot da proficiência em Língua Portuguesa



Fonte: Elaboração do autor.

No Apêndice C, observa-se que apenas o bisserial do item LP54 ficou próximo de zero ($biss_{54}=0,07$), porém para o grupo 3 (Ciclo II 1º ano) esse valor aumentou ficando próximo dos outros dois itens (LP66 e

LP93) que ficaram com os valores de bisserial abaixo de 0,15 ($biss_{66}=0,139$ e $biss_{93}=0,119$), critério proposto pelo INEP (Andrade, Laros & Gouveia, 2010). Após confirmar o gabarito desses três itens, partiu-se para a fase 2 do Bilog-MG de estimativa dos parâmetros. Como na fase 2, todos os itens foram calibrados, então optou-se por mantê-los para a estimativa das proficiências na fase 3.

Dentre os 39 itens disponíveis da Prova Brasil, seis destes foram aplicados como itens de ligação (itens: 02, 14, 15, 16, 32, 34) escolhidos pelos especialistas que consideram os melhores que se adequaram aos alunos do Ciclo II 2º ano, na 2ª Prova Belém de 2014 de Língua Portuguesa, os demais entraram no processo de calibração, mas não consta nas tabelas por não terem sido respondidos pelos alunos da Prova Belém.

Tabela 4.06 - Estatísticas da Proficiência em Língua Portuguesa.

Estatísticas	Ciclo II 2º ano	Ciclo II 1º ano
Média (m)	156,5	139,3
Desvio Padrão	44,5	39,6
Mínimo	72,88	70,186
1º quartil (25%)	121,68	108,13
Mediana (50%)	152,23	135,88
3º quartil (75%)	186,03	165,20
Máximo	304,91	301,14
Número de respostas (n)	5.966	5.305
Média de Alfa de Cronbach (α)	0,830	0,815

Fonte: Elaboração do autor, a partir do R Projects.

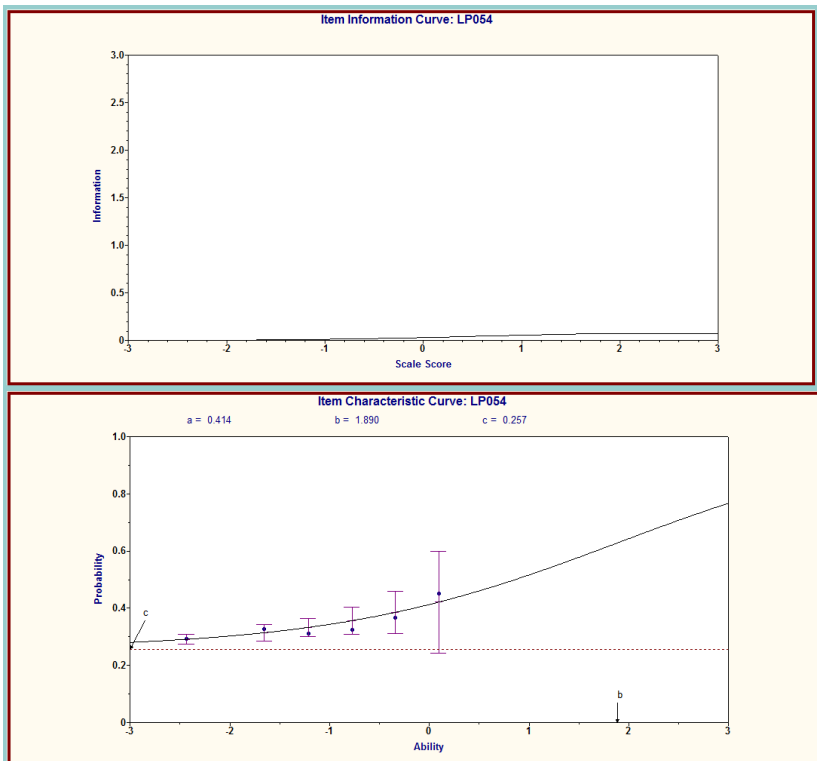
No apêndice D, observa-se que os itens apresentaram boas estimativas, em relação ao parâmetro “a” com boa discriminação ($a > 0,4$ para ogiva normal equivalente na logística $a > 0,7$); em relação ao parâmetro “b” com exceção do item LP54 ($b_{54} = 1,89$), todos os demais itens mostraram valores negativos indicando que os itens têm dificuldade inferior à média zero do grupo de referência (9º ano EF); em relação ao parâmetro “c” os valores ficaram entre 4% e 33%, sendo que todos os itens estando dentro de um valor aceitável, não díspares da probabilidade de acertos ao acaso com quatro alternativas (25%).

O gráfico 09 mostra a curva característica do item LP54 citado no parágrafo anterior. O item apresenta discriminação aceitável ($a_{54} = 0,414$), sendo o mais difícil ($b_{54} = 1,89$) da prova e associado ao maior erro (Erro-padrão $b_{54}=0,71$). Além disso, sua curva de informação mostra pouca contribuição desse item para a informação total iniciando depois

do zero, média do grupo de referência, e distante da região de maior informação dada pelo teste.

O gráfico 10 mostra a Curva Característica e curva de informação do item com a maior informação ao teste (LP98). Pode-se notar que esse item é relativamente “fácil” ($b_{98}=-1,429$), ou seja, a Curva Característica do Item está mais para a esquerda da proficiência adequada para o aluno do 5º ano (200 pontos equivalente a cerca de -0,5 nessa escala), e, portanto, alunos com habilidade inferior ao adequado no 5º ano EF tem probabilidade de acertos superiores a 0,65. O item LP98 junto com o item LP92 e LP93 apresentam os maiores valores para o parâmetro “a” ($a_{92}=2,527$; $a_{93}=1,918$; $a_{98}=2,472$), sendo, portanto, os itens com melhor discriminação na prova, porém são os itens que apresentam os maiores erros associados ao parâmetro “a” (Erro-padrão $a_{92}=0,423$; Erro-padrão $a_{93}=0,437$ e Erro padrão $a_{98}=0,222$).

Gráfico 09 - CCI e CII do item LP54



Fonte: Programa Bilog-MG.

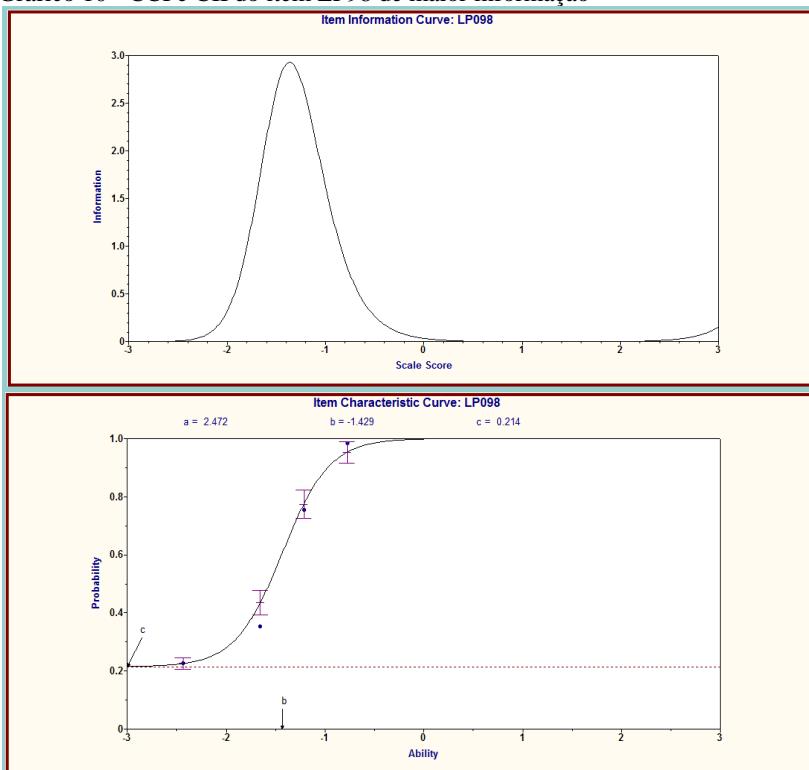
Tabela 4.07 - Itens que chamaram a atenção em Língua Portuguesa.

Item	Bisserial	A	EP* a	B	EP* b	c	EP* c
LP54	0,070	0,414	0,136	1,890	0,710	26%	0,023
LP66	0,139	2,076	0,309	-0,225	0,056	20%	0,009
LP92	0,152	2,527	0,423	-0,839	0,049	29%	0,010
LP93	0,119	1,918	0,437	-0,483	0,099	28%	0,011
LP98	0,465	2,472	0,222	-1,429	0,027	21%	0,012

Fonte: Elaboração do autor, a partir do programa bilog-MG.*EP = Erro-padrão.

Na tabela 4.07, listaram-se alguns itens que chamaram a atenção e os valores dos parâmetros e do coeficiente bisserial encontram-se resumidos. A tabela completa dos valores dos itens pode ser visualizada nos apêndice C e D.

Gráfico 10 - CCI e CII do item LP98 de maior informação



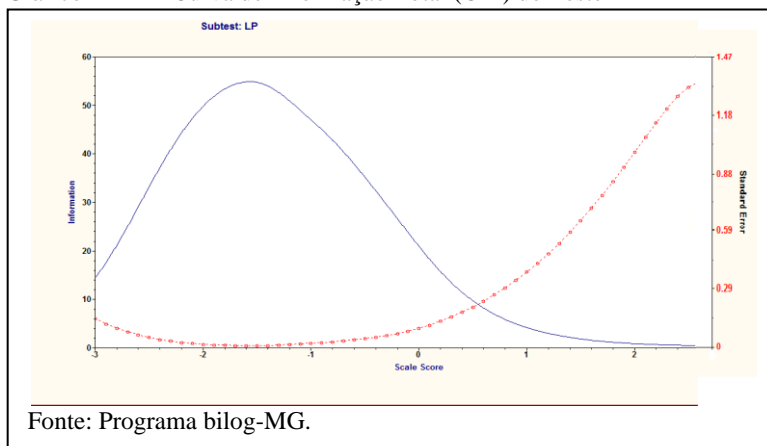
Fonte: produzido pelo programa Bilog-MG.

Observa-se na Curva de Informação Total (CIT) do teste em Língua Portuguesa, gráfico 11, gerada com os dados do presente estudo,

que a linha pontilhada representa o erro padrão associada à CIT, e é inversamente proporcional à linha contínua, ou seja, quanto mais informativa a região do instrumento, menor é o erro de estimativa.

O Gráfico 11 mostra, também, que aproximadamente até próximo do valor 0,5 da escala a curva do erro-padrão da informação é baixa, região em que o instrumento é mais informativo, ou seja, a capacidade de medir a proficiência dos alunos é melhor para os itens posicionados até este ponto.

Gráfico 11 - LP Curva de Informação Total (CIT) do Teste



4.3 INTERPRETAÇÃO DOS ITENS DE LÍNGUA PORTUGUESA

A tabela 4.08 mostra os níveis propostos pela escala SAEB para os alunos do 5º ano do Ensino Fundamental, equivalente aos alunos do Ciclo II 2º ano, iniciando no nível 0 para os alunos com proficiência abaixo de 125 pontos, até o nível 9 para alunos com proficiências iguais ou acima de 325, conforme boletim de desempenho da escola atualizado em junho de 2016 pelo INEP.

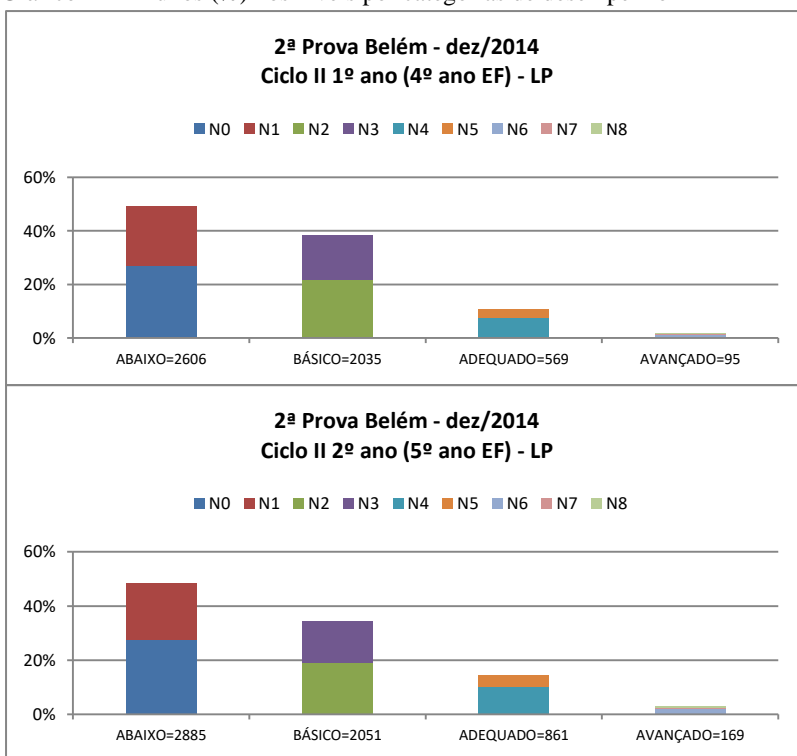
Uma vez que a Prova Brasil não é aplicada para alunos do 4º ano do ensino fundamental, pode-se, conforme proposto pelo SisPAE, considerar os níveis para esses alunos com uma defasagem de 15 pontos em relação aos níveis do 5º ano. A tabela 4.08 mostra os níveis para o 4º ano, variando de $\frac{1}{2}$ desvio padrão.

Tabela 4.08 - Língua Portuguesa: níveis de desempenho 4º e 5º ano

Níveis	Proficiência 5º ano	Proficiência 4º ano
0	Menor que 125	Menor que 110
1	125 a <150	110 a <135
2	150 a <175	135 a <160
3	175 a <200	160 a <185
4	200 a <225	185 a <210
5	225 a <250	210 a <235
6	250 a <275	235 a <260
7	275 a <300	260 a <285
8	300 a <325	285 a <310
9	325 ou mais	310 ou mais

Fonte: INEP, boletim, jun/2016 e SisPAE 2014.

Gráfico 12 - Alunos (%) nos níveis por categorias de desempenho



Fonte: Elaboração do autor.

No Gráfico 12 observam-se as categorias de desempenhos definidas a partir de um desvio padrão (abaixo do básico, básico,

adequado, avançado) e os níveis nomeados por N0, N1, ..., N9 propostos em intervalos de meio desvio padrão. Em ambos os anos, quase metade dos alunos se encontram abaixo do básico, com quase um quarto no nível 0.

Na interpretação da escala encontram-se inclusos os procedimentos: (a) identificação de itens âncora, e (b) apresentação desses itens a um painel de especialistas. Como para o SAEB, um item é considerado âncora em um determinado nível, quando “*o percentual de acerto do item no nível considerado e nos níveis acima dele é maior que 65% e o percentual de acerto do item nos níveis anteriores são menores que 65%*” (INEP/MEC/Brasil, 2002), ou seja, basicamente o primeiro critério apresentado por Andrade, Tavares & Valle (2000) no capítulo anterior. Os itens foram posicionados em ½ desvio padrão, no nível onde a probabilidade de acerto fosse 65%, ou seja, no ponto da escala em que o item atinge exatamente 65%.

Na tabela 4.09, os 38 itens aplicados no Ciclo II 2º ano estão posicionados nos níveis propostos pela escala do SAEB, sendo a seguinte quantidade de itens por nível: N0, 3 itens; N1, 5 itens; N2, 12 itens; N3, 9 itens, N4, 3 itens, N5, 3 itens; e N9=1 item. Nenhum item foi posicionado nos níveis: N6, N7 e N8.

Tabela 4.09 - Posicionamento dos itens de LP - 5º ano (250;50)

Item	Descritor	b	Ponto na escala SAEB c/ probabilidade 65%	Nível (½DP)
LP002	D01	196	210,7	N4=[200,225)
LP014	D02	168	175,1	N3=[175,200)
LP015	D10	220	228,8	N5=[225,250)
LP016	D14	220	228,0	N5=[225,250)
LP032	D04	158	168,5	N2=[150,175)
LP034	D05	175	183,3	N3=[175,200)
LP040	D06	132	141,8	N1=[125,150)
LP041	D01	100	110,2	N0<125)
LP042	D04	120	131,9	N1=[125,150)
LP043	D09	138	153,6	N2=[150,175)
LP044	D15	155	179,2	N3=[175,200)
LP045	D05	116	127,8	N1=[125,150)
LP046	D11	207	212,2	N4=[200,225)
LP047	D03	168	173,9	N2=[150,175)

Item	Descritor	b	Ponto na escala SAEB c/ probabilidade 65%	Nível (½DP)
LP048	D07	203	209,2	N4=[200,225)
LP049	D07	190	187,8	N3=[175,200)
LP050	D12	191	193,1	N3=[175,200)
LP051	D08	120	124,5	N0<125)
LP052	D02	171	173,2	N2=[150,175)
LP053	D13	148	153,1	N2=[150,175)
LP054	D10	354	363,1	N9≥325
LP055	D14	121	132,5	N1=[125,150)
LP056	D09	180	186,6	N3=[175,200)
LP057	D15	163	174,5	N2=[150,175)
LP058	D11	161	179,7	N2=[150,175)
LP059	D03	111	123,3	N0<125)
LP060	D05	136	153,2	N2=[150,175)
LP061	D04	144	153,2	N2=[150,175)
LP062	D01	129	134,6	N1=[125,150)
LP063	D06	172	188,9	N3=[175,200)
LP064	D12	153	153,9	N2=[150,175)
LP065	D08	161	167,1	N2=[150,175)
LP066	D10	238	241,7	N5=[225,250)
LP067	D13	170	172,0	N2=[150,175)
LP068	D14	154	158,3	N2=[150,175)
LP069	D12	156	161,6	N2=[150,175)
LP070	D07	171	181,3	N3=[175,200)
LP071	D02	173	175,9	N3=[175,200)

Fonte: Elaboração do autor.

Na tabela 4.10, os 44 itens aplicados no Ciclo II 1º ano estão posicionados nos níveis propostos pela escala do SAEB, sendo que a quantidade de itens para os níveis foram: N1, 11 itens; N2, 12 itens; N3, 9 itens, N4, 9 itens, N5, 2 itens; e N9=1 item. Nenhum item foi posicionado nos níveis: N0, N6, N7 e N8.

Tabela 4.10 - Posicionamento dos itens de LP - 4º ano (250;50)

Item	Descritor	b	Ponto na escala SAEB c/ probabilidade 65%	Nível ($\frac{1}{2}DP$)
LP041	D01	100	110,2	N1=[110,135)
LP042	D04	120	131,9	N1=[110,135)
LP047	D03	168	173,9	N3=[160,185)
LP048	D07	203	209,2	N4=[185,215)
LP051	D08	120	124,5	N1=[110,135)
LP053	D13	148	153,1	N2=[135,160)
LP054	D10	354	363,1	N9 \geq 360
LP055	D14	121	132,5	N1=[110,135)
LP056	D09	180	186,6	N4=[185,215)
LP059	D03	111	123,3	N1=[110,135)
LP062	D01	129	134,6	N1=[110,135)
LP064	D12	153	153,9	N2=[135,160)
LP069	D12	156	161,6	N3=[160,185)
LP071	D02	173	175,9	N3=[160,185)
LP072	D06	151	163,0	N3=[160,185)
LP073	D01	187	191,4	N4=[185,215)
LP074	D04	115	122,1	N1=[110,135)
LP075	D15	187	190,3	N4=[185,215)
LP076	D11	202	208,0	N4=[185,215)
LP077	D05	142	154,9	N2=[135,160)
LP078	D12	191	204,5	N4=[185,215)
LP079	D07	187	196,3	N4=[185,215)
LP080	D12	122	138,4	N2=[135,160)
LP081	D02	190	197,1	N4=[185,215)
LP082	D02	213	219,8	N5=[215,235)
LP083	D13	177	184,5	N3=[160,185)
LP084	D09	131	136,3	N2=[135,160)
LP085	D03	137	144,9	N2=[135,160)
LP086	D15	138	153,2	N2=[135,160)
LP087	D03	160	169,3	N3=[160,185)

Item	Descritor	b	Ponto na escala SAEB c/ probabilidade 65%	Nível (½DP)
LP088	D05	162	171,6	N3=[160,185)
LP089	D06	112	121,9	N1=[110,135)
LP090	D09	119	124,8	N1=[110,135)
LP091	D04	108	120,0	N1=[110,135)
LP092	D11	204	204,1	N4=[185,215)
LP093	D06	223	224,2	N5=[215,235)
LP094	D10	127	132,8	N1=[110,135)
LP095	D07	162	167,4	N3=[160,185)
LP096	D08	147	150,7	N2=[135,160)
LP097	D10	155	158,4	N2=[135,160)
LP098	D13	171	174,1	N3=[160,185)
LP099	D08	138	147,5	N2=[135,160)
LP100	D14	130	137,6	N2=[135,160)
LP101	D14	131	138,2	N2=[135,160)

Fonte: Elaboração do autor.

Depois de definidos os níveis e identificados os seus itens âncora, cabe aos especialistas da área, no caso Língua Portuguesa, explicar o significado pedagógico das respostas dadas pelos alunos àqueles itens, fazendo uma descrição do que os alunos demonstram saber por meio da análise das respostas dadas aos diferentes itens de cada nível.

Para concluir essa seção, apresenta-se a interpretação pedagógica feita pelo INEP nos seus relatórios de dois itens aplicados na Prova Brasil, que foram utilizados na Prova Belém.

O item LP32 (exemplo 01) apresenta como texto base uma fábula, gênero textual com narrativa, onde animais levam à reflexão do comportamento humano e apresenta uma estrutura simples e de pouca complexidade. Ele avalia a habilidade de inferir uma informação implícita no texto, descritor D4 da matriz de referência da Prova Brasil. De uma forma lúdica, mas nem por isso menos séria, o leitor é levado a perceber as mensagens nas entrelinhas e, portanto, inferir a informação requerida.

Exemplo 01: Item LP32 da Prova Belém – disponível em Brasil (2002)

O cão e o lobo

Um cão passeava pela floresta quando topou com um lobo magro. Aos poucos os dois fizeram amizade.

- Puxa, cachorro! Como você está gordo e bem-tratado...

— É que eu tenho um dono. Meu dono me dá três boas refeições por dia, escova meu pelo, me dá uma casa de madeira... Em troca disso, pede que eu lhe guarde a casa dos assaltantes e lhe faça uns agradados de vez em quando.

— Só isso? Mas deve ser maravilhoso ter um dono — concluiu o lobo.

O cão então convenceu o lobo a acompanhá-lo, certo de que seu dono gostaria de ter mais um animal de estimação.

Os dois andaram por um certo tempo, até que o lobo percebeu uma coleira no cachorro.

— O que é isso? — perguntou o lobo.

— Ah, isto é uma coleira. Às vezes, meu dono se irrita e me prende numa corrente. Mas é por pouco tempo, logo eu estou solto de novo.

O lobo parou, pensou um pouco... e voltou atrás. De longe, ainda falou para o cachorro:

— Não, cachorro. Não sirvo para essa vida. Eu sei que mais vale a liberdade com fome do que o luxo na prisão.

Fábula recontada por Marcia Kupstas, Sete faces da fábula. São Paulo, Moderna, 1993. (Texto adaptado.)

O que aconteceu com o lobo quando soube que o cachorro usava coleira?

- (A) Desistiu da liberdade.
- (B) Desistiu de ter um dono.
- (C) Resolveu conhecer seu dono.
- (D) Resolveu tirar a coleira do cachorro.

Tabela 4.11 - Percentual de acertos do item LP32

Prova	item	Gabarito	A	B	C	D
Brasil	Não informado	B	15%	55%	10%	12%
Belém	LP32	B	20%	50%	14%	15%

Fonte: Relatório SAEB 2001 (INEP/MEC/Brasil, 2002).

Os percentuais de marcação foram próximos, e na mesma ordem, sendo que a maior porcentagem dos alunos marcou a alternativa correta (B) e o menor percentual o distrator (C). O distrator A foi o mais selecionado pelos participantes. O item está classificado no nível,

N2=[150,175). A partir daí, provavelmente os alunos são capazes de inferir informações implícitas em fábulas.

Exemplo 02: Item LP34 da Prova Belém, disponível em devolutivas.

Leia os quadrinhos abaixo e responda.



A quem Hagar está dando ordens?

- (A) Alunos.
- (B) Aniversariantes.
- (C) Marinheiros.
- (D) Soldados.

O texto base, do exemplo 2, é uma tirinha, gênero textual que combina linguagem verbal e não verbal, com balões de formas variadas que servem de suporte ao diálogo entre as personagens. O objetivo comunicativo desse gênero é, em geral, divertir o leitor. O item exige do participante inferir com quem a personagem (Hagar) faz uma interlocução, uma vez que os interlocutores não são mostrados. Para responder corretamente, ele deve reconhecer as palavras do campo semântico da área da navegação, exemplificadas nos comandos da personagem ("Icem a vela!", "Levantar âncoras!" e "Virar a bombordo!"), e relacioná-las à imagem do último quadrinho, em que a personagem segura uma miniatura de barco viking. Desse modo, deduz que os interlocutores são tripulantes imaginários do barco.

Tabela 4.12 - Percentual de acertos do item LP34

Prova	item	Gabarito	A	B	C	D
Brasil	203	C	17%	9%	52%	13%
Belém	LP34	C	20%	13%	47%	19%

Fonte: Elaboração do autor, a partir do Devolutivas INEP.

Os distratores A e D atraíram, respectivamente, 17% e 13% dos respondentes na Prova Brasil, tabela 4.12. O item está classificado no nível, N3=[175,200]. A partir daí, provavelmente, os estudantes são capazes de inferir interlocutores em tira.

4.4 BASE CLONE DE MATEMÁTICA

Da mesma forma que em Língua Portuguesa, só estava acessível os itens de Matemática com os seus parâmetros (a, b, c) calibrados na escala SAEB, sendo necessário simular respostas associadas a tais parâmetros. Assim, realizaram-se simulações com amostras de 10mil, 30mil, 50mil, 80mil e 100 mil indivíduos utilizando os procedimentos descritos na seção 3.3.1, e compararam-se os valores dos parâmetros (a, b, c) estimados pelo Bilog-MG com os valores reais para escolher a melhor amostra.

A escolha da amostra, a ser utilizada, se deu através da análise da linha de tendência expressa na equação linear ($y = \alpha x + \beta$), onde y representava os valores estimados pela amostra e x os valores reais dos parâmetros; e do coeficiente de determinação (R^2), de forma que quanto mais o R^2 se aproxima de 1, mais explicativo é o modelo e melhor ele se ajusta à amostra. Essa análise gráfica foi realizada em cada amostra e para cada um dos parâmetros.

Para auxiliar a comparação das equações das linhas de tendências e do R^2 , organizaram-se em tabelas os valores encontrados. A tabela 4.13, apresenta os indicadores utilizados para avaliar o ajuste da amostra aos valores reais do parâmetro “a” dos itens de Matemática. Observa-se que o ajuste se deu melhor em relação aos dados de 100mil respondentes, seja de acordo com valor de R^2 , seja em relação aos coeficientes da equação, pois o coeficiente linear (β) mais se aproxima de zero e o coeficiente angular (α) de 1.

Tabela 4.13- MT comparação dos dados simulados - parâmetro “a”

Parâmetro “a”	Equação ($\alpha x + \beta$)	α	β	R^2
10mil	0,9614x+0,0715	0,96	0,07	0,95
30mil	0,9931x+0,0238	0,99	0,02	0,98
50mil	0,9608x+0,0498	0,96	0,05	0,99
80mil	0,9855x+0,0222	0,99	0,02	0,99
100mil	1,0056x+0,0059	1,01	0,01	0,99

Fonte: Elaboração do autor.

Na tabela 4.14, observa-se que o parâmetro “b” dos itens de Matemática se ajustou melhor aos dados de 100mil respondentes, de acordo com valor de R^2 ; aos dados da amostra de 80mil, em relação ao coeficiente angular; e aos dados de 30 mil pelo coeficiente linear.

Tabela 4.14 - MT comparação dos dados simulados - parâmetro “b”

Parâmetro “b”	Equação ($\alpha x + \beta$)	α	β	R²
10mil	0,9443x-0,0225	0,94	0,02	0,95
30mil	0,9705x+0,0036	0,97	0,00	0,98
50mil	0,9568x-0,0434	0,96	0,04	0,99
80mil	0,9847x-0,0108	0,98	0,01	0,99
100mil	0,9829x-0,0096	0,98	0,01	0,99

Fonte: Elaboração do autor.

O parâmetro “c” dos itens de Matemática se ajustou melhor em relação aos dados de 100mil respondentes, quer seja de acordo com valor de R², quer seja em relação aos coeficientes da equação, pois o coeficiente linear mais se aproxima de zero e o coeficiente angular de 1. Os valores encontram-se abaixo na tabela 4.15.

Tabela 4.15 - MT Comparação dos dados simulados - parâmetro “c”

Parâmetro “c”	Equação ($\alpha x + \beta$)	α	β	R²
10mil	0,6248x+0,0741	0,62	0,07	0,57
30mil	0,7728x+0,0440	0,77	0,04	0,86
50mil	0,7832x+0,0335	0,78	0,03	0,85
80mil	0,8904x+0,0195	0,89	0,02	0,90
100mil	0,9473x+0,0131	0,95	0,01	0,94

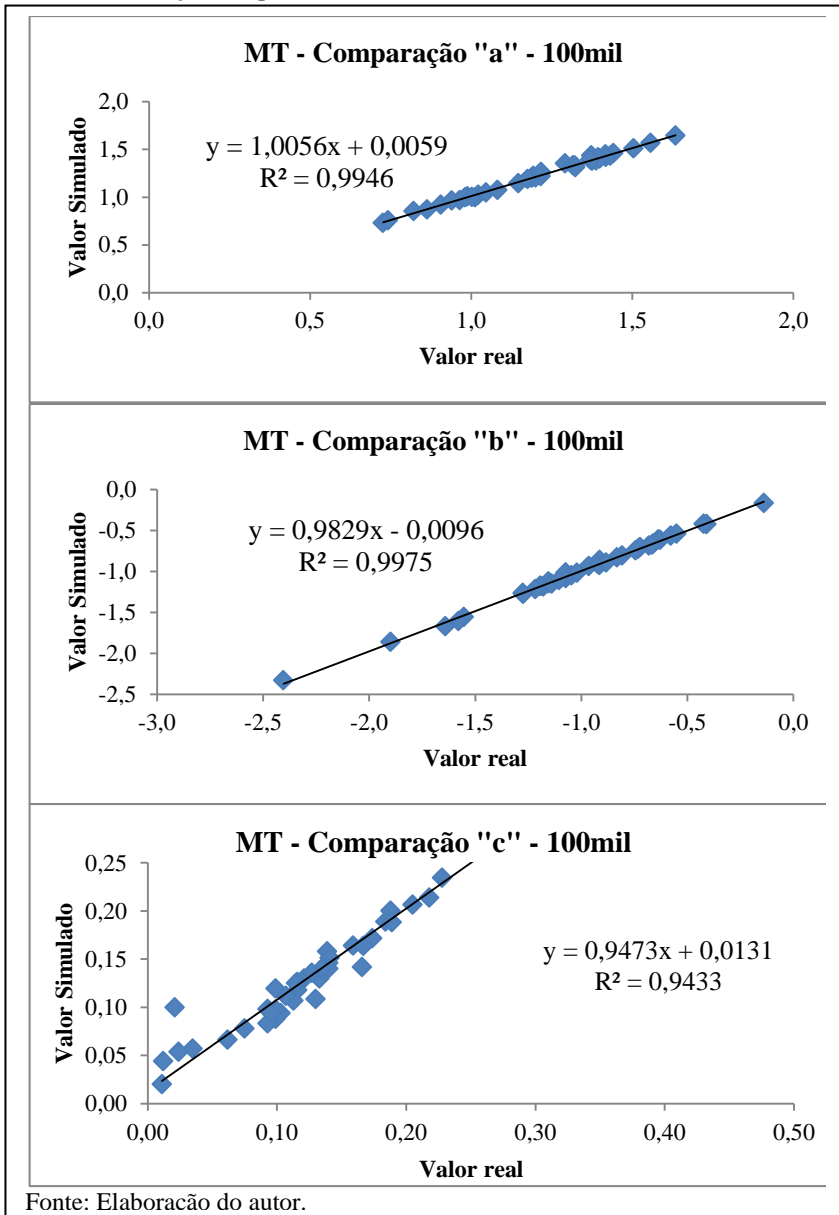
Fonte: Elaboração do autor.

Comparando as várias amostras optou-se para os dados de 100mil respondentes tanto para Língua Portuguesa como para Matemática, pois se ajustava bem em todos os parâmetros e mantinha a mesma quantidade de dados a serem utilizadas no processo de calibração em ambas as disciplinas.

No gráfico 13, pode observar a relação entre os valores reais e os simulados para os parâmetros “a”, “b”, “c” na amostra de 100 mil indivíduos.

Conforme relatório do SisPAE 2014, em Matemática a média de proficiência dos alunos na Semec ficou no nível abaixo do básico, a diferença entre o desempenho dos alunos do 5º ano EF e do 9º ano EF nesse nível é de 50 pontos (5º ano EF = 175 e 9º ano EF = 225) numa escala (250;50), ou seja, de um (01) desvio padrão.

Gráfico 13 - Relação dos parâmetros a, b, c na amostra 100mil - MT.



Assim, utilizando o mesmo procedimento usado em Língua Portuguesa mudou-se a escala (0;1) em que se encontravam os

parâmetros do grupo de referência da 8ª série, para uma escala (-1;1) para um grupo de referência do 5º ano EF.

Fazendo a transformação linear nos parâmetros dos itens do SAEB de um desvio padrão observou-se que os parâmetros mantiveram a mesma precisão, conforme mostrado na Tabela 4.16.

Tabela 4.16 - MT amostra de 100mil (Escala -1,1) grupo de 5º ano

Parâmetro	Equação ($\alpha x + \beta$)	α	β	R^2
A	0,9775x+0,0362	0,98	0,04	0,9933
B	0,9774x+0,0431	0,98	0,04	0,998
C	0,9233x+0,013	0,92	0,01	0,9768

Fonte: Elaboração do autor.

4.5 ESCALA DE MATEMÁTICA

A escala permite distribuir os alunos em categorias de desempenhos qualitativos, em relação ao desempenho adequado esperado. Tanto o SAEB e o SisPAE no 5º ano EF, a partir de 225 ou mais pontos consideram os alunos com desempenho adequado ao seu ano de escolarização em Matemática. Para o 4º ano do ensino, o SisPAE estabelece que os alunos a partir de 210 pontos atingem a aprendizagem adequada. Assim, a classificação dos alunos em Matemática no 4º e 5º ano do Ensino fundamental em 4 categorias de desempenho pode ser expressa na tabela 4.17.

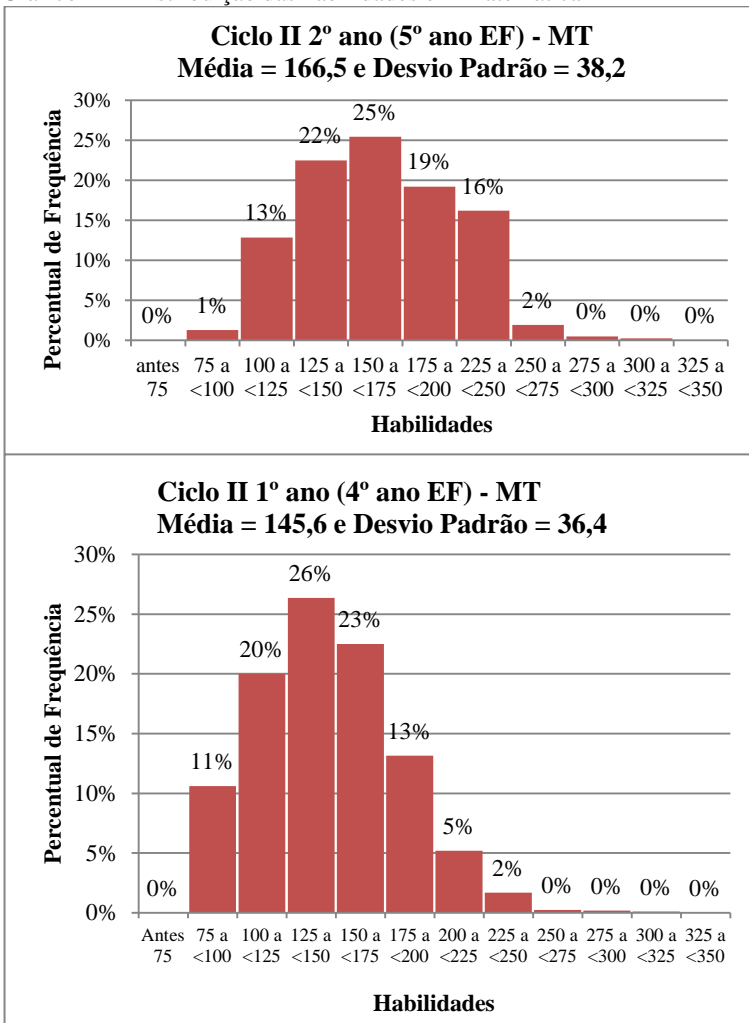
Tabela 4.17 - Categorias de desempenho em Matemática

Categoria	4º ano E.F.	5º ano E.F.
Abaixo do Básico	menor que 160	menor que 175
Básico	160 a <210	175 a <225
Adequado	210 a <260	225 a <275
Avançado	Igual ou maior a 260	Igual ou maior a 275

Fonte: SisPAE 2014 (<https://sispae.vunesp.com.br/Padroes.aspx>, 16/12/2016).

O Gráfico 14 mostra as distribuições das habilidades em Matemática dos alunos do Ciclo II 1º ano e do Ciclo II 2º ano em 2014 na Prova Belém, na escala (250;50). Pode-se observar que a porcentagem de alunos do Ciclo II 2º ano com habilidade maior ou igual a 225 (18%) - aprendizagem adequada no Ciclo II 5º ano EF - foi maior que a porcentagem de alunos do Ciclo II 1º ano com mesma faixa de habilidade (2%), isto mostra uma diferença positiva (16%) entre a porcentagem de alunos com habilidades igual ou maior que 225 de um ano para outro.

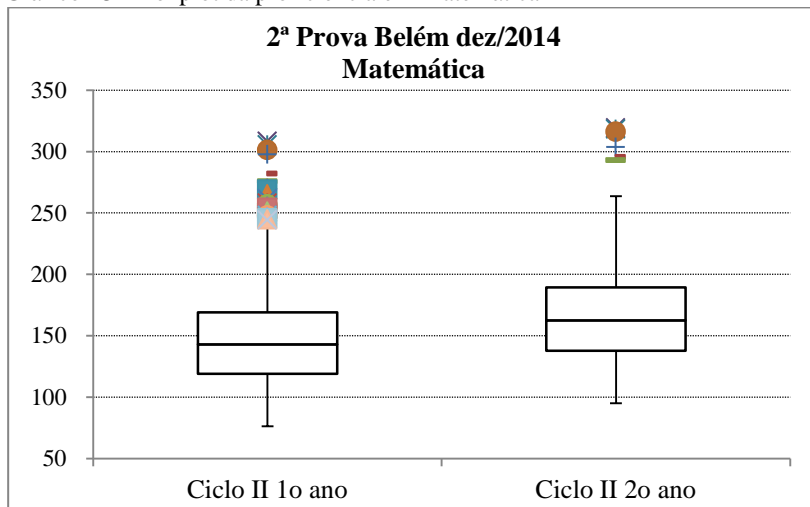
Gráfico 14 - Distribuição das habilidades em Matemática



Fonte: Elaboração do autor.

No Gráfico 15, observam-se as habilidades do grupo do Ciclo II 1º e 2º anos em Matemática, em que metade dos alunos do Ciclo II 1º ano ficaram abaixo de 150 pontos e metade dos alunos do Ciclo II 2º ano ficaram acima de 160 pontos. Os pontos *outliers*, no gráfico *boxplot* do Ciclo II 2º ano, representam 16 alunos, com habilidade concentrada entre 293,1 a 319,4 pontos; já no Ciclo II 1º ano, representam 48 alunos, distribuídos na escala de 244,3 a 308,5 pontos.

Gráfico 15 - Boxplot da proficiência em Matemática



Fonte: Elaboração do autor.

Tabelas 4.18 - Estatísticas da Proficiência em Matemática

Estatísticas	Ciclo II 2º ano	Ciclo II 1º ano
Média (m)	166,5	145,6
Desvio Padrão	38,2	36,4
Mínimo	94,89	76,26
1º quartil (25%)	138,1	118,9
Mediana (50%)	163,1	143,0
3º quartil (75%)	189,9	169,1
Máximo	319,4	308,5
Número de respostas (n)	5.966	5.305
Média de Alfa de Cronbach (α)	0,755	0,755

Fonte: Elaboração do autor.

Na Tabela 4.18 são apresentadas as principais estatísticas para os dados de Matemática na 2ª Prova Belém 2014, onde o *alfa de Cronbach* (α), ficou em 0,755 indicando boa consistência interna. Como já se esperava os alunos do Ciclo II 1º ano têm uma proficiência média (145,6) menor que os alunos do Ciclo II 2º ano (166,5), além disso, as médias são maiores que as medianas indicando uma assimetria positiva para a curva dos dados.

A correlação bisserial dos itens de Matemática indica que quanto maior o seu valor, maior a capacidade do item de discriminar grupos de alunos que construíram determinada proficiência, daqueles que não a

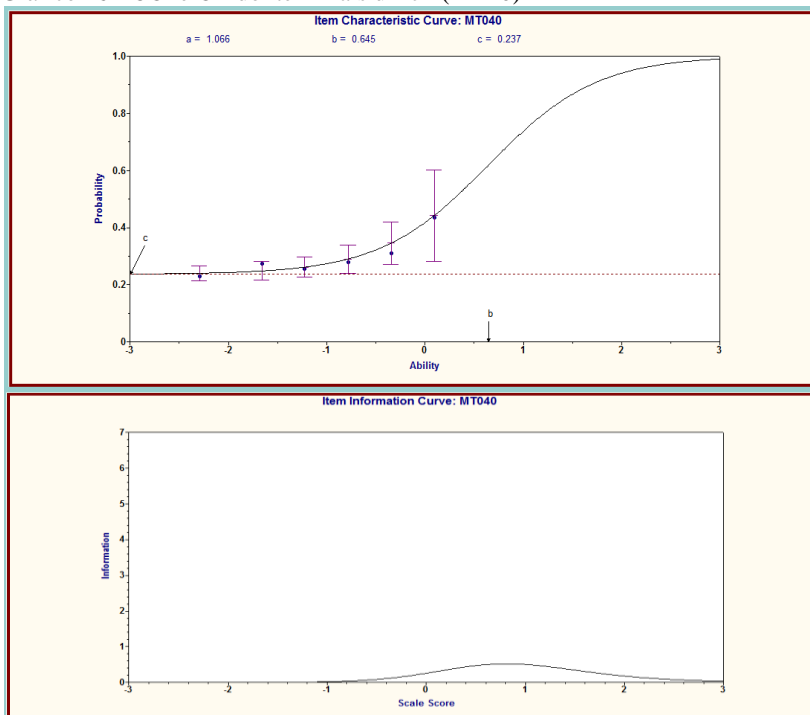
construíram. Observa-se, no apêndice E, que quatro itens não ultrapassaram 0,15, critério proposto pelo INEP (Andrade, Laros & Gouveia, 2010) para os valores bisserial: no item MT49 o valor foi negativo em todos os grupos; e nos itens MT40 e MT57 não chegaram a 0,10; e no item MT102 o valor foi 0,15.

Inicialmente, confirmou-se o gabarito desses quatro itens e houve ajuste em outros três itens com problemas de gabarito (MT46, MT55 e MT66): o item MT55 ficou repetido no bloco MT1B1 (6ª questão igual a 11ª questão), e os itens MT46 e MT66 tiveram troca de gabarito, sendo necessário recodificar as respostas. Logo em seguida, iniciou-se a fase 2 do bilog-MG de estimativa dos parâmetros. Como na fase 2, todos os itens foram calibrados, optou-se por manter os itens e estimar as proficiências.

Dos 39 itens de Matemática disponíveis da Prova Brasil foram aplicados seis de ligação (itens: MT02, MT04, MT09, MT20, MT34, MT35) na 2ª Prova Belém de 2014 do Ciclo II 2º ano, os demais entraram no processo de calibração, mas não constam nas tabelas devido não terem sido aplicados aos alunos. No apêndice F, observa-se que os itens apresentaram boas estimativas em relação ao parâmetro “a”, mostrando boa discriminação; em relação ao parâmetro “b” com exceção dos itens MT40, MT49 e MT59 todos os demais itens mostraram valores negativos, indicando que os itens apresentam dificuldade inferior à média zero do grupo de referência (9º ano EF); em relação ao parâmetro “c” os valores ficaram entre 7% e 44%, sendo que apenas o item MT57 ($c=44\%$) ficou com o valor da probabilidade de acertos ao acaso próximo do dobro da probabilidade de acerto casual com quatro alternativas (25%), isso ocorreu devido o distrator da alternativa D ter atraído poucos alunos (4%) e mais de 80% dos alunos terem marcado o gabarito ($b=47\%$) ou apenas um dos distratores ($c=35\%$).

O gráfico 16 mostra a curva característica (CCI) do item MT40 citado no parágrafo anterior. O item apesar de bisserial baixo (bisserial $MT40=0,092$), apresenta boa discriminação ($a_{40} = 1,066$), e é considerado o item mais difícil ($b_{40} = 0,645$) da prova, e o maior erro (Erro-padrão $b_{40} = 0,273$). Além disso, a curva de informação mostra pouca contribuição desse item para a informação total iniciando depois do zero, média do grupo de referência, e distante da região de maior informação dada pelo teste.

Gráfico 16 - CCI e CII do item mais difícil (MT40)



Fonte: produzido pelo programa bilog-MG.

Os itens MT49, MT94 e MT102 encontram-se entre os itens com os maiores valores para o parâmetro “a” ($a_{49}=2,990$; $a_{94}=3,430$; $a_{102}=2,245$), portanto estão entre os itens com melhor discriminação na prova, porém são os itens que apresentam os maiores erros associados ao parâmetro “a” (Erro-padrão_ a_{49} =0,723; Erro-padrão_ a_{94} =0,529; Erro-padrão_ a_{102} =0,387).

O Gráfico 17 mostra a Curva Característica e curva de informação do item (MT75) com a maior informação ao teste. Pode-se notar que esse item é fácil ($b_{75}=-2,458$), ou seja, a Curva Característica do Item está à esquerda da proficiência adequada para o aluno do 5º ano (225 pontos equivalem aproximadamente a -0,25 nessa escala), e, portanto, alunos com habilidade inferior ao adequado no 5º ano EF tem alta probabilidade de acerto desse item.

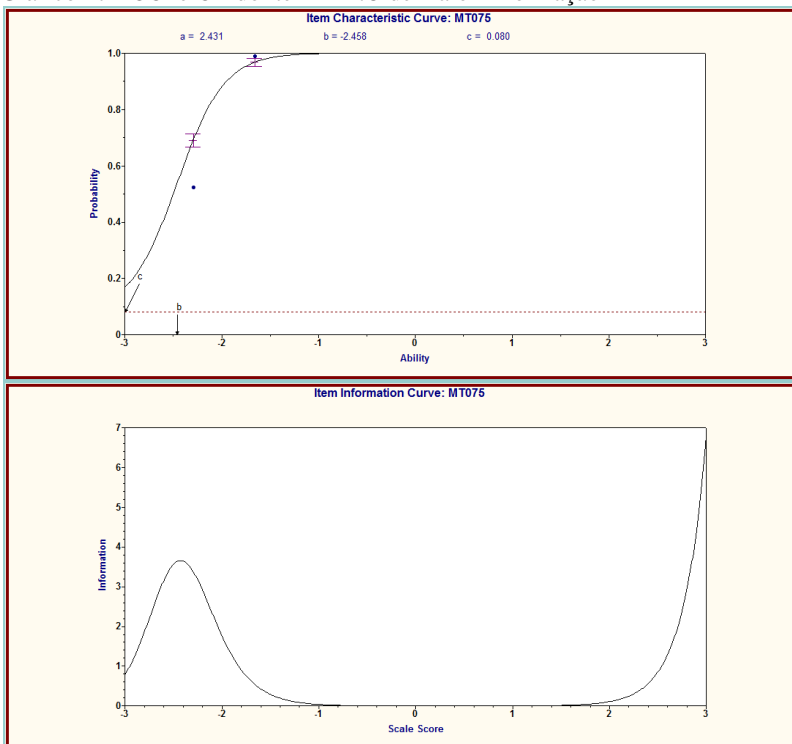
Na tabela 4.19 listaram-se alguns itens que chamaram a atenção e resumiram-se os valores dos parâmetros e do coeficiente bisserial. A tabela completa dos valores pode ser visualizada nos apêndice E e F.

Tabela 4.19 - Itens que chamaram a atenção em Matemática.

Item	bisserial	a	EP* a	b	EP* b	C	EP* c
MT40	0,092	1,066	0,302	0,645	0,273	23,7%	0,013
MT49	-0,037	2,990	0,723	0,141	0,072	19,9%	0,006
MT57	0,079	1,584	0,379	-0,017	0,113	44,4%	0,012
MT94	0,159	3,430	0,529	-0,839	0,035	20,2%	0,009
MT102	0,150	2,245	0,387	-0,636	0,061	21,1%	0,010

Fonte: Programa bilog-MG. *EP: Erro-padrão

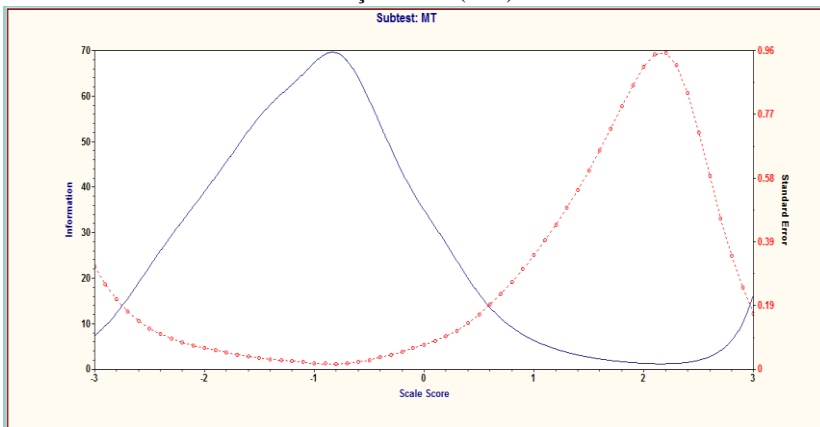
Gráfico 17 - CCI e CII do item MT75 de maior informação



Fonte: programa bilog-MG.

Observa-se na Curva de Informação Total (CIT) do teste em Matemática, gráfico 18, que a partir do valor 0,5 da escala a curva do erro-padrão da informação em pontilhado ultrapassa a curva da informação contínua, informando que o instrumento traz mais informação nas proficiências dos alunos posicionados antes deste ponto.

Gráfico 18 - MT Curva de Informação Total (CIT) do teste



Fonte: Programa Bilog-MG.

4.6 INTERPRETAÇÃO DOS ITENS DE MATEMÁTICA

Interpretar uma escala de desempenho significa escolher alguns pontos na escala que possam descrever as habilidades que os alunos demonstram possuir quando situados em torno desses pontos. No caso, na escala SAEB os níveis foram definidos variando de $\frac{1}{2}$ desvio padrão, iniciando no 5º ano do Ensino Fundamental em Matemática, com nível 0 para os alunos com proficiências menores que 125 até o nível 10 para alunos com proficiências iguais ou acima de 350, conforme boletim de desempenho da escola atualizado em junho de 2016 pelo INEP.

Tabela 4.20 - Matemática: escala de níveis 4º e 5º ano

Níveis	Proficiência 5º ano	Proficiência 4º ano
0	Menor que 125	Menor que 110
1	125 a <150	110 a <135
2	150 a <175	135 a <160
3	175 a <200	160 a <185
4	200 a <225	185 a <210
5	225 a <250	210 a <235
6	250 a <275	235 a <260
7	275 a <300	260 a <285
8	300 a <325	285 a <310
9	325 a <350	310 a < 335
10	350 ou acima	335 ou mais

Fonte: INEP, boletim junho/2016 e Sispae 2014.

Como o Nível de desempenho adequado para a 5º ano do ensino fundamental, em Matemática segundo a escala do SAEB, é o nível 5 e em transição para o 6 (225 a 275 pontos), INEP/MEC/Brasil (2002), na categoria abaixo do básico posiciona-se os níveis 0, 1 e 2.

Uma vez que a Prova Brasil não é aplicada para alunos do 4º ano do ensino fundamental, pode-se conforme proposto pelo SisPAE, considerar os níveis para esses alunos em Matemática com uma defasagem de 15 pontos em relação aos níveis do 5º ano. A tabela 4.20 mostra os níveis propostos para a escala do SAEB em Matemática para o 4º e 5º ano EF, variando de ½ desvio padrão, articulados com o posicionamento do SisPAE.

Na tabela 4.21, os 43 itens aplicados no Ciclo II 1º ano (4º ano EF) estão posicionados nos níveis propostos pela escala do SAEB, sendo a seguinte quantidade de itens por nível: N1, 5 itens; N2, 10 itens; N3, 10 itens, N4, 8 itens, N5, 8 itens; N6, 1 item; e N7, 1 item. Nenhum item foi posicionado nos níveis: N0, N8, N9 e N10.

Tabela 4.21 - Posicionamento dos itens de Matemática - 4º ano

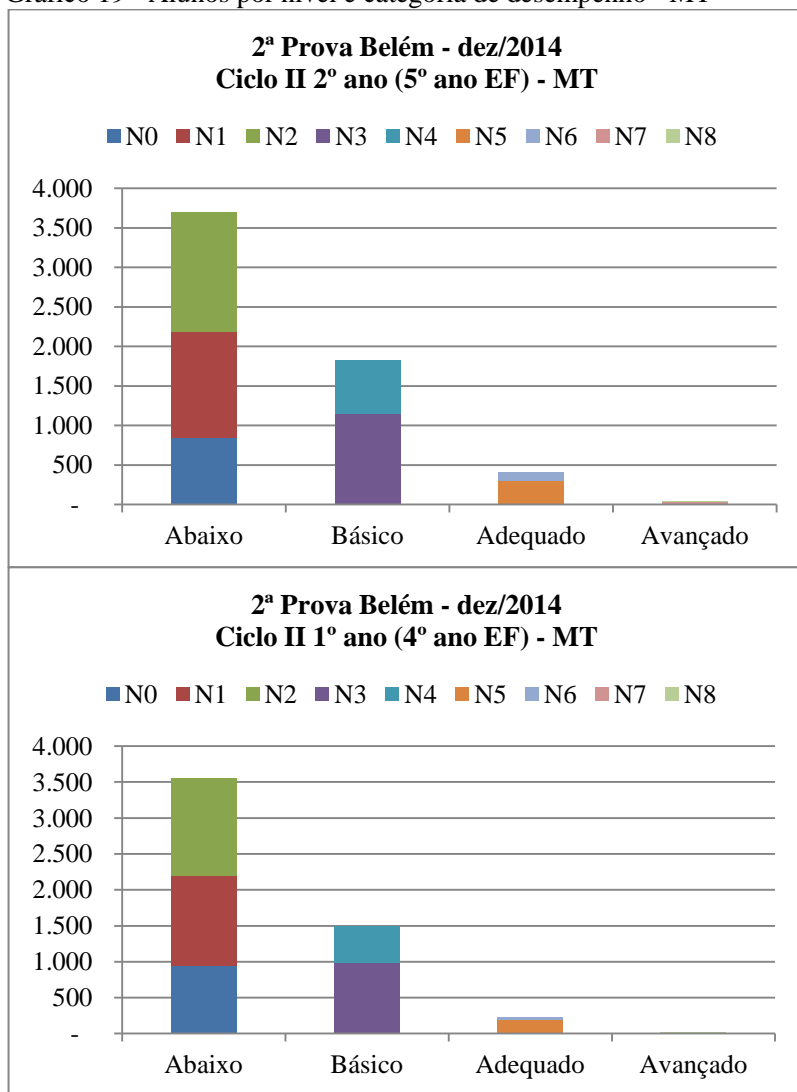
Item	Descritor	b	Ponto na escala SAEB c/ probabilidade 65%	Nível (½DP)
MT041	D08	216	218,3	N5=[210,235)
MT042	D09	209	213,0	N5=[210,235)
MT045	D01	156	166,9	N3=[160,185)
MT046	D07	190	195,4	N4=[185,210)
MT048	D13	232	239,5	N6=[235,260)
MT049	D15	265	269,5	N7=[260,285)
MT054	D14	160	165,2	N3=[160,185)
MT055	D16	152	155,7	N2=[135,160)
MT065	D22	151	157,8	N2=[135,160)
MT066	D24	186	191,6	N4=[185,210)
MT071	D19	147	154,6	N2=[135,160)
MT072	D03	124	136,0	N2=[135,160)
MT073	D01	203	209,7	N4=[185,210)
MT074	D02	174	192,3	N4=[185,210)
MT075	D03	108	115,5	N1=[110,135)
MT076	D06	176	186,8	N4=[185,210)
MT077	D02	221	225,2	N5=[210,235)

Item	Descritor	b	Ponto na escala SAEB c/ probabilidade 65%	Nível ($\frac{1}{2}DP$)
MT078	D27	122	130,4	N1=[110,135)
MT079	D17	190	195,1	N4=[185,210)
MT080	D18	160	163,8	N3=[160,185)
MT081	D19	147	147,9	N2=[135,160)
MT082	D19	157	158,8	N2=[135,160)
MT083	D19	179	180,7	N3=[160,185)
MT084	D20	191	196,8	N4=[185,210)
MT085	D04	167	180,1	N3=[160,185)
MT086	D02	156	166,8	N3=[160,185)
MT087	D11	220	221,7	N5=[210,235)
MT088	D12	149	166,6	N3=[160,185)
MT089	D28	112	123,8	N1=[110,135)
MT090	D27	121	131,4	N1=[110,135)
MT091	D11	212	216,6	N5=[210,235)
MT092	D28	116	133,4	N1=[110,135)
MT093	D27	131	138,3	N2=[135,160)
MT094	D07	206	208,5	N4=[185,210)
MT095	D10	169	178,4	N3=[160,185)
MT096	D13	214	220,3	N5=[210,235)
MT097	D19	169	171,8	N3=[160,185)
MT098	D21	224	232,1	N5=[210,235)
MT099	D13	132	143,9	N2=[135,160)
MT100	D22	147	152,1	N2=[135,160)
MT101	D23	142	141,6	N2=[135,160)
MT102	D25	218	221,1	N5=[210,235)
MT103	D26	164	171,7	N3=[160,185)

Fonte: Elaboração do autor.

O gráfico 19 mostra a quantidade de alunos em cada nível, em relação às categorias qualitativas anteriormente definidas na tabela 4.17.

Gráfico 19 - Alunos por nível e categoria de desempenho - MT



Fonte: Elaboração do autor.

Na tabela 4.22, os 38 itens aplicados no Ciclo II 2º ano (5º ano EF) estão posicionados nos níveis propostos pela escala do SAEB, sendo a seguinte quantidade de itens por nível: N1, 4 itens; N2, 7 itens; N3, 7 itens, N4, 9 itens, N5, 8 itens; N6, 2 item; e N8, 1 item. Nenhum item foi posicionado nos níveis: N0, N7, N9 e N10.

Tabela 4.22 - Posicionamento dos itens de Matemática - 5º ano

Item	Descritor	b	Ponto na escala SAEB c/ probabilidade 65%	Nível (½DP)
MT002	D02	198,6932	189,7	N3=[175,200)
MT004	D06	208,9456	225,9	N5=[225,250)
MT009	D10	208,2139	215,1	N4=[200,225)
MT020	D17	203,3127	209,9	N4=[200,225)
MT034	D23	226,3532	237,0	N5=[225,250)
MT035	D25	178,3004	188,4	N3=[175,200)
MT040	D03	294,9678	300,7	N8=[300,325)
MT041	D08	216,4061	218,3	N4=[200,225)
MT042	D09	209,3273	213,0	N4=[200,225)
MT043	D02	121,2678	133,5	N1=[125,150)
MT044	D06	204,9498	216,8	N4=[200,225)
MT045	D01	156,2016	166,9	N2=[150,175)
MT046	D07	190,3155	195,4	N3=[175,200)
MT047	D27	134,763	145,6	N1=[125,150)
MT048	D13	231,938	239,5	N5=[225,250)
MT049	D15	265,4471	269,5	N6=[250,275)
MT050	D17	200,4533	205,5	N4=[200,225)
MT051	D18	163,3514	173,2	N2=[150,175)
MT052	D19	144,3726	159,0	N2=[150,175)
MT053	D20	210,434	213,1	N4=[200,225)
MT054	D14	159,772	165,2	N2=[150,175)
MT055	D16	152,2271	155,7	N2=[150,175)
MT056	D04	214,3431	224,0	N4=[200,225)
MT057	D05	249,814	238,9	N5=[225,250)
MT058	D11	217,4736	234,5	N5=[225,250)
MT059	D12	252,2085	259,5	N6=[250,275)
MT060	D10	174,873	180,0	N3=[175,200)
MT061	D28	125,4502	136,2	N1=[125,150)
MT062	D27	133,5043	143,2	N1=[125,150)

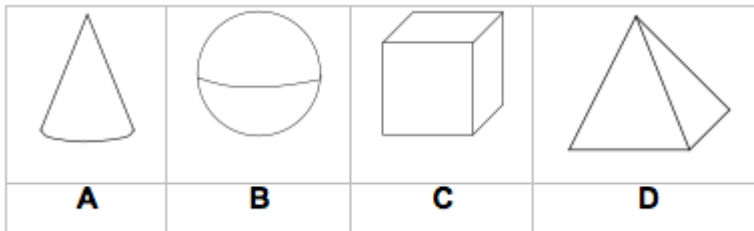
Item	Descritor	b	Ponto na escala SAEB c/ probabilidade 65%	Nível (½DP)
MT063	D07	225,3209	245,8	N5=[225,250)
MT064	D21	224,5731	227,0	N5=[225,250)
MT065	D22	150,912	157,8	N2=[150,175)
MT066	D24	186,2589	191,6	N3=[175,200)
MT067	D25	205,4881	211,3	N4=[200,225)
MT068	D13	189,3452	196,1	N3=[175,200)
MT069	D23	193,5841	198,9	N3=[175,200)
MT070	D26	228,656	232,6	N5=[225,250)
MT071	D19	147,4701	154,6	N2=[150,175)

Fonte: Elaboração do autor.

A seguir dois exemplos de itens comentados e interpretados pelo INEP para ilustrar a interpretação de itens.

Exemplo 01: Item MT02 da Prova Belém, disponível em devolutivas

Vítor gosta de brincar de construtor. Ele pediu para sua mãe comprar blocos de madeira com superfícies arredondadas. Quais dos blocos abaixo a mãe de Vítor poderá comprar?



- (A) A e C
- (B) A e B
- (C) B e D
- (D) B e C

Tabela 4.23 - Percentual de respostas às alternativas do item MT02

Prova	item	Gabarito	A	B	C	D
Brasil	021	B	12%	53%	13%	13%
Belém	MT02	B	13%	59%	13%	14%

Fonte: Elaboração do autor, a partir do relatório SAEB 2001 (INEP, 2002).

O item MT02 apresenta quatro desenhos de sólidos geométricos: um cone, uma esfera, um cubo e uma pirâmide, identificados pelas letras

A, B, C e D, respectivamente. Cada um deles representa um bloco de madeira e é solicitado ao aluno que identifique os que têm superfície arredondada. Quem respondeu acertadamente distinguiu as partes planas das arredondadas nas superfícies de sólidos geométricos e, assim, indicou o cone (figura A) e a esfera (figura B) como os blocos corretos, e ambos aparecem juntos na alternativa B. Os alunos com proficiência a partir de 189,7 pontos tem 65% ou mais probabilidade de acertar esse item.

Mais da metade dos alunos acertaram o item MT02, tabela 4.23. No grupo formado pelos alunos com menor desempenho no teste, esse percentual foi inferior a $\frac{1}{3}$, enquanto que no de maior desempenho, superior a $\frac{4}{5}$. A diferença entre eles revela que o item teve uma alta capacidade de discriminação dos dois grupos. Desde o nível de proficiência 100, o gabarito foi a opção de resposta preferencial dos alunos. Houve uma alta correlação entre o desempenho dos alunos no teste e o acerto do item.

Nenhum distrator atraiu significativamente os alunos e o percentual de escolha das alternativas incorretas A, C e D ficou entorno de 13% nas duas provas. Por outro lado, causa estranheza que os alunos que marcaram a opção A não tenham percebido que a esfera não está incluída nesse par, pois esse sólido, em geral, é reconhecido pelos estudantes como sendo redondo. Talvez isso se deva ao fato de o enunciado solicitar figura com superfícies (partes) arredondadas e a esfera ser toda ela arredondada.

Exemplo 02: Item MT04 da Prova Belém, disponível em INEP (2002).

A distância da casa de Manuela até o ponto de ônibus é maior que 200m e menor que 500m. Essa distância poderá ser de:

- (A) 199 m
- (B) 600 m
- (C) 398 m
- (D) 100 m

Tabela 4.24 - Percentual de respostas às alternativas do item MT04

Prova	Item	Gabarito	A	B	C	D
Brasil	Não informado	C	15%	21%	37%	18%
Belém	MT04	C	16%	28%	36%	19%

Fonte: Elaboração do autor, a partir da Devolutiva do INEP.

Os conceitos requeridos para a resolução deste item são: medidas de comprimento, noções de maior e menor e de intervalo entre números naturais. O aluno tem que raciocinar levando em conta um intervalo em

uma série concreta, encontrando a distância que se enquadra entre os pontos de referência dados: maior que 200 m e menor que 500m, isto é, entre essas duas medidas.

Na tabela 4.24, observa-se que a alternativa errada “B”, possivelmente foi a que mais atraiu o grupo de alunos que não souberam localizar, adequadamente, a distância pedida, levando em conta apenas parte do enunciado (maior que 200). O mesmo se pode supor em relação à escolha das alternativas erradas “A” e “D”, que também respondem a apenas uma parte da questão (menor que 500). Na realidade, constata-se aqui a mesma dificuldade já percebida anteriormente: os alunos têm dificuldade em considerar duas ordens concomitantemente. Os alunos com proficiência a partir de 225,9 pontos tem probabilidade de 65% ou mais de acertar esse item.

CONSIDERAÇÕES FINAIS

O presente estudo teve por objetivo equalizar a Prova Belém ao Sistema de Avaliação de Educação Básica (SAEB) com o auxílio dos modelos matemáticos propostos pela TRI e usando dados simulados para o grupo de referência. Pode-se averiguar que o procedimento de equalização dos itens se mostrou adequado. A equalização propriamente dita permite a comparação entre os estudantes que fizeram a prova Belém e aqueles que fizeram a Prova Brasil. Os resultados permitem concluir que os objetivos estipulados foram alcançados e mostram a eficácia da equalização por meio da TRI.

O uso de banco de dados simulados para o grupo de referência foi uma estratégia adequada para a estimativa dos valores parâmetros dos itens não calibrados e das estimativas das proficiências dos alunos do Ciclo II 1º e 2º anos (equivalentes ao 4º e 5º anos do ensino fundamental).

As médias das proficiências estimadas para os alunos do Ciclo II 1º e 2º anos por meio da Prova Belém, foram para as disciplinas de Língua Portuguesa (139.3/156.5) e Matemática (145.6/166.5) compatíveis com os resultados divulgados pelo Sistema Paraense de Avaliação Educacional - SisPAE a que foram submetidos os mesmos alunos das escolas municipais de Belém em novembro de 2014: Língua Portuguesa (148.8/167.6) e Matemática (150.7/166.5) que também utiliza a escala SAEB como referência.

Diante de todo o exposto, este estudo não teve o propósito de extenuar o assunto; pelo contrário, tendo em vista que muito ainda se pode construir na direção de agregar técnicas, metodologias, percepções e construções teóricas no que se refere à interface da avaliação da educação local e os sistemas nacionais utilizando a TRI.

Em vista do grande número de avaliações realizadas pelos estados e municípios brasileiros, cada um com suas particularidades, as comparações, entre resultados, precisam ser conduzidas as devidas considerações e cuidados para a consolidação do sistema de avaliação da educação pública.

Esse estudo abriu horizontes para a implantação de banco de itens no município de Belém, ação em curso conjuntamente com a Companhia de Informática de Belém (CINBESA), e para equalização de todo o ensino fundamental da rede municipal na escala SAEB, pois além do Ciclo II 1º e 2º anos, também os alunos do Ciclo I 3º ano (3º ano do ensino fundamental) e os alunos do Ciclo IV 1º e 2º anos (equivalente

aos 8º e 9º anos do ensino fundamental) participam de provas aplicadas pela secretaria municipal.

Além disso, a secretaria iniciou a aplicação de teste de produção escrita para os alunos, que poderão também utilizar a TRI através de itens politômicos para construir uma escala que permita avaliar a produção escrita desses alunos.

Esses são alguns desafios que o presente coloca como horizonte, não deixando de lembrar os estudos de fatores associados à proficiência que a SEMEC ainda não realiza.

A integração de métodos quantitativos e qualitativos é fundamental para a análise dos sistemas educacionais, pois, em linhas gerais, os métodos quantitativos localizam, por exemplo, o que acontece no interior das escolas, porém, para explicar o porquê, é necessária a utilização de métodos qualitativos. Logo, o grande desafio para a avaliação passa pela utilização, harmonização e padronização de métodos e técnicas conduzidas de forma otimizada por órgãos do governo, secretarias de educação, universidades e escolas, tendo como objetivo único à melhoria da qualidade do ensino.

REFERÊNCIAS

- ALVES, Maria Bernardete Martins & ARRUDA, Susana Margareth. **Como fazer referências: bibliográficas, eletrônicas e demais formas de documento**. Florianópolis: Universidade Federal de Santa Catarina, Biblioteca Universitária, c2007. Disponível em: <http://www.bu.ufsc.br/design/framerefer.html>. Acesso em: 11 Jan. 2017.
- ANDRADE, D. F. **Comparando desempenhos de grupos de alunos por intermédio da Teoria de Resposta ao Item**. Revista Estudos em Avaliação Educacional, São Paulo: FCC, nº 23, p. 71-92, jan-jun/2001.
- ANDRADE, D. F. & VALLE, R. C. **Introdução à Teoria de Resposta ao Item: conceitos e aplicações**. Revista Estudos em Avaliação Educacional, São Paulo: FCC, nº 18, p. 13-32, 1998.
- ANDRADE, D. F.; TAVARES, H. R. & VALLE, R. C. **Teoria da Resposta ao Item: Conceitos e Aplicações**. In: Simpósio Nacional de Probabilidade e Estatística (SINAPE). Anais... Caxambu: Associação Brasileira de Estatística, 2000.
- ANDRADE, J. M.; LAROS, J. A. & GOUVEIA, V. V. **O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores**. Revista Avaliação Psicológica (Online), Porto Alegre: Instituto Brasileiro de Avaliação Psicológica, v.9, n.3, p. 421-435, dez. 2010.
- ARAÚJO, Carlos Henrique & LUZIO, Nildo. **Avaliação da Educação Básica: em busca da qualidade e equidade no Brasil**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira -INEP, 2005.
- ARAUJO, Eutalia A. C.; ANDRADE, Dalton F. & BORTOLOTTI, Silvana L. V. **Teoria da Resposta ao Item**. Revista da Escola de Enfermagem da USP, São Paulo, v.43, n. spe, p. 1000-1008, 2009.
- BARBETTA, P.; TREVISAN, L.; TAVARES, H. & AZEVEDO, T. **Aplicação da Teoria da Resposta ao Item Uni e Multidimensional**. Revista Estudos em Avaliação Educacional, São Paulo: FCC, v.25, n. 57, p. 280-302, jan./abr. 2014.

BAUER, A.; PIMENTA, C. O.; SOUSA, S. Z. L. & HORTA Neto, J. L. **Avaliação em larga escala em municípios brasileiros: o que dizem os números?** Revista Estudos em Avaliação Educacional, São Paulo: FCC, v.26, n. 62, p. 326-352, maio/ago. 2015.

BORGATTO, A. & ANDRADE, D. **Análise Clássica de Testes com Diferentes Graus de Dificuldade.** Revista Estudos em Avaliação Educacional, São Paulo: FCC, v. 23, n. 52, p. 146-156, maio/ago. 2012.

BROOKE, N; CUNHA, M. A. A. & FALEIROS, M. **A Avaliação Externa como Instrumento da Gestão Educacional nos Estados: Relatório Final.** Belo Horizonte: GAME/FAE/UFMG com o apoio da Fundação Victor Civita, Agosto/2011.

CERDÁ, E. **Psicometria general.** Barcelona: Editorial Herder, 1972.

COELHO, Maria Inês de Matos. **Vinte anos de avaliação da educação básica no Brasil: aprendizagens e desafios.** Revista Ensaio: avaliação e políticas públicas em educação, Rio de Janeiro: CESGRANRIO, v. 16, n. 59, p. 229-258, abr./jun. 2008.

CONDÉ, F. N. **Análise empírica de itens.** Encontro SAEB: Novas perspectivas para avaliação da Educação Básica no Brasil. Brasília: INEP, 2002.

CONDÉ, F. N. & LAROS, J. A. **Unidimensionalidade e a Propriedade de Invariância das Estimativas da Habilidade pela TRI.** Revista Avaliação Psicológica (Online), Porto Alegre: Instituto Brasileiro de Avaliação Psicológica, v.6, n. 2, p. 205-215, dez/2007.

EBEL, Robert L. **Measuring Educational Achievement.** New Delhi: Prentice-Hall of India Private Ltd., 1965.

ERTHAL, Teresa Cristina. **Manual de Psicometria.** Rio de Janeiro: Zahar, 8ª edição, 2009.

FERREIRA, Francisco Fialho Guedes. **Escala de Proficiência para o Enem utilizando Teoria da Resposta ao Item.** 2009, 118f. Dissertação (mestrado em Estatística) - Programa de Pós-Graduação em Matemática e Estatística, Universidade Federal do Pará, Belém, 2009.

FREITAS, L. C.; SORDI, M. R. L.; MALAVASI, M. M. S. & FREITAS, H. C. L. **Avaliação educacional: caminhando pela contramão**. 6ª edição, Petrópolis: Vozes, 2014. (Coleção Fronteiras Educacionais).

GATTI, B. A. **Avaliação de sistemas educacionais no Brasil**. Sísifo - Revista de Ciências da Educação, Lisboa, n. 9, p. 7-18, maio/ago. 2009.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6ª edição São Paulo: Atlas, 2008.

GIMENES, N.; SILVA, V. G.; PRÍNCIPE, L. M.; LOUZANO, P.; MORICONI, G. M. **Além da Prova Brasil: Investimento em Sistemas próprios de Avaliação externa**. Revista Estudos em Avaliação Educacional, São Paulo: FCC, v. 24, n. 55, p. 12-32, abr./ago. 2013.

GOVERNO DO ESTADO DO PARÁ/SEDUC & VUNESP. **SISPAE 2014**. Disponível em: <https://sispae.vunesp.com.br>. Acesso em: 20 out. 2016.

HORTA Neto, João Luiz. **Um olhar retrospectivo sobre a avaliação externa no Brasil: das primeiras medições em educação até o SAEB de 2005**. Revista Ibero-americana de Educación, Organización de Estados Ibero-americanos para a Educação, a Ciência e a Cultura (OEI), v. 42, n. 5, abril de 2007.

INEP/MEC/Brasil. **SAEB 2001 - Relatório Língua Portuguesa**. Brasília: INEP, 2002.

_____. **Relatório Nacional SAEB 2003** - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. – Brasília, INEP, 2006.

_____. **PDE | Prova Brasil- Plano de Desenvolvimento da Educação**. Brasília: MEC, SEB; INEP, 2011.

_____. **Prova Brasil**. Página do Instituto Nacional de Estudos e pesquisas educacionais Anísio Teixeira sobre Prova Brasil/SAEB. Disponível em: <http://provabrasil.inep.gov.br/>. Acesso em: 25 set. 2012.

_____. **Devolutivas Pedagógicas**. Disponível em: <http://devolutivas.inep.gov.br>. Acesso em: 15 dez. 2016.

_____. **Boletim de desempenho da escola.** Disponível em: <http://sistemasprovabrasil.inep.gov.br/provaBrasilResultados/>. Acesso em: 20 out. 2016.

KIM, S. & COHEN, A. S. **A Comparison of Linking and Concurrent Calibration Under Item Response Theory.** Paper presented at the Annual Meeting of the American Educational Research Association. New York, NY, April 8-12, 1996.

KOLEN, Michael J. & BRENNAN, Robert L. **Test Equating, Scaling, and Linking: Methods and Practices.** New York: Springer-Verlag, 1995.

KLEIN, Ruben. **Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB).** Revista Ensaio: avaliação e políticas públicas em educação, Rio de Janeiro: CESGRANRIO, v.11, n. 40, p.283-296, jul./set. 2003.

LOPES, F. L. & VENDRAMINI, C. M. M. **Equalização de provas acadêmicas via Teoria de Resposta ao Item.** Revista Psico-USF (online), Bragança Paulista: Universidade de São Francisco, Programa de Pós-Graduação Stricto Sensu em Psicologia, v. 18, n. 1, p. 141-150, jan./abril 2013.

LOPES, V. V. **Cartografia da avaliação educacional no Brasil.** 2007. Tese (Doutorado em educação) – Programa de Pós-graduação em Educação, Faculdade de Educação, Universidade de São Paulo, São Paulo, 2007.

LUCKESI, Cipriano C. **O que é mesmo o ato de Avaliar a Aprendizagem?** Pátio. Porto alegre: ARTMED. Ano 3, n. 12 fev./abr. 2000. Disponível em: <https://www.nescon.medicina.ufmg.br/biblioteca/imagem/2511.pdf>. Acesso em: 10 mar. 17.

MEDIANO, Zélia Domingues. **Módulos Instrucionais para medidas e avaliação em educação.** Rio de Janeiro: Francisco Alves, 1976.

OLIVEIRA, Ana Paula de Matos. **A prova Brasil como política de regulação da rede pública do Distrito Federal.** 2011. dissertação

(mestrado em educação) - Programa de Pós-graduação em Educação, Faculdade de Educação, Universidade de Brasília, Brasília, 2011.

PADILLA, J. C. **Escalas de Medición**. Bogotá: Corporación Universitaria Unitec. Paradigmas, v.2, n.2, p. 104 -125, julio-diciembre de 2007. Sistema Institucional de Investigación de Unitec (SIU).

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

PASQUALI, L. & PRIMI, R. **Fundamentos da Teoria da Resposta ao Item - TRI**. Revista Avaliação Psicológica (Online), Porto Alegre: Instituto Brasileiro de Avaliação Psicológica, v. 2, n. 2, p. 99-110, 2003.

PESTANA, M. H. & GAGEIRO, J. N. **Análise de Dados para Ciências Sociais. A complementaridade do SPSS**. 6ª edição. Lisboa: Edições Sílabo, 2014.

RIBEIRO, B. B. D. **A função social da avaliação escolar e as políticas de avaliação da educação básica no Brasil dos anos 90: breves considerações**. Revista Inter-Ação, Goiânia: Faculdade de Educação/UFG, v. 2, n. 27, p. 127-142, jul./dez. 2002.

RODRIGUES, Margarida M. M. **Proposta de Análise de itens das provas do SAEB sob a perspectiva pedagógica e a psicométrica**. Revista Estudos em Avaliação Educacional, São Paulo: FCC, v. 17, n. 34, p. 43-78, maio/ago. 2006.

SANTOS, Lilian M. **Desempenho Escolar em Pernambuco: Análise dos Itens e das Habilidades usando Teoria Clássica e TRI**. 2008. Dissertação (mestrado em Estatística) - Pós-graduação em Estatística, Departamento de Estatística/CCEN/UFPE, Recife, 2008.

SILVA, W & SOARES, T. M. **Eficácia dos processos de linkagem na avaliação educacional em larga escala**. Revista Estudos em Avaliação Educacional, São Paulo: FCC, v. 21, n. 45, p. 191-211, jan./abr. 2010.

SOUSA, Sandra Z. & OLIVEIRA, Romualdo P. **Sistemas estaduais de avaliação: uso dos resultados, implicações e tendências**. Cadernos de Pesquisa, São Paulo: Departamento de Administração Escolar e Economia da Educação - FE/EDA/USP, v. 40, n. 141, dez. 2010.

STEVENS, S. S. **On the Theory of Scales of Measurement.** Science, New Series, v. 103, N. 2684, p. 677-680, June 1946. Published by: American for the Advancement of Science.

STUFFLEBEAM, Daniel L. **Empowerment Evaluation, Objectivist Evaluation, and Evaluation Standards: Where the Future of Evaluation Should Not Go and Where It Needs to Go.** American Journal of Evaluation, October 1994; vol. 15, 3: p. 321-338.

TAVARES, Heliton Ribeiro. **Curso TRI.** Belém: Universidade Federal do Pará, Departamento de Estatística. Disponível em: <http://www.ufpa.br/heliton/arquivos/tri/CursoTRI.pdf>. Acesso em: 10 mar. 2015.

UNIVERSIDADE FEDERAL DE SANTA CATARINA. Biblioteca Universitária. **Trabalho acadêmico: guia fácil para diagramação:** formato A5. Florianópolis, 2009. Disponível em: <http://www.bu.ufsc.br/design/GuiaRapido2012.pdf>. Acesso em: 11 abr. 2016.

URBINA, Susana. **Fundamentos da testagem psicológica.** Porto Alegre: Artmed, 2007.

VALLE, R. C. **A construção e a interpretação das escalas de conhecimento.** Revista Estudos em Avaliação Educacional, São Paulo: FCC, n. 23, p. 71-92, jan-jun/2001.

VIANNA, H. M. **Avaliações nacionais em larga escala: análises e propostas.** Revista Estudos em Avaliação Educacional, São Paulo: FCC, nº 27, p. 41-76, jan-jun/2003.

_____. **Fundamentos de um programa de avaliação educacional.** Revista Estudos em Avaliação Educacional, FCC: São Paulo, n. 28, p. 23-38, jul./dez. 2003.

VITAL, J. F. & SOARES, T. M. **Análise do comportamento de métodos para equalização de testes educacionais: estudo do caso de equiparação dos resultados do SIMAVE/2001 com os do SAEB/97.** Simpósio Brasileiro de Pesquisa Operacional, Rio de Janeiro/RJ, 8 a 11 de novembro de 2002.

APÊNDICE A - ITENS NOS BLOCOS DE LÍNGUA PORTUGUESA

Itens da Prova Belém - Ciclo II 2º ano - Língua Portuguesa

Bloco 1 - Língua Portuguesa - Ciclo II 2º ano

Item	Id	Enunciado	Desc.	GAB
LP40	228	O texto TRATA, principalmente, da	D06	C
LP41	229	Os dois amigos andavam por	D01	B
LP42	230	O homem que estava caído, ao fingir-se de morto para se salvar, mostrou que era	D04	D
LP43	550	A FINALIDADE do texto é	D09	D
LP02	680*	De acordo com essas instruções, quem socorre uma pessoa vítima de desmaio deve:	D01	D
LP34	685*	A quem Hagar está dando ordens?	D05	C
LP32	686*	O que aconteceu com o lobo quando soube que o cachorro usava coleira?	D04	B
LP44	386	O tema VIOLÊNCIA NA ESCOLA é tratado	D15	B
LP45	401	No TEXTO 1, o menino MAIOR está	D05	C
LP46	492	No texto, a frase que apresenta uma OPINIÃO é	D11	B
LP47	494	A expressão "entrar em parafuso" indica que o Galo Cucuduco estava	D03	C

* Item SAEB (calibrado)

Bloco2 - Língua Portuguesa - Ciclo II 2º ano

Item	Id_item	Enunciado	Desc.	GAB
LP48	211	Na INTRODUÇÃO da história, a pequena pastora estava	D07	C
LP49	215	No DESFECHO da história,	D07	B
LP50	216	No trecho "... foi alegremente levá-las a Jesus" a palavra destacada dá ideia de	D12	A
LP51	217	A CAUSA da menina ter ficado triste foi o fato de	D08	D
LP14	697*	A expressão vê-la se refere à	D02	D
LP16	699*	O poeta escreveu a palavra "horizonte" com letras separadas para	D14	B
LP15	700*	Nessa fábula, a frase "parecia um Gato gigante" revela a voz	D10	D
LP52	628	No texto "Luis da Câmara Cascudo e o folclore brasileiro", o termo "Esses hábitos" diz respeito ao costume de	D02	B
LP53	519	A parte mais ENGRAÇADA da piada é quando	D13	D
LP54	422	No primeiro quadrinho, o trecho "Tchau, mãe!" indica uma linguagem	D10	C
LP55	423	No primeiro quadrinho, o ponto de exclamação (!) presente no trecho "Nada disso!" indica que a mãe do menino está falando	D14	B

* Item SAEB (calibrado)

Bloco3 - Língua Portuguesa - Ciclo II 2º ano

Item	Id_item	Enunciado	Desc.	GAB
LP56	220	A finalidade do TEXTO 2 é	D09	A
LP57	221	O TEXTO 1 e o TEXTO 2 são, respectivamente,	D15	D
LP58	522	No TEXTO 2, o verso “Deve ser bem protegida” indica que o autor está	D11	C
LP59	247	A expressão “fama internacional” quer dizer que Papai Noel	D03	C
LP32	686*	O que aconteceu com o lobo quando soube que o cachorro usava coleira?	D04	B
LP34	685*	A quem Hagar está dando ordens?	D05	C
LP02	680*	De acordo com essas instruções, quem socorre uma pessoa vítima de desmaio deve:	D01	D
LP60	184	No último quadro da tirinha "No Ver-o-Peso...", Dona Dica está	D05	D
LP61	431	O caboclo disse que estava dando banho na minhoca porque	D04	D
LP62	433	No texto, quando o caboclo ouviu a voz do policial, ele ficou	D01	A
LP63	529	O TEMA principal do texto "Combate: mutirão de dengue visitou 16 mil residência" é	D06	B

* Item SAEB (calibrado)

Bloco 4 - Língua Portuguesa - Ciclo II 2º ano

Item	Id_item	Enunciado	Desc.	GAB
LP64	242	No trecho “Joãozinho estava no sofá com seu pai...”, as palavras em destaque dão ideia de	D12	A
LP65	631	A desidratação pode levar à morte quando	D08	D
LP66	155	O trecho que apresenta linguagem INFORMAL é	D10	B
LP67	152	Há traços de HUMOR no trecho	D13	D
LP68	154	O trecho que indica que o personagem está fazendo uma PERGUNTA é	D14	A
LP69	384	O trecho que dá a ideia de TEMPO é	D12	A
LP14	697*	A expressão vê-la se refere à	D02	D
LP16	699*	O poeta escreveu a palavra "horizonte" com letras separadas para	D14	B
LP15	700*	Nessa fábula, a frase "parecia um Gato gigante" revela a voz	D10	D
LP70	169	No DESFECHO da história,	D07	B
LP71	171	No trecho "Ele começou a gritar bem alto", a palavra sublinhada refere-se ao	D02	C

* Item SAEB (calibrado)

Itens da Prova Belém - Ciclo II 1º ano - Língua Portuguesa

Bloco 1 - Língua Portuguesa - Ciclo II 1º ano

item	Id_item	Enunciado	Desc.	Gab
LP72	554	O TEMA central do texto é	D06	B
LP73	556	As ilustrações do primeiro livro de Luiz foram feitas por seu	D01	C
LP59	247**	A expressão "fama internacional" quer dizer que Papai Noel	D03	C
LP74	248	O texto sugere que o Papai Noel é	D04	A
LP47	494**	A expressão "entrar em parafuso" indica que o Galo Cucuduco estava	D03	C
LP41	229**	Os dois amigos andavam por	D01	B
LP42	230**	O homem que estava caído, ao fingir-se de morto para se salvar, mostrou que era	D04	D
LP75	218	Ao comparar o TEXTO 1 com o TEXTO 2, pode-se afirmar que	D15	C
LP56	220**	A finalidade do TEXTO 2 é	D09	A
LP76	219	No TEXTO 1, o trecho "...Nesta época as pessoas deveriam se preocupar mais com a união e o amor entre todos" É	D11	B
LP77	84	No último quadro, Dona Dica está	D05	D

** Item comum ao Ciclo II 2º ano

Bloco 2 - Língua Portuguesa - Ciclo II 1º ano

item	Id_item	Enunciado	Desc.	Gab
LP48	211**	Na INTRODUÇÃO da história, a pequena pastora estava	D07	C
LP51	217**	A CAUSA da menina ter ficado triste foi o fato de	D08	D
LP78	355	No trecho "Parece um gigante na cozinha!..." a palavra destacada dá ideia de	D12	C
LP79	586	Considerando que o CLÍMAX determina o desfecho da história, podemos encontrá-lo no trecho	D07	C
LP80	367	No texto acima, o trecho "no chinelo da menina" indica	D12	A
LP53	519**	A parte mais ENGRAÇADA da piada é quando	D13	D
LP81	649	No texto, a palavra castanha tem o sentido de uma	D03	C
LP82	555	No trecho "Muito surpresa, ela lhe disse...", a palavra em destaque refere-se	D02	B
LP83	506	Na piada, há traços de HUMOR quando	D13	D
LP54	422**	No primeiro quadrinho, o trecho "Tchau, mãe!" indica uma linguagem	D10	C
LP55	423**	No primeiro quadrinho, o ponto de exclamação (!) presente no trecho "Nada disso!" indica que a mãe do menino está falando	D14	B

** Item comum ao Ciclo II 2º ano

Bloco 3 - Língua Portuguesa - Ciclo II 1º ano

Item	Id_item	Enunciado	Desc.	Gab
LP84	504	O cartaz tem a FINALIDADE de	D09	C
LP85	655	A palavra "desembuche" significa que o Saci quer que a bruxa	D03	C
LP86	158	Os gêneros dos textos 1 e 2 são	D15	A
LP87	157	No texto 1, o sentido da expressão "morta de fadiga" indica que a mãe está	D03	B
LP88	159	De acordo com o texto 2,	D05	C
LP89	523	O cartaz TRATA do combate	D06	A
LP90	598	A FINALIDADE do texto é	D09	B
LP62	433**	No texto, quando o caboclo ouviu a voz do policial, ele ficou	D01	A
LP91	166	Segundo o texto 1, o futebol traz	D04	C
LP92	167	No texto 2, o trecho que expressa uma OPINIÃO é	D11	B
LP93	12	O ASSUNTO principal do texto é	D06	B

** Item comum ao Ciclo II 2º ano

Bloco 4 - Língua Portuguesa - Ciclo II 1º ano

Item	Id_item	Enunciado	Desc.	Gab
LP94	592	A linguagem dos DOIS textos é parecida com a usada pelos	D10	D
LP95	654	Na narrativa, o trecho "O urubu cansou de esperar e foi embora" corresponde	D07	D
LP96	385	A CAUSA do castigo do índio foi	D08	B
LP69	384**	O trecho que dá a ideia de TEMPO é	D12	A
LP97	240	As palavras "perai" e "cara" são empregadas normalmente em	D10	C
LP64	242**	No trecho "Joãozinho estava no sofá com seu pai...", as palavras em destaque dão ideia de	D12	A
LP98	238	No texto, observa-se HUMOR no trecho	D13	B
LP99	79	O MOTIVO da sobrinha refrescar-se na água do rio foi	D08	D
LP100	496	Os pontos de exclamação (!), no final da fala do menino, indicam que ele está	D14	C
LP101	173	No trecho "Socorro! Socorro! Lobo! Lobo!" O ponto de exclamação indica	D14	B
LP71	171**	No trecho "Ele começou a gritar bem alto", a palavra sublinhada refere-se ao	D02	C

** Item comum ao Ciclo II 2º ano

APÊNDICE B - ITENS NOS BLOCOS DE MATEMÁTICA

Itens da Prova de Matemática - Ciclo II 2º ano - Matemática

Bloco 1 - Matemática - Ciclo II 2º ano

Item	Id	Enunciado	Desc.	GAB
MT40	527	O polígono que tem todos os LADOS PARALELOS é o	D03	B
MT41	261	De acordo com o cartaz 2, a duração do filme corresponde a	D08	C
MT42	263	O tempo entre o início da primeira sessão do filme e o início da segunda é de	D09	B
MT43	316	A figura ao lado corresponde à PLANIFICAÇÃO do SÓLIDO GEOMÉTRICO representado na alternativa	D02	B
MT02	702*	Vitor gosta de brincar de construir. Ele pediu para sua mãe comprar blocos de madeira com superfícies arredondadas. Quais dos blocos abaixo a mãe de Vitor poderá comprar?	D02	B
MT04	703*	A distância da casa de Manuela até o ponto de ônibus é maior que 200m e menor que 500m. Essa distância poderá ser de:	D06	C
MT09	704*	Paulo precisa trocar uma nota de 100 reais por notas de valores menores. Essa troca pode ser feita por	D10	C
MT44	318	Dos objetos em destaque, o que tem a capacidade de 5 ml, aproximadamente, é o da alternativa	D06	A
MT45	516	Das praias destacadas no mapa da Ilha do Mosqueiro, a mais próxima da ponte Sebastião Oliveira é a	D01	D
MT46	475	Sabendo que Joca comprou 1 kg de café, a alternativa que indica EXATAMENTE os pacotes que comprou é	D07	C**
MT47	449	Dentre os bairros indicados na tabela a seguir, o que possui população cuja maioria é formada por MULHERES é	D27	D

* Item SAEB (calibrado); **Gabarito ajustado

Bloco 2 - Matemática - Ciclo II 2º ano

Item	Id	Enunciado	Desc.	GAB
MT48	266	O número 1625 é formado por	D13	C
MT49	267	A alternativa que corresponde a uma decomposição correta do número 1625 é	D15	D
MT50	305	Maria fez a conta de SUBTRAÇÃO a seguir para saber quanto receberia de troco se pagasse, com uma cédula de R\$ 100,00, uma guirlanda que custava R\$ 35,00. O resultado correto da conta é	D17	A
MT51	306	Uma comerciante fez a conta de MULTIPLICAR a seguir para saber quantas bolas de Natal havia no estoque de sua loja. Sabendo que no estoque havia 128 pacotes com 6 bolas em cada, o resultado correto da conta é	D18	B
MT20	705*	A diferença entre 20 000 e 16 226 é	D17	C

Item	Id	Enunciado	Desc.	GAB
MT34	707*	Ângelo foi fazer compras, encontrou algumas promoções e comprou: uma calça por R\$35,80, uma camisa por R\$15,30 e um par de tênis por R\$42,39. Ao sair da loja observou que em sua carteira ainda restavam R\$8,60. Ângelo possuía antes das compras	D23	D
MT35	714*	Ricardo e Júlio resolveram verificar suas alturas. Ricardo descobriu que mede 1,56m e que Júlio mede 1,09m. Quanto Júlio tem que crescer para alcançar a altura que Ricardo tem hoje?	D25	D
MT52	307	Lúcia comprou 43 bolas de Natal, 15 anjos e 27 sinos para enfeitar a recepção do prédio em que mora. O total de enfeites natalinos comprados foi	D19	C
MT53	308	O preço da árvore de Natal em uma das lojas do Ver o Peso é R\$ 580,00. Sabendo que este valor pode ser dividido, sem juros, em 5 parcelas iguais, o valor de cada parcela é	D20	D
MT54	463	Sabendo que a reta numerada, abaixo, representa a linha do tempo de um determinado período e que a seta indica o ANO DE INAUGURAÇÃO do Theatro da Paz, a alternativa correspondente a esse ano é	D14	C
MT55	518	O número 616, decomposto em sua forma polinomial, corresponde a	D16	A

* Item SAEB (calibrado)

Bloco 3 - Matemática - Ciclo II 2º ano

Item	Id	Enunciado	Desc.	GAB
MT56	253	Os QUADRILÁTEROS que formam a Bandeira do Brasil são	D04	B
MT57	300	Observe a Bandeira de Belém representada na malha a seguir. A alternativa que corresponde à AMPLIAÇÃO correta da bandeira é	D05	B
MT02	702*	Vitor gosta de brincar de construtor. Ele pediu para sua mãe comprar blocos de madeira com superfícies arredondadas. Quais dos blocos abaixo a mãe de Vitor poderá comprar?	D02	B
MT58	303	O PERÍMETRO da fachada é	D11	A
MT59	304	A ÁREA da fachada é	D12	C
MT04	703*	A distância da casa de Manuela até o ponto de ônibus é maior que 200m e menor que 500m. Essa distância poderá ser de:	D06	C
MT09	704*	Paulo precisa trocar uma nota de 100 reais por notas de valores menores. Essa troca pode ser feita por	D10	C
MT60	578	Para pagar sua conta de água, Simone usou uma cédula de R\$ 50,00, uma cédula de R\$ 5,00 e duas moedas de R\$ 0,25. Qual o valor da conta de água de Simone?	D10	D

Item	Id	Enunciado	Desc.	GAB
MT61	312	O gráfico, a seguir, indica o levantamento dos tipos de brinquedos mais solicitados pelas meninas no período do Natal. Os dois brinquedos MAIS solicitados são	D28	B
MT62	311	Leia na tabela a seguir os preços de alguns enfeites natalinos. Os dois produtos MAIS caros da tabela são	D27	A
MT63	27	A largura média do rio Guamá, transformada em metros, é	D07	C

* Item SAEB (calibrado)

BlocoMT_C22_B4

Item	Id	Enunciado	Desc.	GAB
MT64	309	Foram consumidos 0,3 de um bolo. Este número decimal corresponde ao número fracionário da alternativa	D21	A
MT65	273	Observe a imagem a seguir. O comprimento da fita natalina indicado pela reta numerada é	D22	B
MT66	274	Observe a imagem da lasanha representada ao lado. A fração que corresponde à parte consumida da lasanha é	D24	C**
MT67	310	Na Praça do Relógio, em Belém, a altura do relógio é 11,91 m e a do poste de iluminação é 8,75 m. Desta forma, o relógio é MAIS alto que o poste	D25	B
MT20	705*	A diferença entre 20 000 e 16 226 é	D17	C
MT34	707*	Ângelo foi fazer compras, encontrou algumas promoções e comprou: uma calça por R\$35,80, uma camisa por R\$15,30 e um par de tênis por R\$42,39. Ao sair da loja observou que em sua carteira ainda restavam R\$8,60. Ângelo possuía antes das compras	D23	D
MT35	714*	Ricardo e Júlio resolveram verificar suas alturas. Ricardo descobriu que mede 1,56m e que Júlio mede 1,09m. Quanto Júlio tem que crescer para alcançar a altura que Ricardo tem hoje?	D25	D
MT68	528	No número 895, o algarismo 9 ocupa a ORDEM das	D13	B
MT69	650	Observe os enfeites para árvore de Natal que estão à venda em uma loja da cidade. O valor a ser pago por 5 bolas, 4 sinos e 3 anjinhos é	D23	A
MT70	483	O percentual de MENINOS que formam o coral é	D26	B
MT71	470	No ônibus havia 44 pessoas sentadas e 49 em pé. O total de pessoas que formava o grupo era	D19	D

* Item SAEB (calibrado); ** Gabarito ajustado.

Itens da Prova de Matemática - Ciclo II – 1º ano – Matemática

Bloco 1 - Matemática - Ciclo II 1º ano

item	Id	Enunciado	Desc.	GAB
MT72	97	A figura plana com exatamente TRÊS ângulos internos é o:	D03	B
MT41	261**	De acordo com o cartaz 2, a duração do filme corresponde a	D08	C
MT42	263**	O tempo entre o início da primeira sessão do filme e o início da segunda é de	D09	B
MT73	25	O mapa, a seguir, representa parte do bairro da Campina, em Belém. O percurso, destacado no mapa, indica o caminho feito por uma pessoa que foi do Mercado de Ferro (Ponto P) até o Arquivo Público (Ponto Q). A alternativa que corresponde a esse percurso é	D01	C
MT45	516**	Das praias destacadas no mapa da Ilha do Mosqueiro, a mais próxima da ponte Sebastião Oliveira é a	D01	D
MT74	95	O SÓLIDO GEOMÉTRICO denominado CILINDRO é o da alternativa	D02	B
MT75	288	Dos polígonos representados abaixo, o que tem 5 LADOS é o	D03	B
MT76	102	Dos objetos em destaque, o que tem a capacidade de 1 L, aproximadamente, é o da alternativa	D06	B
MT77	562	O DADO a seguir tem a forma de um POLIEDRO porque possui	D02	C
MT46	475**	Sabendo que Joca comprou 1 kg de café, a alternativa que indica EXATAMENTE os pacotes que comprou é	D07	C*
MT78	278	Leia na tabela a seguir os preços de alguns enfeites natalinos. O produto MAIS caro da tabela é	D27	B

* Gabarito ajustado; ** Item comum ao Ciclo II 2º ano

Bloco 2 - Matemática - Ciclo II 1º ano

Item	Id	Enunciado	Desc.	GAB
MT48	266**	O número 1625 é formado por	D13	C
MT49	267**	A alternativa que corresponde a uma decomposição correta do número 1625 é	D15	D
MT79	268	Maria fez a conta de SUBTRAÇÃO a seguir para saber quanto receberia de troco se pagasse, com uma cédula de R\$ 50,00, uma guirlanda que custava R\$ 35,00. O resultado correto da conta é	D17	A
MT80	269	Uma comerciante fez a conta de MULTIPLICAR a seguir para saber quantas bolas de Natal havia no estoque de sua loja. Sabendo que no estoque havia 58 pacotes com 6 bolas em cada, o resultado correto da conta é	D18	B
MT81	270	Lúcia comprou 43 bolas de Natal e 27 sinos para enfeitar a recepção do prédio em que mora. O total	D19	C

Item	Id	Enunciado	Desc.	GAB
		de enfeites natalinos comprados foi		
MT55*	518**	O número 616, decomposto em sua forma polinomial, corresponde a	D16	A
MT82	651	Em uma pequena fábrica foram confeccionados Panetones de três tipos: 430 de chocolate, 380 de frutas e 250 de doce de leite. O total de panetones fabricados foi de	D19	C
MT83	652	Sabendo que o valor da conta de água da casa de João, em fevereiro, foi de 98 reais e em março foi de 120 reais, quanto pagou A MAIS neste mês?	D19	A
MT84	271	O preço da árvore de Natal em uma das lojas do Ver o Peso é R\$ 488,00. Sabendo que este valor pode ser dividido, sem juros, em 4 parcelas iguais, o valor de cada parcela é	D20	D
MT54	463**	Sabendo que a reta numerada, abaixo, representa a linha do tempo de um determinado período e que a seta indica o ANO DE INAUGURAÇÃO do Theatro da Paz, a alternativa correspondente a esse ano é	D14	C
MT55*	518**	O número 616, decomposto em sua forma polinomial, corresponde a	D16	A

*Item Duplicado; ** Item comum ao Ciclo II 2º ano.

Bloco 3 - Matemática - Ciclo II 1º ano

Item	Id	Enunciado	Desc.	GAB
MT85	99	Das figuras planas abaixo, o QUADRILÁTERO é o	D04	D
MT86	250	Os SÓLIDOS GEOMÉTRICOS, representados nas imagens a seguir, são	D02	D
MT87	581	O perímetro do banheiro é	D11	C
MT88	582	A malha quadriculada representa o terreno que Joaquim comprou em Mosqueiro. Sabendo que o lado de cada quadrinho mede 1m, qual a ÁREA do terreno?	D12	C
MT89	279	O gráfico, a seguir, indica o levantamento dos tipos de brinquedos mais solicitados pelas meninas no período do Natal. A alternativa que indica o brinquedo MAIS solicitado é	D28	B
MT90	6	O bairro MAIS populoso é	D27	A
MT91	264	O PERÍMETRO da fachada é	D11	A
MT92	594	Observe o gráfico abaixo e marque a alternativa que corresponde à região que concentra o MENOR volume de água doce no Brasil.	D28	A
MT93	593	Qual o continente que concentra o MENOR percentual de água do planeta?	D27	C
MT94	583	A tabela a seguir indica a quantidade de copos com água que 3 pessoas bebem diariamente. Sabendo	D07	C

Item	Id	Enunciado	Desc.	GAB
		que a capacidade do copo é 250 ml, quem costuma beber 1 litro de água por dia é		
MT95	584	Num supermercado, Seu Carlos comprou 1 garrafão de 20 litros de água mineral por R\$ 4,50. Ao efetuar o pagamento, recebeu de troco uma cédula 10 reais, uma cédula de 5 reais e uma moeda de 50 centavos. A cédula que ele usou para pagar a sua compra foi de	D10	C

Sem item comum com o Ciclo II 2º ano

Bloco 4 - Matemática - Ciclo II 1º ano

Item	Id	Enunciado	Desc.	GAB
MT96	596	Uma torneira pingando por dois meses desperdiça 2.760 litros de água. Este número possui quantas centenas?	D13	C
MT97	233	Quantas porções de caruru foram vendidas A MAIS que as de maniçoba?	D19	B
MT98	272	Foram consumidos 0,2 de um bolo. Este número decimal corresponde ao número fracionário da alternativa	D21	D
MT65	273**	Observe a imagem a seguir. O comprimento da fita natalina indicado pela reta numerada é	D22	B
MT71	470**	No ônibus havia 44 pessoas sentadas e 49 em pé. O total de pessoas que formava o grupo era	D19	D
MT99	524	O feirante que vendeu menos de 12 centenas de bananas foi	D13	A
MT100	100	A distância, aproximada, entre a Ilha do Fortinho e a Ilha Jararaquinha é	D22	A
MT101	476	Observe o sino para árvore de Natal que está à venda em uma loja da cidade. O valor a ser pago por 5 sinos é	D23	A
MT66*	274**	Observe a imagem da lasanha representada ao lado. A fração que corresponde à parte consumida da lasanha é	D24	C
MT102	29	Em 18 de março de 2013, a maré do rio Guamá atingiu 2,9 m. Quanto faltou para que ela alcançasse o seu nível máximo?	D25	B
MT103	345	Sabendo que na bandeja, a seguir, há 20 brigadeiros, a quantidade que equivale a 50% deste total é	D26	A

* Gabarito ajustado; ** Item comum ao Ciclo II 2º ano

APÊNDICE C - BISSERIAIS EM LÍNGUA PORTUGUESA

NM_Item	CIH1ºano	CIH2ºano	Multi	NM_Item	CIH1ºano	CIH2ºano	Multi
LP002		0,528	0,698	LP068		0,593	0,593
LP014		0,466	0,840	LP069	0,510	0,604	0,554
LP015		0,473	0,750	LP070		0,525	0,525
LP016		0,274	0,642	LP071	0,402	0,605	0,505
LP032		0,573	0,845	LP072	0,386		0,386
LP034		0,425	0,704	LP073	0,338		0,338
LP040		0,669	0,669	LP074	0,626		0,626
LP041	0,636	0,641	0,640	LP075	0,318		0,318
LP042	0,616	0,667	0,644	LP076	0,261		0,261
LP043		0,478	0,478	LP077	0,499		0,499
LP044		0,376	0,376	LP078	0,298		0,298
LP045		0,637	0,637	LP079	0,310		0,310
LP046		0,352	0,352	LP080	0,451		0,451
LP047	0,508	0,599	0,558	LP081	0,334		0,334
LP048	0,291	0,421	0,375	LP082	0,225		0,225
LP049		0,366	0,366	LP083	0,452		0,452
LP050		0,410	0,410	LP084	0,596		0,596
LP051	0,644	0,657	0,653	LP085	0,609		0,609
LP052		0,461	0,461	LP086	0,482		0,482
LP053	0,664	0,679	0,676	LP087	0,448		0,448
LP054	0,102	0,049	0,070	LP088	0,474		0,474
LP055	0,574	0,651	0,618	LP089	0,581		0,581
LP056	0,316	0,381	0,353	LP090	0,621		0,621
LP057		0,552	0,552	LP091	0,560		0,560
LP058		0,421	0,421	LP092	0,152		0,152
LP059	0,658	0,602	0,630	LP093	0,119		0,119
LP060		0,501	0,501	LP094	0,602		0,602
LP061		0,552	0,552	LP095	0,477		0,477
LP062	0,623	0,624	0,619	LP096	0,564		0,564
LP063		0,477	0,477	LP097	0,492		0,492

NM_Item	CII1ºano	CII2ºano	Multi	NM_Item	CII1ºano	CII2ºano	Multi
LP064	0,502	0,541	0,521	LP098	0,465		0,465
LP065		0,610	0,610	LP099	0,648		0,648
LP066		0,139	0,139	LP100	0,539		0,539
LP067		0,494	0,494	LP101	0,585		0,585

Fonte: Elaboração do autor, com os valores do bilog-MG.

APÊNDICE D - ESTIMATIVAS DOS PARÂMETROS - LP

Parâmetro: discriminação (a), dificuldade (b), acerto casual (c), SE (Erro).

Item	A	SE a	b	SE b	c	SE c
LP002	0,966	0,000	-0,978	0,000	11%	0,000
LP014	1,256	0,000	-1,486	0,000	19%	0,000
LP015	1,442	0,000	-0,540	0,000	14%	0,000
LP016	1,058	0,000	-0,551	0,000	19%	0,000
LP032	1,192	0,000	-1,667	0,000	14%	0,000
LP034	0,963	0,000	-1,361	0,000	20%	0,000
LP040	1,477	0,081	-2,145	0,038	10%	0,023
LP041	1,732	0,062	-2,718	0,024	5%	0,018
LP042	1,478	0,051	-2,361	0,023	4%	0,014
LP043	0,823	0,060	-2,039	0,089	13%	0,041
LP044	0,583	0,050	-1,722	0,119	11%	0,042
LP045	1,365	0,066	-2,438	0,038	7%	0,024
LP046	1,234	0,128	-0,778	0,051	23%	0,015
LP047	1,578	0,083	-1,496	0,024	17%	0,011
LP048	1,588	0,112	-0,860	0,030	17%	0,008
LP049	1,293	0,134	-1,089	0,052	33%	0,019
LP050	1,407	0,130	-1,077	0,043	26%	0,016
LP051	1,925	0,086	-2,361	0,026	19%	0,019
LP052	1,272	0,109	-1,433	0,051	27%	0,022
LP053	2,275	0,100	-1,845	0,017	16%	0,011
LP054	0,414	0,136	-1,890	0,710	26%	0,023
LP055	1,328	0,056	-2,334	0,034	10%	0,022
LP056	0,965	0,075	-1,262	0,048	23%	0,019
LP057	1,143	0,080	-1,580	0,046	12%	0,023
LP058	0,729	0,061	-1,612	0,089	12%	0,036
LP059	1,453	0,052	-2,516	0,026	5%	0,018
LP060	0,899	0,055	-2,065	0,065	9%	0,032
LP061	1,205	0,080	-1,922	0,052	16%	0,028
LP062	1,737	0,079	-2,192	0,027	18%	0,018
LP063	0,857	0,067	-1,412	0,061	11%	0,026
LP064	1,586	0,092	-1,753	0,031	29%	0,016
LP065	1,578	0,105	-1,606	0,035	19%	0,018
LP066	2,076	0,309	-0,225	0,056	20%	0,009
LP067	1,368	0,109	-1,460	0,046	26%	0,021
LP068	1,490	0,101	-1,751	0,041	21%	0,022
LP069	1,554	0,079	-1,704	0,026	19%	0,014

Item	A	SE a	b	SE b	c	SE c
LP070	1,113	0,085	-1,429	0,049	16%	0,022
LP071	2,088	0,118	-1,402	0,021	22%	0,010
LP072	0,839	0,091	-1,788	0,092	18%	0,040
LP073	1,414	0,163	-1,144	0,049	22%	0,016
LP074	1,810	0,108	-2,447	0,035	13%	0,026
LP075	1,358	0,163	-1,151	0,052	24%	0,017
LP076	1,203	0,170	-0,867	0,072	22%	0,016
LP077	1,059	0,083	-1,955	0,056	12%	0,029
LP078	0,719	0,108	-1,077	0,097	17%	0,034
LP079	0,898	0,121	-1,150	0,074	19%	0,027
LP080	0,908	0,062	-2,323	0,070	10%	0,036
LP081	1,301	0,151	-1,086	0,053	19%	0,016
LP082	1,443	0,222	-0,664	0,080	18%	0,012
LP083	1,532	0,135	-1,330	0,037	14%	0,014
LP084	1,658	0,116	-2,161	0,039	19%	0,025
LP085	1,605	0,106	-2,053	0,034	13%	0,021
LP086	0,994	0,074	-2,024	0,059	10%	0,030
LP087	1,079	0,107	-1,637	0,058	17%	0,028
LP088	1,172	0,104	-1,605	0,047	15%	0,022
LP089	1,397	0,084	-2,513	0,046	10%	0,031
LP090	1,748	0,114	-2,377	0,039	17%	0,028
LP091	1,268	0,075	-2,573	0,051	10%	0,033
LP092	2,527	0,423	-0,839	0,049	29%	0,010
LP093	1,918	0,437	-0,483	0,099	28%	0,011
LP094	1,647	0,115	-2,229	0,041	19%	0,027
LP095	1,562	0,133	-1,589	0,037	20%	0,017
LP096	1,862	0,141	-1,864	0,032	22%	0,019
LP097	1,615	0,137	-1,724	0,038	24%	0,020
LP098	2,472	0,222	-1,429	0,027	21%	0,012
LP099	1,670	0,102	-2,025	0,029	9%	0,018
LP100	1,277	0,096	-2,185	0,054	17%	0,032
LP101	1,488	0,101	-2,163	0,041	16%	0,025

Fonte: Elaboração do autor, com os valores do bilog-MG.

APÊNDICE E - VALORES BISSERIAIS EM MATEMÁTICA

NM_Item	Bisserial C21	Bisserial C22	BisserialMulti
MT002		0,432	0,708
MT004		0,431	0,655
MT009		0,453	0,823
MT020		0,367	0,81
MT034		0,386	0,757
MT035		0,457	0,849
MT040		0,092	0,092
MT041	0,217	0,322	0,275
MT042	0,208	0,322	0,27
MT043		0,46	0,46
MT044		0,398	0,398
MT045	0,426	0,492	0,458
MT046	0,46	0,533	0,497
MT047		0,519	0,519
MT048	0,249	0,272	0,262
MT049	-0,046	-0,028	-0,037
MT050		0,435	0,435
MT051		0,421	0,421
MT052		0,496	0,496
MT053		0,354	0,354
MT054	0,459	0,57	0,508
MT055	0,508	0,578	0,543
MT056		0,369	0,369
MT057		0,079	0,079
MT058		0,263	0,263
MT059		0,209	0,209
MT060		0,552	0,552
MT061		0,509	0,509
MT062		0,514	0,514
MT063		0,272	0,272

NM_Item	Bisserial C21	Bisserial C22	BisserialMulti
MT064		0,285	0,285
MT065	0,457	0,484	0,457
MT066	0,299	0,465	0,382
MT067		0,411	0,411
MT068		0,464	0,464
MT069		0,412	0,412
MT070		0,298	0,298
MT071	0,512	0,511	0,507
MT072	0,491		0,491
MT073	0,284		0,284
MT074	0,289		0,289
MT075	0,6		0,6
MT076	0,362		0,362
MT077	0,181		0,181
MT078	0,542		0,542
MT079	0,351		0,351
MT080	0,465		0,465
MT081	0,507		0,507
MT082	0,479		0,479
MT083	0,299		0,299
MT084	0,357		0,357
MT085	0,415		0,415
MT086	0,446		0,446
MT087	0,167		0,167
MT088	0,362		0,362
MT089	0,549		0,549
MT090	0,542		0,542
MT091	0,213		0,213
MT092	0,448		0,448
MT093	0,597		0,597

NM_Item	Bisserial C21	Bisserial C22	BisserialMulti
MT094	0,159		0,159
MT095	0,447		0,447
MT096	0,269		0,269
MT097	0,493		0,493
MT098	0,226		0,226
MT099	0,453		0,453
MT100	0,551		0,551
MT101	0,478		0,478
MT102	0,15		0,15
MT103	0,406		0,406

Fonte: Elaboração do autor, com os valores do bilog-MG.

APÊNDICE F - ESTIMATIVAS DOS PARÂMETROS - MT

Parâmetro: Discriminação (a), Dificuldade (b), acerto casual (c) e SE (erro)

Item	A	SE a	B	SE b	c	SE c
MT002	1,374	0,0	-0,913	0,0	40,9%	0,0
MT004	0,906	0,0	-0,731	0,0	9,3%	0,0
MT009	1,558	0,0	-0,744	0,0	16,7%	0,0
MT020	1,442	0,0	-0,831	0,0	18,4%	0,0
MT034	1,370	0,0	-0,422	0,0	10,7%	0,0
MT035	1,316	0,0	-1,275	0,0	12,7%	0,0
MT040	1,066	0,302	0,645	0,2733	23,7%	0,013
MT041	1,941	0,179	-0,629	0,0327	25,9%	0,008
MT042	1,212	0,112	-0,759	0,0423	25,3%	0,014
MT043	1,094	0,068	-2,282	0,0733	12,7%	0,047
MT044	1,060	0,107	-0,819	0,0518	14,8%	0,022
MT045	1,141	0,063	-1,689	0,0406	13,1%	0,023
MT046	2,218	0,127	-1,081	0,0187	16,4%	0,008
MT047	1,420	0,079	-2,037	0,0429	9,6%	0,030
MT048	1,243	0,135	-0,363	0,0577	19,3%	0,010
MT049	2,990	0,723	0,141	0,0716	19,9%	0,006
MT050	1,618	0,145	-0,898	0,0358	20,8%	0,015
MT051	1,050	0,090	-1,515	0,0692	19,1%	0,036
MT052	1,138	0,064	-1,870	0,0459	7,5%	0,027
MT053	2,135	0,212	-0,729	0,0325	23,6%	0,012
MT054	1,698	0,085	-1,613	0,0240	18,1%	0,015
MT055	1,888	0,093	-1,750	0,0236	21,0%	0,016
MT056	1,079	0,119	-0,656	0,0550	17,4%	0,020
MT057	1,584	0,379	-0,017	0,1132	44,4%	0,012
MT058	0,643	0,100	-0,597	0,1113	16,7%	0,042
MT059	0,949	0,174	0,004	0,1201	21,7%	0,019
MT060	1,977	0,133	-1,331	0,0271	18,4%	0,016
MT061	1,429	0,077	-2,206	0,0442	9,1%	0,033
MT062	1,461	0,086	-2,054	0,0462	12,1%	0,033
MT063	0,601	0,092	-0,462	0,1149	14,4%	0,039
MT064	1,307	0,168	-0,484	0,0582	26,4%	0,016
MT065	1,351	0,072	-1,774	0,0360	17,7%	0,022
MT066	1,329	0,092	-1,154	0,0328	22,0%	0,015
MT067	1,396	0,135	-0,815	0,0419	20,4%	0,017
MT068	1,347	0,114	-1,092	0,0419	19,0%	0,021
MT069	1,256	0,120	-1,019	0,0486	22,3%	0,023

Item	A	SE a	B	SE b	c	SE c
MT070	1,520	0,190	-0,412	0,0529	22,8%	0,013
MT071	1,417	0,071	-1,847	0,0331	15,1%	0,021
MT072	1,312	0,082	-2,210	0,0474	10,1%	0,032
MT073	1,095	0,158	-0,880	0,0702	20,7%	0,022
MT074	0,624	0,086	-1,408	0,1379	15,5%	0,053
MT075	2,431	0,131	-2,458	0,0247	8,0%	0,023
MT076	0,963	0,110	-1,356	0,0672	16,7%	0,032
MT077	1,759	0,313	-0,538	0,0804	21,3%	0,011
MT078	1,701	0,102	-2,258	0,0368	11,0%	0,029
MT079	1,571	0,181	-1,094	0,0444	21,2%	0,016
MT080	1,601	0,138	-1,621	0,0383	22,1%	0,022
MT081	1,888	0,147	-1,858	0,0356	25,7%	0,023
MT082	2,038	0,168	-1,663	0,0312	25,8%	0,019
MT083	1,021	0,139	-1,316	0,0736	26,3%	0,033
MT084	1,668	0,181	-1,078	0,0415	18,8%	0,014
MT085	1,038	0,097	-1,519	0,0533	11,4%	0,027
MT086	1,141	0,098	-1,704	0,0516	13,3%	0,029
MT087	1,662	0,313	-0,608	0,0864	26,3%	0,013
MT088	0,784	0,078	-1,808	0,0961	12,5%	0,045
MT089	1,583	0,085	-2,389	0,0346	7,1%	0,026
MT090	1,584	0,090	-2,268	0,0357	8,3%	0,026
MT091	1,542	0,242	-0,733	0,0715	22,2%	0,014
MT092	1,051	0,060	-2,328	0,0476	6,5%	0,028
MT093	1,878	0,112	-2,120	0,0296	11,0%	0,023
MT094	3,430	0,529	-0,839	0,0350	20,2%	0,009
MT095	1,272	0,120	-1,470	0,0439	14,4%	0,023
MT096	1,978	0,267	-0,699	0,0548	13,6%	0,009
MT097	2,830	0,233	-1,457	0,0224	20,9%	0,013
MT098	1,458	0,238	-0,560	0,0853	16,3%	0,012
MT099	1,122	0,082	-2,099	0,0610	12,1%	0,037
MT100	1,896	0,133	-1,844	0,0304	17,3%	0,021
MT101	1,605	0,138	-1,952	0,0500	28,2%	0,031
MT102	2,245	0,387	-0,636	0,0606	21,1%	0,010
MT103	1,147	0,117	-1,563	0,0577	18,3%	0,031

Fonte: Elaboração do autor, com os valores do bilog-MG.

ANEXO I - MATRIZES DE LÍNGUA PORTUGUESA- MRA

(5º ano do Ensino Fundamental - SAEB)

A matriz de referência de Língua Portuguesa é composta por seis tópicos, relacionados a habilidades dos estudantes. Dentro de cada tópico há um conjunto de descritores ligados às competências desenvolvidas.

Tópico I. Procedimentos de Leitura

D01 - Localizar informações explícitas em um texto.

D03 - Inferir o sentido de uma palavra ou expressão.

D04 - Inferir uma informação implícita em um texto.

D06 - Identificar o tema de um texto.

D11 - Distinguir um fato da opinião relativa a esse fato.

Tópico II. Implicações do Suporte, do Gênero e /ou do Enunciador na Compreensão do Texto

D05 - Interpretar texto com auxílio de material gráfico diverso (propagandas, quadrinhos, foto, etc.).

D09 - Identificar a finalidade de textos de diferentes gêneros.

Tópico III. Relação entre Textos

D15 - Reconhecer diferentes formas de tratar uma informação na comparação de textos que tratam do mesmo tema, em função das condições em que ele foi produzido e daquelas em que será recebido.

Tópico IV. Coerência e Coesão no Processamento do Texto

D02 - Estabelecer relações entre partes de um texto, identificando repetições ou substituições que contribuem para a continuidade de um texto.

D07 - Identificar o conflito gerador do enredo e os elementos que constroem a narrativa.

D08 - Estabelecer relação causa /consequência entre partes e elementos do texto.

D12 - Estabelecer relações lógico-discursivas presentes no texto, marcadas por conjunções, advérbios, etc.

Tópico V. Relações entre Recursos Expressivos e Efeitos de Sentido.

D13 - Identificar efeitos de ironia ou humor em textos variados.

D14 - Identificar o efeito de sentido decorrente do uso da pontuação e de outras notações.

Tópico VI. Variação Linguística

D10 - Identificar as marcas linguísticas que evidenciam o locutor e o interlocutor de um texto.

ANEXO II - MATRIZES DE MATEMÁTICA - MRA

(5° ano do Ensino Fundamental - SAEB)

A matriz de referência de Matemática é composta por quatro temas, relacionados a habilidades desenvolvidas pelos estudantes. Dentro de cada tema há um conjunto de descritores ligados às competências desenvolvidas.

Tema I. Espaço e Forma

D01 - Identificar a localização /movimentação de objeto em mapas, croquis e outras representações gráficas.

D02 - Identificar propriedades comuns e diferenças entre poliedros e corpos redondos, relacionando figuras tridimensionais com suas planificações.

D03 - Identificar propriedades comuns e diferenças entre figuras bidimensionais pelo número de lados, pelos tipos de ângulos.

D04 - Identificar quadriláteros observando as posições relativas entre seus lados (paralelos, concorrentes, perpendiculares).

D05 - Reconhecer a conservação ou modificação de medidas dos lados, do perímetro, da área em ampliação e /ou redução de figuras poligonais usando malhas quadriculadas.

Tema II. Grandezas e Medidas

D06 - Estimar a medida de grandezas utilizando unidades de medida convencionais ou não.

D07 - Resolver problemas significativos utilizando unidades de medida padronizadas como km/m/cm/mm, kg/g/mg, l/ml.

D08 - Estabelecer relações entre unidades de medida de tempo.

D09 - Estabelecer relações entre o horário de início e término e /ou o intervalo da duração de um evento ou acontecimento.

D10 - Num problema, estabelecer trocas entre cédulas e moedas do sistema monetário brasileiro, em função de seus valores.

D11 - Resolver problema envolvendo o cálculo do perímetro de figuras planas, desenhadas em malhas quadriculadas.

D12 - Resolver problema envolvendo o cálculo ou estimativa de áreas de figuras planas, desenhadas em malhas quadriculadas.

Tema III. Números e Operações /Álgebra e Funções

D13 - Reconhecer e utilizar características do sistema de numeração decimal, tais como agrupamentos e trocas na base 10 e princípio do valor posicional.

D14 - Identificar a localização de números naturais na reta numérica.

D15 - Reconhecer a decomposição de números naturais nas suas diversas ordens.

D16 - Reconhecer a composição e a decomposição de números naturais em sua forma polinomial.

D17 - Calcular o resultado de uma adição ou subtração de números naturais.

D18 - Calcular o resultado de uma multiplicação ou divisão de números naturais.

D19 - Resolver problema com números naturais, envolvendo diferentes significados da adição ou subtração: juntar, alteração de um estado inicial (positiva ou negativa), comparação e mais de uma transformação (positiva ou negativa).

D20 - Resolver problema com números naturais, envolvendo diferentes significados da multiplicação ou divisão: multiplicação comparativa, ideia de proporcionalidade, configuração retangular e combinatória.

D21 - Identificar diferentes representações de um mesmo número racional.

D22 - Identificar a localização de números racionais representados na forma decimal na reta numérica.

D23 - Resolver problema utilizando a escrita decimal de cédulas e moedas do sistema monetário brasileiro.

D24 - Identificar fração como representação que pode estar associada a diferentes significados.

D25 - Resolver problema com números racionais expressos na forma decimal envolvendo diferentes significados da adição ou subtração.

D26 - Resolver problema envolvendo noções de porcentagem (25%, 50%, 100%).

Tema IV. Tratamento da Informação

D27 - Ler informações e dados apresentados em tabelas.

D28 - Ler informações e dados apresentados em gráficos (particularmente em gráficos de colunas).