

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO PROFISSIONAL EM GESTÃO E
AVALIAÇÃO DA EDUCAÇÃO PÚBLICA**

TIAGO GIOSEFFI DE SOUZA

**Estudo estatístico e pedagógico de itens de Matemática do CAEd na
área de números e operações, álgebra e funções do Proeb.**

JUIZ DE FORA
2020

Tiago Gioseffi de Souza

**Estudo estatístico e pedagógico de itens de Matemática do CAEd na
área de números e operações, álgebra e funções do Proeb.**

Dissertação apresentada ao Programa de Pós-graduação Profissional em Gestão e Avaliação da Educação Pública, da Faculdade de Educação, da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de Mestre em Gestão e Avaliação da Educação Pública.

Orientador: Dr. Cassiano Caon Amorim

Juiz de Fora

2020

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

GIOSEFFI DE SOUZA , TIAGO .

Estudo estatístico e pedagógico de itens de Matemática do CAEd na área de números e operações, álgebra e funções do Proeb. / TIAGO GIOSEFFI DE SOUZA . -- 2020.

158 f. : il.

Orientador: Cassiano Caon Amorim

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Educação/CAEd. Programa de Pós-Graduação em Gestão e Avaliação da Educação Pública, 2020.

1. Elaboração de itens. 2. Formação de elaboradores. 3. Matemática. I. Caon Amorim , Cassiano , orient. II. Título.

TIAGO GIOSEFFI DE SOUZA

**Estudo estatístico e pedagógico de itens de Matemática do CAEd na
área de números e operações, álgebra e funções do Proeb.**

Dissertação apresentada ao Programa de Pós-graduação Profissional em Gestão e Avaliação da Educação Pública, da Faculdade de Educação, da Universidade Federal de Juiz de Fora, como requisito parcial à obtenção do título de Mestre em Gestão e Avaliação da Educação Pública.

Aprovada em 09 de dezembro de 2020.

BANCA EXAMINADORA



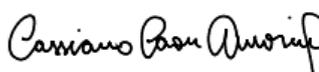
Dr. Cassiano Caon Amorim - Orientador

Universidade Federal de Juiz de Fora



Dra. Sandra Machado de Souza Lima

Universidade Federal Fluminense



Dr. Marco Aurélio Kistemann Júnior

Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Dizer que a caminhada até aqui não foi fácil seria clichê, pois todos sabemos que para alcançar nossos objetivos, não é fácil. O mais importante, não é apenas se orgulhar por superar todos os caminhos tortuosos, frustrações, medos e angústias ao longo do caminho, mas sim, recordar de quem esteve conosco e nos ajudou a superar tudo isso. É certo que a cada dia temos de provar nossas capacidades diante das situações que enfrentamos, mas é preciso lembrar que ninguém aprende a caminhar sozinho, e que a vida se torna bela e mais simples, praticando diariamente a empatia, e principalmente compartilhando o que temos de melhor. E por grande felicidade, em meu caminho estive com pessoas que se doaram, mesmo que em um pequeno instante, fazendo uma grande diferença para que este trabalho se tornasse realidade.

Em primeiro lugar agradeço a Deus que sempre está presente em tudo e em todas as coisas em nossas vidas, iluminando não só os caminhos, mas também nossos pensamentos, que sempre são voltados para o bem.

Aos meus pais, Osvaldo e Maria Helena, que foram incansáveis em todo o meu percurso escolar.

A Simone, por ter estado o tempo todo ao meu lado nessa caminhada me apoiando.

Ao Prof. Dr. Cassiano Caon, meu orientador e a Profa. Dra. Mônica da Motta, minha agente de suporte, por me darem todo o apoio acadêmico.

Aos membros da banca Profa. Dra. Sandra Machado de Souza Lima e Prof. Dr. Marco Aurélio Kistemann Júnior que gentilmente se disponibilizaram e deram suas colaborações nesta pesquisa.

A todos e todas as colegas da turma de 2018 do PPGP, em especial, Ana Paula Paiva, Bruna Lopes, Danyele Wiesel, Elizabeth Resende, Flávia Ferreira, Lethycia Lopes, Luciana Bortolucci, Rachel Finamore e Verônica Dutra, grandes companheiras do início ao fim.

A todos os companheiros e companheiras da equipe de Matemática da CIA do CAEd, por colaborarem com minha pesquisa.

Aos colegas Ailton Galvão e Clayton Furtado da CME do CAEd por me ajudarem no ponto inicial de toda pesquisa.

A todos e todas da equipe do PPGP, que sempre foram excepcionais em tudo o que se prontificam em realizar.

Ao CAEd por permitir a realização desse estudo, contribuindo para meu aprimoramento profissional.

A todos e todas que direta ou indiretamente contribuíram para realização dessa pesquisa.

“O conhecimento serve para encantar as pessoas, não para humilhá-las.”
(Mário Sérgio Cortella, s/d, p. s/p)

RESUMO

O caso de gestão a ser estudado irá discutir itens eliminados nas avaliações do Programa de Avaliação da Rede Pública de Educação Básica (PROEB) que compõe o Sistema Mineiro de Avaliação da Educação Pública (SIMAVE). Os objetivos definidos para este estudo foram analisar pedagógica e psicometricamente esses itens buscando identificar possíveis vieses adotados no processo de elaboração de itens da equipe de Matemática na Coordenação de Instrumentos de Avaliação (CIA) do pelo Centro de Políticas Públicas e Gestão da Avaliação (CAEd). Assumimos como hipótese que ocorram elaborações com a mesma estrutura de itens, ocasionando repetição de instrumentos de avaliação com os mesmos padrões, não enriquecendo em qualidade o banco. Outro aspecto que levamos em consideração são os poucos estudos e investimentos na formação do elaborador na CIA, que acaba replicando os mesmos padrões já existentes no CAEd. Para tanto, empregamos, como metodologia de pesquisa, um levantamento documental dos quantitativos de itens aplicados a partir de planilhas com os dados estatísticos dos itens e de montagem de testes do CAEd, além do referencial bibliográfico fundamentando a discussão em torno do processo de elaboração de itens e matrizes de referência. Como instrumentos utilizamos itens eliminados da área temática de Números e Operações, Álgebra e Funções da disciplina de Matemática, utilizados nas avaliações dos anos de 2014 a 2018 no PROEB. A partir dos dados gerados na pesquisa de campo, foi possível constatar alguns padrões adotados no processo de elaboração dos instrumentos, fundamentando a construção de um plano de ação voltado para a formação do elaborador e novos caminhos para pesquisa com a avaliação educacional em larga escala.

Palavras-Chave: elaboração de itens, formação de elaboradores, Matemática.

ABSTRACT

The management case to be studied will discuss items eliminated in the evaluations of the Evaluation Program of the Public Basic Education Network (PROEB) that make up the Minas Gerais System of Evaluation of Public Education (SIMAVE). The objectives defined for this study were to pedagogically and psychometrically analyze these items in order to identify possible biases adopted in the process of preparing items of the Mathematics team in the Coordination of Assessment Instruments (CIA) by the Center for Public Policies and Assessment Management (CAEd). We assume as a hypothesis that elaborations with the same structure of items occur, causing repetition of assessment instruments with the same standards, not enriching the bank in quality. Another aspect that we take into account is the few studies and investments in the training of the writer at the CIA, which ends up replicating the same standards that already exist in CAEd. To do so, we used, as a research methodology, a documentary survey of the quantities of items applied from spreadsheets with the statistical data of the items and the assembly of CAEd tests, in addition to the bibliographic reference supporting the discussion around the process of preparing reference items and matrices. As instruments we used items eliminated from the thematic area of Numbers and Operations, Algebra and Functions of the Mathematics discipline, used in the assessments from 2014 to 2018 in PROEB. From the data generated in the field research, it was possible to verify some standards adopted in the process of elaborating the instruments, justifying the construction of an action plan aimed at training the educator and new paths for research with educational evaluation on a large scale.

Keywords: item elaboration, educator training, mathematics.

LISTA DE ILUSTRAÇÃO

Figura 1	Estrutura de um item.....	27
Figura 2	Etapas do processo de elaboração de itens.....	29
Figura 3	Regras para redação de um item de Matemática	30
Figura 4	Processo de validação de itens da ETS.....	33
Figura 5	Comparando TCT e TRI.....	63
Figura 6	Curva Característica do Item j.....	64
Figura 7	Exemplo de ponto na CCI.....	65
Figura 8	Exemplo de relatório de item na CIA.....	69
Figura 9	Identificação dos parâmetros CCI.....	71
Figura 10	Modelo ajustado X observações empíricas do item M110104E4.....	77
Quadro 1	Etapas para elaboração de um item de Matemática do guia do INEP	30
Quadro 2	Observações levantadas sobre os descritores D02, D34 e D81 (PROEB 2014 a 2018)	48
Quadro 3	Limitações da TCT X Vantagens da TRI	62
Quadro 4	Passos para estimação dos parâmetros – Ajuste	76
Quadro 5	Item M120389H6	52
Quadro 6	Item M120390H6	85
Quadro 7	Item M120189H6	90

Quadro 8	Item M090047H6	94
Quadro 9	Item M100093H6	97
Quadro 10	Item M1D26I0117	101
Quadro 11	Item M040046H6	104
Quadro 12	Item M090235H6	109
Quadro 13	Item M050057H6	114
Quadro 14	Aspectos identificados na análise dos itens	117
Quadro 15	Achados e propostas de ação	124
Quadro 16	Proposta de ação I	126
Quadro 17	Proposta de ação II	128
Quadro 18	Proposta de ação III.....	129

LISTA DE TABELAS

Tabela 1 – Matriz de referências do PROEB.....	39
Tabela 2 – Quantitativo de itens eliminados por descritor – (PROEB 2014 a 2018)	48
Tabela 3 – Quantitativo de itens eliminados por área avaliada	50
Tabela 4 – Itens eliminados da disciplina de Matemática da área de Números e Operações, Álgebra e Funções, no PROEB entre os anos de 2014 e 2018.....	51
Tabela 5 – Itens eliminados na área de Números e Operações, Álgebra e Funções, no PROEB entre os anos de 2014 e 2018 com menor percentual de acerto.....	55
Tabela 6 – Classificação de acordo com a discriminação do item.....	71
Tabela 7 – Classificação de dificuldade do item	72

LISTA DE ABREVIATURAS

APA	Associação Americana de Pesquisa Educacional
BIB	Blocos Incompletos Balanceados
BNCC	Base Nacional Comum Curricular
BNI	Banco Nacional de Itens
CAEd	Centro de Políticas Públicas e Avaliação da Educação
CAP	Coordenação de Análise e Publicações
CCG	Coordenação de Casos de Gestão
CCI	Curva Característica do Item
CCM	Coordenação de Cursos e Material Didático
CDS	Coordenação de Desenvolvimento de Sistemas
CIA	Coordenação de Instrumentos de Avaliação
CIS	Coordenação de Implantação de Sistemas
CME	Coordenação de Medidas Educacionais
COA	Coordenação de Operações de Avaliação
CPA	Coordenação de Projetos Especiais de Avaliação
CPD	Coordenação de Processamento de Documentos
CPV	Coordenação de Produção Visual
CQI	Controle de Qualidade da Informação
CSA	Coordenação de Secretaria Acadêmica e Administração
CSU	Coordenação de Suporte
DIF	Differential item functioning (Funcionamento Diferencial do Item)
ETS	Educational Testing Service
IDEB	Índice de Desenvolvimento da Educação Básica
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais
MCS	Método dos Campos Semânticos
MEC	Ministério da Educação e Cultura
NEAP	National Assessment of Educational Progress
NO/AF	Números e Operações/Álgebra e Funções
PAAE	Programa de Avaliação da Aprendizagem Escolar
PCN	Parâmetros Curriculares Nacionais
PPGP	Programa de Pós-Graduação Profissional em Gestão e Avaliação da Educação Pública
PROALFA	Programa de Avaliação da Alfabetização
PROEB	Programa de Avaliação da Rede Pública de Educação Básica
SAEB	Sistema de Avaliação da Educação Brasileira
SEE/MG	Secretaria do Estado de Educação de Minas Gerais
SIMAVE	Sistema Mineiro de Avaliação da Educação Pública
TCT	Teoria Clássica dos Testes
TRI	Teoria de Resposta ao Item
UAV	Unidade de Avaliação

UFJF	Universidade Federal de Juiz de Fora
UFO	Unidade de Formação
USG	Unidade de Sistemas de Gestão

SUMÁRIO

1	INTRODUÇÃO	14
2	ESTUDO DA ÁREA DE NÚMEROS E OPERAÇÕES, ÁLGEBRA E FUNÇÕES E A ELIMINAÇÃO DE ITENS NO PROEB	19
2.1	BREVE HISTÓRICO DAS POLÍTICAS DE AVALIAÇÃO NO BRASIL	19
2.2	DESENHO E CONTEXTO ATUAL DO CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO	23
2.3	OS INSTRUMENTOS DE AVALIAÇÃO EM LARGA ESCALA	25
2.3.1	A elaboração de itens	25
2.3.2	A Matriz de Referência e suas relações com o currículo e o ensino	35
2.4	APRESENTANDO AS EVIDÊNCIAS DO CASO	45
3	A PSICOMETRIA NOS ITENS (A ESTATÍSTICA NOS ITENS) E ANÁLISE DOS ITENS ELIMINADOS	57
3.1	MÉTODOS DE CORREÇÃO DE TESTES	57
3.1.1	Teoria Clássica dos Testes - T.C.T.	58
3.1.2	Teoria de Resposta ao Item – T.R.I.	60
3.1.3	Parâmetros Estatísticos	68
3.1.4	O que significa um item eliminado	75
3.1.5	Procedimentos e instrumentos metodológicos	79
3.2	ANÁLISE DOS ITENS	80
4	PLANO DE AÇÃO EDUCACIONAL	123
4.1	EIXO 1 – ESTIMULAR A PESQUISA COM ITENS NO INTUITO DE REDUZIR AS ELIMINAÇÕES E OTIMIZAR O TRABALHO DA ELABORAÇÃO	125
4.1.1	Levantamento do quantitativo de itens eliminados por áreas temáticas e produção de pesquisa	126
4.1.2	Possibilitar caminhos para produção de novos itens e realizar pré-testes	127
4.2	EIXO 2 – APRIMORAR OS ASPECTOS FORMATIVOS DA EQUIPE	129
5	CONSIDERAÇÕES FINAIS	132
	REFERÊNCIAS	134
	APÊNDICE A – Questionário de pesquisa acadêmica	139
	APÊNDICE B – Plano de ação educacional proposto	153

1 INTRODUÇÃO

Esta dissertação, desenvolvida no Programa de Mestrado em Gestão e Avaliação da Educação Pública da Universidade Federal de Juiz de Fora, tem por finalidade analisar pedagógica e psicometricamente nove itens que foram eliminados pela metodologia da Teoria da Resposta ao Item (TRI) em testes de proficiência aplicados pelo Centro de Políticas Públicas e Gestão da Avaliação (CAEd) nas últimas cinco edições do Programa de Avaliação da Rede Pública de Educação Básica (PROEB), projeto que compõe o Sistema Mineiro de Avaliação da Educação Pública (SIMAVE) no estado de Minas Gerais.

Considerando o vínculo que tem-se com Coordenação de Instrumentos de Avaliação (CIA) do CAEd, e atuando como analista de instrumentos de avaliação de Matemática, verificou-se que uma das principais atribuições dessa função é a elaboração de itens para composição de testes de proficiência. Portanto, a partir de itens que apresentaram eliminação nesses testes, essa pesquisa pretende verificar se existe algo em comum, ou até mesmo vícios desenvolvidos no processo de elaboração desses itens, que possam estar contribuindo para eliminação estatística dos mesmos.

Sobre este aspecto, a interpretação estatística de cada um desses instrumentos foi fundamental para essa análise. Outro ponto que cabe salientar, é sobre a possibilidade de tais itens poderem contribuir para a formação dos elaboradores do centro de políticas públicas, sob a perspectiva pedagógica. Identificou-se ainda que um item eliminado pela TRI pode contribuir significativamente para interpretações do nível de aprendizagem dos estudantes avaliados, auxiliando a esquadriñar possíveis linhas de raciocínio adotadas por eles, o que permite cada vez mais diagnósticos precisos, possibilitando ainda, novos tipos de distratores e diversos outros elementos que compõe um item. A partir da análise psicométrica, um item pode ser eliminado por três motivos estatísticos, que são: ajuste, bisserial ou por comportamento diferencial do item (DIF).¹

Dessa forma, estatisticamente, tais itens podem não funcionar em determinadas populações ou na rede avaliada, no entanto, a partir dos dados estatísticos, conjectura-se ser possível resgatar informações sobre a aprendizagem

¹ Ajuste, bisserial e DIF serão abordados mais adiante nesse trabalho.

dos alunos, ou seja, sob o ponto de vista pedagógico, o item tem muito a contribuir tanto para a formação interna dos analistas do CAEd, como ferramenta para tomada de decisões e estratégias de intervenção.

Nesse contexto de avaliações em larga escala, o Centro de Políticas Públicas e Avaliação da Educação, da Universidade Federal de Juiz de Fora (CAEd/UFJF), está totalmente imerso, atuando em diversos estados e municípios, operacionalizando maneiras de mensurar o rendimento de estudantes e as condições de oferta da educação básica nas redes públicas.

Sendo assim, um estudo estatístico e pedagógico aprofundado de itens eliminados pode possibilitar-nos aprimorar nossas práticas de trabalho, auxiliando principalmente na elaboração de testes cada vez mais adequados para as populações a serem avaliadas, ou seja, buscar equilíbrio entre proficiência, dificuldade do item e população avaliada.

Conjugando os objetivos do CAEd para com a avaliação em larga escala no Brasil, e tendo em vista a formação continuada e o aperfeiçoamento técnico das equipes de elaboração, essa pesquisa se propõe a colocar em pauta a discussão de nove itens eliminados do conteúdo de Matemática, da área de Números e Operações, Álgebra e Funções ao longo de cinco edições da matriz do PROEB, projeto de atuação mais longo do CAEd, e ainda, o que mais se aproxima da realidade do pesquisador enquanto professor de Matemática que das redes municipal e estadual do município de Juiz de Fora/MG, se aproximando das duas realidades, ou seja, de um lado a visão enquanto professor em sala de aula, e de outro, como colaborador da instituição avaliadora.

Salienta-se ainda a possibilidade de implementação de novos processos de análises pedagógicas a partir de itens eliminados, na expectativa de contribuir para a formação da equipe de elaboração na Coordenação de Instrumentos de Avaliação (CIA), ou demais centros de pesquisa em avaliação que tenham interesse em implementar novas práticas de trabalho, no que tange a qualidade do banco de itens e a interpretação da aprendizagem das redes avaliadas contratantes.

Foi possível identificar que a área de Números e Operações, Álgebra e Funções foi a que revelou maior quantitativo de itens com eliminação no período observado. Tais itens serão apresentados, analisados e discutidos individualmente, sob as perspectivas psicométrica e pedagógica, considerando os dados estatísticos e as

práticas diárias dos elaboradores da equipe de Matemática na Coordenação de Instrumentos de Avaliação (CIA).

Conforme aponta Ribeiro e Cury (2018), a Álgebra é um dos pilares da Matemática

[...] e objeto de pesquisa desde que a humanidade se debruçou sobre a realidade para construir seu conhecimento, chegando às abstrações que permitem novas visões sobre cada conceito criado. Assim, deveria ser explorada desde os anos iniciais do ensino, pois dela faz parte um conjunto de processos e pensamentos que têm origem em experiências com números, padrões, entes geométricos e análise de dados. (RIBEIRO & CURY, 2018, p.5).

Ponte (2006) afirma que na maioria dos países, Números e Álgebra são dois temas fundamentais da Matemática escolar. Os Números têm um papel decisivo nas aprendizagens matemáticas nos primeiros anos de escolaridade e a Álgebra surge como um tema matemático fundamental a partir dos anos intermediários. Quem não tiver uma capacidade razoável de trabalhar com números e suas operações e de entender e usar a linguagem abstrata da Álgebra, fica seriamente limitado nas suas opções escolares e profissionais e no seu exercício da cidadania democrática.

A Base Nacional Comum Curricular (BNCC), organiza e aponta de maneira ampla, as cinco áreas temáticas da disciplina de Matemática que devem ser trabalhadas pelos professores em sala de aula, abordando os objetivos e a importância de se trabalhar cada uma delas, que são: Números, Álgebra, Geometria, Grandezas e Medidas, e Probabilidade e Estatística.

Sobre a área temática de números, a BNCC apresenta como objetivo, o desenvolvimento do pensamento numérico, que implica o conhecimento de maneiras de quantificar atributos de objetos e de julgar e interpretar argumentos baseados em quantidades. No processo da construção da noção de número, os alunos precisam desenvolver, entre outras, as ideias de aproximação, proporcionalidade, equivalência e ordem, noções fundamentais da Matemática. Para essa construção, é importante propor, por meio de situações significativas, sucessivas ampliações dos campos numéricos. No estudo desses campos numéricos, devem ser enfatizados registros, usos, significados e operações (BRASIL, 2017).

Já a unidade temática Álgebra, por sua vez, objetiva o desenvolvimento de um tipo especial de pensamento – pensamento algébrico – que é essencial para utilizar

modelos matemáticos na compreensão, representação e análise de relações quantitativas de grandezas e, também, de situações e estruturas matemáticas, recorrendo ao uso de letras e outros símbolos. Para esse desenvolvimento, é necessário que os alunos identifiquem regularidades e padrões de sequências numéricas e não numéricas, estabeleçam leis matemáticas que expressem a relação de interdependência entre grandezas em diferentes contextos, bem como criar, interpretar e transitar entre as diversas representações gráficas e simbólicas, para resolver problemas através de equações e inequações, com compreensão dos procedimentos utilizados. As ideias matemáticas fundamentais vinculadas a essa unidade são: equivalência, variação, interdependência e proporcionalidade. Em síntese, essa unidade temática deve enfatizar o desenvolvimento de uma linguagem, o estabelecimento de generalizações, a análise da interdependência de grandezas e a resolução de problemas por equações ou inequações (BRASIL, 2017).

Assim, ao se realizar o levantamento do número total de itens que apresentaram episódios de eliminação no PROEB, entre os anos de 2014 e 2018, a área de Números e Álgebra revelou o total de vinte e oito itens, sendo o maior quantitativo dentre todas as áreas cobertas pela Matriz de Referência, se tornando o primeiro estímulo a realização dessa pesquisa. Cabe ressaltar que essas duas áreas da Matemática são bastante exploradas e estudadas por diversos educadores e estudiosos, como PONTES (2008) e ARAÚJO (2008). Esta última aponta que o cenário atual do ensino de Álgebra no Brasil é reflexo de várias fases de sua evolução, e uma breve revisão do ensino dessa área e de resultados de pesquisa torna-se necessária para se compreender o que hoje acontece na sala de aula.

Portanto, um dos caminhos que adotados para termos essa percepção, é através desse estudo, em que busca-se entender possíveis padrões adotados pela equipe de elaboração que possam estar contribuindo para a eliminação desses itens. Quais são os motivos pedagógicos que ocasionaram a eliminação e de que maneira ela pode contribuir para a formação da equipe de elaboração no CAEd? Para atingirmos esse objetivo, apresentaremos nove itens com seus respectivos motivos de eliminação, que são bisserial, ajuste e DIF, na expectativa de contribuir para a formação e aprimoramento da equipe de elaboração e contribuir para composição de um banco de itens de qualidade. Um banco de itens confiável, com itens válidos, apresentando bons parâmetros é essencial para um centro de pesquisa e avaliação, como é o CAEd. Os impactos positivos que isso pode causar, seja no âmbito da

pesquisa educacional e até mesmo financeiro para a instituição, nos mobiliza para este estudo.

Dessa forma, tem-se como questão de pesquisa a seguinte pergunta: **Como podemos intervir no fluxo de trabalho da elaboração para que ocorram cada vez menos eliminações de itens?**

Esta pergunta nos estimula investigar se a forma atual de realização do trabalho da equipe de Matemática na CIA é feita de forma adequada aos referenciais teóricos sobre a elaboração e construção de instrumentos de avaliação.

Assim, no primeiro capítulo deste estudo sobre a área de Números e Operações, Álgebra e Funções do PROEB, será realizado um breve histórico sobre a avaliação em larga escala no Brasil, seu surgimento, propósitos e processo de desenvolvimento. Se apresentarão também, o desenho atual do CAEd, como um centro de políticas públicas no cenário educacional nacional, suas atividades e dinâmica de trabalho. Serão abordados a elaboração de itens, bem como os itens eliminados, possibilitando-nos analisar como são desenvolvidos os itens de Matemática na CIA, além de abordarmos a Matriz De Referência e as evidências sobre a produção e eliminação de itens no capítulo I.

No segundo capítulo, será discutida a psicometria dos itens, os métodos de correção clássica e da teoria de resposta ao item (TRI), a análise dos itens e reflexões sobre a eliminação de itens em relação aos parâmetros estatísticos. A discussão teórica, a metodologia e análise dos dados também serão tratados nesse capítulo.

No capítulo 3, apontaram-se possíveis propostas de intervenção no processo de elaboração de itens e mudanças na Matriz De Referência, assim como, na dinâmica de capacitação e formação de elaboradores no CAEd.

2 ESTUDO DA ÁREA DE NÚMEROS E OPERAÇÕES, ÁLGEBRA E FUNÇÕES E A ELIMINAÇÃO DE ITENS NO PROEB

Nesse estudo sobre a área de Números e Operações, Álgebra e Funções do PROEB, será realizado primeiramente neste capítulo, um breve histórico das políticas de avaliação educacional no Brasil, além de como se dá o atual desenho e funcionamento do CAEd, e a elaboração de itens na CIA. Ainda neste capítulo serão discutidas as matrizes de referência, em específico a da avaliação do PROEB, que compõe o SIMAVE no estado de Minas Gerais, finalizando com relatos que propõe a presente dissertação.

2.1 BREVE HISTÓRICO DAS POLÍTICAS DE AVALIAÇÃO NO BRASIL

Há quase três décadas iniciou-se no Brasil o processo de avaliações em larga escala na educação brasileira, fortemente influenciado pela promulgação da Constituição Federal de 1988 (BRASIL, 1988). Os norte-americanos já vivenciavam experiências com avaliações educacionais desde o ano de 1969 com a *National Assessment of Educational Progress* (NEAP), o que inspirou o Brasil, desde a década de 80, a criar, pelo Ministério da Educação e Cultura (MEC), através do seu Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), o Sistema de Avaliação da Educação Brasileira (SAEB), cujo formato e concepção foram bem semelhantes ao sistema americano, como, por exemplo, o desenho amostral e a estruturação das provas (PONTES, 2012). Sua primeira edição ocorreu em 1990, avaliando de forma amostral turmas de 1.º, 3.º, 5.º e 7.º anos do ensino fundamental, nas disciplinas de Língua Portuguesa, Ciências Naturais, Redação e Matemática, esta última que será a discutida nesse trabalho.

O SAEB, desde então, é aplicado a cada dois anos e é composto por um conjunto de avaliações que permite realizar um diagnóstico da educação básica brasileira e de alguns fatores que permitem interferir no desempenho do estudante, fornecendo um indicativo de qualidade do ensino ofertado, por meio de provas elaboradas a partir de matrizes de referências e questionários (INEP, 2019).

Em sua terceira edição, no ano de 1995, foi adotada na construção do teste a Teoria da Resposta ao Item (TRI), permitindo inovações tecnológicas a qual se configuram e se consolidaram em seu desenho atual, permitindo estimular habilidades

dos alunos, independentemente do conjunto de itens respondidos² (BONAMINO, 2012)

Desde 2005, o governo federal, por meio do INEP, utiliza a média combinada, obtida nas disciplinas de Língua Portuguesa e Matemática nas avaliações do SAEB ou da Prova Brasil, como um dos elementos para o cálculo do principal indicador de qualidade da educação nacional, o Índice de Desenvolvimento da Educação Básica (IDEB).

Assim, o SAEB é considerado um marco na avaliação educacional brasileira, que despertou o interesse de estados e municípios a criarem seus sistemas próprios de avaliação em larga escala para suas respectivas redes de ensino, na expectativa de utilizarem o modelo desenvolvido na esfera federal como forma de garantir estratégias para aprendizagem de seus alunos e observar a qualidade de oferta da educação.

Minas Gerais foi um dos estados pioneiros na criação de sistemas próprios de avaliação em larga escala, já apresentando seus primeiros passos desde a década de 90, em que a Secretaria do Estado de Educação – SEE/MG, instituiu através da resolução n.º 6.908, de 18 de janeiro de 1992, o Programa de Avaliação da Escola Pública de Minas Gerais. Com o decorrer dos tempos, Minas Gerais continuou se aprimorando e ganhando destaque na avaliação de sua rede de ensino. Foi até que entre os anos de 1999 e 2002, quando o estado era governado por Itamar Franco, tendo à frente da SEE/MG o então secretário Murilo Avelar Hingel, que instituiu através da resolução n.º 14 de 03 de fevereiro de 2000, o Sistema Mineiro de Avaliação da educação Básica – SIMAVE. Posteriormente, através da resolução n.º 14 de 03 de fevereiro de 2000, substituída em seguida pela resolução n.º 104 de 14 de julho de 2000, Hingel instituiu o Programa de Avaliação da Educação Básica – PROEB, com amplo apoio do CAEd no seu desenvolvimento e elaboração.

Com o firme propósito de criar um sistema de ensino mais justo e inclusivo, no qual as hipóteses de aprendizagem sejam iguais para todos os estudantes, o SIMAVE é um programa que visa diagnosticar o desempenho dos alunos em diferentes áreas do conhecimento e níveis de escolaridade, bem como subsidiar a implementação, a (re)formulação e o monitoramento de políticas educacionais, contribuindo ativamente para a melhoria da qualidade da educação no estado. Desde que foi instituído, o

² A TRI será abordada adiante nas próximas seções deste trabalho.

sistema já avaliou milhares de estudantes em diferentes etapas de escolaridade da educação básica (MINAS, 2011).

Além do PROEB, o SIMAVE é composto por outros dois programas, cobrindo quase em sua totalidade as etapas de escolarização no estado de Minas Gerais, que é o Programa de Avaliação da Alfabetização (PROALFA) e o Programa de Avaliação da Aprendizagem Escolar (PAAE). O PROALFA avalia os estudantes do 2.º e 3.º anos do Ensino Fundamental em Língua Portuguesa. Sua primeira aplicação foi realizada em 2005 de forma amostral, e desde então, o programa passou por sucessivas ampliações, como a ocorrida no ano de 2017 em que Matemática passou a fazer parte das disciplinas avaliadas. Já o PAAE se trata de um sistema on-line que permite avaliar a aprendizagem a partir da geração de provas e emissão de relatórios de desempenho por turma, que podem ser impressas e corrigidas na própria escola. O PAAE tem a finalidade de identificar intervenções pedagógicas que se tornam necessárias de maneira imediata, em duas aplicações ao longo do ano letivo, uma no início do primeiro semestre como diagnóstica das turmas e outra no final do ano letivo para certificação da aprendizagem e possíveis avanços, possibilitando a avaliação planejada dos professores. Desenvolve também formação de educadores por meio de oficinas presenciais e a distância sobre temas como produção de itens e produção de texto (AVALIAR, 2019).

O SIMAVE ainda conta com estratégias para apropriação dos resultados das avaliações por professores, gestores, alunos e comunidade, de maneira assegurar que eles sejam utilizados como indicativos da qualidade educacional.

Com esse propósito, a SEE/MG, em parceria com o Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF), apresenta uma ampla política de divulgação e apropriação dos resultados do Proeb/Proalfa. Revistas para os gestores e professores, cartazes personalizados com os dados de cada unidade escolar, material para oficinas de estudo e vídeos educativos compõem uma série de produtos distribuídos às escolas municipais e estaduais de Minas Gerais. Esses produtos foram elaborados sob três importantes princípios: o de informar os resultados do SIMAVE, o de subsidiar as ações de intervenção pedagógica e o de fornecer indicadores para a elaboração de ações de gestão. Com essa ação, a SEE/MG presta contas à sociedade acerca da qualidade dos serviços educacionais desenvolvidos em sua rede de ensino. (SIMAVE, 2011, p. 8).

As revistas trazem por objetivo divulgar os resultados da avaliação externa em larga escala, detalhando suas etapas. São apresentados os fundamentos da avaliação: a Matriz de Referência, que traz as habilidades avaliadas pelo teste; a composição dos cadernos de teste; a diferença entre Teoria da Resposta ao Item (TRI) e Teoria Clássica dos Testes (TCT); a estrutura da Escala de Proficiência, com seus Domínios e Competências; os Padrões de Desempenho Estudantil, acompanhados de itens exemplares (SIMAVE, 2011).

Dessa forma, o SIMAVE se compõe em três avaliações (PROALFA, PROEB e o PAAE), porém, esta pesquisa se restringirá apenas as avaliações do PROEB, que se dedica a avaliar os estudantes dos anos finais do ensino fundamental e médio. O PROEB teve sua primeira edição no ano 2000, com avaliações de Língua Portuguesa e Matemática. Em 2001, avaliou os conteúdos de Ciências Humanas e Ciências da Natureza, em 2002 avaliou apenas o conteúdo de Língua Portuguesa, já em 2003, apenas Matemática, em 2004 e 2005 não realizou essas avaliações retornando em 2006, avaliando sempre os conteúdos de Língua Portuguesa e Matemática (SOARES, 2011, p. 63).

Para que diagnósticos sejam estabelecidos, é preciso avaliar. Não há melhoria na qualidade da educação que seja possível sem que processos de avaliação que acompanhem, continuamente, os efeitos das políticas educacionais propostas para tal fim (SIMAVE, 2011).

Através do SIMAVE, o estado de Minas Gerais segue em busca de oferecer uma educação de qualidade, prestando contas à sociedade através desse sistema de avaliação.

2.2 DESENHO E CONTEXTO ATUAL DO CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO

O Centro de Políticas Públicas e Avaliação da Educação (CAEd) é uma instituição vinculada a Universidade Federal de Juiz de Fora (UFJF), que tem por finalidade operacionalizar, elaborar e desenvolver programas estaduais e municipais de educação, destinados a mensurar o rendimento de estudantes das escolas públicas. O CAEd também cria e promove cursos de formação, qualificação e aprimoramento aos profissionais da Educação de diversos estados do Brasil, além de desenvolver softwares para a gestão de escolas públicas com o objetivo de

modernizar a gestão educacional. Oferece ainda apoio para o desenvolvimento de projetos educacionais promovidos a partir de iniciativas privadas (CAEd, 2019).

Conforme apresentado no site da instituição, o CAEd se organiza em cinco unidades que posteriormente se subdividem em coordenações. Essas unidades são: Unidade de Pesquisa, Unidade de Avaliação, Unidade de Formação, Unidade de Sistemas de Gestão e Unidade de Administração.

Tocantins (2018, p. 39) denota as subdivisões e atuação das coordenações existentes nas unidades de pesquisa, de avaliação, formação e da unidade de sistemas de gestão.

A Unidade de Pesquisa se ocupava do desenvolvimento de programas de pesquisas sobre os objetos e processos relacionados ao CAEd e à política de avaliação. Organizava-se nos seguintes programas: o Programa 1: Avaliação, Currículo e Desenvolvimento Profissional de Gestores e Professores da Educação Básica; o Programa 2: Equidade, Políticas e Financiamento da Educação Pública; o Programa 3: Modelos, Instrumentos e Medidas Educacionais; e o Programa 4: Gestão, Avaliação e Reforma da Educação Pública.

A Unidade de Avaliação (UAV) é responsável por todo o desenvolvimento das avaliações em larga escala, desde o planejamento até a aplicação e divulgação dos resultados para as redes.

Para tanto, organizava-se nas seguintes coordenações: a Coordenação de Análise e Publicações (CAP), a Coordenação de Instrumentos de Avaliação (CIA), a Coordenação de Medidas Educacionais (CME), a Coordenação de Operações de Avaliação (COA), a Coordenação de Processamento de Documentos (CPD), a Coordenação de Produção Visual (CPV) e a Coordenação de Projetos Especiais de Avaliação (CPA). (TOCANTINS, 2018, p. 39).

Já a Unidade de Formação (UFO) é responsável pelo desenvolvimento do aperfeiçoamento profissional dos diversos atores envolvidos no cenário educacional, tais como os gestores, secretários, especialistas educacionais e principalmente os professores.

Para a estruturação desses cursos, a UFO mantinha as seguintes coordenações: a Geral da Unidade de Formação, a do Programa de Pós-Graduação Profissional em Gestão e Avaliação da Educação Pública, a de Casos de Gestão (CCG), a de Cursos e Material Didático (CCM) e a de Secretaria Acadêmica e Administração (CSA). (TOCANTINS, 2018, p. 39).

O CAEd ainda conta com a Unidade de Sistemas de Gestão (USG), que fica a cargo de solucionar e desenvolver programas e softwares que auxiliem na gestão e a execução de tarefas dentro do âmbito escolar das redes de ensino. A USG é subdividida em outras quatro coordenações, que são: Coordenação de Desenvolvimento de Sistemas (CDS), Controle de Qualidade da Informação (CQI), Coordenação de Suporte (CSU) e Coordenação de Implantação de Sistemas (CIS).

No ano de 2016, Tocantins (2018, p. 40), aponta que houve uma reestruturação no CAEd, ocasionando realojamento e reorganização de algumas equipes internas, não gerando extinção de nenhuma das estruturas já existentes. Porém, a partir do referido ano, o CAEd passou a atuar sobre duas frentes, caracterizadas pelo fazer e o controle. O fazer se refere desde os primeiros encaminhamentos até o fechamento dos contratos com as redes de ensino, até a aplicação dos testes. O controle envolve a devolutiva dos dados, pautada na comunicação do CAEd com a rede.

Assim, o CAEd busca o

(i) o desenvolvimento de tecnologias que possibilitem a produção e publicação integradas de dados de pesquisa e dados administrativos e o uso intensivo de dispositivos digitais na avaliação e monitoramento dos sistemas educacionais; (ii) a promoção de uma abordagem sistêmica da avaliação e do monitoramento dos sistemas educacionais. (CAED, s.d, p. 8).

A atuação do CAEd também se estende ao desenvolvimento de tecnologias digitais no campo da avaliação e da gestão educacional, com o objetivo de modernizar e otimizar os trabalhos no ambiente educacional. Dessa forma, o CAEd busca se especializar cada vez mais no ramo da avaliação em larga escala, desenvolvendo instrumentos de avaliação que se adequem a atual conjuntura da sociedade, aferindo com fidedignidade as diferentes competências e habilidades das distintas áreas de conhecimento, possibilitando ao estudante a garantia do direito à aprendizagem (TOCANTINS, 2018, p. 44).

O CAEd traz um relacionamento de longa data com o SIMAVE, estando juntos desde a criação do PROEB no ano 2000. Todo esse sistema de avaliação foi desenvolvido no Centro de Políticas Públicas em parceria com o governo do Estado de Minas Gerais, tendo como finalidade diagnosticar o desempenho dos alunos em diferentes áreas do conhecimento e níveis de escolaridade, bem como subsidiar a

implementação, a (re)formulação e o monitoramento de políticas educacionais, contribuindo ativamente para a melhoria da qualidade da educação no Estado (SIMAVE, 2009).

2.3 – OS INSTRUMENTOS DE AVALIAÇÃO EM LARGA ESCALA

Nesta seção serão tratados nos dois tópicos seguintes, a estrutura e o processo de elaboração de um item e o que é uma Matriz de Referência, trazendo em destaque a matriz do PROEB. Como parte do referencial teórico nesta seção serão trazidos alguns aspectos sobre a elaboração baseado nos guias de elaboração e revisão de itens do CAEd e do INEP, além de alguns complementos contidos nos *Standarts* da Associação Americana de Pesquisa Educacional (APA) e nas Diretrizes para Revisão e Equidade na Avaliação da Educational Testing Service (ETS). Já sobre as matrizes de referência serão abordar seu Educacional conceito, estrutura, sua utilidade para os sistemas de avaliação, e ainda, traremos como exemplo a matriz do PROEB.

2.3.1 A Elaboração de Itens

Elaborar um bom item que vá compor uma avaliação em larga escala não é uma tarefa simples, que exige bom conhecimento não só da área específica, mas dos objetivos ao qual se deseja avaliar. De acordo com o guia de elaboração e revisão de itens do INEP, um item é a unidade básica de um instrumento para coleta de dados, que pode ser uma prova, um questionário etc.

Um item é a unidade básica observável de qualquer teste que geralmente contém uma afirmação que provoca ou exige uma resposta por parte do avaliando. Essa resposta recebe um valor numérico específico, quase sempre 1 quando está correta e 0 quando está errada, mas pode também ser alocada em uma escala de valores indo do baixo ao alto. (HARDEN, 2017 apud HALADYNA, 2004).

Nos testes educacionais, item pode ser considerado sinônimo de questão, termo mais popular e utilizado com frequência nas escolas. A elaboração de itens, conforme aponta o guia de elaboração de itens do CAEd, está listado na segunda etapa no processo de implementação de uma avaliação externa, porém, em conjunto a este processo nessa etapa, estão incluídos o estabelecimento e utilização de uma

Matriz de Referência, a pré-testagem dos itens e a análise estatística e pedagógica desses instrumentos. Mas neste tópico vamos dar ênfase ao processo de elaboração, considerando as orientações existentes no guia de elaboração de itens do CAEd, do INEP, nos *Standarts* da Associação Americana de Pesquisa Educacional (APA) e nas Diretrizes para Revisão e Equidade na Avaliação da Educational Testing Service (ETS).

Os itens para composição de testes podem ser de dois tipos: de resposta livre ou objetivos. No entanto, trata-se de itens objetivos, cujo tipo mais comum que compõem os testes de avaliação educacional, exige dos avaliados a busca por uma única alternativa correta, geralmente construídos sob o padrão de múltipla escolha. Para se elaborar itens de qualidade, cada instituição segue alguns padrões, que se tornam semelhantes em alguns aspectos em sua maioria. O perfil do elaborador interfere em grande parte desse processo, os quais estão destacados a seguir:

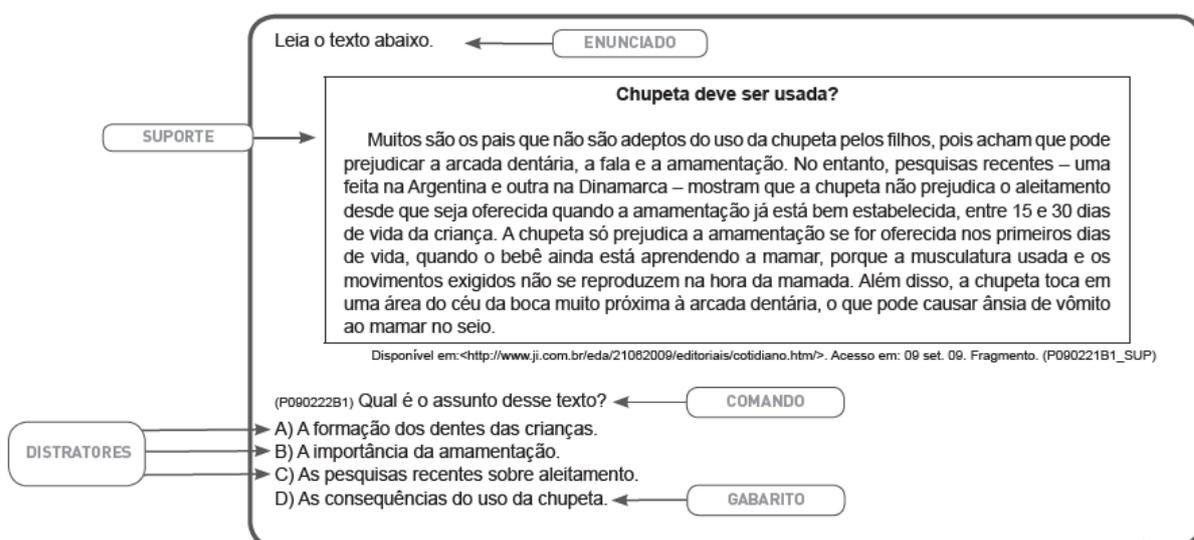
1. O elaborador deve ter domínio da área de conhecimento a ser avaliada. Isso significa que ele precisa entender o conteúdo escolar como um meio para se desenvolverem habilidades e competências. Trata-se de explorar, conforme abordado nos PCNs, o conteúdo nas suas dimensões “conceitual, factual, procedimental e atitudinal”, de modo a levar o estudante a mobilizar seus recursos cognitivos.
2. O elaborador deverá entender os processos de desenvolvimento e aprendizagem que caracterizam os estudantes para os quais o item será construído. Isso significa que o professor-elaborador deve estar familiarizado com os prováveis níveis de desenvolvimento cognitivo e educacional, a fim de ajustar a complexidade e o grau de dificuldade dos itens de modo apropriado ao padrão das alternativas de resposta.
3. O elaborador deve ter o domínio da linguagem verbal utilizada pelos estudantes para quem o teste será construído. Ele deve, além de conhecer o significado das palavras e usá-las, ser habilidoso no seu emprego, de modo a fazer com que elas expressem o desejado da maneira mais simples possível.
4. O elaborador deve ter a habilidade de utilizar as técnicas de escrever itens. Para isso, é preciso que esteja familiarizado com os diversos tipos de teste e com suas possibilidades e limitações. Além disso, deve conhecer as características gerais de bons itens e precisa estar consciente dos erros comumente cometidos. (CAEd, 2008, p.18).

Assim como COSTA (2014), defende-se que para exercer a função como elaborador de itens, é interessante que se recrute e treine profissionais com experiência docente. A experiência docente e o uso da linguagem adequada aos avaliados são dois fatores indispensáveis ao elaborador, que deve ser capaz de antever a dificuldade que os avaliados poderão ter durante a resolução do item.

A elaboração de um item se inicia orientada por uma Matriz de Referência, composta por descritores que avaliam as habilidades ou competências que são esperadas, associando conteúdos curriculares e operações mentais desenvolvidas pelos estudantes (CAEd, 2008).

O passo seguinte é a atenção quanto a estrutura do item, que é composta basicamente por quatro componentes: enunciado, suporte, comando e alternativas. A Figura 1 a seguir ilustra cada um desses componentes utilizando o exemplo de um item de Língua Portuguesa.

Figura 1 – Estrutura de um item



Fonte: CAEd (2014).

O **enunciado** é o estímulo para que o estudante mobilize os recursos cognitivos necessários para solucionar uma situação problema. Nele devem estar contidas todas as informações indispensáveis para que se atinja o sucesso na tarefa. O **suporte** já não se torna obrigatório em todo item, uma vez que existem habilidades que ao serem avaliadas, eles não se tornam necessários. Porém, para algumas habilidades específicas, principalmente na disciplina de Matemática, o suporte torna-se indispensável, atuando em conjunto com o enunciado, como estímulo para resolução do item, podendo ser apresentado na forma de um texto, figura, imagem ou outra categoria de recurso. Já o **comando**, a grosso modo, seria a pergunta do item. Ele deve ser uma instrução clara e objetiva do que o avaliado precisa realizar, podendo ser na forma de uma interrogação ou mesmo uma frase a ser completada pela alternativa correta.

As **alternativas** são apresentadas como uma lista de opções, que se divide entre o gabarito, que se trata da única alternativa correta e os distratores. Este último indica alternativas incorretas a resolução do problema, que devem ser elaborados coerentemente com a situação-problema proposta, aparentando serem corretas e que envolvam situações hipotéticas de raciocínio em busca da solução (INEP, 2010 apud HALADYNA, 2004).

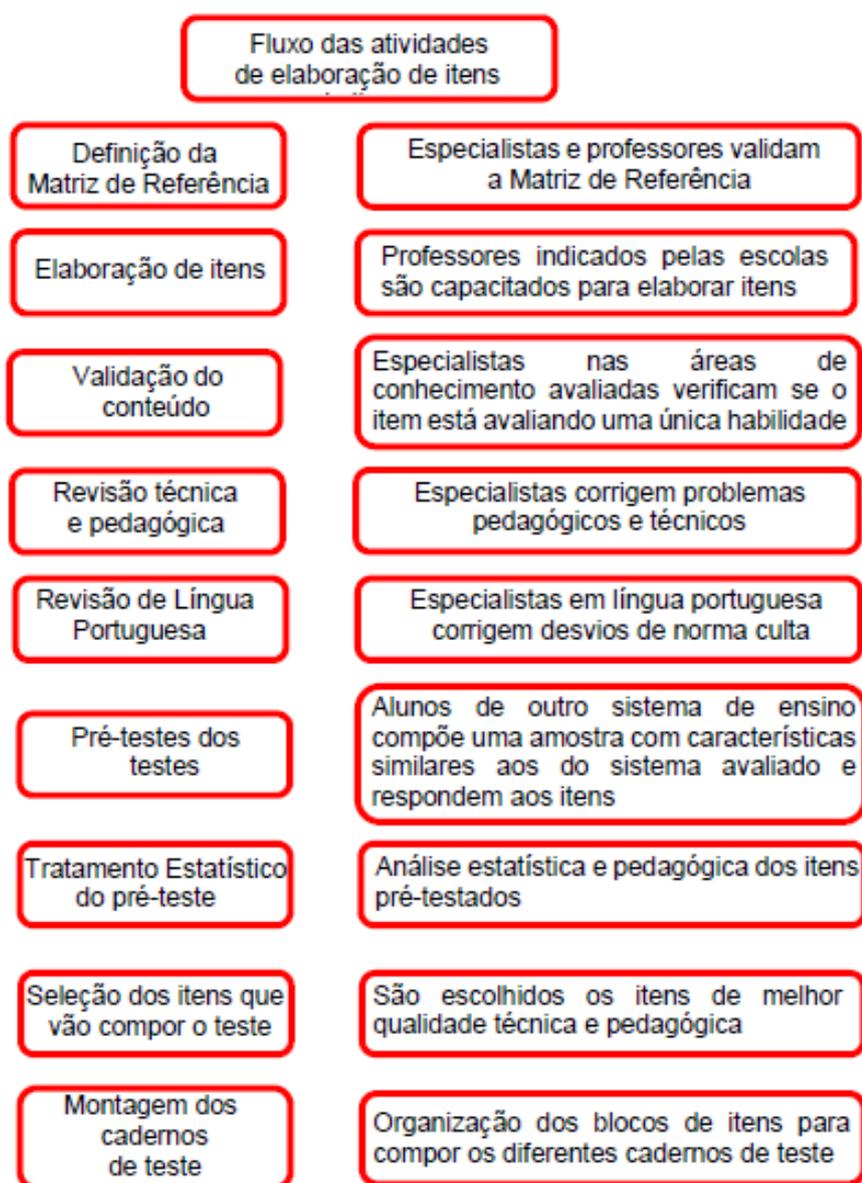
Um distrator deve buscar sempre retratar a dificuldade do avaliado em relação à habilidade proposta, não sendo construído de forma que induza ao erro, como “pegadinhas”, mas que traduza de fato a dificuldade do estudante em resolver a situação proposta. Cada distrator deve ser justificado quando o item é construído, demonstrando o possível caminho adotado pelo estudante durante a resolução, e ainda, fornecendo elementos que permitem verificar a plausibilidade dos distratores no processo posterior de revisão técnico-pedagógica. É interessante que haja plausibilidade e harmonia entre os distratores, de forma que nenhum deles se destaque em relação aos demais e nem ao gabarito.

Isso significa que o distrator plausível deve retratar hipóteses de raciocínio utilizadas na busca da solução da situação-problema apresentada. Como consequência, se esse distrator retrata uma dificuldade real do participante com relação à habilidade, não devem ser criadas situações capazes de induzi-lo ao erro. A utilização de erros comuns observados em situação de ensino-aprendizagem costuma aumentar a plausibilidade dos distratores. Por outro lado, aqueles que retratam erros grosseiros ou alternativas absurdas, dentro ou não do contexto do item, tendem a induzir a identificação da alternativa correta. (INEP, 2010, p. 11).

O item deve ser estruturado de modo que se configure uma unidade de proposição e contemple uma única habilidade da Matriz de Referência. Para tanto, devem ser observadas a coerência e a coesão entre enunciado, suporte, comando e alternativas, de modo que haja uma articulação entre elas e se explicita uma única situação-problema e uma abordagem homogênea de conteúdo (INEP, 2010).

No CAEd, os itens seguem um fluxo de elaboração, passando por diversas etapas. A Figura 2 a seguir, demonstra as etapas de elaboração de um item na CIA:

Figura 2 – Etapas do processo de elaboração de itens



Fonte: CAEd (2008)

Assim como no CAEd, o guia de revisão e elaboração de itens do INEP também destaca a primeira etapa do processo de elaboração de itens a partir da seleção de uma habilidade da Matriz de Referência, em seguida apresentando uma lista com 15 tópicos a serem seguidos. O quadro 1 a seguir lista as etapas apresentadas no guia do INEP para elaboração de um item de Matemática:

Quadro 1 – Etapas para elaboração de um item de Matemática do guia do INEP

1. Selecione uma habilidade da Matriz de Referência.
2. Construa a situação-problema, atentando-se para a realidade cotidiana do público-alvo.
3. Dê preferência a fontes primárias, originais e sem adaptações, caso utilize textos-base referenciados, de acordo com as normas da ABNT.
4. Utilize, preferencialmente, textos que abordem temas atuais e sejam adequados ao público-alvo.
5. Evite a utilização de textos muito extensos, levando em consideração o tempo de leitura do item durante a realização do exame.
6. Elabore itens inéditos (não publicados, divulgados ou utilizados em sala de aula).
7. Elabore, sempre que possível, até três itens abordando o(s) mesmo(s) texto(s)-base, desde que os itens contemplem diferentes habilidades.
8. Evite abordagens de temas que suscitem polêmicas.
9. Evite utilizar ou redigir texto-base, enunciado e alternativas que possam induzir o participante do teste ao erro (“pegadinhas”).
10. Elabore o enunciado: utilize termos impessoais como: “considere-se”, “calcula-se”, “argumenta-se” etc.; não utilize termos como: “falso”, “exceto”, “incorreto”, “não”, “errado”; não utilize termos absolutos como: “sempre”, “nunca”, “todo”, “totalmente”, “absolutamente”, “completamente”, “somente” etc.; não utilize sentenças como: “Pode-se afirmar que”, “É correto afirmar que” etc.
11. Construa as alternativas: com paralelismo sintático e semântico, extensão equivalente e coerência com o enunciado; independentes umas das outras, de maneira que não sejam excludentes, negando informações do texto, nem semanticamente muito próximas; dispostas de maneira lógica (sequência narrativa, alfabética, crescente/decrescente etc.); evite repetição de palavras que aparecem no enunciado; evite alternativas demasiadamente longas; não use: “todas as anteriores”, “nenhuma das anteriores”; o gabarito deve estar exposto de forma clara, ser a única alternativa correta e não deve ser mais atrativo que os distratores; os distratores não devem ser absurdos em relação à situação-problema apresentada.
12. Pontue as alternativas e observe as regras para sua redação, de acordo com a área de conhecimento:

Figura 3 – Regras para redação de um item de Matemática

MATEMÁTICA E SUAS TECNOLOGIAS			
FORMATO	REGRA	EXEMPLO	EXCEÇÃO
ALTERNATIVAS QUE IMPLEMENTAM A SENTENÇA DO ENUNCIADO	Usar ponto final em cada alternativa e, no caso de palavra ou oração, iniciar alternativa com minúscula	Número inteiro, número decimal, número acompanhado de unidade de medida, potência, porcentagem, palavra, oração	Quando se tratar de fração, equação, sistema, fórmula, expressão algébrica, figura
ALTERNATIVAS QUE RESPONDEM A UMA INTERROGAÇÃO	Não usar ponto final nas alternativas	Número inteiro, número decimal, número acompanhado de unidade de medida, potência, fração, equação, sistema, porcentagem, figura	Quando se tratar de palavra e oração, as quais devem ser iniciadas por maiúscula

Fonte: INEP(2010)

13. Construa as justificativas para as alternativas com argumentação consistente que explique o erro ou o acerto, de maneira que não sejam tautológicas.
14. Indique o nível de dificuldade estimada do item (fácil, médio ou difícil) com base na sua experiência docente.
15. Considere o tempo médio de três minutos para resolução do item.

Fonte: INEP (2010)

Nesse processo inclui-se ainda mais uma etapa, que caracterizamos por etapa de validação do item. Pasquali (2017) aponta que a validade ocupa uma posição central na teoria da medida, constituindo-se um parâmetro fundamental e indispensável e que atualmente, é definida como a medida em que as evidências empíricas embasam as interpretações e os usos propostos para o teste. Os estudos sobre validade datam desde 1900, e atualmente está concentrada na validade de construto ou de traços latentes.

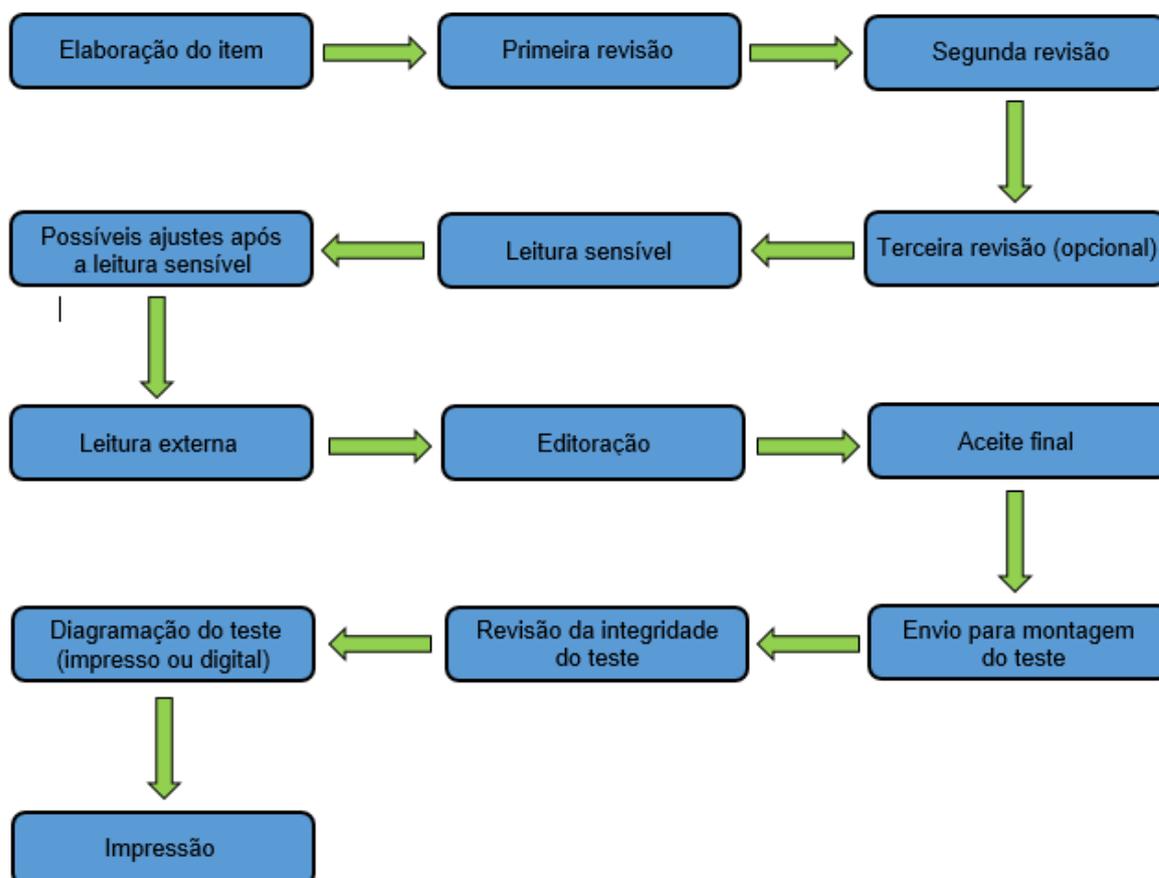
A validade de construto diz respeito ao grau em que o instrumento, neste caso o item, é capaz de distinguir as habilidades que se propõe a avaliar, ou seja, validar um construto é o mesmo que verificar se um item avalia adequadamente o que se deseja avaliar (TOFFOLI, et al., 2016 apud PASQUALI, 2007).

No INEP, o processo de validação é constituído por 6 etapas, que são:

1. Revisão técnico-pedagógica.
2. Aceite do Inep.
3. Revisão linguística.
4. Pré-testagem (Teoria Clássica dos Testes – TCT e Teoria de Resposta ao Item – TRI).
5. Leitura pedagógica com base nos parâmetros psicométricos.
6. Inserção no Banco Nacional de Itens (BNI). (INEP, 2010, p. 15).

Já o processo de validade de itens, defendido pela ETS, apresenta-se em 13 etapas, apresentados na figura 4 a seguir.

Figura 4 – Processo de validação de itens da ETS



Fonte: Elaborado pelo autor (2020)

Durante o processo de validação, após cada uma das revisões, o item pode ser aprovado ou receber pequenas modificações do revisor. Caso a mudança necessária seja mais complexa, o item retorna ao elaborador, podendo ainda ser rejeitado, caso não esteja de acordo com os padrões estabelecidos pela instituição avaliadora.

O revisor geralmente é um profissional com mais experiência no processo de elaboração, que possui autonomia para realizar pequenas modificações que julgar necessárias em um item recém elaborado.

Durante a troca de mensagens entre revisor e produtor, que, em alguns casos, é feita de maneira que ambos não sejam identificados, cada um deles deve, em um exercício de empatia, tentar se colocar no lugar do outro. O revisor deve ter o cuidado de registrar as inadequações de forma polida. Reciprocamente, o autor deve receber as críticas respeitosamente e, aceitando-as ou não, responder de maneira educada. A prática mostra que podem surgir conflitos entre produtores e revisores quando estas formalidades são abandonadas,

pois algumas pessoas podem interpretar discordâncias como se fossem ofensas pessoais, se entenderem que as críticas foram construídas com o intuito de minimizar a sua capacidade intelectual. (COSTA, 2014, p. 56).

Como trata-se de itens para avaliação educacional, todos eles devem ser pré-testados e calibrados, antes de serem enviados ao banco de itens, podendo também ser excluídos nessa etapa, mesmo após terem percorrido por todo esse longo processo de elaboração. A pré-testagem permite obter uma análise estatística detalhada do item, sob a perspectiva técnica e ainda pedagógica, demonstrando de fato a qualidade do instrumento.

Aproximadamente metade dos itens sobrevive após todas as atividades relacionadas à produção de itens, aos procedimentos de revisão e validação. Utilizando a TRI, os itens são classificados com três parâmetros, chamados de: discriminação, dificuldade e possibilidade de acerto ao acaso. Estando qualificado com estes parâmetros, finalmente o item passa a compor um banco, de onde os itens com diversos níveis de dificuldade podem ser selecionados para integrar um teste (COSTA, 2014).

Enfim, a elaboração de itens é um processo complexo, que demanda tempo e uma numerosa e experiente equipe, para que se possa ser cumprida a maior parte dessas etapas, o que certamente contribui para a composição de um banco de qualidade. Considerando experiência do pesquisador como analista, na elaboração e revisão de itens, o estudo constante dos documentos e leituras que norteiam o trabalho de elaboração torna-se fundamental, como uma maneira de se manter a formação continuada do profissional. Percebe-se ainda que existem pequenas diferenças no processo de elaboração, que se altera de acordo com o guia explorado, mas que de modo geral, é semelhante em alguns tópicos estratégicos. Entende-se que todo o trabalho de uma equipe de elaboração e revisão de um banco de itens deve ser orientada por documentos como os guias do CAEd, do INEP e o da APA, acordados antecipadamente e discutidos constantemente.

Haladyna (2004) comenta que a construção dos distratores é uma das partes mais difíceis para ser produzida em um item, e que certamente pode ser amenizada com os acordos e discussões dos guias de elaboração de cada instituição, de forma que toda uma equipe se mobilize, debatendo coletivamente em busca de novas possibilidades para realização do trabalho.

2.3.2 A Matriz de Referência e Suas Relações Com o Currículo e o Ensino

Após definir-se a população a ser avaliada, a etapa seguinte no processo de implementação de uma avaliação em larga escala é a construção de uma Matriz de Referência. Ela também se configura na primeira etapa no processo de elaboração de itens, funcionando como o documento norteador para a elaboração, definindo quais serão as habilidades e competências avaliadas em cada área de conhecimento.

Macedo (2005) argumenta que competência pode ser entendida de três modos diferentes, como condição prévia do sujeito, herdada ou adquirida, funcionando como capacidade de um organismo. Competência está definida no Inep (2010) como sendo a capacidade de mobilização de recursos cognitivos, socioafetivos ou psicomotores, estruturados em rede, com vistas a estabelecer relações com e entre objetos, situações, fenômenos e pessoas para resolver, encaminhar e enfrentar situações complexas. Já as habilidades decorrem das competências adquiridas ao longo do processo de aprendizagem, em que um indivíduo apresenta o conhecimento necessário para realizar determinada tarefa.

Essas habilidades se apresentam nas matrizes na forma de descritores.

As Matrizes de Referência são compostas por um conjunto de descritores, os quais contemplam dois pontos básicos do que se pretende avaliar: o conteúdo programático a ser avaliado em cada período de escolarização e o nível de operação mental necessário para a habilidade avaliada. Tais descritores são selecionados para compor a Matriz, considerando-se aquilo que pode ser avaliado por meio de itens de múltipla escolha. A Matriz de Referência para avaliação de Matemática tem como eixo a habilidade de resolver problemas contextualizados em consonância com os Parâmetros Curriculares Nacionais. Os temas selecionados – espaço e forma, grandezas e medidas, números e operações e tratamento da informação - representam conteúdos com base nos quais são elaborados descritores que expressam habilidades em Matemática. (CAEd, 2008, p. 14).

Elas são elaboradas a partir de estudos provindos dos currículos adotados pelas redes, em consonância com o conteúdo de livros didáticos e debates de educadores e demais especialistas em educação. Nesse processo são selecionadas habilidades que podem ser avaliadas, através de testes padronizados, de maneira que

fiquem de acordo com a etapa de escolaridade, quanto ao conteúdo e ao grau de dificuldade.

Neste ponto, cabe salientar que Moreira e Candau (2007) elencam que o currículo escolar é associado a distintas concepções, como simplesmente conteúdo a serem ensinados, experiências de aprendizagem escolar a serem vividas pelos alunos, planos pedagógicos elaborados por professores, escolas ou sistemas educacionais, objetivos a serem alcançados por meio do processo de ensino e processos de avaliação que terminam por influir nos conteúdos e nos procedimentos selecionados nos diferentes graus da escolarização. No entanto, o currículo, deve ser entendido como

[...] as experiências escolares que se desdobram em torno do conhecimento, em meio a relações sociais, e que contribuem para a construção das identidades de nossos/as estudantes. Currículo associa-se, assim, ao conjunto de esforços pedagógicos desenvolvidos com intenções educativas. (Moreira; Candau, 2007, p. 18).

Os autores ainda complementam dizendo que

[...] é por intermédio do currículo que as “coisas” acontecem na escola. No currículo se sistematizam nossos esforços pedagógicos. O currículo é, em outras palavras, o coração da escola, o espaço central em que todos atuamos, o que nos torna, nos diferentes níveis do processo educacional, responsáveis por sua elaboração. O papel do educador no processo curricular é, assim, fundamental. Ele é um dos grandes artífices, queira ou não, da construção dos currículos que se materializam nas escolas e nas salas de aula. (Moreira; Candau, 2007, p. 19).

Assim, existem muitas críticas e polêmicas sobre Matrizes de Referência das avaliações em larga escala, e uma delas é o fato de muitas redes de ensino adotarem os descritores de seus sistemas de avaliação como se fossem seus próprios currículos, prática essa que não é recomendável.

Bonamino e Souza (2012) descrevem que existem evidências nacionais e internacionais que, devido às avaliações em larga escala terem tomado dimensões que trazem maior responsabilização para a gestão escolar nos resultados, ocasionou ameaças ao cumprimento do currículo nas redes. A prática principal neste quesito é o que se chama de **ensinar para o teste**, que é simplesmente o fato de professores

concentrarem suas atenções com os estudantes, na aprendizagem dos tópicos mais explorados nessas avaliações. Nossa experiência em formação de docentes em diferentes redes de ensino apontam para uma afirmação de há uma prática em se explorar a Matriz de Referência e seus descritores como se fosse o currículo. Essa prática é adotada principalmente nas áreas de linguagens e Matemática, possibilitando a defasagem do cumprimento mínimo do currículo, que possui cobertura ampla sobre cada um dos conteúdos a serem ministrados, o que afeta diretamente no conhecimento escolar.

Desta forma, entende-se, assim como Moreira e Candau (2007) que, conhecimento escolar é uma construção específica da esfera educativa, não como uma simplificação de conhecimentos produzidos fora da escola. O conhecimento escolar possui características próprias, que o difere de outras formas de conhecimento.

É difícil discordar da alegação de que as avaliações em larga escala lidam com uma visão estreita de currículo escolar diante do que as escolas se propõem como objetivos para a formação de seus estudantes. Também é complexo o uso de testes padronizados para aferir objetivos escolares relacionados a aspectos não cognitivos. O problema decorre do fato de os currículos escolares possuírem múltiplos objetivos, ao passo que as medidas de resultados utilizadas pelas avaliações em larga escala tipicamente visam a objetivos cognitivos relacionados à leitura e à Matemática. Essa não é exatamente uma limitação das avaliações, mas demanda atenção para riscos relativos ao estreitamento do currículo, os quais podem acontecer quando há uma interpretação distorcida do significado pedagógico dos resultados da avaliação. (BONAMINO; SOUSA, 2012, p. 383).

Rabelo (2013, p. 178) fala que o vulto exagerado com que a avaliação ganhou nos últimos anos, deixa os professores em constante angústia, gerando repetição dos modelos praticados pelas avaliações educacionais externas em sala de aula, sem que, de fato, compreendam efetivamente as sutilezas que a norteiam e embasam todo o processo de avaliação.

No entanto, deve-se salientar que a base para criação de uma Matriz De Referência são os currículos. Um currículo sempre é fundamentado a partir de documentos como os Parâmetros Curriculares Nacionais (PCN's) e atualmente a Base Nacional Comum Curricular (BNCC). Portanto, temos uma hierarquia que deve ser respeitada ao manipularmos tais documentos que norteiam partes específicas do

processo educacional, não dando total soberania as Matrizes de Referência dos sistemas de avaliação. Elas devem ser entendidas como um “recorte” do currículo, e não como tal.

As matrizes de referência são elaboradas sem a pretensão de esgotar o repertório das habilidades necessárias ao pleno desenvolvimento do aluno. Portanto, não devem ser entendidas como habilidades únicas a serem trabalhadas em sala de aula. Sua finalidade é balizar a criação de itens dos testes, o que as distingue das propostas curriculares, estratégias de ensino e diretrizes pedagógicas. (SIMAVE, 2011, p. 17).

Os descritores são originados da associação entre os conteúdos existentes nos currículos e as operações mentais esperadas que seja desenvolvida pelo estudante, possibilitando a demonstração de suas habilidades.

É a matéria-prima para a elaboração dos itens. Como o próprio nome sugere, constituem uma sumária “descrição” das habilidades esperadas ao final de cada período escolar avaliado. Implicam, como fundamento, aspectos conceituais ou teóricos relacionados às diferentes áreas do conhecimento a serem avaliadas. A função dos descritores é, portanto, indicar as habilidades que serão objeto de avaliação, no conjunto de itens que compõem o teste. (SIMAVE, 2011, pág. 19).

Dessa forma, em um teste, cada descritor corresponde a apenas uma habilidade da Matriz De Referência, e subseqüentemente, cada item do teste deve se relacionar apenas a um único descritor.

No Brasil, as primeiras Matrizes de Referência para avaliação foram apresentadas pelo Sistema Nacional de Avaliação da Educação Básica (SAEB). Desde então, essas matrizes vêm sendo utilizadas como base para as avaliações realizadas pelos estados e municípios brasileiros que possuem seus próprios programas de avaliação em larga escala. Em Minas Gerais, as Matrizes de Referência para avaliação do SIMAVE/PROEB também foram elaboradas tendo por base as habilidades presentes nas matrizes do SAEB. (SIMAVE, 2011, p.17).

Na estrutura de uma Matriz De Referência, os descritores são organizados de acordo com suas áreas temáticas. Toma-se como exemplo a Matriz De Referência de Matemática do PROEB, que atualmente é composta por 86 descritores divididos em quatro áreas conforme pode ser visto na tabela 1 a seguir.

	D15 Utilizar o Teorema de Tales na resolução de problemas.			X		
	D16 Utilizar semelhança de triângulos na resolução de problemas.				X	
	D17 Interpretar geometricamente os coeficientes da equação de uma reta.				X	X
	D18 Utilizar o cálculo da distância entre dois pontos no plano na resolução de problemas.					X
	D19 Determinar a equação de uma reta a partir de dois pontos dados ou de um ponto e sua inclinação.					X
	D20 Reconhecer, dentre as equações do segundo grau com duas incógnitas, as que representam circunferências.					X
	D21 Utilizar o Teorema de Euler para determinar o número de faces, de vértices ou de arestas de poliedros convexos.					X
	D22 Utilizar a lei dos senos ou a lei dos cossenos na resolução de problemas.					X

5EF 7EF 9EF 1EM 3EM

GANDEZAS E MEDIDAS	D23 Executar a medição de grandezas por meio de medidas convencionais ou não.	X				
	D24 Utilizar conversão entre unidades de medida, na resolução de problemas.	X	X	X		
	D25 Utilizar conversão entre unidades de medidas de tempo na resolução de problemas.	X				
	D26 Reconhecer horas em relógios digitais e/ou analógicos.	X				
	D27 Corresponder o horário de início e de término com o intervalo de duração de um evento ou acontecimento.	X				
	D28 Utilizar o cálculo da medida do perímetro de uma figura bidimensional na resolução de problemas.	X	X	X	X	X
	D29 Utilizar o cálculo da medida da área de figuras bidimensionais na resolução de problemas.	X	X	X	X	X
	D30 Utilizar o cálculo da medida de área da superfície dos principais sólidos geométricos na resolução de problemas.					X
	D31 Corresponder cédulas e/ou moedas.	X				
	D32 Utilizar o cálculo da medida de volume/capacidade na resolução de problemas.			X	X	X

5EF 7EF 9EF 1EM 3EM

NÚMEROS E OPERAÇÕES / ÁLGEBRA E FUNÇÕES		5EF	7EF	9EF	1EM	3EM
	D33 Reconhecer características do sistema de numeração decimal.	X				
	D34 Corresponder números reais a pontos da reta numérica.	X	X	X	X	X
	D35 Executar cálculos com números naturais.	X				
	D36 Executar cálculos com números inteiros.		X			
	D37 Executar cálculos com números racionais.	X				
	D38 Utilizar números naturais, envolvendo diferentes significados da adição ou subtração, na resolução de problemas.	X				
	D39 Utilizar números naturais, envolvendo diferentes significados da multiplicação ou divisão, na resolução de problemas.	X				
	D40 Utilizar números naturais, envolvendo diferentes significados das operações, na resolução de problemas.		X			
	D41 Utilizar números inteiros, envolvendo diferentes significados das operações, na resolução de problemas.		X			
	D42 Corresponder diferentes representações de um número racional.	X	X	X	X	
	D43 Reconhecer fração como representação associada a diferentes significados.		X	X		
	D44 Utilizar números racionais, expressos na forma decimal, envolvendo diferentes significados da adição e/ou subtração, na resolução de problemas.	X				
	D45 Utilizar números racionais, expressos na forma decimal, envolvendo multiplicação ou divisão, na resolução de problemas.	X				
	D46 Utilizar números racionais, envolvendo diferentes significados das operações, na resolução de problemas.		X	X		
	D47 Identificar composições ou decomposições de números naturais.	X				
D48 Executar expressões numéricas com números racionais.		X				
D49 Executar expressões numéricas com números reais.			X	X		

D50 Utilizar porcentagem na resolução de problemas.	X	X	X	X	X
D51 Utilizar relações de proporcionalidade entre duas ou mais grandezas na resolução de problemas.		X	X	X	X
D52 Utilizar o princípio multiplicativo de contagem na resolução de problemas.		X			
D53 Utilizar mínimo múltiplo comum ou máximo divisor comum entre números naturais na resolução de problemas.		X			
D54 Identificar uma equação ou inequação polinomial do 1º grau que expressa um problema.		X	X		
D55 Identificar um sistema de equações do 1º grau que expressa um problema.			X		X
D56 Executar algoritmo de resolução de uma equação polinomial de 1º grau.		X			
D57 Executar algoritmo de resolução de uma inequação polinomial de 1º grau.		X			
D58 Executar algoritmo de resolução de um sistema linear de duas equações polinomiais de 1º grau, com duas incógnitas.			X		
D59 Utilizar equação ou inequação polinomial de 1º grau na resolução de problemas.		X	X	X	
D60 Determinar o conjunto solução de um sistema de equações lineares.					X
D61 Utilizar sistema de equações polinomiais de 1º grau com duas incógnitas na resolução de problemas.			X		
D62 Executar o cálculo do valor numérico de uma expressão algébrica.		X	X		
D63 Determinar o conjunto solução de uma equação do 2º grau.			X		
D64 Utilizar equação polinomial de 2º grau na resolução de problemas.			X		
D65 Utilizar equação ou inequação polinomial de 2º grau na resolução de problemas.				X	
D66 Identificar zeros, regiões de crescimento e de decréscimo ou máximos e mínimos de uma função a partir de seu gráfico.				X	X
D67 Corresponder uma função polinomial do 1º grau a seu gráfico.				X	X
D68 Corresponder uma função polinomial de 2º grau a seu gráfico.				X	X
D69 Utilizar função polinomial de 2º grau na resolução de problemas.				X	

	D70 Utilizar as coordenadas do vértice de uma função polinomial de 2º grau na resolução de problemas de máximo ou mínimo.				X	X
	D71 Corresponder um polinômio fatorado por meio de polinômios de 1º grau às suas raízes.				X	X
	D72 Corresponder uma função exponencial a seu gráfico.				X	X
	D73 Utilizar função exponencial na resolução de problemas.				X	X
	D74 Corresponder uma função logarítmica a seu gráfico.					X
	D75 Corresponder uma função trigonométrica a seu gráfico.					X
	D76 Utilizar juros simples ou juros composto na resolução de problemas.				X	
	D77 Utilizar propriedades de progressões aritméticas na resolução de problemas.					X
	D78 Utilizar propriedades de progressões geométricas na resolução de problemas.					X
	D79 Determinar o seno, o cosseno ou a tangente de arcos no círculo trigonométrico.					X
	D80 Utilizar métodos de contagem na resolução de problemas.					X
	D81 Utilizar noções de probabilidade na resolução de problemas.					X

5EF 7EF 9EF 1EM 3EM

TRATAMENTO DA INFORMAÇÃO	D82 Identificar informações a partir de dados dispostos em tabelas.	X				
	D83 Identificar dados apresentados por meio de gráficos.	X				
	D84 Utilizar dados apresentados em tabelas ou gráficos na resolução de problemas.		X	X	X	X
	D85 Corresponder listas e/ou tabelas simples a gráficos.		X	X	X	X
	D86 Utilizar medidas de tendência central na resolução de problemas.				X	

Fonte: <http://simave.educacao.mg.gov.br/proeb/matrizes-de-referencia/>

Em geral, os descritores das Matrizes de Referência de Matemática são organizados por blocos de cada área temática, sendo elas: Espaço e Forma, Grandezas e Medidas, Números e Operações/Álgebra e Funções, e Tratamento da Informação. Cada uma dessas áreas é contemplada com descritores, que abrangem de forma geral o currículo que guia o ensino na rede envolvida no sistema de avaliação.

Nessa pesquisa, serão tratados os itens eliminados em vários descritores pertencentes a área temática de Números e Operações/Álgebra e Funções (NO/AF) da matriz do PROEB que, como pudemos observar na figura 1, é contemplada com 49 descritores, codificados sequencialmente de D33 a D81, que circundam e aferem esta importante área da Matemática. Os descritores contidos nessa área temática são fundamentais na avaliação, principalmente das séries dos anos iniciais, uma vez que, abordam conhecimentos que envolvem a utilização de números no dia a dia na forma de códigos ou valores pelos estudantes.

Até a 5.ª série 6.º ano, para que o estudante aprenda os significados dos números, deve-se partir de contextos significativos envolvendo, por exemplo, o reconhecimento da existência de números naturais, números racionais e outros, e de suas representações e classificações, como primos, compostos, pares, ímpares, etc. (CAEd, 2009, p. 65).

No ensino fundamental, estão relacionadas ao tema, atividades como a resolução de situações-problema envolvendo contagem, medidas e significados de operações, a partir de estratégias individuais de cálculo; leitura e escrita de números naturais e racionais; ordenação de números naturais e racionais na forma decimal, interpretação do valor posicional de cada uma das ordens; resolução de problemas numéricos envolvendo as operações fundamentais.

Já no ensino médio, a área temática continua complementando habilidades exploradas no ensino fundamental, implementando o uso de equações e inequações de 1.º e 2.º grau na resolução de problemas; correspondências de funções a seus gráficos; situações com expressões numéricas; polinômios; trigonometria; porcentagem e Matemática financeira.

Na seção seguinte vamos apresentar as evidências que fundamentam este trabalho, falando de maneira mais detalhada, a relação do pesquisador com o caso de gestão e os itens que serão analisados. Vamos ainda nesta seção, discorrer sobre

alguns aspectos dos itens que serão analisados neste estudo, além do percurso realizado para escolha desses itens.

2.4 APRESENTANDO AS EVIDÊNCIAS DO CASO

Atuando como auxiliar e analista de instrumentos de avaliação, a principal atribuição do pesquisador é centrada na produção de instrumentos para composição dos testes de proficiência realizados pelo Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF). Diariamente o pesquisador atua na elaboração, correção e revisão de itens que são produzidos na equipe de Matemática da Coordenação de Instrumentos de Avaliação (CIA) e esporadicamente se reúnem para momentos formativos com as equipes de pesquisa e interações com integrantes da Coordenação de Medidas e Educacionais - (CME).

Em uma dessas interações com a equipe de medidas, pôde-se aproximar dos dados estatísticos gerados após a aplicação dos testes realizados pelo CAEd. Sendo assim, passou-se a ter contato com os dados gerados a partir da utilização da TRI, observando os percentuais de acerto, de procura pelas alternativas, bisseriais, ou seja, o comportamento de forma geral dos itens nos testes.

Um dos detalhes que se mostrou mais atrativo foi a eliminação de itens, algo que, entendeu-se, possuir extrema importância para as equipes da CIA, e principalmente, para a formação dos elaboradores. Até então já se atuava há dois anos e meio no CAEd, e nunca se pôde verificar o comportamento dos autorais em um teste ou pré-teste, sem nenhum *feedback*, a fim de se orientar para verificar se estava-se trilhando os caminhos corretos na elaboração desses instrumentos.

Avançando na discussão, percebeu-se que a eliminação de itens em um teste pode ser causada por três diferentes fatores. O primeiro fator apresentado ocorreu devido a uma das alternativas erradas, que chamamos de distrator, ser mais procurada pelos estudantes ao invés do gabarito, ocasionando a eliminação por motivo de **bisserial**. Os outros dois fatores que se destacaram e despertaram tal curiosidade são as eliminações causadas por motivos estatísticos, que são os **ajustes** realizados e o **Differential Item Functioning** - (DIF), que traduzindo, significa comportamento diferencial do item.

Nesse momento formativo, foram apresentados os dados estatísticos de itens recém elaborados no banco em um pré-teste realizado na rede pública de ensino no

Estado de Minas Gerias nos anos finais (5EF, 9EF e 3EM), aguçando ainda mais a curiosidade sobre o desempenho de itens no estado citado. Abordou-se que cerca de 10% dos itens que compõem os blocos dos testes de proficiência em Matemática do CAEd são eliminados após serem submetidos a análise realizada através da TRI, o que levou a reflexão sobre atuação do pesquisador na elaboração e o potencial de informações que esses instrumentos eliminados podem trazer.

Com intenção de aprofundar os estudos nessa vertente, debruçou-se sobre um dos projetos avaliados pelo CAEd em Minas Gerais, o Programa de Avaliação da Rede Pública de Educação Básica (PROEB), que compõe o SIMAVE.

A escolha por esse projeto se justifica pelo fato de ser um dos projetos com maior amplitude e consolidação de atuação do CAEd, e ainda, o mais longo em toda sua trajetória, sem contar que se trata do projeto ao qual o pesquisador tem mais proximidade levando em consideração sua atuação como professor de Matemática nas redes municipal e estadual no município de Juiz de Fora/MG, possibilitando ao mesmo uma visão sobre os itens em uma via de mão dupla, ou seja, como elaborador e professor.

Nessa investigação trabalhou-se com um recorte temporal de 5 anos, composto pelas edições de 2014 a 2018 do PROEB, em razão do período mais próximo de sua atuação tanto no CAEd quanto nas redes públicas de ensino, pois nesse período, no banco de itens poderia haver ainda um bom quantitativo de itens de sua autoria possibilitando, principalmente, uma melhor reflexão sobre a atuação do mesmo.

Buscou-se por todos os itens utilizados nesse período e se identificou que em todos os testes do PROEB, de 2014 a 2018, foram mobilizados 1571 itens de Matemática. Desse total, retirando-se todas as duplicidades de itens existentes, pois, nesse quantitativo, existiam itens que foram utilizados em mais de uma edição, ou até mesmo em duas etapas diferentes em um único ano, fez com que esse total se reduzisse para 797 itens. Na busca por itens que apresentaram pelo menos um episódio de eliminação ao longo do período observado, identificou-se que dos 1571 itens, houve 98 eliminações, que retirando-se novamente as duplicidades, contabilizou-se em um total de 74 itens mobilizados eliminados, sendo deste total, 45 elaborados no CAEd e os 29 restantes caracterizados como **itens âncora**, que são fornecidos pelo INEP para calibragem da escala de proficiência dos testes, pois no CAEd, a escala é equivalente à escala SAEB.

Os itens denominados âncoras auxiliam ordenando os desempenhos dos alunos de acordo com a escala de proficiência, do nível mais baixo ao mais alto em um continuum (KLEIN, 2015).

Interpretar a escala significa escolher alguns pontos ou níveis da escala e descrever os conhecimentos e habilidades que os alunos demonstraram possuir quando situados em torno desses pontos. A metodologia para interpretação das escalas inclui dois procedimentos principais: identificação de itens âncoras e a apresentação desses itens a um painel de especialistas. (KLEIN, 2015, p. 190).

Para Klein (2009), considerando-se um intervalo de 25 pontos na escala de proficiência, um item será âncora se:

- O número de alunos no nível que respondeu ao item for maior que 50.
- O percentual de acerto do item nos níveis anteriores é menor que 65%.
- O percentual de acerto do item no nível considerado e nos níveis acima é maior que 65%.
- O ajuste da curva é bom. (KLEIN, 2015, p. 192).

Assim, todo item considerado como âncora possui bom desempenho dentro do teste, porém, nos estudos realizados diagnosticaram-se itens com essa denominação, fornecidos pelo INEP, eliminados na população aferida em Minas Gerais.

Neste caso, esses itens não puderam ser utilizados para esta pesquisa dentro do âmbito do Mestrado Profissional, uma vez que são itens que foram utilizados em avaliações do SAEB, e que o CAEd deve preservar o sigilo dos mesmos.

Dessa forma, inicialmente se estabeleceu o critério de realizar um estudo de pelo menos três itens que fossem elaborados no CAEd, dos três descritores que apresentassem maior número de itens do total 98 itens eliminados encontrados, revelando os seguintes dados da tabela 2 a seguir.

Tabela 2 – Quantitativo de itens eliminados por descritor – (PROEB 2014 a 2018)

D01	3		D15	0		D29	4		D43	0		D57	2		D71	2
D02	7		D16	0		D30	0		D44	2		D58	0		D72	4
D03	0		D17	3		D31	2		D45	0		D59	2		D73	0
D04	0		D18	2		D32	0		D46	3		D60	1		D74	1
D05	2		D19	0		D33	4		D47	1		D61	0		D75	1
D06	2		D20	1		D34	8		D48	1		D62	1		D76	1
D07	0		D21	0		D35	1		D49	0		D63	0		D77	3
D08	0		D22	0		D36	0		D50	3		D64	2		D78	0
D09	0		D23	0		D37	0		D51	3		D65	0		D79	0
D10	1		D24	1		D38	0		D52	0		D66	0		D80	1
D11	0		D25	0		D39	0		D53	0		D67	1		D81	6
D12	0		D26	0		D40	1		D54	1		D68	2		D82	0
D13	1		D27	0		D41	0		D55	0		D69	0		D83	0
D14	0		D28	2		D42	2		D56	0		D70	3		D84	1
															D85	3
															D86	0

Fonte: Elaborado pelo autor. (2020)

Feito este levantamento, constatou-se que os descritores D02, D34 e D81, foram os que mais apresentaram itens eliminados no período investigado. No entanto, tal critério não foi bem sucedido, ao se identificar tais itens eliminados nos referidos descritores, levantando as seguintes observações descritas no quadro 2 a seguir.

Quadro 2 – Observações levantadas sobre os descritores D02, D34 e D81 - (PROEB 2014 a 2018)

Descritor	Observações
D02	Das 7 ocorrências identificadas, foram utilizados apenas 5 itens. Desses, 3 são distintos e foram elaborados no CAEd. Os outros 2 são itens âncora, sendo um utilizado no 7EF em 2017 e o outro utilizado nas edições de 2015, 2017 e 2018 no 3EM.

D34	<p>Das 8 ocorrências identificadas, foram utilizados apenas 4 itens, todos de origem do INEP, servindo de parâmetro para escala, da seguinte forma:</p> <p>Item 1: Utilizado 3 vezes, sendo a primeira no 7EF em 2015, a segunda no 5EF em 2016 e a terceira no 7EF em 2017.</p> <p>Item 2: Utilizado 3 vezes, sendo a primeira no 7EF em 2015, a segunda no 7EF em 2017 e a terceira no 5EF em 2018.</p> <p>Item 3: Utilizado apenas uma vez no 1EM na edição do ano de 2015.</p> <p>Item 4: Utilizado apenas uma vez no 1EM na edição do ano de 2015.</p> <p><i>Obs.: Optou-se aqui por identificar os itens por Item 1, Item 2, Item 3 e Item 4 para que não seja revelado a codificação desses itens.</i></p>
D81	<p>Das 6 ocorrências identificadas, foram utilizados apenas dois itens de origem do INEP, sendo o primeiro utilizado no 3EM das edições de 2015, 2016 e 2017 e o segundo também utilizado no 3EM nas edições de 2015, 2016 e 2018.</p>

Fonte: Elaborado pelo autor. (2020)

Com esse levantamento, pôde-se observar que o quantitativo de itens não foi favorável a pesquisa, devido ao fato de apenas 3 itens serem de procedência do CAEd, os quais se pretende investigar. No entanto, nesse primeiro momento, é possível verificar que alguns desses itens, que foram fornecidos pelo INEP, vem sendo constantemente utilizados nos testes do PROEB, apresentando recorrentes eliminações e até então, nada havia sido levantado sobre a validade dos mesmos. Neste ponto entende-se que caberia ao CAEd questionar o INEP sobre a validade desses instrumentos, e ainda, que as equipes de montagem e elaboração dos testes na CIA não os reutilizassem. Certamente esses itens foram elaborados especificamente para o público do SAEB pelos elaboradores do INEP, ou seja, não são itens específicos para os projetos avaliados no CAEd. Salienta-se ainda a possibilidade de aprendizagem e treinamento das equipes da CIA envolvidas com esse tipo de levantamento, possibilitando o debate sobre os desafios e possibilidades desses instrumentos nos testes do CAEd.

Em razão do baixo quantitativo de itens, decidiu-se então adotar um segundo critério de escolha dos itens para serem estudados, utilizando como parâmetro, as áreas temáticas da Matriz de Referência do PROEB, filtrando os itens de acordo com

o pertencimento a cada uma delas. Como já visto na seção anterior, a Matriz de Referência do PROEB é composta por descritores que englobam quatro grandes áreas temáticas da Matemática, classificadas como **Espaço e Forma, Grandezas e Medidas, Números e Operações, Álgebra e Funções e Tratamento da Informação**.

Através desse novo critério adotado, foi possível separar e classificar todos os 74 itens eliminados de acordo com a área temática que o mesmo contemplou da Matriz de Referência, favorecendo quantitativamente a pesquisa. A tabela 3, apresenta o quantitativo de itens, sem duplicidades, que apresentaram eliminações ao longo das edições apontadas nessa dissertação, de acordo com a área de cobertura da matriz de referência do PROEB, e a instituição responsável por sua elaboração.

Tabela 3: Quantitativo de itens eliminados por área avaliada

Espaço e Forma	Grandezas e Medidas	Números e Operações / Álgebra e Funções	Tratamento da Informação
Origem de elaboração do Item			
CAEd: 13	CAEd:4	CAEd:28	CAEd: 0
INEP/SAEB: 4	INEP/SAEB :3	INEP/SAEB:20	SAEB: 2
Total: 17	Total:7	Total:48	Total: 2

Fonte: Elaborado pelo autor. (2020)

De acordo com o segundo critério de busca, a área de cobertura da matriz que apresentou um quantitativo favorável de itens para realização da pesquisa foi Números e Operações/Álgebra e Funções (NO/AF). Foram identificados 28 itens elaborados pela equipe de Matemática na CIA, conforme apresentado na tabela 4, com seus respectivos motivos de eliminação e dados estatísticos.

Tabela 4: Itens eliminados da disciplina de Matemática da área de Números e Operações, Álgebra e Funções, no PROEB entre os anos de 2014 e 2018

PROEB Ano	Etapa	Item	Desc.	Gab.	Acerto	Biss.	Percentuais por Resposta - TCT					Branco / Nulo	Bisseriais por Alternativa					Motivo de Eliminação	Habilidade
							A	B	C	D	E		A	B	C	D	E		
2015	7EF	M070108G5	D59	D	0,246	0,134	0,195	0,238	0,315	0,246		0,006	-0,100	-0,080	0,041	0,134		BISSERIAL	Utilizar equação ou inequação polinomial de 1º grau na resolução de problema.
2015	7EF	M070107G5	D57	B	0,289	0,164	0,247	0,289	0,329	0,117		0,019	-0,046	0,164	-0,071	-0,043		BISSERIAL	Executar algoritmo de resolução de uma inequação polinomial de 1º grau.
2015	1EM	M100060E4	D76	C	0,190	0,147	0,234	0,343	0,190	0,160	0,067	0,005	-0,054	0,218	0,147	-0,270	-0,263	BISSERIAL	Utilizar juros simples ou juros composto na resolução de problemas.
2015	1EM	M1D26I0117	D71	E	0,116	0,120	0,266	0,158	0,172	0,283	0,116	0,005	-0,074	-0,054	-0,113	0,135	0,120	BISSERIAL	Corresponder um polinômio fatorado por meio de polinômios de 1º grau às suas raízes.
2015	1EM	M100063E4	D71	C	0,139	0,101	0,073	0,088	0,139	0,620	0,076	0,005	-0,247	-0,166	0,101	0,183	-0,192	BISSERIAL	Corresponder um polinômio fatorado por meio de polinômios de 1º grau às suas raízes.
2015	3EM	M110104E4	D77	D	0,753	0,632	0,032	0,051	0,097	0,753	0,065	0,002	-0,439	-0,454	-0,429	0,632	-0,356	AJUSTE	Utilizar propriedades de progressões aritméticas na resolução de problemas.
2016	5EF	M050057H6	D33	A	0,674	0,592	0,674	0,102	0,086	0,131		0,007	0,592	-0,407	-0,387	-0,337		DIF	Reconhecer características do sistema de numeração decimal.
	9EF				0,505	0,487	0,505	0,188	0,147	0,155		0,005	0,487	-0,250	-0,261	-0,261		DIF	Reconhecer características do sistema de numeração decimal.
2016	9EF	M050060H6	D47	A	0,837	0,686	0,837	0,054	0,041	0,064		0,005	0,686	-0,536	-0,553	-0,463		DIF	Identificar composições ou decomposições de números naturais.
	9EF	M090046H6	D64	C	0,142	0,027	0,167	0,406	0,142	0,279		0,007	-0,006	0,194	0,027	-0,225		BISSERIAL	Utilizar equação polinomial de 2º grau na resolução de problema.
2016	9EF	M090047H6	D64	B	0,137	0,041	0,538	0,137	0,206	0,113		0,005	0,305	0,041	-0,292	-0,226		BISSERIAL	Utilizar equação polinomial de 2º grau na resolução de problema.

2016	3EM	M120189H6	D75	E	0,138	0,191	0,213	0,182	0,208	0,257	0,138	0,003	0,100	-0,088	-0,068	-0,084	0,191	AJUSTE	Corresponder uma função trigonométrica a seu gráfico.
2016	3EM	M120138H6	D80	C	0,181	0,188	0,220	0,370	0,181	0,111	0,114	0,004	-0,266	-0,006	0,188	-0,027	0,200	AJUSTE	Utilizar métodos de contagem na resolução de problemas.
2016	3EM	M100093H6	D67	B	0,090	0,115	0,204	0,090	0,183	0,363	0,157	0,003	-0,283	0,115	-0,059	0,155	0,088	BISSERIAL	Corresponder uma função polinomial de 1º grau a seu gráfico.
2016	3EM	M120085A8	D74	C	0,208	0,191	0,071	0,384	0,208	0,191	0,142	0,003	-0,100	-0,048	0,191	-0,058	0	AJUSTE	Corresponder uma função logarítmica a seu gráfico.
2017	7EF	M090061H6	D59	C	0,352	0,143	0,072	0,237	0,352	0,331		0,008	-0,214	-0,072	0,143	0,011		BISSERIAL	Utilizar equação ou inequação polinomial de 1º grau na resolução de problema.
2017	7EF	M080057H6	D46	A	0,221	0,130	0,221	0,505	0,161	0,108		0,005	0,130	0,005	-0,105	-0,054		BISSERIAL	Utilizar números racionais, envolvendo diferentes significados das operações, na resolução de problemas.
2017	7EF	M090183H6	D42	B	0,165	0,142	0,058	0,165	0,337	0,434		0,006	-0,247	0,142	0,036	-0,037		BISSERIAL	Corresponder diferentes representações de um número racional.
2017	7EF	M081180E4	D57	B	0,305	0,155	0,196	0,305	0,327	0,157		0,015	-0,085	0,155	-0,033	-0,053		BISSERIAL	Executar algoritmo de resolução de uma inequação polinomial de 1º grau.
2017	1EM	M120389H6	D72	D	0,173	0,198	0,134	0,189	0,237	0,173	0,263	0,005	-0,082	-0,016	-0,139	0,198	0,056	AJUSTE	Corresponder uma função exponencial a seu gráfico.
	3EM				0,169	0,177	0,087	0,184	0,216	0,169	0,341	0,003	-0,094	-0,076	-0,146	0,177	0,095	AJUSTE	Corresponder uma função exponencial a seu gráfico.
2017	1EM	M120390H6	D72	A	0,129	0,178	0,129	0,419	0,114	0,182	0,150	0,007	0,178	0,150	-0,187	-0,137	-0,083	AJUSTE	Corresponder uma função exponencial a seu gráfico.
	3EM				0,122	0,160	0,122	0,509	0,083	0,142	0,141	0,002	0,160	0,106	-0,170	-0,148	-0,063	AJUSTE	Corresponder uma função exponencial a seu gráfico.
2017	1EM	M110233H6	D70	B	0,224	0,176	0,118	0,224	0,234	0,219	0,198	0,006	-0,192	0,176	-0,044	0,001	0,008	BISSERIAL	Utilizar as coordenadas do vértice de uma função polinomial de 2º grau na resolução de problemas de máximo ou mínimo.
	3EM				0,240	0,203	0,099	0,240	0,250	0,248	0,159	0,005	-0,082	0,203	-0,096	-0,008	-0,059	AJUSTE	Utilizar as coordenadas do vértice de uma função polinomial de 2º grau na resolução de problemas de máximo ou mínimo.
2017	1EM	M090235H6	D42	C	0,378	0,555	0,059	0,267	0,378	0,282		0,013	-0,197	-0,288	0,555	-0,247		DIF	Corresponder diferentes representações de um número racional.
2017	3EM	M120911E4	D68	D	0,447	0,364	0,205	0,110	0,126	0,447	0,109	0,004	-0,213	-0,228	-0,186	0,364	0,001	AJUSTE	Corresponder uma função polinomial de 2º grau a seu gráfico.

2017	3EM	M120909E4	D68	B	0,568	0,434	0,143	0,568	0,110	0,121	0,054	0,004	-0,138	0,434	-0,269	-0,293	-0,253	AJUSTE	Corresponder uma função polinomial de 2º grau a seu gráfico.
2017	3EM	M100208H6	D77	E	0,144	0,124	0,121	0,145	0,198	0,388	0,144	0,003	-0,267	-0,205	-0,178	0,323	0,124	BISSERIAL	Utilizar propriedades de progressões aritméticas na resolução de problemas.
2018	9EF	M040046H6	D50	C	0,346	0,560	0,043	0,526	0,346	0,081		0,004	-0,420	-0,328	0,560	-0,226		DIF	Utilizar porcentagem na resolução de problema.
2018	9EF	M090560H6	D46	A	0,210	0,037	0,210	0,298	0,257	0,230		0,004	0,037	0,034	-0,105	0,051		BISSERIAL	Utilizar números racionais, envolvendo diferentes significados das operações, na resolução de problemas.
2018	3EM	M090058H6	D51	A	0,657	0,598	0,657	0,174	0,083	0,078		0,008	0,598	-0,393	-0,401	-0,346		AJUSTE	Utilizar relações de proporcionalidade entre duas ou mais grandezas na resolução de problema.

Fonte: Elaborado pelo autor. (2020)

Para crivar ainda mais os estudos realizados, uma vez que obteve-se um bom quantitativo de itens, conforme apresentado na tabela 4 anteriormente, optou-se por seleccionar, dentre os 28 itens localizados, os três itens que apresentassem o menor percentual de acerto de cada um dos três motivos de eliminação apresentados, a fim de analisar concomitantemente a estrutura pedagógica desses itens, com seus respectivos dados estatísticos. Ou seja, interpretou-se pedagógica e estatisticamente seus desempenhos, associando com as práticas diárias de elaboração na CIA, no sentido de detectar possíveis rumos adotados no processo de elaboração de instrumentos de avaliação.

Dessa forma, os itens seleccionados, de acordo com o menor percentual de acerto, em cada um dos motivos de eliminação, estão listados na tabela 5 a seguir, com seus respectivos dados.

Tabela 5: Itens eliminados na área de Números e Operações, Álgebra e Funções, no PROEB entre os anos de 2014 e 2018 com menor percentual de acerto.

PROEB Ano	Etapa	Item	Desc.	Gab.	Acerto	Biss.	Percentuais por Resposta - TCT					Branco / Nulo	Bisseriais por Alternativa					Motivo de Eliminação	Habilidade
							A	B	C	D	E		A	B	C	D	E		
2017	1EM	M120389H6	D72	D	0,173	0,198	0,134	0,189	0,237	0,173	0,263	0,005	-0,082	-0,016	-0,139	0,198	0,056	AJUSTE	Corresponder uma função exponencial a seu gráfico.
	3EM				0,169	0,177	0,087	0,184	0,216	0,169	0,341	0,003	-0,094	-0,076	-0,146	0,177	0,095	AJUSTE	Corresponder uma função exponencial a seu gráfico.
2017	1EM	M120390H6	D72	A	0,129	0,178	0,129	0,419	0,114	0,182	0,150	0,007	0,178	0,150	-0,187	-0,137	-0,083	AJUSTE	Corresponder uma função exponencial a seu gráfico.
	3EM				0,122	0,160	0,122	0,509	0,083	0,142	0,141	0,002	0,160	0,106	-0,170	-0,148	-0,063	AJUSTE	Corresponder uma função exponencial a seu gráfico.
2016	3EM	M120189H6	D75	E	0,138	0,191	0,213	0,182	0,208	0,257	0,138	0,003	0,100	-0,088	-0,068	-0,084	0,191	AJUSTE	Corresponder uma função trigonométrica a seu gráfico.
2016	9EF	M090047H6	D64	B	0,137	0,041	0,538	0,137	0,206	0,113		0,005	0,305	0,041	-0,292	-0,226		BISSERIAL	Utilizar equação polinomial de 2º grau na resolução de problema.
2016	3EM	M100093H6	D67	B	0,090	0,115	0,204	0,090	0,183	0,363	0,157	0,003	-0,283	0,115	-0,059	0,155	0,088	BISSERIAL	Corresponder uma função polinomial do 1º grau a seu gráfico.
2015	1EM	M1D26I0117	D71	E	0,116	0,120	0,266	0,158	0,172	0,283	0,116	0,005	-0,074	-0,054	-0,113	0,135	0,120	BISSERIAL	Corresponder um polinômio fatorado por meio de polinômios de 1º grau às suas raízes.
2018	9EF	M040046H6	D50	C	0,346	0,560	0,043	0,526	0,346	0,081		0,004	-0,420	-0,328	0,560	-0,226		DIF	Utilizar porcentagem na resolução de problema.
2017	1EM	M090235H6	D42	C	0,378	0,555	0,059	0,267	0,378	0,282		0,013	-0,197	-0,288	0,555	-0,247		DIF	Corresponder diferentes representações de um número racional.
2016	5EF	M050057H6	D33	A	0,674	0,592	0,674	0,102	0,086	0,131		0,007	0,592	-0,407	-0,387	-0,337		DIF	Reconhecer características do sistema de numeração decimal.
	9EF				0,505	0,487	0,505	0,188	0,147	0,155		0,005	0,487	-0,250	-0,261	-0,261		DIF	Reconhecer características do sistema de numeração decimal.

Fonte: Elaborado pelo autor. (2020)

Nesse sentido, esta dissertação, que tem por objetivo analisar esses 9 itens da área de Números e Operações, Álgebra e Funções, verificá-los individualmente analisando os motivos de eliminação de cada um, a fim de observar possíveis vieses e colocá-los em discussão considerando os referenciais teóricos e as experiências da equipe de elaboração. Nesta análise, serão levados em consideração os aspectos estatísticos, a partir de uma análise psicométrica concomitantemente com os aspectos pedagógicos de cada item, na tentativa de buscar novos métodos para aprimorar as práticas no processo de elaboração, possibilitando elevar a qualidade do banco de itens do CAEd.

Outro aspecto que cabe ser levantado por esta pesquisa é a possibilidade de implantação de um processo de análise a partir de itens eliminados, na tentativa de colaborar tanto para a formação da equipe de elaboração, quanto para interpretação da aprendizagem das redes contratantes. O intuito é propor novos modelos de estudos e pesquisas sobre a qualidade do item no CAEd, de maneira que permita maior interação entre as equipes internas envolvidas no processo avaliativo, auxiliando desde a formação do elaborador até a divulgação dos resultados, oportunizando análises pedagógicas muito mais complexas e detalhadas sobre a aprendizagem dos alunos.

A TRI permite a compreensão de quais competências e habilidades presentes e avaliadas na Matriz de Referência que o estudante domina. Enquanto profissionais atuantes no campo da avaliação educacional, deve-se refletir também sobre tudo aquilo que estudantes demonstram não dominar ou aprender, de acordo com os testes.

3 A PSICOMETRIA NOS ITENS (A ESTATÍSTICA NOS ITENS) E ANÁLISE DOS ITENS ELIMINADOS

Neste capítulo se colocará em debate alguns autores especialistas em métodos de correção de testes para avaliações em larga escala, além de apresentarmos as imagens e discussões ao nível pedagógico e psicométrico dos itens selecionados para esta pesquisa. Portanto, a TCT e a TRI serão abordadas de maneira detalhada, trazendo consigo os três motivos de eliminação de itens analisados nesse trabalho. Buscou-se entender pedagógica e estatisticamente o que significa um item eliminado em um teste, e ainda, o que eles podem contribuir para formação de uma equipe de elaboração. Para isso, buscou-se por autores que entendem do assunto, e relatos de colaboradores de instituições avaliadoras que atuam com avaliação educacional em larga escala.

3.1 MÉTODOS DE CORREÇÃO DE TESTES

Os resultados obtidos em grande parte das avaliações são expressos apenas por seus escores brutos ou padronizados, fazendo com que a Teoria Clássica dos Testes (TCT), ainda seja tradicionalmente utilizada em alguns processos de avaliação e seleção de indivíduos no Brasil (ANDRADE et al, 2000).

De acordo com Sartes e Souza-Formigone (2013), a maioria dos testes que se utilizam da TCT são longos e redundantes, gerando medidas influenciadas pelas características da população avaliada, sendo essa uma de suas maiores limitações na sua usabilidade. No Brasil, essa prática passou a ser repensada a partir de 1995, ano em que se passou a utilizar a Teoria da Resposta ao Item (TRI) nas avaliações do SAEB, possibilitando superar algumas dessas limitações. Atualmente, os estudos sobre a TRI vem tomando grande proporção na área da psicologia e no cenário educacional brasileiro, principalmente por aqueles que atuam diretamente no campo da avaliação educacional em larga escala. A TRI surgiu a partir da necessidade de se romper as limitações em se avaliar somente através de percentuais de acertos ou escores dos testes, e ainda, da dificuldade de se comparar diferentes resultados em testes e em situações diversas (KLEIN, 2003).

Fontanive et al. (2007) demonstram preocupação com a divulgação de resultados das avaliações dos sistemas escolares aos diferentes públicos. Devido a

termos e ao pertencimento de áreas muito técnicas do ramo da psicologia e da Matemática, a compreensão de tais métodos de correção se tornam abstrusos para grande parte da comunidade escolar, tornando necessárias formações e trocas de experiências, a fim de que tais resultados e métodos se tornem mais tangíveis a todos. Sustentado por essas premissas, buscou-se apresentar nesse capítulo, conceitos e exemplos sobre a TCT e a TRI de maneira bem simples, tornando sua percepção mais trivial e com um breve comparativo entre esses dois métodos. Os parâmetros da TRI e maneiras de interpretar seus resultados em um item também serão abordados nessa seção, de maneira que o leitor compreenda como a TRI apresenta o desempenho de um instrumento de avaliação. A interpretação de resultados por meio das curvas características também será tratada aqui. Ainda neste capítulo, será abordada a metodologia e a eliminação de itens, de forma teórica e prática, utilizando os itens eliminados escolhidos segundo os critérios adotados na seção anterior deste trabalho, por meio de uma dinâmica diferenciada, em que é apresentado todo o histórico de utilização dos itens no PROEB, seus dados estatísticos e imagem. Nessa seção, será dado destaque as experiências e relatos de colaboradores da equipe de Matemática da CIA do CAEd sobre os itens propostos para análise.

3.1.1 Teoria Clássica de Testes – TCT

A TCT é um método de correção de testes desenvolvido por volta de 1930, a partir dos estudos do psicólogo inglês Charles Edward Spearman, que elaborou também inúmeros trabalhos na área da estatística, marcando de certa forma, o início da psicometria. Posteriormente este método foi sendo desenvolvido, principalmente com as contribuições do norte-americano Louis Leon Thurstone, também psicólogo que utilizou das concepções de Spearman para desenvolver na TCT a chamada análise fatorial. A análise fatorial foi uma importante contribuição para a TCT, uma vez que ela tem como finalidade, analisar de forma multivariada a dimensionalidade dos instrumentos de medida, abrangendo um conjunto de técnicas estatísticas para descrever as variáveis iniciais em um teste, a partir de variáveis hipotéticas obtidas como função das variáveis originais (SARTES e SOUZA-FORMIGONI, 2013).

A TCT leva em consideração o escore total do teste, ou seja, a pontuação total que um indivíduo pode obter nesse teste. Nesse tipo de método o avaliado deve realizar tarefas, ou no caso da avaliação educacional, resolver itens, de forma que a

sua pontuação obtida será o resultado da soma de todas as tarefas realizadas com sucesso, sendo esses pontos brutos ou ponderados.

Klein (2013, p. 37) aponta que a TCT possui um modelo para se chegar ao resultado observado de um indivíduo no teste, que leva em consideração o escore verdadeiro e o erro de medida, sendo que este último não depende da habilidade do aluno, através da relação:

$$X = T + E,$$

em que **X** equivale à nota do estudante no teste (observado); **T** é o escore real (total) ou verdadeiro e igual ao erro de medida, que se trata de uma variável aleatória.

Sartes e Souza-Formigoni (2013, p. 243) relatam que a análise de itens na TCT é um processo que visa selecionar itens de boa qualidade de um conjunto, muito das vezes maior que o necessário, a fim de avaliar duas de suas características, que são: **a dificuldade e a discriminação**. A dificuldade de um item, se tratando do tipo politômico na TCT, é a proporção de respostas a uma categoria de escolha, ou seja, é o quociente entre o número de estudantes que acertaram pelo número total de respondentes, ou ainda, pela média das respostas de todos os indivíduos. O índice de dificuldade pode variar entre 0 (zero) e 1 (um), em que o menor significa que ninguém respondeu o item corretamente, e no maior, todos o acertam. Já a discriminação em um item deve, em qualquer método de correção, diferenciar avaliados que possuem um bom nível de conhecimento daqueles que ainda necessitam de suporte pedagógico. Na TCT, a discriminação consiste em diferenciar o grupo que obteve baixa ou alta pontuação no teste, podendo ser divididos em dois grupos, sendo o primeiro, o grupo dos 27% que obtiveram as maiores pontuações, e o segundo, dos 27% que obtiveram as menores pontuações no teste. Na TCT, um item se torna mais discriminativo quanto maior for seu valor no teste, podendo variar de -1 a +1 (SARTES e SOUZA-FORMIGONI, 2013).

Outros dois estudos que são realizados na TCT e acabam se tornando pré-requisitos para sua execução, são os estudos de validade e fidedignidade do teste envolvendo os itens.

Pasquali (2017) afirma que a TCT também se preocupa com a validade de seus itens e testes, de forma a legitimar a validade de um instrumento, segundo o conceito de erro de estimação. O erro de estimação, significa o quanto o escore observado de

um indivíduo no teste se afasta do escore verdadeiro, que pode ser obtido através da seguinte relação:

$$EE = S_c \sqrt{1 - r_{TC}^2}$$

em que **EE** é o erro de estimação; **S_c** é o desvio padrão da medida do critério e **r_{TC}²** é o coeficiente de validade, isto é, a correlação entre o teste e o critério (PASQUALI, 2017).

Toffoli et al. (2016, p. 347) mencionam que o conceito de validade vem sendo proposto e modificado desde os anos de 1920, mas que atualmente, devido a sua simplicidade de entendimento, o conceito dado por Kelley (1927) é muito utilizado, e diz: “Um teste é válido se mede o que foi proposto a medir”.

Já a fidedignidade está totalmente relacionada com a discriminação dos itens. A fidedignidade de um teste é maior quando seus itens possuem um alto índice discriminativo (CAMPOS, 2011).

Diversos autores listam vários pontos fracos na TCT, porém, Sartes e Souza-Formigoni (2013, p. 243), fundamentado nas concepções de Embretson e Reise (2000) e Hambleton e Slater (1997), diz que a

[...]principal limitação da TCT refere-se ao fato de que todas as medidas são dependentes da amostra dos indivíduos que responderam ao instrumento. Isto significa que as avaliações do teste são válidas somente se a amostra for representativa ou se o instrumento for utilizado em outra amostra com características semelhantes.

Portanto, isto leva a grandes questionamentos quanto ao uso exclusivo da TCT atualmente nos testes educacionais, mas que também não se pode desmerecer suas possibilidades de interpretação de desempenho de estudantes.

3.1.2 Teoria da Resposta ao Item - TRI

Esta seção tratará especificamente sobre a teoria da resposta ao item, abordando sucintamente a trajetória histórica, e sua importância como método de correção de testes no atual cenário educacional brasileiro. Procurou-se retratar a TRI com simplicidade, apresentando toda sua estrutura funcional de forma que ela possa ser entendida facilmente por todos que atuam na educação. Buscou-se e entendeu-

se que, o trabalho de professores e demais profissionais, que atuam de forma direta ou indireta com a avaliação em larga escala, se torne mais lúcido e proveitoso com esta abordagem.

Retratando brevemente a história que concebe a teoria da resposta ao item, Pasquali e Primi (2003) apontam que o responsável mais direto pela origem da TRI moderna é fruto dos estudos e aplicações de Frederic Lord entre os anos de 1952 e 1953, e que a

TRI foi sendo elaborada aos poucos desde os anos 50 por vários autores, embora suas raízes remontem há mais de uma década anterior. Entre estes precursores se encontram os trabalhos de Richardson (1936), comparando os parâmetros dos itens obtidos pela teoria clássica da Psicometria com os moldes que hoje usa a TRI; os trabalhos de Lawley (1943, 1944), indicando alguns métodos para estimar os parâmetros dos itens, os quais se afastavam da teoria clássica e os trabalhos de Tucker (1946), que parece ter sido o primeiro a utilizar a expressão curva característica do item – Item Characteristic Curve, ICC - que constitui um conceito chave na TRI. Também deve ser mencionada a contribuição de Lazarsfeld (1950), que introduziu o conceito de traço latente, ainda que no contexto da medida das atitudes, conceito novamente que se constituiu num parâmetro chave da nova TRI. (PASQUALI e PRIMI, 2003, p. 101).

A TRI vem tomando grande importância no cenário educacional brasileiro, principalmente por ser o método de correção adotado no Exame Nacional do Ensino Médio (ENEM). É importante ressaltar que uma teoria não substitui a outra, apesar de a TCT e a TRI serem métodos que levam em consideração fatores distintos, mas que surgiu a partir da necessidade de superar as limitações da apresentação de resultados somente através de percentuais de acertos ou escores dos testes, e ainda, da dificuldade de comparar resultados de diferentes testes em diversas situações (KLEIN, 2003).

Andrade et al. (2000) aponta que a TRI propõe modelos para os traços latentes, ou seja, o que se deseja avaliar nos indivíduos e que não podem ser analisados diretamente. Uma de suas vantagens sobre a TCT é que ela permite maior comparabilidade entre populações, desde que ambas tenham sido submetidas a testes que contenham itens comuns.

Pasquali e Primi (2003) trazem algumas limitações contidas na teoria clássica, que se tornam elementos dificultadores em sua aplicação, fazendo com que a TRI seja um método de correção mais preciso e procurado. Porém, a TRI não substitui completamente a TCT, que se faz necessária funcionando como método

complementar. Comungando também com os relatos de Sartes e Souza-Formigoni (2013), com relação ainda a algumas limitações da TCT, elaboramos o quadro 3 a seguir com alguns destaques, a fim de compararmos ambos os métodos.

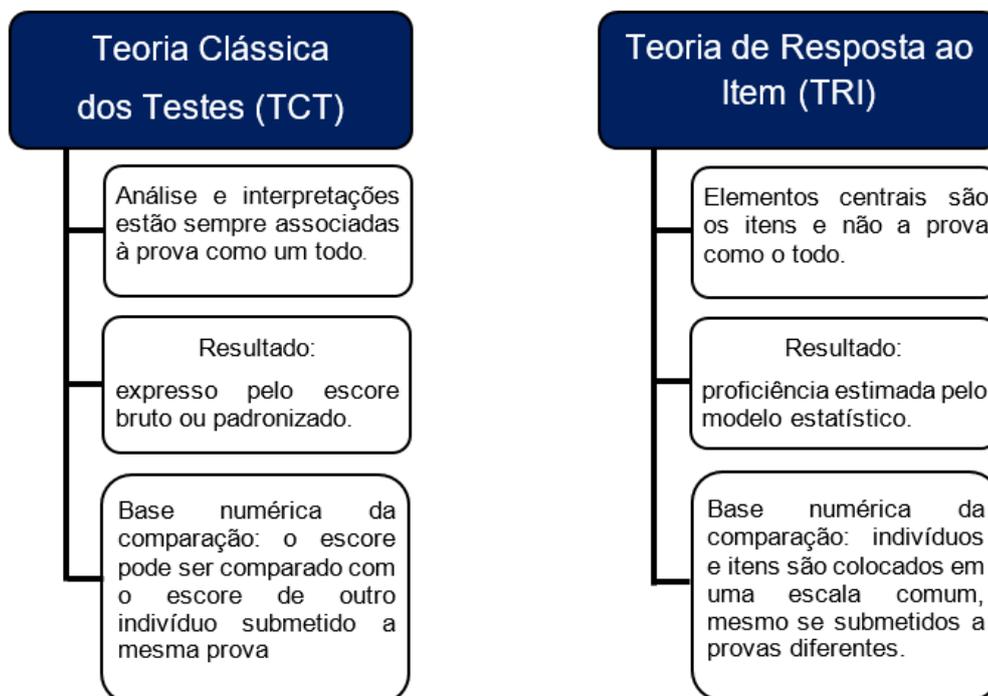
Quadro 3 – Limitações da TCT X Vantagens da TRI

Limitações da TCT	Vantagens da TRI
Todas as medidas são totalmente dependentes da amostra de indivíduos que responderam ao instrumento, tornando a avaliação válida somente se a amostra for representativa e com características semelhantes.	O cálculo da proficiência do indivíduo não depende da amostra e do conjunto de itens utilizados.
Testes diferentes geram resultados diferentes para uma mesma amostra. Considerando-se os índices de dificuldade e discriminação de um instrumento, que seja aplicado em um mesmo grupo duas vezes, este pode gerar resultados distintos, de acordo com a proficiência de seus avaliados.	O cálculo dos parâmetros dos itens não depende da amostra de indivíduos avaliada. Neste caso, um mesmo instrumento pode ser aplicado para grupos de indivíduos diferentes, que os resultados não serão influenciados por esse fator.
Se um mesmo instrumento for utilizado em testes diferentes, os resultados não são expressos em uma mesma escala, fazendo-se necessário uma forma de equalização para compará-los.	Neste método, cada item é alocado em um ponto da escala de proficiência, fazendo com que o item seja acertado por um indivíduo que possua proficiência suficiente para acertá-lo de acordo com seu nível nessa escala. Dessa forma, um item que esteja fixado em um ponto mais baixo da escala, seja acertado por um indivíduo com proficiência baixa, enquanto um item localizado nos pontos mais elevados dessa mesma escala, seja acertado por indivíduos de maior proficiência. Todos os itens do teste são colocados em uma única escala.
A fidedignidade neste método fica a cargo do escore verdadeiro e da variância. Ambos precisam ser iguais, atuando de forma paralela, em duas aplicações diferentes para uma mesma amostra.	A fidedignidade não depende do paralelismo exigido na TCT, em que os resultados do escore verdadeiro e a variância sejam sempre os mesmos.
Neste método, a variância dos erros de medida de todos os indivíduos é a mesma.	A variância dos erros de medida não são os mesmos para todos os indivíduos.

Fonte: Elaborado pelo autor (2020)

Outros aspectos que podem ser comparados entre a TCT e a TRI, são elucidados por Rabelo (2013), conforme a figura 5 a seguir.

Figura 5 – Comparando TCT e TRI



Fonte: Rabelo (2013, p. 128)

Portanto, a TRI considera item por item do teste, funcionando como um modelo matemático que leva em consideração a probabilidade de resposta a um item, modelando a proficiência, ou seja, a habilidade do aluno.

A TRI muda o foco de análise da prova como um todo para a análise de cada item. A TRI é 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. Quanto maior a proficiência do aluno, maior a probabilidade de ele acertar o item. (KLEIN, 2003, p. 127).

Sartes e Souza-Formigoni (2013) dizem que a TRI parte da suposição de que existe no indivíduo um traço latente relativo a uma de suas características, como, por exemplo, seu nível de conhecimento em Matemática. Ou seja, o traço latente nesse caso se refere ao nível de conhecimento ou habilidade que o indivíduo possui sobre o

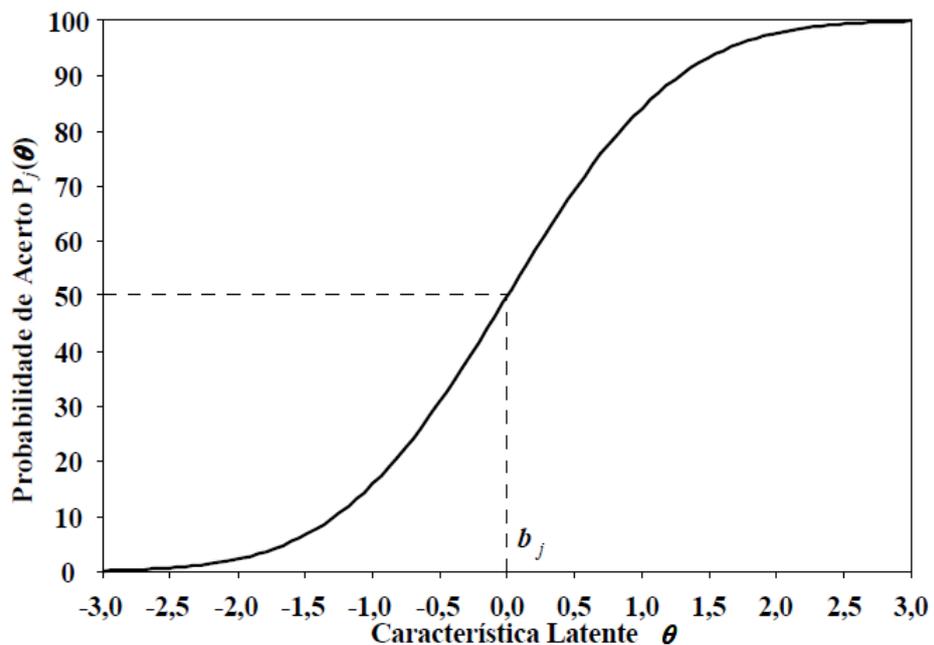
conteúdo avaliado, que muito das vezes pode ser caracterizado, em outros contextos, com o construto.

Dessa forma, a TRI permite estimar se o indivíduo responderá positivamente ou não a um item, sendo que para cada um desses itens será também estimada uma probabilidade de resposta.

Tal probabilidade é determinada pelas características (parâmetros) de cada item por meio de uma função Matemática, tipicamente logística. O primeiro passo para aplicação da TRI é a estimação dos parâmetros dos itens, que pode ser feita por diferentes modelos estatísticos. (SARTES e SOUZA-FORMIGONI, 2013, p. 244).

Como tratou-se de itens do tipo dicotômico, ou seja, aqueles que possuem apenas duas possibilidades de resposta (certo ou errado), o modelo matemático que é utilizado para representar o comportamento do item é através da curva característica do item (CCI), que exemplifica de forma clara que quanto maior for a habilidade do avaliado naquilo que lhe foi proposto, maior será a sua chance de acerto. Na figura 6 pode-se observar um exemplo de CCI de um item denominado item j.

Figura 6 – Curva Característica do Item j.

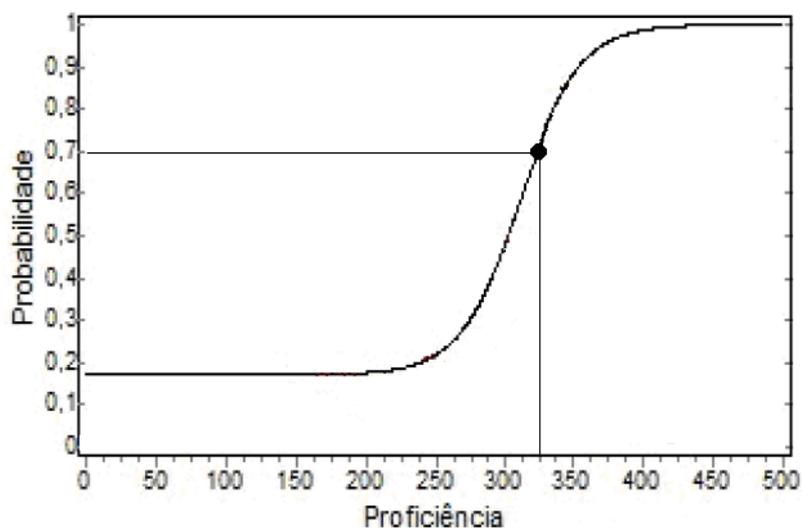


Fonte: Fletcher, 2010.

De acordo com a CCI do exemplo da figura 6, é possível observar que a curva apresenta a probabilidade positiva de acerto do item, indicado no eixo vertical, em função da característica latente do indivíduo, apresentado no eixo horizontal. Ou seja, o eixo vertical representa a probabilidade de acerto do item, enquanto o eixo horizontal representa teoricamente a proficiência do indivíduo avaliado. Cada ponto da curva é associado a um par ordenado, semelhante ao indicado na figura 6, em que um indivíduo com característica latente igual a 0,0 (zero), possui 50% de probabilidade de acertar o item. A escala do eixo horizontal apresentada na figura 6, nem sempre é vista dessa forma, pois usualmente utiliza-se a escala de 0 (zero) a 500 (quinhentos) pontos, semelhante à escala SAEB.

Na figura 7 a seguir, é possível observar um ponto destacado sobre uma CCI, que servirá de exemplo para que se entenda a relação existente entre os dois eixos do gráfico, agora com escala do eixo horizontal do modo usual, de 0 (zero) a 500 (quinhentos) pontos.

Figura 7 – Exemplo de Ponto na CCI.



Fonte: Elaborado pelo autor (2020)

Neste caso, a CCI da figura 7 representa um item em que, um indivíduo com aproximadamente 325 pontos de proficiência, tenha em torno de 70% de probabilidade de acertá-lo. Dessa mesma forma, pode-se corresponder os demais pontos da curva, indicando para cada proficiência, um percentual de probabilidade de acerto para os indivíduos.

Portanto, a TRI mensura tanto a proficiência dos alunos quanto o grau de dificuldade dos itens componentes das provas numa mesma escala de proficiência. Além disso, ela permite uma maior comparabilidade entre os resultados dos testes ao longo dos anos, em substituição aos modelos estatísticos clássicos, que não permitem comparar testes diferentes realizados por populações distintas.

A TRI possui, em sua aplicabilidade alguns modelos logísticos de parametrização, porém, os que mais se destacam atualmente são os modelos logísticos de um, dois ou três parâmetros para itens dicotômicos. Como o próprio nome diz, cada um desses modelos leva em consideração um, dois ou três parâmetros, que são os conhecidos parâmetros **A**, **B** e **C** da TRI.

O modelo logístico de três parâmetros mantém todas as características dos modelos de um e dois parâmetros, porém acrescenta a possibilidade de se avaliar a probabilidade de o sujeito acertar o item ao acaso, ou seja, sem que tenha habilidade suficiente. (COUTO e PRIMI, 2011, p. 6).

O parâmetro **A** se refere a **discriminação** do item, que é o mesmo que identificar o quanto o item discrimina os bons respondentes daqueles que ainda apresentam dificuldades com o conteúdo proposto. Na CCI, o parâmetro A do item é representado pela inclinação da curva exatamente no ponto de inflexão, que é onde a probabilidade de resposta correta é de 50%, habitualmente indicada pela letra do alfabeto grego teta (θ).

O poder de discriminação de um item é a característica que lhe permite avaliar a proficiência desse aluno, ou seja, para acertar um item, o aluno deve apresentar uma proficiência mínima que, se não a possuir, a probabilidade de ele acertar o item é baixa. (MORAES, 2017, p. 44).

O segundo parâmetro é o **B**, que diz respeito a **dificuldade** do item, que se relaciona com o percentual de indivíduos que respondem corretamente o item.

Neste caso, quanto maior o número de pessoas que acertam o item, mais fácil este é, e vice-versa. Por isso, este parâmetro é conhecido como a *dificuldade* do item. (MORAES, 2017, p. 44).

Já o parâmetro **C** corresponde a probabilidade de **acerto casual**, que muitos caracterizam como “chute”. Esse parâmetro representa a possibilidade de um

indivíduo acertar o item, mesmo que não tenha o devido conhecimento para solucioná-lo corretamente.

Nessa perspectiva, compreende-se por que a elaboração de um item é tão importante. Ela pode tanto comprometer um pré-teste quanto colaborar com os aspectos cognitivos que envolvem a ação dos avaliados frente a esse instrumento. A elaboração de um bom teste padronizado exige conhecimento dos parâmetros dos itens. (MORAES, 2017, p.44).

Conforme conferido por Couto e Primi (2013), as variáveis contidas nas funções que possibilitam calcular a TRI através dos modelos logísticos de um, dois ou três parâmetros são as mesmas, porém, no modelo de três parâmetros, em que é considerado os três parâmetros (A, B e C), o valor de **C** coincide com o valor de $P_i(\theta)$, para um valor de $\theta = -\infty$, com a seguinte relação:

$$P_i(\theta) = C_i + (1 - C_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}, \text{ onde,}$$

P_i: probabilidade de se acertar ao item i dado um determinado valor de θ ;

θ : valor do traço latente ou variável que se estiver medindo;

b_i: índice de dificuldade do item i;

e: base dos logaritmos neperianos que vale 2,72;

D: constante de aproximação aos valores da curva normal acumulada.

Para que todo esse processo ocorra de maneira construtiva, é preciso associar os aspectos teóricos e os empíricos. Os aspectos teóricos correspondem desde a elaboração dos itens até a sua aplicação, já fazendo referência a pré-testagem dos mesmos. Com a pré-testagem é possível captar informações e gerar os dados estatísticos, possibilitando a análise empírica, de maneira a verificar se tal item discrimina e avalia de forma concisa seus avaliados.

Na TRI, um item muito fácil, que alunos bons e com déficit de aprendizagem o acertem, significa que o item possui baixa discriminação, ou seja, não é um bom item, que acaba sendo eliminado do teste. A TRI também entende que um aluno que tenha acertado itens muito difíceis e errado os fáceis, pode imprimir uma situação de acerto ao acaso, o que não gera notas elevadas a seu favor. Ou seja, a TRI avalia em um

continuum, diagnosticando qual o comportamento geral de um indivíduo em todo teste.

Corroborando com Couto e Primi (2013), é de se concordar que a construção de medidas é uma tarefa muito complexa e laboriosa e que se faz fundamental para que o profissional em educação, possa ter em mãos, instrumentos de qualidade auxiliando na avaliação dos estudantes.

Desta forma acredita-se que seja desejável aos construtores e usuários de teste um conhecimento, mesmo que elementar das características dos modelos da TRI que permita, se não utilizar o modelo em pesquisas de desenvolvimento de instrumentos, ao menos compreender os conceitos, quando são explanados em trabalhos alheios. (COUTO e PRIMI, 2013, p. 14).

Dessa forma, considera-se que conhecendo, mesmo que superficialmente, alguns propósitos da TRI, suas características e principalmente, sabendo interpretar os dados por ela gerados, o trabalho de elaboradores, revisores e principalmente dos professores, que atuam diretamente com os estudantes avaliados seja, indubitavelmente mais acurado.

Na seção seguinte, se discutirá sobre os parâmetros estatísticos nos itens, além da importância em saber interpretar tais dados, como uma maneira de aproximação ainda mais da teoria da resposta ao item.

3.1.3 Parâmetros Estatísticos

Cada vez mais as avaliações educacionais vêm assumindo um papel estratégico nos processos de mudança e nas reformas educacionais. Nesse sentido, como já mencionado no início da seção 3.1 deste trabalho, Fontanive et al. (2007), salientam sobre a importância da compreensão dos métodos de correção utilizados nas avaliações em larga escala, principalmente no momento da divulgação de resultados, que atinge diretamente os atores administrativos das redes de ensino e, os professores, que trabalham diariamente com os estudantes em sala de aula. No entanto, esses argumentos levam à reflexão sobre a importância dos colaboradores do CAEd, principalmente os que atuam no banco de itens, e de outras instituições que trabalham com a elaboração de instrumentos e testes, serem conscientes no momento de leitura dos dados estatísticos dos itens.

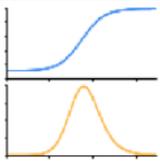
O primeiro e mais importante passo para aplicação da TRI, conforme elucidado por Pereira (2004), é a estimação dos parâmetros dos itens, que caracterizam os modelos de resposta ao item. Todo processo estatístico realizado no momento de pré-testagem, ou até mesmo na aplicação definitiva do teste, é de fato muito complexo.

Dessa forma, entende-se que os parâmetros estatísticos dos itens e da avaliação como um todo, são importantes para que sejam levantadas informações sobre a eficiência e qualidade da educação, tornando evidentes possíveis falhas e orientando estratégias de intervenção. Sobre a eficiência e qualidade da avaliação, a instituição responsável pela elaboração do construto, tem papel fundamental nesse processo. Os parâmetros estatísticos têm grande importância na interpretação pedagógica das redes, sobretudo, por meio de seus itens.

Nesse ponto, corroborando com Tocantins (2018), entendemos que os profissionais que sejam inseridos no contexto de uma instituição ligada a pesquisa com avaliação educacional em larga escala, seja submetido a um processo de ação formativa em um espaço formal, uma vez que, em seu período de formação como profissional da área específica, principalmente a docente, as graduações não abordarem o tema da avaliação em larga escala com profundidade.

Buscando minimizar tais dificuldades de interpretação, procurou-se compreender, mesmo que de modo superficial e bem habitual, os principais parâmetros em um item após ser pré-testado ou efetivamente aplicado. Na figura 8 a seguir, podemos observar um relatório para exemplo dos dados gerados de um item na CIA.

Figura 8 – Exemplo de relatório de item na CIA

Item:		Descritor: D49		Gabarito: A		Suporte: NENHUM		Bloco: 1		Pos.: 5		Curva
Parametro da TRI			Análise Clássica			Percentuais por opção de resposta						
A	B	C	Perc. de Acerto	Bisserial	Ancoragem	B/N	A	B	C	D	E	
0,0116	345,9	0,10723	0,317	0,505	368,149	0,004	0,317	0,174	0,171	0,116	0,219	
OBS:						Biss A	Biss B	Biss C	Biss D	Biss E		
						0,505	-0,036	-0,189	-0,257	-0,231		

Fonte: Paula (2018, pág. 62).

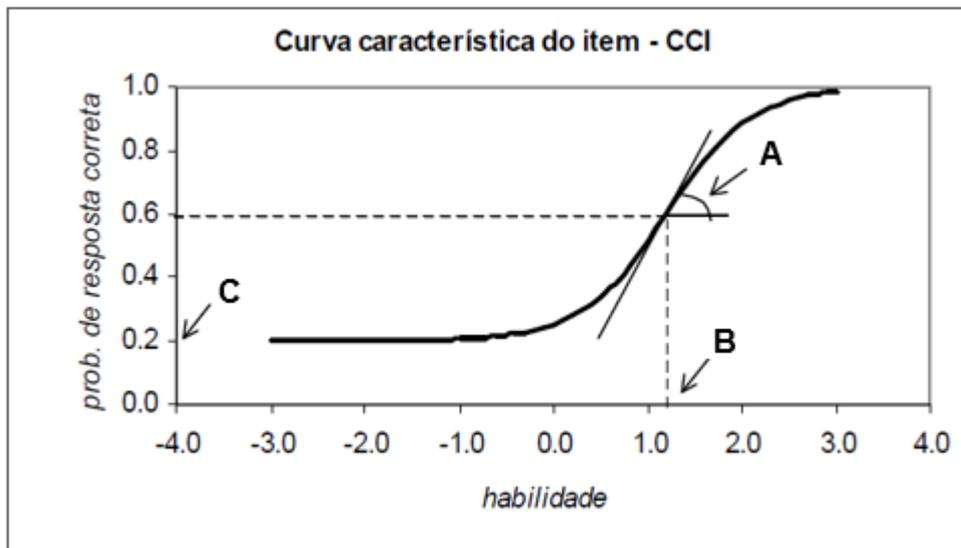
Para identificação e codificação do item, na primeira linha é possível identificar seu código (neste exemplo o código consta em branco, que seria ao lado da palavra

item), o descritor ao qual ele pertence, o código do suporte, caso o item possua o auxílio de uma imagem, por exemplo; e, por último, o bloco ao qual pertenceu e sua referida posição no mesmo. Em um teste os itens são organizados em blocos, chamados de Blocos Incompletos Balanceados (BIB). O BIB é uma técnica utilizada na montagem dos cadernos de testes, funcionando como uma espécie de rodízio de itens em cadernos diferentes.

Os Blocos Incompletos Balanceados são muito úteis nos processos de avaliação pois permitem que conciliemos o grande número de itens necessários ao estudo da performance da sala de aula com o pequeno número de itens que podemos submeter aos alunos individualmente. Os Blocos Incompletos Balanceados têm a propriedade de distribuir os itens de forma uniforme entre o conjunto dos alunos desde que observados alguns cuidados durante a aplicação dos testes. A combinação da Teoria de Resposta ao Item com os Blocos Incompletos Balanceados nos permite obter informações precisas sobre a performance do aluno e da sala de aula simultaneamente. O uso conjunto destas mesmas técnicas também pode ser utilizado para que um banco de itens testados e calibrados cresça num ritmo muito maior e consistente do que seria possível com outras metodologias. (BEKMAN, 2001, p. 134).

A partir da segunda linha, é possível observar os parâmetros estatísticos do referido item. Os parâmetros A, B e C da TRI são os primeiros que podem ser consultados nesse relatório. Como já mencionado na seção 3.1.2, esses três parâmetros se referem a discriminação (A), dificuldade (B) e a probabilidade de acerto ao acaso (C), porém, nesta seção será um pouco mais aprofundado. No entanto, a CCI pode informar graficamente os parâmetros psicométricos dos itens, conforme o modelo logístico utilizado, neste caso, o de três parâmetros (Pasquali, 1996, 2000). Na figura 9 a seguir, verifica-se com mais clareza cada um desses parâmetros.

Figura 9 – Identificação dos parâmetros CCI



Fonte: Andrade et al. (2000, pág. 11)

Na CCI, fica evidente que o parâmetro de discriminação do item, identificado por “A” na figura 9, corresponde ao ângulo de inclinação no ponto de inflexão da curva em relação ao eixo da habilidade (horizontal). Tecnicamente, Andrade et al. (2000), diz que o parâmetro A é obtido a partir da derivada da curva no ponto de inflexão, fazendo com que, quanto maior o ângulo de inclinação, maior é a capacidade discriminativa do item, ou seja, diferir estudantes com baixa ou alta habilidade.

A tabela 6 a seguir, é um apontamento realizado por Rabelo (2013), classificando o fator discriminativo de itens em função de seu valor de parâmetro A.

Tabela 6 – Classificação de acordo com a discriminação do item.

Valor do parâmetro A	Fator discriminativo
$a = 0$	Nenhuma
$0 < a \leq 0,0035$	Muito baixa
$0,0035 < a \leq 0,0065$	Baixa
$0,0065 < a \leq 0,0135$	Moderada
$0,0135 < a \leq 0,0170$	Alta
$a > 0,0170$	Muito alta

Fonte: Rabelo (2013, p. 138)

O parâmetro de dificuldade (B), identificado por “**B**” na figura 9, é medido na mesma escala de habilidade do item, correlacionada ao ponto de inflexão da curva. Dessa forma, quanto maior for o valor de B, maior será a dificuldade do item e vice-versa.

Assim como na discriminação, Rabelo (2013) registra, conforme apresenta a tabela 7 a seguir, uma classificação de dificuldade do item, conforme seu valor de parâmetro B, em testes que a escala média seja de 500 pontos e o desvio padrão igual a 100.

Tabela 7 – Classificação de dificuldade do item

Classificação	Distribuição Esperada	Dificuldade do Item
Muito Fácil	10%	$b \leq 372$
Fácil	20%	$372 < b \leq 448$
Mediano	40%	$448 < b \leq 551$
Difícil	20%	$551 < b \leq 627$
Muito Difícil	10%	$b > 627$

Fonte: Rabelo (2013, p. 134)

Rabelo (2013), fundamentado em Pasquali (2003), aponta que é recomendado que se tenha uma distribuição normal de itens de acordo com a sua classificação de dificuldade em um teste, de forma que, 10% sejam de itens pertencentes a cada uma das faixas extremas (muito fácil e muito difícil), 20% dos itens seja de cada uma das faixas seguintes (fácil e difícil), e por último que, 40% seja de itens da faixa central (medianos).

Ainda nessa primeira etapa, tem-se o parâmetro C, identificado por “**C**” no eixo vertical da figura 9, correspondendo a probabilidade de acerto ao acaso, por estudantes de baixa proficiência (habilidade). Para itens com 5 alternativas, o valor de parâmetro C ideal é recomendado que seja no máximo igual a 0,20, o que corresponde a 20% (vinte por cento), equivalendo à probabilidade de acerto de uma das alternativas. Em itens com 4 alternativas, essa probabilidade aumenta para 0,25 também sendo a probabilidade de acerto de uma das alternativas.

A parte referente a análise clássica do relatório da figura 8 contém duas informações, que são o percentual de acerto do item e a correlação bisserial. O percentual de acerto é um cálculo simples, que se refere ao percentual de indivíduos, em relação ao total da população avaliada, que acertou o item, ou seja, marcaram a alternativa correta. Já a correlação bisserial está relacionada com a análise elaborada sobre o “**percentual por opção de resposta**” de cada uma das alternativas item. Na prática, o valor da bisserial de um item sempre é igual ao valor da correlação bisserial da alternativa que corresponde ao gabarito.

Conforme apontado por Pereira (2004), a correlação bisserial são medidas estatísticas que medem a correlação do resultado de um item em particular do teste com o resultado do teste, ou seja, pelo score bruto ou total.

A correlação bisserial atua como uma ferramenta central na análise de itens, e Gomes (2014, p. 21) aponta que é esperado um coeficiente maior que 0,30, pois valores abaixo ou bem próximos de zero (negativos e próximos a zero), indica que os indivíduos com bom desempenho no teste estão buscando alternativas erradas, ou seja, errando a questão (o item). Esse é primeiro ponto que pode demonstrar a ineficiência de um item em um teste.

Dessa forma, a correlação bisserial é obtida a partir da seguinte relação:

$$\rho_{pb} = \frac{S_p - S}{\sigma} \sqrt{\frac{p}{1-p'}}$$

em que:

S_p é a nota média no teste para os indivíduos que acertaram o item;

S é a nota média no teste para todos os indivíduos;

σ é o desvio padrão das notas obtidas no teste por todos os indivíduos;

p é a proporção de acertos no item.

Em relação às alternativas, Gomes (2014) salienta que a correlação bisserial também pode ser calculada para cada uma das alternativas.

Para isso, basta substituir na equação, a nota média no teste para os indivíduos que acertaram o item pela nota média no teste dos indivíduos que marcaram o item. (GOMES, 2014, p. 21).

Dessa forma, a correlação bisserial é uma medida que afere a capacidade de discriminação do item em relação ao resultado do teste. A bisserial é capaz de associar o desempenho de um estudante em um item, e no teste. Dessa forma, é interessante que, apenas a alternativa referente ao gabarito apresente bisserial positiva, caso contrário, se além desta, outra alternativa apresentar correlação bisserial positiva, significa que estudantes com boa proficiência estão sendo atraídos por esta alternativa. Portanto, caso duas alternativas de um item apresentem correlação bisserial positiva, pode-se estar de frente a um instrumento com problemas de elaboração. Este é um dos fatores, como já mencionado anteriormente, que move a realização desta pesquisa com itens. Assim, as informações sobre o percentual por opção de resposta auxiliam no cálculo das bisseriais por alternativa.

A próxima informação do relatório diz respeito ao ponto de **ancoragem** do item. A ancoragem corresponde a localização do item na escala de proficiência, de modo que seu valor indique a quantidade de habilidade necessária para que indivíduo da amostra avaliada acerte o item. A ancoragem é um processo fundamental na construção das escalas de proficiência, como apontado por Klein (2009) na seção 2.4 deste trabalho.

A informação seguinte contida no relatório do item na CIA é o percentual de respondentes que deixaram o item em branco ou anularam a resposta, indicado por **B/N** na figura 8. Por último, temos dois pequenos gráficos à direita do relatório, que se trata da CCI e da curva de informação do item. A curva de informação é o modo gráfico, por meio de uma curva normal, de indicar o quanto um item contém de informação psicométrica para a medida da habilidade, atuando como um termômetro do item ao longo da escala de habilidade, mostrando a importância dos três parâmetros sobre o montante de informação do item (ANDRADE et al. 2000).

Uma observação levantada por Klein (2003) é que

[...] uma das propriedades importantes da TRI é o fato dos parâmetros dos itens e as proficiências dos indivíduos serem invariantes. Tanto os parâmetros dos itens obtidos de grupos diferentes de alunos testados quanto os parâmetros de proficiência baseados em grupos diferentes de itens são invariantes, exceto pela escolha de origem e escala. Graças a essa propriedade, a TRI, associada a outros procedimentos estatísticos, permite comparar alunos, estimar a distribuição de proficiências da população e subpopulações e ainda a monitorar os progressos de um sistema educacional. (KLEIN, 2003, p. 127).

Dessa forma, pressupõe-se que todas as informações estatísticas do item contribuem não só para o progresso dos sistemas educacionais, mas também para o desenvolvimento de novas pesquisas dentro de todos os segmentos e áreas da educação. No entanto, é necessário que se saiba interpretar o que as avaliações trazem de devolutiva, principalmente aqueles que atuam diretamente com a elaboração de instrumentos de avaliação.

Na seção seguinte se aprofundará nas justificativas de cada um dos motivos de eliminação de itens apontados, sobre uma perspectiva teórica e de experiências a partir de relatos da equipe de elaboração da CIA.

3.1.4 O Que Significa Um Item Eliminado

Nesta seção será abordado o que de fato eliminou estatisticamente os itens definidos para esta pesquisa. Como já mencionado anteriormente, os motivos de eliminação de itens no CAEd são caracterizados por ajuste, DIF ou por correlação bisserial. Dessa forma, busca-se abordar tais assuntos com mais naturalidade, uma vez que as análises que envolvem os motivos estatísticos de eliminação são processos matemáticos muito complexos, e que, em sua maioria, utiliza-se o auxílio de softwares computacionais para identificá-los. Portanto, não interessa neste trabalho apresentar de maneira minuciosa as fórmulas, relações matemáticas ou outros aspectos muito aprofundados sobre cada um dos motivos que rodeiam a análise de dados dessa pesquisa, mas possibilitar o entendimento de forma mais sucinta e objetiva.

Primeiramente, se abordará as eliminações com motivo de ajuste. O processo de estimação dos parâmetros, comumente chamado de **ajuste**, ou **ajuste ao modelo**, trata de uma relação entre os aspectos empíricos e teóricos de um item. De acordo com Andriola (2009), a lógica desse procedimento é encontrar valores para os parâmetros que maximizem a probabilidade de ocorrência das respostas dos sujeitos aos itens. Isso porque, de acordo com os modelos logísticos mais utilizados, é preciso encontrar a variável latente de cada sujeito. Matematicamente, esse processo ocorre através da “**máxima verossimilhança**”, considerando-se inicialmente valores estimados para o item, que são ajustados por uma série de interações, que são calculadas com o auxílio de softwares computacionais, pois seria uma tarefa incomensurável diante de sua complexidade e necessidade de otimização de tempo.

Para Couto e Primi (2011), o ajuste ao modelo ocorre segundo alguns passos, que serão descritos no quadro 4 a seguir.

Quadro 4 - Passos para estimação dos parâmetros – Ajuste

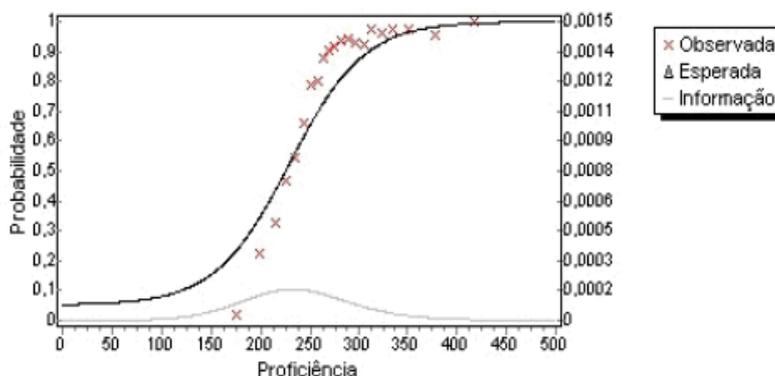
1º passo	Elaboração de itens	É o passo para definição do que se deseja avaliar, de forma que o instrumento se adapte adequadamente a população que se deseja avaliar.
2º passo	Pré-testagem	Deve-se aplicar os instrumentos a uma amostra representativa para a qual se construiu. Neste passo o objetivo é estimar os parâmetros da psicometria clássica e verificar a unidimensionalidade dos itens. Nessa fase já é possível, em alguns casos, verificar qual o modelo que melhor se ajusta aos dados.
3º passo	Estimação dos parâmetros	Nesta fase aplica-se o método da máxima verossimilhança, em busca de valores que aumentam a probabilidade de ocorrência de dados. Como os parâmetros são desconhecidos, inicia-se um processo iterativo de cálculos, assumindo valores hipotéticos para os itens, derivados da psicometria clássica. Aqui também se estima a habilidade do sujeito avaliado, considerando-se os valores provisórios, para que, através das interações, se chegue a parâmetros mais precisos, de forma que ocorra o ajuste da curva empírica com a curva teórica.

Fonte: Elaborado pelo autor a partir de Couto & Primi (2011)

Couto e Primi (2011) apontam ainda que o ajuste pode ocorrer de duas formas distintas. A primeira é conhecida como estimação condicional, em que os valores dos parâmetros do item eram fixos e a habilidade do sujeito obtida através de sucessivas estimações, até que se ajustem os padrões empíricos do item. A segunda é chamada de estimação conjunta, que é o método em que são calculados, ao mesmo tempo, os parâmetros do item e a proficiência do sujeito.

Na prática, podemos utilizar a CCI do item M110104E4 para representar graficamente a sua eliminação por motivo de ajuste no PROEB de 2015, de acordo com a figura 10, a seguir.

Figura 10 – Modelo ajustado X observações empíricas do item M110104E4



Fonte: Acervo de dados da pesquisa.

Nesta curva característica, estão representados o comportamento esperado do item, por meio da linha sólida em preto, o comportamento empírico, em vermelho, e a curva de informação mais abaixo em cinza. Observe que neste caso, tem-se a representação de um modelo já ajustado, e uma discordância entre as curvas empírica e teórica. Nessa situação, o operador do software estatístico, que auxilia no cálculo dos parâmetros dos itens, realizou a aproximação possível com a curva teórica, esgotando-se recursos estatísticos para um melhor ajuste da curva empírica. Dessa forma, este item que foi aplicado no 3EM do PROEB de 2015, foi descartado por motivo de ajuste, pois sua curva empírica não possui equivalência com a teórica, mesmo após as interações estatísticas. Resumindo, um item é eliminado por motivo de ajuste, quando não é possível ajustar seus dados empíricos ao teórico.

O segundo motivo que entra em destaque na eliminação de itens é o comportamento diferencial do item. Conforme aponta Soares et al. (2005), a análise de DIF consiste em verificar se um item tem ou não o mesmo comportamento para indivíduos pertencentes dois grupos distintos, porém, com mesma habilidade cognitiva. Outro ponto que Soares et al. (2005) salienta, é que, em geral, com a análise de DIF deseja-se analisar se um item apresenta graus de dificuldade diferentes para subgrupos da população que tem o mesmo nível de conhecimento.

Fundamentado nos estudos da ETS, Soares et al. (2005) afirma que grande parte dos motivos pelo qual o DIF ocorre em itens ainda é desconhecido, mas que pode ser causado por três fatores, que são a **familiaridade** com o conteúdo do item, o **interesse** pessoal pelo conteúdo e a **reação emocional negativa**. Percebe-se que estes três fatores se relacionam diretamente com o conteúdo abordado no item, em

que alguns casos, os indivíduos podem se sentir atraídos e por ter facilidade em desenvolvê-lo, ou ainda, se sentir desmotivado, caso não seja de seu agrado.

Outro ponto também elucidado por Soares et al. (2005), é que, no caso do Brasil, um país de dimensões continentais, o DIF pode se destacar por motivos regionais, culturais da população, aspectos sociais, entre outros. Dessa forma, cabe ressaltar sobre a importância e o cuidado que se deve ter no processo de elaboração de itens, restando cautela quanto ao uso de enunciados ou ilustrações que podem remeter a temas regionais, causando mais familiaridade para uns do que para outros. Características da linguagem como expressões ou gírias locais, fatos ocorridos em determinada região, nomes ou palavras que possam ser mais familiares para determinadas populações devem ser evitados.

No entanto, embora em grau elevado o DIF possa prejudicar a comparabilidade dos resultados, quando moderado e localizado em poucos itens, o DIF além de, praticamente, não afetar a proficiência produzida pode, se devidamente analisado, trazer informações importantes sobre diferenças curriculares e diferenças sócio-culturais, por exemplo, entre as regiões. (SOARES et al., 2005, p.482).

Dessa forma, conjectura-se que a análise de itens com eliminações por DIF podem se tornar uma grande aliada das equipes de elaboração, no que tange não só a preparação de novos membros das equipes, mas também para formação continuada, em busca de desenvolvimento de novas concepções no trabalho.

Por último, se destacam as eliminações de itens por motivo da bisserial. Como já mencionado na seção 3.1.3 desse trabalho, a correlação bisserial se relaciona com o percentual de procura de cada uma das alternativas do item. Para fins de estudo sobre itens, são apresentados no relatório de cada um, as bisseriais por alternativa. A bisserial se trata de um instrumento de fácil visualização e identificação de possíveis incoerências em um item, uma vez que, apenas a alternativa que corresponda ao gabarito do item, deve apresentar correlação bisserial positiva.

Desta forma, é possível analisar quais alternativas podem estar se tornando mais atrativas a uma determinada população, evitando conseqüentemente, fatores que influenciem a eliminação do item, no momento de sua construção. Couto e Primi (2011) alertam sobre o rigor que se deve ter no processo de construção de itens, associando que, construir instrumentos de avaliação é semelhante ao processo de se escrever uma novela, pois, se seguir rigorosamente todas as regras levasse a boas novelas, então todos seriam excelentes romancistas. No entanto, reiterando e

combinando esses argumentos aos propósitos desse trabalho, que é utilizar itens eliminados para aprimoramento profissional, destaca-se o quanto esses fatores de eliminação podem auxiliar na busca pela qualidade do item, e nortear o trabalho de tantos profissionais que atuam nesse campo ainda pouco conhecido.

3.1.5 Procedimentos e Instrumentos Metodológicos

Nesta seção, entraremos no universo metodológico que envolve esta pesquisa, a começar pelo tipo de estudo, que se caracteriza por um estudo de caso. Considerando os argumentos de Yin (2015), um estudo de caso é usado em muitas situações para contribuir com o conhecimento geral dos fenômenos individuais, grupais, organizacionais, sociais, políticos e relacionados. Dessa forma, se justifica os motivos pela compreensão da logística da elaboração de itens na CIA, principalmente da atuação da equipe de Matemática, verificando seu alinhamento com as definições que norteiam o processo de elaboração contidos neste trabalho.

A abordagem é qualitativa, uma vez que se está envolvido em uma pesquisa que necessita da opinião e experiência de membros de equipes de trabalho, relatórios estatísticos e um quantitativo significativo de itens. Yin (2016) aponta cinco características de uma pesquisa qualitativa, que são: estudar o significado da vida das pessoas, nas condições da vida real; representar as opiniões e perspectivas das pessoas de um estudo; abranger as condições contextuais em que as pessoas vivem; contribuir com revelações sobre conceitos existentes ou emergentes que podem ajudar a explicar o comportamento social humano; e esforçar-se por usar múltiplas fontes de evidências em vez de basear em um única fonte (YIN, 2016).

Portanto, embasados nos dados estatísticos dos itens, busca-se evidências para realizar um alinhamento entre as práticas de elaboração e as opiniões e experiências de colaboradores da equipe de Matemática, na tentativa de verificar se o trabalho realizado está de acordo com as práticas recomendadas e se estas não contribuem para a eliminação de itens. Como consideramos que os itens eliminados podem contribuir para a iniciação e formação continuada de elaboradores, trazem experiências sobre o aprimoramento do trabalho, por meio de evidências, busca-se contribuir para elevar ainda mais a qualidade do trabalho realizado no processo de avaliação educacional.

Como instrumento de pesquisa, se utilizará uma entrevista na forma de revisão coletiva de itens com todos os integrantes da equipe de Matemática, responsáveis pela elaboração e revisão de itens na CIA, que atualmente funciona com três colaboradores na elaboração e dois na revisão, contabilizando o total de cinco colaboradores na equipe. As entrevistas serão realizadas em momentos diferentes, com os colaboradores da elaboração e da revisão, oportunizando a verificação do alinhamento do trabalho realizado na equipe de Matemática, desde a produção até a finalização do item. Será colocada uma dinâmica diferenciada de revisão coletiva do item, apresentando seus dados estatísticos, histórico de utilização e sua imagem, sincronicamente com perguntas específicas sobre o conhecimento de itens.

O tipo de pesquisa se enquadra dentro de alguns modelos, como a pesquisa bibliográfica, em que se busca o auxílio de literaturas que norteiam o processo de elaboração de itens. A análise documental, que foi feita através dos relatórios estatísticos gerados no CAEd, e a pesquisa de campo, realizada por meio de entrevistas/questionários, busca compreender a relação dos membros da equipe de Matemática com o trabalho de elaboração de instrumentos para avaliação.

Da natureza, tem-se o propósito de apresentar uma pesquisa aplicada, com intuito de gerar conhecimento para aplicação de práticas de trabalho, dentro da CIA no CAEd ou outra instituição que seja do ramo da avaliação educacional em larga escala.

3.2 ANÁLISE DOS ITENS

Esta pesquisa tem como objetivo identificar possíveis vieses adotados no processo de elaboração de itens da equipe de Matemática na Coordenação de Instrumentos de Avaliação (CIA) do CAEd, que possam estar influenciando na eliminação de itens. Para avançar na discussão, nesta seção, serão trazidas imagens e análises sobre cada um dos itens que foram selecionados conforme o critério de escolha estabelecido. Para exibir o problema levantado, destaca-se que foram escolhidos três itens de cada um dos motivos de eliminação (ajuste, bisserial e DIF), levando em consideração os três itens com menor percentual de acerto em cada um desses motivos. Isto é, após identificados todos os itens eliminados em cada um dos motivos de eliminação, foram selecionados os três que apresentaram o menor percentual de acerto. Ainda nesta seção, será apresentado o diagnóstico de alguns

membros da equipe de Matemática da CIA produzidos durante a pesquisa de campo, dialogando sobre aspectos da elaboração e colocando os itens em discussão.

Portanto, além de um parecer do pesquisador, esse diagnóstico foi realizado por cinco especialistas da equipe de Matemática da CIA, identificados como Entrevistado 1 (**E1**), Entrevistado 2 (**E2**), Entrevistado 3 (**E3**), Entrevistado 4 (**E4**) e Entrevistado 5 (**E5**). Esses especialistas foram interpelados em cada um desses itens sobre três questões, a saber: (1) A partir do histórico de utilização do item no PROEB e seus respectivos dados estatísticos, qual ou quais motivos podem ter ocasionado a eliminação do item? (2) Sobre os distratores, eles são coerentes? Você reconhece alguma prática comum de elaboração na construção desses distratores? (3) Observando os dados estatísticos, como você avalia o comportamento do item no teste?

No processo educacional de ensino e de aprendizagem, observar é algo imprescindível na conduta do professor/mediador. As respostas construídas pelos estudantes podem revelar fatos significativos, funcionando como uma maneira de investigar o que um indivíduo está construindo cognitivamente sobre o que se está ensinando. Por vez, na sala de aula, fortuitamente o tempo não se torna hábil para se possa desenvolver determinados conteúdos e realizar tais observações, fazendo com que essas análises se tornem favoráveis nas avaliações em larga escala.

Ribeiro e Cury (2018) apontam que a análise de questões de Matemática feitas por estudantes de qualquer nível de ensino deve também ser objeto de atenção dos professores. Nesse sentido, afirmam que

[...] discussões acerca da forma como essas questões foram resolvidas, buscando entender as causas dos erros e dificuldades, podem levar os docentes a usarem os erros como trampolins para a aprendizagem. (RIBEIRO e CURY, 2018, p. 52).

Por motivo de ajuste, tem-se três itens que foram utilizados no PROEB nos anos de 2016 e 2017, avaliando habilidades que exigem do estudante, associar funções a seus gráficos. Dois deles foram utilizados no mesmo ano em duas etapas distintas do ensino médio, como se pode ver detalhadamente a seguir.

Quadro 05 – Item M120389H6

Observações: Item aplicado no PROEB de 2017 e eliminado no 1EM e 3EM por motivo de ajuste.											
Habilidade avaliada: D72 – Corresponder uma função exponencial a seu gráfico.											
Dados estatísticos na(s) ocorrência(s) com eliminação: 1EM – PROEB 2017											
Código: M120389H6			Descritor: D72			Gabarito: D					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,173	0,198		0,005	0,134	0,189	0,237	0,173	0,263
Obs.: Ajuste							Biss. A	Biss. B	Biss. C	Biss. D	Biss. E
							-0,082	-0,016	-0,139	0,198	0,056
3EM – PROEB 2017											
Código: M120389H6			Descritor: D72			Gabarito: D					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,169	0,177		0,003	0,087	0,184	0,216	0,169	0,341
Obs.: Ajuste							Biss. A	Biss. B	Biss. C	Biss. D	Biss. E
							-0,094	-0,076	-0,146	0,177	0,095
Imagem do Item: (M120389H6) Observe abaixo o esboço do gráfico de uma função exponencial $f: \mathbb{R} \rightarrow (-3, \infty)$.											
A representação algébrica dessa função é											
A) $f(x) = 2x - 1$.											
B) $f(x) = x^2 - 2$.											
C) $f(x) = \left(\frac{1}{2}\right)^x - 3$.											
D) $f(x) = 2^x - 3$.											
E) $f(x) = 3^x - 2$.											
										Ativar Araraçá	

Fonte: CAEd/UFJF.

O descritor D72 exige do estudante a capacidade de associar corretamente o gráfico de uma função exponencial dada, a sua lei de formação (forma algébrica). Neste caso, o elaborador apresentou um gráfico e perguntou de forma direta, com auxílio de funções como alternativas, qual delas é a função do gráfico apresentado. No entanto, nesse tipo de item é muito comum o elaborador adotar algumas técnicas

para elaboração em larga escala de itens, como, por exemplo, utilizar pontos em destaque no gráfico para construir as alternativas.

Fica visível que o ponto de intersecção do gráfico com o eixo das ordenadas foi explorado nas alternativas "B" e "E", como se fossem o coeficiente linear de funções de primeiro grau. Nas duas utilizações desse item, a alternativa "E" apresentou bisserial positiva, o que indica que bons alunos foram atraídos para essa alternativa. Conseqüentemente a alternativa "E" foi a mais procurada pelos estudantes, em que possivelmente o número três tenha sido resultado de uma ligação da coordenada 3 utilizada no eixo das abscissas em destaque com a variável "x" da função exponencial, e a coordenada -2, sendo associada, grosso modo, a um coeficiente linear. Nessa situação podemos concluir que bons alunos foram atraídos, possivelmente, por uma conceituação utilizada em outros tipos de função, como por exemplo na função afim (1.º grau), no entanto, este artifício de construção para o item fez com que a alternativa "E" fosse a mais procurada pelos estudantes, hipoteticamente se tornando um dos motivos de eliminação do item.

De acordo com as análises obtidas na pesquisa de campo, o Entrevistado 1 (E1), aponta que:

Neste item temos duas bisseriais positivas, ou seja, bons respondentes deram como resposta as alternativas D e. (ENTREVISTADO 1, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Já o Entrevistado 2 (E2), em relação ao mesmo item, relata que:

Muitos alunos que obtiveram bons resultados no teste, se confundiram e marcaram a alternativa E como resposta. Acho os distratores coerentes. Geralmente colocam-se expressões em que pelo menos um dos pontos do gráfico faça sentido. Não avaliou claramente, visto que bons alunos marcaram a E. Esse tipo de item não é fácil. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 3 (E3) também chama a atenção quanto aos resultados apresentados pela bisserial, dizendo que:

Uma possibilidade para a eliminação deste item pode estar relacionada a bisserial positiva da alternativa E, mostrando que "bons" estudantes estão sendo atraídos para ela. Uma das formas que os

estudantes resolvem esse tipo de item é pela substituição de pontos do gráfico nas alternativas, o que já indica uma análise das mesmas, o que é inadequado na concepção de item. Percebe-se também uma tentativa Na construção dos distratores A e B de identificar os estudantes que não conhecem a curva do gráfico e a lei de formação de uma função exponencial, e que foram atraídos ao encontrar uma equivalência na substituição de algum ponto. Acredito que os "bons" estudantes que foram atraídos para o distrator E, conhecem o gráfico e a forma da função exponencial, porém, podem indicar uma defasagem na operação de potenciação, também muito comum, entendendo que 3 elevado a 0 é igual a 0. Já vi em muitos itens esse tipo de construção dos distratores, porém o mais indicado na construção dos distratores é que se baseiem em erros pautados na construção da habilidade avaliada, e não em erros que teoricamente já deveriam ter sido superados em etapas anteriores. Apesar do item, em partes, não avaliar de fato o que se propõe, ele traz uma riqueza em sua análise, podendo inclusive denunciar defasagens anteriores que não foram superadas. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

As análises realizadas pelo Entrevistado 4 (E4) apontam uma observação quanto a estratégia utilizada no distrator da alternativa C, dizendo:

Outra alternativa além do gabarito com bisserial positiva. O distrator E foi mais marcado que o gabarito. Apesar de coerentes, o distrator da alternativa C destoou um pouco dos demais por conter a base fracionária e o da alternativa E chamou atenção pelos dados do gráfico, uma vez que essa habilidade, pela minha experiência, muitas vezes não é plenamente consolidada pelos alunos. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 5 (E5) se reporta aos dados estatísticos de outras utilizações do item, realizando um comparativo entre as aplicações no 1.º e 3.º ano do ensino médio, chamando a atenção quanto aos percentuais de acerto, e ainda, quanto a estrutura dos distratores, dizendo:

Quanto ao histórico, percebo que o percentual de acerto do item praticamente se manteve constante e baixo (~17%), mesmo sendo aplicado em etapas distintas. Isso é um indicativo que o item possa ter algum problema ou que de fato os estudantes não conhecem, ou não estão sabendo associar o gráfico de funções exponenciais à representação algébrica, uma vez que os estudantes do 1EM e do 3EM não estão acertando. Analisando os distratores, não percebo incoerência na construção, pelo contrário, o estudante que reconhece o gráfico de uma função exponencial crescente, poderia inclusive eliminar 3 das 5 alternativas, e isso não acontece quando percebemos os percentuais de marcação do gabarito e da alternativa E. De fato, a alternativa E atraiu mais respondentes, talvez por ser a única

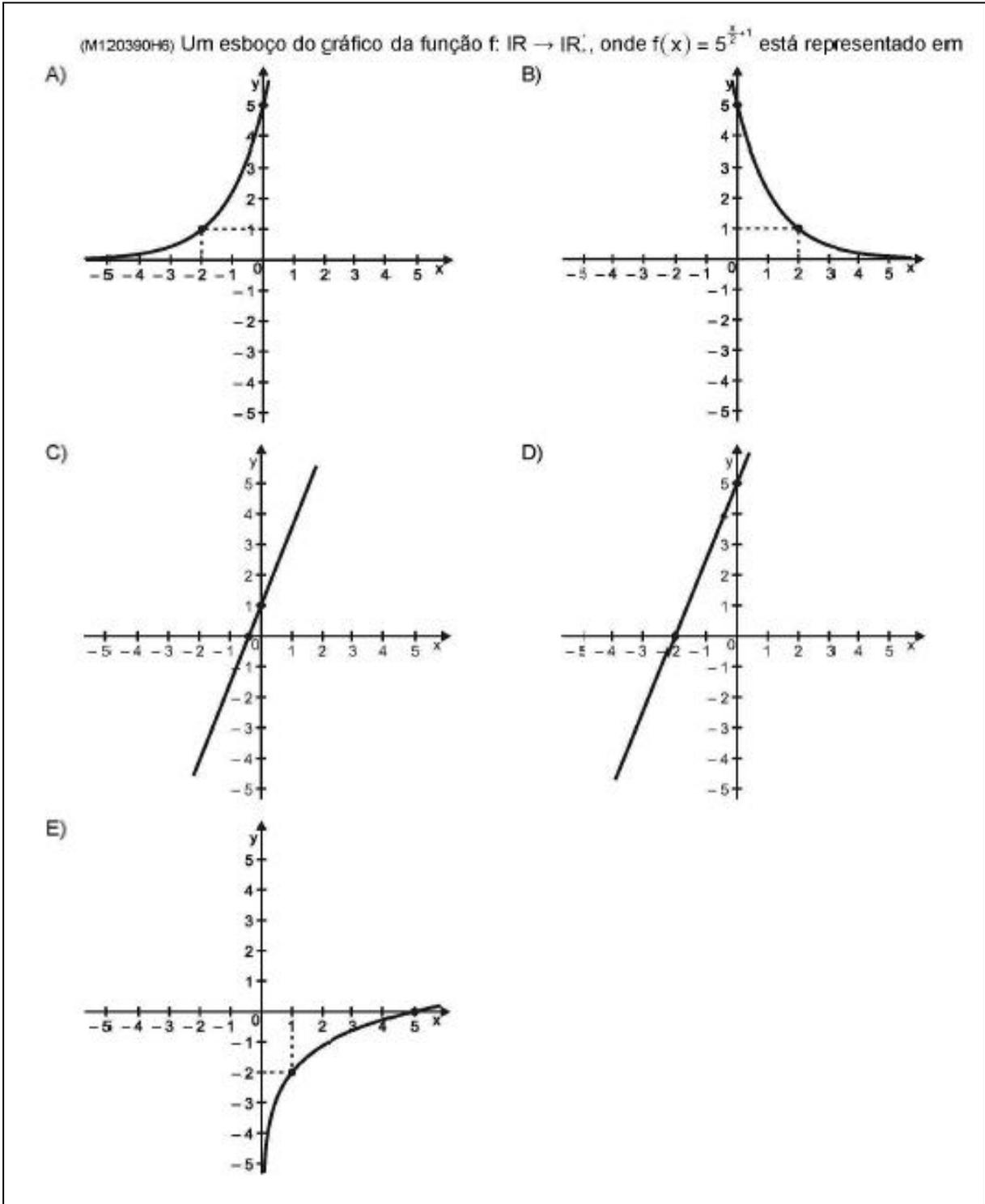
exponencial crescente que apresenta os valores assinalados no gráfico. É uma prática comum na elaboração. O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa E, também é positiva e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelo distrator E. (ENTREVISTADO 5, Questionário de Pesquisa Acadêmica, 31 jul. 2020)

Nesse primeiro item cabe lembrar que, a eliminação se deu por motivo de ajuste, ou seja, estatisticamente, não foi possível um ajuste adequado ao modelo esperado. É possível observar a partir das análises dos entrevistados, que, em unanimidade, se pautaram nas bisseriais das alternativas como um dispositivo que possibilita compreender o comportamento do item, deixando claro que, a bisserial é uma ferramenta indispensável ao se analisar itens, elucidando o trabalho do elaborador mesmo com outros motivos de eliminação.

No quadro 6, temos o segundo item eliminado por motivo de ajuste, avaliando o mesmo descritor do item no quadro 5, porém, com uma proposta inversa.

Quadro 06 – Item M120390H6

Observações: Item aplicado no PROEB de 2017 e eliminado no 1EM e 3EM por motivo de ajuste.											
Habilidade avaliada: D72 – Corresponder uma função exponencial ao seu gráfico.											
Dados estatísticos na(s) ocorrência(s) com eliminação:											
1EM – PROEB 2017											
Código: M120390H6			Descritor: D72			Gabarito: A					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,129	0,178		0,007	0,129	0,419	0,114	0,182	0,150
Obs.:						Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
						0,178	0,150	-0,187	-0,137	-0,083	
3EM – PROEB 2017											
Código: M120390H6			Descritor: D72			Gabarito: A					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,122	0,160		0,002	0,122	0,509	0,083	0,142	0,141
Obs.:						Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
						0,160	0,106	-0,170	-0,148	-0,063	
Imagem do Item:											



Fonte: CAEd/UFJF.

Apesar de se tratar de um item com o mesmo descritor do anterior, neste caso o elaborador optou por explorar a habilidade de forma contrária, apresentando diretamente a lei de formação da função exponencial, dando como alternativas o gráfico de algumas funções diversas, dentre elas, afim e logarítmicas. É possível visualizar que a base 5 da lei de formação não foi explorada apenas na alternativa "C",

o que pode ter justificado o baixo percentual de procura pela referida alternativa em ambas as aplicações em 2017. Aparentemente, neste formato, o item com esse descritor se comportou melhor, apesar de a alternativa “B” ser bastante procurada pelos estudantes nas duas fases aplicadas (1EM e 3EM). Avaliando pedagogicamente, aparentemente os estudantes associaram a lei de formação dada, que é o de uma função exponencial, alternando entre as alternativas “A” e “B”, sendo a primeira o gabarito do item. Podemos supor que os estudantes associaram corretamente a lei de formação ao gráfico de uma função exponencial, porém, ainda não reconhecem o critério que a distingue de uma função crescente e decrescente, razão pela qual supostamente as duas primeiras alternativas foram as mais procuradas por estudantes de boa proficiência, que é possível verificar através das bisseriais positivas.

Sob o olhar dos especialistas, o Entrevistado 1 aponta que

Assim como no item anterior, aqui temos duas bisseriais positivas, ou seja, bons respondentes deram como resposta as alternativas A e B. Mas compreendo, neste caso, que tal prática ainda é comum, visto que tais alternativas são similares, uma corresponde a um gráfico crescente e outro decrescente. (ENTREVISTADO 1, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 2 relata que os alunos ficaram em dúvida entre as alternativas “A” e “B”. E quando questionado sobre os distratores, se eles são coerentes, e ainda, se reconhecia alguma prática comum de elaboração desses distratores. Sua resposta foi:

Sim. Geralmente, tenta-se apresentar gráficos de funções mais comuns aos alunos. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O E2 ainda aborda sobre os pré-requisitos que são necessários o aluno desenvolver para consolidar esta habilidade, argumentando que o item

Foi pouco acertado. Além disso, a maioria marcou a opção B e os demais ficaram igualmente divididos entre A, C, D e. Não é um item muito fácil, tem pré-requisitos que o estudante precisaria dominar. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

As análises realizadas pelo Entrevistado 3, possivelmente foi parcialmente prejudicada, a partir de seu relato de dificuldade em visualizar a lei de formação da função dada. Porém, tal análise trouxe aspectos imprescindíveis sobre o item. Em um de seus argumentos, aponta a possibilidade de reflexão do gráfico correto em relação ao eixo das ordenadas, uma sugestão de construção de distrator, que aparentemente é usual na elaboração. Outro ponto destacado por E3 é sobre a possibilidade de uma defasagem aflorada nos estudantes do 1EM persistir no 3EM.

Podemos seguir nesse item uma linha de análise bastante semelhante ao item anterior. Uma possibilidade para a eliminação deste item pode estar relacionada a bisserial positiva da alternativa B, mostrando que "bons" estudantes estão sendo atraídos para ela. Não é possível visualizar a lei de formação apresentada no item, mas imagino que o erro da alternativa B se baseie em algum erro de sinal, considerando a reflexão do gabarito em relação ao eixo das ordenadas. Também já vi em muitos itens esse tipo de construção dos distratores, porém o mais indicado na construção dos distratores é que se baseiem em erros pautados na construção da habilidade avaliada. O comportamento do item nas duas etapas avaliadas mostra um padrão semelhante, mostrando que as defasagens que levaram os "bons" estudantes do 1EM a considerar como correta a alternativa B permaneceram nos estudantes do 3EM.(ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Sob a perspectiva do E4, fatores pedagógicos podem ter influenciado na eliminação do item, conectando parcialmente sua análise com a do E3 em que

[...] novamente temos duas bisseriais positivas. O distrator da alternativa B atraiu mais da metade dos respondentes. Julgo os distratores plausíveis e acredito que essa eliminação tem mais a ver com questões pedagógicas do que técnicas. Demonstra que os estudantes, quando se trata de função exponencial, tentam buscar no gráfico os dados que levam à solução. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O E5 neste item também aborda o percentual de acerto nas duas aplicações (1EM e 3EM), que, de certa forma, esperava-se um melhor desempenho do item no 3EM, dizendo:

Quanto ao histórico, percebo que o percentual de acerto do item praticamente se manteve constante e baixo (~12%), mesmo sendo aplicado em etapas distintas. Observe que é esperado que pelo menos no 3EM esse percentual seja maior do que foi no 1EM. Analisando os distratores, não percebo uma incoerência na construção. O que me chama a atenção é o fato do contradomínio da função ser os reais não

negativos e mesmo assim cerca de 40% (total) dos respondentes marcaram uma das 3 alternativas que isso não acontece (C, D ou E). A alternativa B chamou a atenção dos respondentes, mais do que o gabarito, talvez por apresentar um expoente fracionário os estudantes têm associado ao gráfico decrescente. O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa B também é positiva e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelo distrator B. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Assim, uma das observações mais tratadas neste item foi quanto utilização do distrator da alternativa B, sendo a utilização do gráfico de uma função exponencial decrescente. Sob o olhar dos especialistas, aparentemente o item não possui problemas de elaboração, e que são comuns as práticas adotadas na construção dos distratores utilizados. Supostamente podemos concluir que a eliminação de tal item, pode ter ocorrido por defasagem no processo de ensino aprendizagem adotado na rede de Minas Gerais, por se tratar de uma habilidade não simples do campo da Álgebra e funções. Como podemos ver, foram duas eliminações de itens com o mesmo descritor, nas mesmas etapas avaliadas e em um mesmo período, ou seja, o ano de 2017.

Ribeiro e Cury (2018) alertam sobre as orientações contidas nos Parâmetros Curriculares Nacionais (PCNs) quanto ao ensino de funções e equações, argumentando esse documento, orienta que as funções não sejam estudadas isoladamente, mas inseridas em um contexto de conhecimentos diversificados dentro da própria Matemática ou em outras áreas que podem ser envolvidas.

O critério central é o da contextualização e da interdisciplinaridade, ou seja, é o potencial de um tema permitir conexões entre diversos conceitos matemáticos e entre diferentes formas de pensamento matemático, ou, ainda, a relevância cultural do tema, tanto no que diz respeito às suas aplicações dentro ou fora da Matemática, como à sua importância histórica no desenvolvimento da própria ciência. (BRASIL, 2002, p. 43).

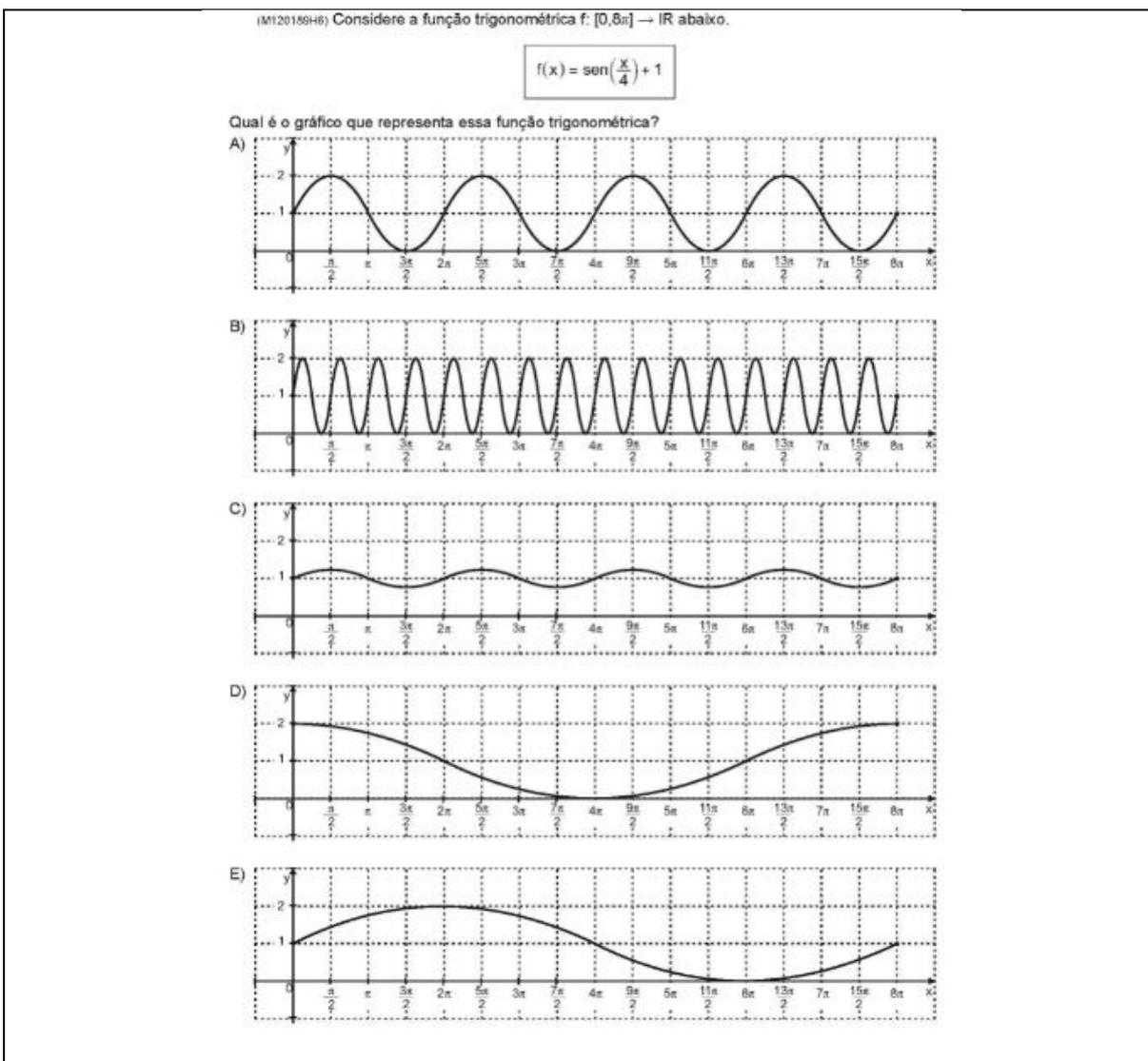
Outro tema que os PCNs abordam sobre a aprendizagem da Matemática é no desenvolvimento das habilidades que envolvem os conhecimentos sobre a trigonometria, orientando que seu estudo

[...]esteja ligado às aplicações, evitando-se o investimento excessivo no cálculo algébrico das identidades e equações para enfatizar os aspectos importantes das funções trigonométricas e da análise de seus gráficos. (BRASIL, 2002, p. 43).

No entanto, nesse estudo, um dos itens eliminados por motivo de ajuste que se enquadrou dentro dos critérios estabelecidos, envolve as funções trigonométricas, conforme se vê no quadro 7 a seguir.

Quadro 07 – Item M120189H6

Observações: Item aplicado no PROEB de 2016 eliminado no 3EM por motivo de ajuste.											
Habilidade avaliada: D75 – Corresponder uma função trigonométrica a seu gráfico.											
Dados estatísticos na(s) ocorrência(s) com eliminação:											
Código: M120189H6			Descritor: D75			Gabarito: E					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,138	0,191		0,003	0,213	0,182	0,208	0,257	0,138
Obs.:	Ajuste						Biss. A	Biss. B	Biss. C	Biss. D	Biss. E
							0,100	-0,088	-0,068	-0,084	0,191
Imagem do Item:											



Fonte: CAEd/UFJF.

Nesse descritor o estudante deve corresponder uma função trigonométrica (dada) ao seu gráfico. Nessa situação, o gabarito foi a alternativa menos procurada, apesar da biserial apresentar um percentual positivo. No entanto, ainda através da biserial, bons estudantes foram atraídos pela alternativa “A”, e através do percentual de procura das alternativas, a opção “D” foi a que apresentou o maior. No contexto da elaboração, esse descritor é abordado de forma bem direta, incluindo como alternativas, gráficos de diversas funções trigonométricas. Porém, é comum que se use diversificações da mesma função pedida, como, por exemplo, a função principal (seno ou cosseno), diferenciando o gráfico em detalhes.

Assim, o E1 salienta que os motivos de eliminação desse item, podem estar relacionados aos mesmos motivos dos itens anteriores, tendo como foco as bisseriais das alternativas.

O E2 já atenta quanto ao percentual de acerto do item, dizendo que:

Os alunos não tiveram muito critério, marcaram, quase de maneira igual uma alternativa. Os distratores são coerentes. Geralmente, procura-se pelas funções parecidas a que é apresentada no suporte. Item pouco acertado. Os alunos menos marcaram o gabarito. Mas, de modo geral, foi distribuída a marcação entre as alternativas, sem muito critério, ou considerando apenas um. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Já o E3, aponta que:

Da mesma forma que nos itens anteriores, esse também apresenta duas alternativas com bisseriais positivas, mostrando que o distrator A chamou a atenção dos "bons" estudantes. Muito complicado ficar tentando descobrir em que erro foi baseado um distrator, mas claramente o distrator da letra A chamou a atenção de muitos "bons" estudantes, o que não deveria acontecer. O item apresentou um comportamento inadequado. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

As observações levantadas pelo E4 também são pautadas no comportamento retilíneo das alternativas, dizendo que:

Os alunos se distribuíram entre as alternativas. Novamente existem duas bisseriais positivas e um distrator chamando mais atenção do que o gabarito. Acredito mais uma vez que as razões sejam mais técnicas do que pedagógicas uma vez que os estudantes escolhem marcar a alternativa que lembra a função seno padrão, demonstrando que talvez não saibam lidar com as composições das funções trigonométricas. Julgo que os distratores estão de acordo com a habilidade avaliada. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Enquanto o E5 apresenta uma possibilidade de criação de distrator, chamando a atenção para os mesmos, dizendo que:

Em linhas gerais, percebo que o percentual de acerto do item é baixo (~13%), a bisserial é menor do que 0,3 e que houve uma distribuição "homogênea" na marcação das alternativas, e isso pode ter ocasionado a eliminação do item. Analisando os distratores, todos são compostos por funções trigonométricas que de fato podem causar dúvida, talvez por isso apresentem percentual de marcação semelhante. Uma prática comum, por exemplo, é abordar a função sem suas variações, como $f(x)=\text{sen}(x)+1$ apresentada na alternativa A.

O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse

caso, a bisserial da alternativa A também é positiva e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelo distrator A. (ENTREVISTADO 5, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Os PCNs (2002, p. 44) evidenciam que o ensino de funções trigonométricas não deva ser exaustivos em cálculos, de maneira que sejam pautados na resolução de problemas que envolvam medições, em especial, o cálculo de distâncias inacessíveis. Essa sugestão é pensada de maneira que nem todo estudante prosseguirá seus estudos na área das ciências exatas.

Na BNCC do ensino médio, a competência 4 específica em seu tópico que o aluno deve

[...] compreender e utilizar, com flexibilidade e fluidez, diferentes registros de representação matemáticos (algébrico, geométrico, estatístico, computacional etc.), na busca de solução e comunicação de resultados de problemas, de modo a favorecer a construção e o desenvolvimento do raciocínio matemático. (BRASIL, 2017, p. 530).

Porém, consultando as habilidades que devem ser desenvolvidas pelo aluno nessa competência, identificou-se a habilidade EM13MAT404, que orienta

[...] identificar as características fundamentais das funções seno e cosseno (periodicidade, domínio, imagem), por meio da comparação das representações em ciclos trigonométricos e em planos cartesianos, com ou sem apoio de tecnologias digitais. (BRASIL, 2017, p. 531).

Assim, é possível suspeitar que o descritor D75 (Corresponder uma função trigonométrica a seu gráfico) da matriz de referência do PROEB, pode estar se tornando um complicador dentro do universo avaliativo, o que requer atenção não somente da elaboração, mas dos propósitos que se deseja avaliar com esse descritor da matriz de referência.

Nas eliminações que tiveram como motivo a bisserial, os três itens selecionados segundo o critério estipulado nesse trabalho, foram utilizados no PROEB nos anos de 2015 e 2016. Nos quadros 08, 09 e 10 colocaremos em discussão tais itens.

O primeiro se trata de um item que envolve as equações de segundo grau, em que o estudante deve modelar e encontrar a solução de uma equação a partir de uma

situação contextualizada, de modo que esta solução satisfaça as condições do contexto.

Quadro 08 – Item M090047H6

Observações: Item aplicado no PROEB de 2016 eliminado no 9EF por bisserial.											
Habilidade avaliada: D64 – Utilizar equação polinomial de 2º grau na resolução de problema.											
Dados estatísticos na(s) ocorrência(s) com eliminação:											
Código: M090047H6			Descritor: D64			Gabarito: B					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,137	0,041		0,005	0,538	0,137	0,206	0,113	
Obs.:						Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
						0,305	0,041	-0,292	-0,226		
Imagem do Item:											
<p>(M090047H6) Um vendedor viajou a trabalho para visitar um cliente em outra cidade. Em um determinado ponto da estrada, ele fez uma parada para fazer algumas ligações e depois seguiu sua viagem. A distância percorrida por ele dessa parada até seu destino foi igual à metade do quadrado da distância percorrida da origem da viagem até o momento da parada. A distância total percorrida pelo vendedor até seu destino foi de 364 quilômetros.</p> <p>Quantos quilômetros esse vendedor percorreu após a parada até chegar ao seu destino?</p> <p>A) 182 B) 338 C) 364 D) 728</p>											

Fonte: CAEd/UFJF.

Itens envolvendo equações, principalmente as de segundo grau, tecnicamente podem ser considerados difíceis de elaborar, devido às possibilidades de contextualização. Na maioria dos livros didáticos, problemas envolvendo equações do segundo grau, apresentam contextos que necessitam de outras habilidades na Matemática, principalmente a geometria.

Os contextos mais utilizados nos livros didáticos envolvem o conceito de área, fazendo com que nas situações problema, seja explorada mais de uma habilidade no estudante, o que foge a realidade de item em uma avaliação em larga escala. Ou seja, estaríamos avaliando dois conteúdos ao mesmo tempo em um único item, que precisa ser construído pautado em um único descritor da Matriz de Referência.

A BNCC orienta que

[...] a Geometria não pode ficar reduzida a mera aplicação de fórmulas de cálculo de área e de volume nem a aplicações numéricas imediatas de teoremas sobre relações de proporcionalidade em situações relativas a feixes de retas paralelas cortadas por retas secantes ou do teorema de Pitágoras. A equivalência de áreas, por exemplo, já

praticada há milhares de anos pelos mesopotâmios e gregos antigos sem utilizar fórmulas, permite transformar qualquer região poligonal plana em um quadrado com mesma área (é o que os gregos chamavam “fazer a quadratura de uma figura”). Isso permite, inclusive, resolver geometricamente problemas que podem ser traduzidos por uma equação do 2.º grau. (BRASIL, 2017, p. 272).

Este fato talvez possa ser um dos fatores que justifique a tenacidade dos livros didáticos em explorar áreas de figuras geométricas na resolução de problemas envolvendo equações do 2.º grau.

Dessa forma, a busca por distratores no processo de elaboração se torna uma tarefa ainda mais difícil, fazendo com que a busca por itens com bons parâmetros envolvendo este assunto, se torne muito inusitada. É o caso do item contido no quadro 08, em que é perceptível que seus distratores não foram muito bem pautados na habilidade avaliada, mas com operações envolvendo a manipulação de dados contidos no enunciado.

O distrator da alternativa “A”, por exemplo, foi procurado por mais da metade dos estudantes com boa proficiência, sendo que sua justificativa pode estar associada a apenas uma divisão por 2 da distância total percorrida de 364 quilômetros informada no enunciado do item. As alternativas “C” e “D”, possivelmente podem ser justificadas, respectivamente pelo próprio valor da quilometragem dada no enunciado e a sua multiplicação por 2.

Considerando essas justificativas, nenhuma habilidade envolvendo as equações de segundo grau, foram utilizadas nos distratores, apenas manipulações aritméticas com os próprios dados informados.

Sob a perspectiva dos especialistas da equipe de Matemática, este fato chamou a atenção de apenas um dos entrevistados.

Para o Entrevistado 1:

Este item apresenta duas bisseriais positivas. Além disso, trata-se de um item de equação do segundo grau, em que o estudante deve modelar uma equação para solucioná-lo, o que não é algo simples. Pelo que vejo dos itens eliminados, é comum tal habilidade estar presente, sobretudo nesses casos que exige do estudante modelar algebricamente o problema. (ENTREVISTADO 1, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 2 condena o tamanho do enunciado e está de acordo com a estrutura dos distratores, como podemos observar em sua resposta sobre os fatores que podem ter influenciado na eliminação do item.

O enunciado ser um pouco extenso, com informações que exigiam a algebrização por parte do aluno. Os distratores são coerentes. O estudante apenas manipular os dados do problema ou confundir quadrado com raiz, e ainda, quadrado com o dobro ou metade. Os alunos, possivelmente, não montaram ou não souberam montar a equação e resolvê-la. Marcaram mais as alternativas que se referiam à metade, que era uma das informações do problema, e um dos dados informados. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 3 chama atenção quanto ao distrator da alternativa A através do comportamento da bisserial, dizendo que:

Da mesma forma que nos itens anteriores esse também apresenta duas alternativas com bisseriais positivas, porém nesse caso a maior bisserial não é a do gabarito, ou seja, essa alternativa se comportou como gabarito. A escolha dos "bons" estudantes pelo distrator A pode indicar uma escolha massiva da primeira alternativa, ou então uma grande defasagem na habilidade avaliada, visto que o estudante que marca como correta a metade da distância total percorrida mostra um desentendimento do problema sugerido, se apegando a apenas algumas informações apresentadas. O que costuma ser comum na construção de distratores em itens que também avaliam a resolução de problemas. O item apresentou um comportamento inadequado. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 4 também chama atenção quanto ao distrator da alternativa A, informando que:

Nesse caso houve um distrator que chamou atenção de mais da metade dos estudantes, se tornando o segundo com bisserial positiva esse distrator atraiu estudantes que não equacionaram o problema. Apenas dividiram a distância por dois. Acredito que esse item também foi condenado pela dificuldade na modelagem. Creio que o distrator seja plausível e que o problema não seja muito técnico uma vez que o gabarito não foi nem o segundo mais marcado. Caso ele não estivesse lá, parece que os estudantes optariam pelo 364. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 5 também considera, assim como o Entrevistado 1, que itens envolvendo a modelagem de equações do segundo grau, por parte do estudante, seja uma habilidade difícil na avaliação.

Acredito que esse item tem um baixo percentual de acerto pelo fato de abordar uma modelagem de equação de 2.º grau, e por exigir do respondente uma boa manipulação algébrica, já que envolve raiz quadrada de 2916. A eliminação desse item deve estar associada a uma baixa discriminação, uma vez que a bisserial está muito baixa, quase nula. Analisando os distratores, percebo que eles poderiam explorar outras possibilidades de erro, que envolvessem erro na modelagem ou até mesmo na interpretação da situação-problema, não apenas uma mera manipulação dos dados do problema. O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa A também é positiva e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelo distrator A. E isso ao me ver é um problema, até mesmo pedagógico, uma vez que, possivelmente, os respondentes apenas efetuaram a divisão de 364 por 2. (ENTREVISTADO 5, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

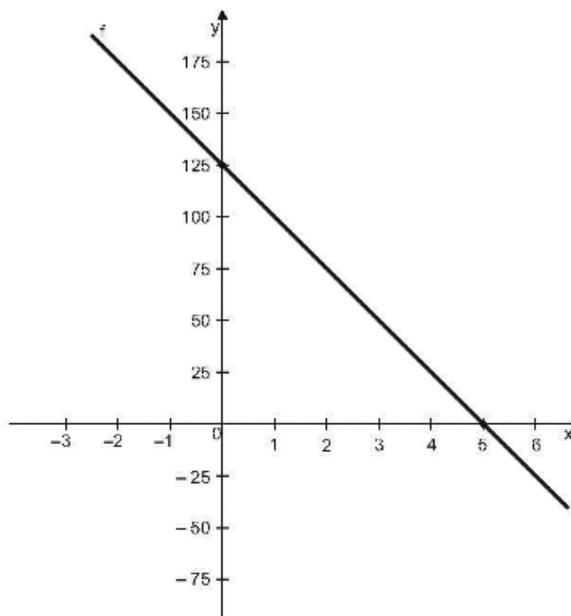
De acordo com as análises dos especialistas, foi claro que todos se voltaram para alternativa A, e que poucos se reportaram sobre os distratores não estarem associados diretamente com o propósito do descritor. Como dito, itens dessa habilidade são complexos, tanto para o elaborador quanto para o estudante, fazendo jus a reflexões e experimentações, quanto ao seu modo de construção e aplicação.

Ainda sobre as eliminações pela bisserial, o item do quadro 09 trata de uma habilidade que exige do estudante associar uma função polinomial do 1.º grau a seu gráfico.

Quadro 09 – Item M100093H6

Observações: Item aplicado no PROEB de 2016 eliminado no 3EM por bisserial.											
Habilidade avaliada: D67 – Corresponder uma função polinomial do 1º grau a seu gráfico.											
Dados estatísticos na(s) ocorrência(s) com eliminação:											
Código: M100093H6			Descritor: D67			Gabarito: B					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,090	0,115		0,003	0,204	0,090	0,183	0,363	0,157
Obs.:	Eliminação: Bisserial					Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
	D67_ Corresponder uma função polinomial do 1º grau a seu gráfico.					-0,283	0,115	-0,059	0,155	0,088	
Anos_Etapas: 2016_3EM											
Outras utilizações no PROEB com aprovação: Não consta											
Imagem do Item:											

(M100093H6) O gráfico abaixo representa uma função polinomial do 1º grau $f: \mathbb{R} \rightarrow \mathbb{R}$.



Qual é a lei de formação dessa função?

- A) $f(x) = -125x + 5$
- B) $f(x) = -25x + 125$
- C) $f(x) = -5x + 125$
- D) $f(x) = 5x - 125$
- E) $f(x) = 25x + 125$

Fonte: CAEd/UFJF.

O verbo corresponder contido no texto do descritor D67 (D67 - Corresponder uma função polinomial do 1.º grau a seu gráfico) pode ser explorando em ambos os sentidos. Neste caso o elaborador optou por apresentar primeiramente o gráfico exigindo do estudante a capacidade de associá-lo a sua lei de formação, sendo que a ideia inversa também é válida para esse tipo de situação.

Isso faz com que essa habilidade não seja muito complexa no processo de elaboração, devido à gama de possibilidades que podem ser exploradas, e ainda, pelo fato de não exigir contextualização.

Considerando o processo de ensino aprendizagem, estudos como o de Tenório et al. (2014), apontam que o estudante, nesse conteúdo busca métodos práticos para identificar os gráficos através dos coeficientes angular e linear de uma função polinomial do 1.º grau.

Assim, o elaborador geralmente recorre a esse tipo de comportamento do aluno para explorar a confecção dos distratores, apresentando variações dos coeficientes

dentre as alternativas, como pode ser observado no item M100093H6. Esse tipo de estrutura proporcionou três bisseriais positivas no item (a do próprio gabarito e mais duas alternativas), e ainda, fazendo com que o gabarito seja a alternativa menos procurada pelos estudantes, gerando um percentual de acerto em torno de aproximadamente 9%.

Alguns desses fatores chamou a atenção dos especialistas que analisaram o item, como, por exemplo o Entrevistado 1, que diz:

Note que tal item apresenta um baixo percentual de acerto, além disso duas alternativas apresentam bisseriais iguais e positivas, o que indica que bons respondentes se sentiram atraídos pela alternativa A. (ENTREVISTADO 1, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Já o Entrevistado 2, comenta que os fatores que podem ter influenciado a eliminação do item pode ser o desconhecimento em relação à habilidade. Sobre os distratores alega que

Sim, são coerentes. Geralmente, apresenta-se expressões algébricas com os dados do gráfico e que tenham algum ponto que satisfaça ao gráfico. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Em sua análise, ele finaliza dizendo que

A alternativa que eles menos marcaram foi o gabarito. Optaram mais pelas alternativas que tinham algum número diretamente relacionado aos que apareceram no gráfico. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul.2020)

O Entrevistado 3, aponta que o item teve um comportamento inadequado, conforme sua análise, que diz:

O item apresenta 3 bisseriais positivas, sendo uma delas maior que a do gabarito. O distrator mais marcado e comumente utilizado é aquele em que o estudante utiliza os pontos de interseção da função com os eixos, como coeficientes, salvo algumas alterações de sinal. O outro distrator que apresentou uma bisserial positiva se pautou em um erro de sinal, considerado para a análise de crescimento/decrescimento da função, também muito utilizado. O item apresentou comportamento inadequado. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Sob a perspectiva do Entrevistado 4, o item está de acordo, apresentando um breve relato sobre possibilidades de resolução por parte do estudante, dizendo:

Considero o item tecnicamente bom. Os distratores estão de acordo, porém, o gabarito foi o menos marcado, gerando três bisseriais positivas. Acredito que, mais uma vez, o item entrega que os estudantes procuram a resolução nos dados do item, demonstrando que, de fato estão cometendo erros dentro da habilidade. Uma pequena minoria está a par do procedimento para a obtenção da resposta. Em termos de técnica de elaboração, não consigo ver como esse item poderia ser diferente do apresentado. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Por último, o Entrevistado 5 aponta a possibilidade de existir algum problema pedagógico no item, ou mesmo com a população avaliada, dizendo:

Esse item apresentou outras duas bisseriais positivas além do gabarito. Sendo assim, ou o item tem de fato um problema pedagógico, ou os estudantes desconhecem de fato a habilidade. Analisando os distratores, percebo que eles exploram manipulações de sinais, mas me chamou a atenção o distrator D ter sido muito marcado, cerca de 36%. Isso revela claramente que muitos estudantes não sabem sequer associar o gráfico de uma função do 1.º grau decrescente ao sinal do coeficiente angular. Acho que é comum abordar essas manipulações de sinais e considerar o coeficiente como sendo a abscissa e a ordenada dos pontos de interseção com o eixo b.

O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, as bisseriais das alternativas D e E também foram positivas e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelos distratores D e E. (ENTREVISTADO 5, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Por meio dos relatos e do relatório estatístico do item, foi possível verificar seu baixo desempenho na avaliação, razões pelo qual foi descartado em 2016. Vale ressaltar sobre a estrutura adotada na elaboração desse item, que boa parte dos especialistas considerou seus distratores coerentes, sendo que praticamente em sua totalidade, foram utilizados artifícios como os pontos de interseção da reta com os eixos no gráfico, os coeficientes, e ainda o decrescimento da reta.

Através deste item, pode-se perceber que as estratégias utilizadas em sua construção, podem ter sido fatores que influenciaram em sua eliminação, apesar de que, não podemos desconsiderar a possibilidade de total desconhecimento do

assunto pela população avaliada. Dessa forma, estamos de frente a uma habilidade que requer maior atenção e discussão em todos os aspectos.

O terceiro e último item eliminado pela bisserial está apresentado no quadro 10 a seguir. Nele, a habilidade estimula o estudante a corresponder um polinômio fatorado as suas raízes, conforme especifica o descritor D71 (D71 – Corresponder um polinômio fatorado por meio de polinômios de 1º grau as suas raízes).

Quadro 10 – Item M1D26I0117

Observações: Item aplicado no PROEB de 2015, sendo eliminado no 1EM por bisserial e aprovado no 3EM.											
Habilidade avaliada: D71 – Corresponder um polinômio fatorado por meio de polinômios de 1º grau as suas raízes.											
Dados estatísticos na(s) ocorrência(s) com eliminação:											
Código: M1D6I0117			Descritor: D71			Gabarito: E					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,116	0,120		0,005	0,266	0,158	0,172	0,283	0,116
Obs.:	Bisserial					Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
						-0,074	-0,054	-0,113	0,135	0,120	
Dados estatísticos na(s) ocorrência(s) sem eliminação:											
Código: M1D6I0117			Descritor: D71			Gabarito: E					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
0,02894	420,5595	0,09734	0,115	0,264	429,844	0,004	0,307	0,149	0,150	0,275	0,115
Obs.:	Ocorrência sem eliminação no 3EM					Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
						-0,073	-0,101	-0,121	0,085	0,264	
Imagem do Item:											
(M1D26I0117) As raízes de um polinômio $p(x)$ são, 0, 1 e 2.											
Uma das expressões que representa esse polinômio na forma fatorada é											
A) $p(x) = (x - 1)(x + 2)$											
B) $p(x) = (x - 2)(x + 1)$											
C) $p(x) = 4(x + 2)(x - 1)$											
D) $p(x) = x(x + 2)(x + 1)$											
E) $p(x) = x(x - 2)(x - 1)$											

Fonte: CAEd/UFJF.

É interessante observar que este item foi aplicado em duas etapas no ano de 2015, sendo aprovado na aplicação ao 3EM, mesmo contendo duas bisserias positivas. Este é um típico item que pode nos transparecer quanto ao comportamento bisserial. Observe que no 1EM, os percentuais de marcação das alternativas se mantêm bastante semelhantes no 3EM, porém, a bisserial diversifica a população avaliada que opta nas alternativas. Isto é, na aplicação do 1EM, o distrator da

alternativa “D”, se tornou um complicador, enquanto no 3EM, o mesmo distrator se manteve com bisserial positiva, porém, em menor percentual. Isso fez com que os estudantes de melhor proficiência reduzissem em aproximadamente 5% no 3EM).

O Entrevistado 1, chama atenção quanto a possibilidade de a eliminação ter se dado por motivo de comportamento diferencial (DIF), apesar de, estatisticamente, a eliminação ter sido configurada por motivo bisserial.

Era bom para analisarmos este item conhecer a ancoragem do item em sua primeira aplicação, pois este pode ser um caso em que o item tenha sido eliminado pelo seu comportamento diferencial (DIF), visto que as bisseriais entre as duas aplicações se diferem um pouco. As alternativas D e E, que apresentaram bisseriais positivas atraíram bons respondentes, e o distrator D é o mais comum de utilizarmos nestes itens, usados até hoje em nossas elaborações, quando o estudante confunde os sinais das raízes na forma fatorada de um polinômio. (ENTREVISTADO 1, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Já o Entrevistado 2, argumenta que o fator que causou a eliminação do item possa ser o desconhecimento da habilidade envolvida por parte dos alunos. Analisando as alternativas, afirma serem coerentes, e que é uma comum de elaboração alternar entre os sinais das raízes para montagem dos distratores. Ele finaliza sua análise dizendo que se trata de

um item relativamente fácil, mas que os estudantes marcaram sem nenhum critério as alternativas. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Assim como o Entrevistado 1, os demais entrevistados também se pautaram sobre as duas aplicações do item, como pode ser vista na fala do Entrevistado 3.

No 1.º quadro a discrepância entre as bisseriais positivas é menor que no segundo, o que mostra que no 1EM o distrator se comportou como gabarito e no 3EM o distrator apenas chamou mais atenção que os demais distratores. O distrator que gerou essas discrepâncias de bisserial mostra que o estudante que o marcou entende que na forma fatorada a raiz do polinômio é o valor que acompanha x . Distrator também muito comum. O item apresentou além do comportamento inadequado da bisserial no primeiro quadro, também apresentou um comportamento diferente em grupos diferentes, o que também não deveria acontecer. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul.2020).

O Entrevistado 4, argumenta que a bisserial do distrator da alternativa "D" superou a do gabarito na aplicação em que não foi eliminado (3EM). Tal fato pode se conectar com a fala do entrevistado 1, que trouxe uma observação sobre a construção desse distrator, dizendo que é o mais comum nas práticas de elaboração.

Nesse item, quando eliminado, parece que a bisserial da alternativa D superou a do gabarito. As estatísticas de quando não foi eliminado demonstram que a habilidade torna o item de alto nível na escala, ou seja, muito difícil para os alunos. A maioria desconsidera o zero como raiz e outros apesar de considerar o zero como raiz, troca o sinal dos fatores. Não vejo problemas técnicos com os distratores. Trata-se de uma habilidade bem direta. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul.2020).

O Entrevistado 5, cogita a possibilidade de haver algum problema pedagógico, observando também, o comportamento do distrator da alternativa "D", dizendo que

Esse item apresentou no 1EM outra bisserial positiva além do gabarito e igual a 0,120. Sendo assim, ou o item tem de fato um problema pedagógico, ou os estudantes desconhecem de fato a habilidade. Analisando os distratores, percebo que eles exploram manipulações de sinais, relacionados às raízes do polinômio. O distrator D chamou a atenção dos respondentes por apresentar apenas os valores do enunciado. A presença desse distrator e o seu percentual de marcação mostram que os estudantes não conhecem essa "tarefa"/"habilidade". É usual a presença desse distrator na elaboração. O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa D também foi positiva e a bisserial do gabarito, embora não tenha sido maior que 0,3, foi próximo e portanto, "aceitável". (ENTREVISTADO 5, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Nesse item, é possível observar a partir das análises dos especialistas, que é comum adotar como estratégia de elaboração de distratores alternar os sinais das raízes nos polinômios dados como alternativas. De fato é uma habilidade sem muitas dificuldades, bem direta e simples, e que como esperado, apresentou um melhor desempenho no 3EM. Esse restabelecimento do item no 3EM, supostamente pode funcionar como um sinalizador de que a rede tenha apresentado avanços no ensino, e que de certa forma, esse item tenha atuado como um indicador de progresso.

Por motivo de DIF, os três itens que serão analisados, foram utilizados no PROEB dos anos de 2016, 2017 e 2018. O primeiro item avalia uma das habilidades

matemáticas mais utilizadas em nosso cotidiano que é a porcentagem. O descritor D50 (D50 – Utilizar porcentagem na resolução de problemas) tem como propósito explorar as capacidades de efetuar cálculos utilizando porcentagem, em situações em que um cenário (situação) é proposto ao estudante.

Quase sempre os elaboradores optam por construir situações envolvendo descontos ou acréscimos em percentuais dando inicialmente um valor de referência que deverá ser trabalhado com o valor percentual que é informado ou solicitado. A situação problema apresentada no quadro 11 a seguir, se trata de um item aplicado no PROEB nas etapas do 5EF e 9EF, nos anos de 2016 e 2018.

Quadro 11 – Item M040046H6

Observações: Item aplicado no PROEB nos anos de 2016 e 2018 para o 5EF e 9EF. Apresentou apenas uma eliminação em 2018 no 9EF por DIF.											
Habilidade avaliada: D50 – Utilizar porcentagem na resolução de problema.											
Dados estatísticos na(s) ocorrência(s) com eliminação:											
Código: M040046H6			Descritor: D50			Gabarito: C					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,346	0,560		0,004	0,043	0,526	0,346	0,081	
Obs.:	Eliminado DIF – 9EF (2018)					Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
							-0,420	-0,328	0,560	-0,226	
Dados estatísticos na(s) ocorrência(s) sem eliminação:											
Código: M040046H6			Descritor: D50			Gabarito: C					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial	318,99740	B/N	A	B	C	D	E
0,03056	311,2	0,12672	0,164	0,436		0,012	0,061	0,572	0,164	0,190	
Obs.:	5EF (2016)					Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
							-0,302	0,039	0,436	-0,296	
Código: M040046H6			Descritor: D50			Gabarito: C					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial	318,99740	B/N	A	B	C	D	E
0,03056	311,2	0,12672	0,332	0,570		0,005	0,045	0,546	0,332	0,072	
Obs.:	9EF (2016)					Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
							-0,399	-0,332	0,570	-0,245	
Código: M040046H6			Descritor: D50			Gabarito: C					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial	325,2	B/N	A	B	C	D	E
0,03205	317,6	0,11854	0,168	0,363		0,009	0,067	0,586	0,168	0,170	
Obs.:	5EF (2018)					Biss. A	Biss. B	Biss. C	Biss. D	Biss. E	
							-0,355	0,092	0,363	-0,279	
Imagem do Item:											
Carlos comprou um ventilador que estava anunciado por R\$ 85,00. Ele pagou à vista e recebeu um desconto de 10% nessa compra.											
Qual foi o valor que Carlos pagou por esse ventilador?											
A) R\$ 10,00											
B) R\$ 75,00											
C) R\$ 76,50											
D) R\$ 85,00											

Fonte: Elaborado pelo autor.

A situação problema propõe ao estudante determinar o valor final de um produto, após ser submetido a um desconto de 10% sob seu valor original. A estratégia adotada nos distratores, aparentemente não envolve erros de cálculos ancorados na própria porcentagem, mas em erros pautados em subtração de valores e interpretações equivocadas por parte do estudante.

Supostamente a alternativa “A” é o próprio percentual informado, enquanto a alternativa “B” é o resultado da subtração do percentual informado do valor original do produto ($85 - 10 = 75$). Já o distrator da alternativa “D”, é o próprio valor do produto, o que podemos pressupor que seja uma alternativa procurada por estudantes que apenas efetuaram a leitura do item e buscaram pelo valor do produto, sem qualquer tipo de compreensão.

Claro que não podemos descartar que isso possa acontecer, porém, é esperado que os distratores envolvam algum tipo de erro dentro da própria habilidade avaliada, e não apenas uma repetição. Evidente que a experiência em sala de aula pode auxiliar e muito no momento da elaboração, principalmente dos distratores.

Sobre o ensino de porcentagem, a BNCC aponta que

[...] para que aprofundem a noção de número, é importante colocá-los diante de problemas, sobretudo os geométricos, nos quais os números racionais não são suficientes para resolvê-los, de modo que eles reconheçam a necessidade de outros números: os irracionais. Os alunos devem dominar também o cálculo de porcentagem, porcentagem de porcentagem, juros, descontos e acréscimos, incluindo o uso de tecnologias digitais. (BRASIL, 2017, p. 269).

Estudos como o realizado por Lopes (2013), em que houve a investigação através da produção de significados sobre a aprendizagem da porcentagem, podem auxiliar na pesquisa sobre o que os estudantes constroem cognitivamente durante a execução de uma tarefa.

Lopes (2013) constatou que alguns estudantes, em sua pesquisa, tentam resolver porcentagens através de regras mecânicas, não demonstrando entender as operações que estão envolvidas, mas sempre as mantendo na sua lógica, fato que pudesse tornar válido, sob a perspectiva de item, os distratores do item no quadro 10.

Outro aspecto que é preciso observar no histórico de aplicações, é o fato da única eliminação do item ter se dado no ano mais recente e a na etapa concluinte do ensino fundamental (9EF).

O distrator " B", foi salientado em unanimidade pelos entrevistados, alegando ser uma prática comum dos estudantes a subtração do percentual no valor informado. Assim, o Entrevistado 1 diz que

Analisando os parâmetros A, B e C do item nota-se que eles são próximos, então tal item não deve ter sido eliminado pelo critério do DIF. Um fator que pode ter sido o da eliminação deste item corresponde ao fato de em algumas aplicações duas alternativas apresentarem bisseriais positivas, o que indica que alguma alternativa possa ter chamado mais a atenção do que outras. Neste caso trata-se da alternativa B, que é muito comum de ser utilizada quando o aluno desconhece o significado do símbolo "%". (ENTREVISTADO 1, Questionário de Pesquisa Acadêmica, 31 jul.2020).

O Entrevistado 2 faz uma análise mais sucinta sobre o item, justificando a eliminação do item dizendo que os estudantes não conhecem o conceito de porcentagem. Julga que os distratores são coerentes, dizendo que é comum apresentar ao estudante alternativas em que confunde desconto com acréscimo e porcentagem com o valor absoluto. O E2 finaliza sua análise dizendo que é

Um item relativamente fácil, mas que mostrou que os estudantes, a maioria deles, procurou pela resposta do desconto absoluto, não reconhecendo o conceito de porcentagem. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Mesmo o motivo de eliminação sendo o de comportamento diferencial (DIF), a bisserial das alternativas ainda é o dispositivo de informação do item mais procurado para análise dos entrevistados. Porém, o Entrevistado 3 buscou relações do comportamento do item nas aplicações apresentadas, dizendo que

O item apresenta comportamento semelhante em mesma etapa, porém, diferente em etapas diferentes. No 5EF vemos uma maior atração dos estudantes "bons" para a alternativa B. Esse comportamento diferencial, evidencia um distrator também muito usual, aquele em que o estudante soma ou subtrai, de acordo com o contexto, a porcentagem ao invés de fazer o cálculo da mesma. O item apresentou além do comportamento inadequado da bisserial no 5EF, também apresentou um comportamento diferente em grupos diferentes, o que também não deveria acontecer. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul.2020).

O mesmo ocorreu na análise do Entrevistado 4, em que a bisserial também foi um dos focos de análise para compreensão do comportamento do item no teste, concluindo que existe um desconhecimento do assunto por parte dos estudantes, e não foi possível identificar fatores que possam ter contribuído para a eliminação, dizendo:

Vejo que mais da metade dos estudantes escolheu a alternativa B demonstrando não estar a par do cálculo das porcentagens, mas não consegui identificar o porquê o item teria sido eliminado. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Entendemos que em uma eliminação por DIF, que se trata do comportamento diferencial do item em diferentes grupos, a análise pedagógica seja realizada utilizando-se o histórico de aplicação do item (aprovadas e eliminadas) no mesmo grupo, visando entender o comportamento diferencial.

Normalmente, quando se têm vários grupos diferentes, pode-se escolher um dos grupos como sendo o grupo de referência, por alguma razão em particular, e realizar a análise comparativa do comportamento do item nos demais grupos em relação ao comportamento neste grupo. Mas também é possível que se deseje realizar a análise comparativa entre todos os grupos. (SOARES et al., 2005, p. 484).

O Entrevistado 5, retoma quanto a preocupação no histórico do item, alegando não compreender a eliminação. Ele ainda chama a atenção quanto a estrutura de apresentação das alternativas, pois, somente uma delas foi composta de um valor com centavos, o que poderia se tornar um atrativo para os estudantes.

Não compreendi a eliminação do item, analisando apenas o histórico apresentado, sem os parâmetros pela TRI, uma vez que ele se comportou de forma semelhante em 2016 (9EF). Talvez fosse necessário apresentar os escores dos diferentes grupos que compõe o teste. Analisando os distratores, o distrator B chamou a minha atenção, pois mais de 50% dos respondentes consideram que um desconto de 10% é de R\$ 10,00. Os distratores são de naturezas diferentes e plausíveis. Talvez, para compor, e para o próprio gabarito não chamar a atenção, o que não aconteceu, apresentaria outra alternativa com centavos. É usual a presença desses distratores na elaboração. O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa B também foi positiva e a bisserial do gabarito, embora não tenha sido maior que

0,3, foi próximo e portanto "aceitável" (9EF de 2016). (ENTREVISTADO 5, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Foi possível verificar a partir das análises realizadas nesse item que os elaboradores concordam com a estrutura do item, e que a eliminação se dá apenas pelo desconhecimento do estudante. Também é possível observar essas análises, não somente no item do quadro 10, mas extensivo aos anteriores, são baseadas sobre as bisseriais. Não são levantadas outras suspeitas pelos entrevistados, que possam estar contribuindo para a eliminação. Apesar de os motivos de eliminação serem informados, eles não foram relevantes em nenhuma das análises dos especialistas até o momento.

Dando continuidade, assim como os dois primeiros itens dessa análise (Quadro 4 e Quadro 5) eram do mesmo descritor, o próximo (Quadro 11) tem uma relação muito grande com o item anterior (Quadro 10). Apesar de não pertencerem ao mesmo descritor, são habilidades que se passam em um mesmo conteúdo em sala de aula, quando é apresentado o conjunto dos números racionais.

Dentre os critérios recomendados para avaliação nos PCNs um deles é utilizar os diferentes significados e representações dos números naturais, inteiros, racionais e das operações envolvendo esses números, para resolver problemas, em contextos sociais, matemáticos ou de outras áreas do conhecimento. (PCN, 1998, p. 76)

Por meio deste critério o professor verifica se o aluno é capaz de comparar e ordenar números naturais, inteiros e racionais; reconhecendo suas diferentes formas de expressão como fracionária, decimal e percentual; representar na forma decimal um número racional expresso em notação fracionária; efetuar cálculos envolvendo adição, subtração, multiplicação, divisão e potenciação; escolher adequadamente os procedimentos de cálculo (exato ou aproximado, mental ou escrito) em função dos contextos dos problemas, dos números e das operações envolvidas. (PCN, 1998, p. 76).

Já na BNCC,

A expectativa em relação a essa temática é que os alunos resolvam problemas com números naturais e números racionais cuja representação decimal é finita, envolvendo diferentes significados das operações, argumentem e justifiquem os procedimentos utilizados para a resolução e avaliem a plausibilidade dos resultados encontrados. No tocante aos cálculos, espera-se que os alunos desenvolvam diferentes estratégias para a obtenção dos resultados,

sobretudo por estimativa e cálculo mental, além de algoritmos e uso de calculadoras. (BRASIL, 2017, p. 268).

Os PCNs apontam que, embora o contato com as representações fracionárias sejam bem menos frequentes em situações do cotidiano, seu estudo se justifica por ser fundamental para o desenvolvimento de conteúdos matemáticos que serão estudados posteriormente, como proporções, equações e cálculos algébricos. (PCN, 1998, p. 103)

A familiaridade do aluno com as diferentes representações dos números racionais (representação fracionária, decimal, percentual) pode levá-lo a perceber qual delas é mais utilizada ou adequada para expressar um resultado. Numa situação em que se deve comunicar um aumento de salário é mais frequente dizer, por exemplo, que o acréscimo no salário foi de 12% ($\frac{12}{100}$) do que de ($\frac{3}{25}$). (PCN, 1998, p. 103).

Portanto, no item a seguir, essa habilidade é explorada de forma bem direta, como veremos no quadro 12 abaixo.

Quadro 12 – Item M090235H6

Observações: Item aplicado no PROEB de 2017 para o 7EF e 1EM. Apresentou uma eliminação por DIF no 1EM.											
Habilidade avaliada: D42 – Corresponder diferentes representações de um número racional.											
Dados estatísticos na(s) ocorrência(s) com eliminação:											
Código: M090235H6			Descritor: D42			Gabarito: C					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,378	0,555		0,013	0,059	0,267	0,378	0,282	
Obs.:	Eliminação: DIF Anos_Etapas: 2017_1EM						Biss. A	Biss. B	Biss. C	Biss. D	Biss. E
							-0,197	-0,288	0,555	-0,247	
Dados estatísticos na(s) ocorrência(s) sem eliminação:											
Código: M090235H6			Descritor: D42			Gabarito: C					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
0,02803	318,1	0,15202	0,212	0,385	325,5	0,007	0,056	0,283	0,212	0,442	
Obs.:	7EF – PROEB 2017						Biss. A	Biss. B	Biss. C	Biss. D	Biss. E
							-0,251	-0,137	0,385	-0,076	
Imagem do Item:											
(M090235H6) A representação decimal do número $\frac{25}{2}$ é											
A) 0,08.											
B) 2,25.											
C) 12,5.											
D) 25,2.											

Fonte: CAEd/UFJF.

O item do quadro 12 foi aplicado no PROEB para as etapas do 7EF e do 1EM no ano de 2017. É interessante observar, através das bisseriais por alternativa e dos percentuais de busca por alternativa a justificativa da eliminação por DIF. Apesar da aplicação aprovada ter se dado com percentual de acerto bem menor (aproximadamente 21%), e ainda, na etapa do fundamental, era de se esperar um percentual de acerto bem maior no ensino médio. De fato foi (aproximadamente 37%), mas não como esperado. No entanto, a distribuição das bisseriais entre as alternativas se deu de forma bem distinta, cabendo observar a atuação dos distratores nesse momento. Aparentemente, a composição estrutural dos distratores se dá da seguinte forma: (A) É o resultado da divisão do denominador 2 pelo numerador 25; (B) Junção do numerador e denominador, atribuindo uma casa decimal a direita conforme a apresentação do denominador 2; (C) Gabarito; (D) Junção do numerador e denominador, atribuindo duas casas decimais a direita, conforme apresentação do numerador 25. Observe que os distratores “B” e “D” se deram a partir de composições com os elementos da fração, aguçando a atenção dos estudantes do 1EM, se tornando as duas opções mais procuradas, depois do gabarito, por estudantes de melhor proficiência. A bisserial da alternativa “A”, que se trata da divisão $\frac{\text{denominador}}{\text{numerador}}$, invés de $\frac{\text{numerador}}{\text{denominador}}$, foi em torno de 6% menos procurada na aplicação no ensino médio, porém, se manteve próxima em percentual de marcações. Fato interessante, e que desperta curiosidade, é que esperávamos o comportamento do ensino médio no ensino fundamental, sendo que ocorreu justamente o oposto a partir da observação das bisseriais. Pela TCT, o percentual de procura pelas alternativas “B” e “D” no ensino fundamental, evidentemente foram maiores, mas o fato é que, os melhores estudantes do ensino médio, foram atraídos justamente para essas alternativas, o que não era esperado.

O fato é que, sim, parte-se das observações das bisseriais, e dos percentuais de procura da população pelas alternativas oferecidas no item consciente da eliminação por DIF, o que leva a um questionamento muito maior do que apenas salientar que os estudantes não sabem corresponder uma fração a sua representação na forma decimal, e ainda, que no ensino fundamental, os estudantes saibam fazer, e se esquecem de como realizar tal tarefa no ensino médio.

Enfim, pressupõe-se que dessa forma se possa ter evidenciado o comportamento diferencial desse item (Quadro 12) nas duas aplicações, e que buscase por possíveis explicações.

Outro aspecto que é preciso destacar é sobre o número de alternativas oferecidas neste item. Itens com quatro alternativas são produzidos para estudantes do ensino fundamental, enquanto, com cinco alternativas, são direcionados a estudantes avaliados do ensino médio. Sobre este ponto, é visível que tal item foi construído com propósitos de avaliação voltados exclusivamente ao ensino fundamental, sendo este aplicado no 1EM, possivelmente possibilitando uma discussão acerca dos fatores que possam ter contribuído para sua eliminação, apesar de tal comparação ser necessária no processo avaliativo.

Evidente que não se pode comparar em conhecimento, alunos do 7EF com os do 1EM em determinados conteúdos, mas corresponder um número racional a sua forma fracionária se torna possível nessas duas etapas, e com este item, fica perceptível uma ascensão muito baixa por parte dos alunos do ensino médio, uma vez que houve um aumento de aproximadamente 16% no percentual de acerto desta aplicação no ensino médio.

Na pesquisa de campo, o Entrevistado 1 comenta que é comum itens dessa habilidade apresentar eliminações, e possivelmente apresenta familiaridade com os distratores, ao se referir sobre um deles.

Esta é uma habilidade muito comum de se ver itens sendo eliminados, ainda mais quando se tem uma alternativa em que o numerador da fração corresponde à parte inteira do número e o denominador corresponde à parte decimal. Vale destacar que tal formato que o item foi utilizado já não corresponde aos padrões atuais de elaboração, quando o enunciado está junto ao comando. (ENTREVISTADO 1, Questionário de Pesquisa Acadêmica, 31 jul.2020).

O E1 chama a atenção quanto ao padrão utilizado no comando do item, em que o mesmo deve ser separado do enunciado, algo que não foi relevante para os demais entrevistados.

O entrevistado 2 realizou sua análise pautado nos percentuais de busca pelas alternativas, considerando que os números apresentados no enunciado possam ter funcionado como chamariz.

Quando interpelado sobre o que possa ter influenciado na eliminação, o E2 diz que:

Os estudantes marcaram mais as alternativas em que aparecia um dos números do enunciado. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul.2020).

Sobre os distratores, o E2 argumenta que são coerentes, e que é uma prática comum a modelagem a partir da composição do numerador com o denominador. Ele finaliza sua análise dizendo que o item é

Relativamente fácil, mas que se apresentou de forma difícil, visto que os estudantes apenas procuram as alternativas onde havia um dos números do problema. Não reconhecendo conceito de fração e número decimal. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Já o Entrevistado 3 identifica o comportamento diferencial, argumentando que houve maior procura da alternativa “D” pelos alunos do 7EF, e ainda, comenta sobre a usabilidade de um dos distratores.

O item apresenta comportamento diferente entre as etapas, sendo o distrator D mais atrativo para o 7EF. Esse comportamento diferencial, evidencia um distrator também muito usual, aquele em que o estudante utiliza o numerador da fração como parte inteira e o denominador como parte decimal, sendo bem mais atrativo para a etapa 7EF. O item apresentou nesse caso apenas um comportamento diferente entre as etapas avaliadas, o que também é inadequado. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

É importante ressaltar que, das análises obtidas na pesquisa de campo, esta do E3 foi a primeira em que foram considerados o motivo de eliminação do item, e ainda, a que levantou suspeitas sobre um dos distratores.

O Entrevistado 4 não conseguiu identificar e argumentar neste item, fatores que possam ter contribuído na eliminação, e levanta hipóteses sobre os distratores, considerando-o adequado diante das técnicas de elaboração, dizendo:

Ao meu ver o item estaria de acordo com as técnicas e não consigo identificar um motivo pelo qual teria sido eliminado. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul.2020).

O entrevistado 5 também alega não compreender a eliminação o item, alegando não ser possível realizar uma boa inferência sem os parâmetros da TRI na ocorrência com eliminação. Ele também aborda sobre a regularidade quanto a construção e utilização do distrator da alternativa “D”, que foi muito procurada pelos estudantes do 7EF.

Não compreendi a eliminação do item, analisando apenas o histórico apresentado, sem os parâmetros pela TRI. Apenas com a análise clássica não é possível fazer tal inferência. Analisando os distratores, percebo que eles estão abordando o que é usual na técnica de elaboração de itens e no 7EF, me chamou a atenção o percentual de marcação da alternativa D, cerca de 44% dos respondentes apenas acrescentaram a vírgula entre os números de numerador e denominador. O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso de fato ocorreu e, embora a alternativa D tenha sido muito marcada, ela não foi marcada por respondentes que tiveram uma boa "nota" (escore) no teste. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

É possível verificar a partir desse item que existem práticas que são comuns na elaboração de instrumentos para essa habilidade. Foi possível também identificar que não existe, em sua maioria, preocupação com o construto no decorrer das análises da pesquisa de campo, e que a eliminação ocorre, conforme apontado pelos especialistas, quase a totalidade, por defasagem na aprendizagem dos estudantes. Quanto a essa habilidade, existe a possibilidade de distratores como o das alternativas “B” e “D” estarem se tornando complicadores para os estudantes, uma vez que se justificam apenas pela composição de dados apresentados no enunciado, gerando motivação para estudos muito mais aprofundados sobre este tema.

Por fim, tem-se um último item analisado nas eliminações por DIF, que desperta curiosidade quanto ao que ele se propõe avaliar e aos seus resultados. O descritor D33 (D33 – Reconhecer características do sistema de numeração decimal) propõe ao estudante identificar a ordem de algarismos em um número a partir do sistema de numeração decimal. O quadro 13, apresenta um item utilizado no PROEB no ano de 2016, aplicado nas etapas do 5EF e do 9EF.

Quadro 13 – Item M050057H6

Observações: Item aplicado no PROEB de 2016 para o 7EF e 9EF, sendo eliminado nessas duas ocorrências.
--

Habilidade avaliada: D33 – Reconhecer características do sistema de numeração decimal.											
Dados estatísticos na(s) ocorrência(s) com eliminação:											
Código: M050057H6			Descritor: D33			Gabarito: A					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,674	0,592		0,007	0,674	0,102	0,086	0,131	
Obs.: Eliminado DIF – 5EF_PROEB_2016							Biss. A	Biss. B	Biss. C	Biss. D	Biss. E
							0,592	-0,407	-0,387	-0,337	
Código: M050057H6			Descritor: D33			Gabarito: A					
Parâmetros da TRI			Análise Clássica		Ancoragem	Percentuais por opção de resposta					
A	B	C	Perc. de Acerto	Bisserial		B/N	A	B	C	D	E
			0,505	0,487		0,005	0,505	0,188	0,147	0,155	
Obs.: Eliminado DIF – 9EF_PROEB_2016							Biss. A	Biss. B	Biss. C	Biss. D	Biss. E
							0,487	-0,250	-0,261	-0,261	

Imagem do Item:

(M050057H6) Observe o número no quadro abaixo.

275 349

Qual é o algarismo que ocupa a ordem das dezenas de milhar nesse número?

A) 7
B) 5
C) 4
D) 2

Fonte: CAEd/UFJF.

Uma das orientações contidas na BNCC sobre a área temática de números para estudantes do ensino fundamental é que:

Nessa fase espera-se também o desenvolvimento de habilidades no que se refere à leitura, escrita e ordenação de números naturais e números racionais por meio da identificação e compreensão de características do sistema de numeração decimal, sobretudo o valor posicional dos algarismos. Na perspectiva de que os alunos aprofundem a noção de número, é importante colocá-los diante de tarefas, como as que envolvem medições, nas quais os números naturais não são suficientes para resolvê-las, indicando a necessidade dos números racionais tanto na representação decimal quanto na fracionária. (BRASIL, 2017, p. 268).

Dessa forma, o descritor D33 não exige tanto uma contextualização para ser avaliado em uma avaliação em larga escala, podendo ser trabalhado de forma bem direta em sua maioria, e como foi apresentado nesse item (quadro 13).

Na estrutura dos distratores é possível observar que todas as alternativas tratam de algarismos pertencentes a outras ordens do número informado. Como se trata apenas de um reconhecimento (identificação) de uma ordem, não é exigido a realização de uma operação com os algarismos informados na elaboração de distratores.

De acordo com os relatórios das duas aplicações do item, visualmente se torna difícil compreender sua eliminação, pois o mais esperado é que apenas a bisserial do gabarito seja positiva, e isso ocorreu em ambas as aplicações.

No entanto, seu percentual de acerto foi maior no 5EF, sendo em torno de 67%, que não era de esperar, uma vez que se deseja um melhor percentual na etapa mais além (9EF), que ficou próximo dos 50%.

O Entrevistado 1 observa que as bisseriais do item ficaram bem próximas nas aplicações, o que se pode discordar, observando principalmente a bisserial da alternativa B. Ele diz que:

Analisando tal item não consigo identificar possíveis motivos para tal item ser eliminado, pois apresentam bisseriais próximas nas duas aplicações. (ENTREVISTADO 1, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 2 inicialmente observa o maior percentual de acerto do item no 5EF. Alega que os distratores são coerentes, e que é comum apresentar Algarismos de outras ordens do número como alternativas. Ele finaliza sua análise, dizendo que

Esse item foi mais bem acertado no 5EF do que no 9EF, quando, teoricamente, não pode haver esse comportamento diferencial. Além disso, os estudantes do 9EF, depois do gabarito, marcaram qualquer alternativa, sem critério. Enquanto que no 5EF eles ficaram mais em dúvida entre B e D. (ENTREVISTADO 2, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Já o entrevistado 3 faz uma análise bem sucinta sobre item, porém, aparentemente fundamentando-se pelo motivo de eliminação, dizendo que:

O item apresenta comportamento diferenciado em cada etapa. Os distratores apresentados são comumente utilizados, pautados nas diferentes ordens do número. O item apresentou um comportamento diferente entre as etapas avaliadas, o que é inadequado. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

O Entrevistado 4 chama atenção quanto ao nível de complexidade do item, alegando que ele pode ter se tornando tão fácil para a população avaliada, que pode ter se tornando ineficaz na avaliação. Na análise do entrevistado 4 ele diz:

Única coisa que penso é que, como a maioria dos estudantes (mais de 50%) marcaram a alternativa correta, o item pode ter se tornando ineficaz para diferenciar os alunos que sabem daqueles que não

sabem. Tornou-se um item que não diferencia bem os estudantes. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Esse fato é importante e realmente pode acontecer. Como já mencionado em seções anteriores, um item com baixa discriminação pode se tornar ineficaz dentro da avaliação. Todos os itens que são eliminados nas avaliações realizadas pelo CAEd, os parâmetros da TRI (A, B e C) não são informados no relatório estatístico, tornando difícil uma melhor compreensão pedagógica do que se passou nessa aplicação.

Também sob este mesmo argumento, o Entrevistado 5 realiza sua análise, apontando que possivelmente a discriminação desse item tenha sido baixa.

Não compreendi a eliminação do item, analisando apenas o histórico apresentado, sem os parâmetros pela TRI. Suponho que esse item tenha sido tão fácil para a etapa que ele não discriminou os respondentes. Analisando os distratores, percebo que eles estão abordando o que é usual na técnica de elaboração de itens e o que é possível (os outros Algarismos que compõe o número em questão). O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e que as demais sejam negativas e isso de fato ocorreu. Suponho que esse item tenha sido tão fácil para a etapa que ele não discriminou os respondentes. (ENTREVISTADO 4, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Nesta análise, é possível perceber o quão simples é a habilidade avaliada, e o nível de complexidade que obtemos ao analisá-la. Inicialmente é quase incompreensível entender o porquê uma habilidade tão simples venha a ser eliminada em teste, mas que pode ser desvendada a partir da análise de seus dados estatísticos.

Enfim, o item realmente se demonstra fácil para os estudantes, podendo se tornar ineficiente na avaliação. Possivelmente, não existem problemas com a elaboração desse instrumento, o que leva a conclusão de que, mesmo a partir de uma eliminação estatística, é possível ter boas conclusões sobre avaliação como um todo, como, por exemplo, de que a rede avaliada tenha atingido certa consolidação sobre este tema.

Isso leva a entender que dentro do universo avaliativo, é possível obter informações relevantes e levantar diagnósticos, mesmo a partir de dados que são desfavoráveis para a avaliação.

Até aqui, foi possível identificar que sim, existem práticas comuns na elaboração de itens, e que são frequentemente utilizadas, o que pode contribuir para

um banco com itens semelhantes em vários aspectos, podendo eles serem positivos ou negativos.

No quadro 14, a seguir, foram levantados os principais aspectos observados na análise dos itens, envolvendo as práticas de elaboração adotadas conforme observado.

Quadro 14 - Aspectos identificados na análise dos itens

Item M120389H6
<ul style="list-style-type: none"> – O distrator da alternativa “E” foi bastante apontado pelos entrevistados (Observado através da bisserial, que foi positiva e atraiu bons estudantes). – Como distratores, são propostas funções, em que pelo menos um dos pontos do gráfico faça sentido.
Item M120390H6
<ul style="list-style-type: none"> – O distrator da alternativa “B” foi o mais apontado pelos entrevistados e procurado pelos estudantes (apesar de ser considerado um excelente distrator). – É comum apresentar gráficos de funções mais comuns aos estudantes (funções afim e quadrática) como proposta de distratores, além de opções envolvendo funções inversas. – Possíveis fatores pedagógicos, com relação aos estudantes avaliados, possivelmente podem justificar a eliminação – Se trata de um item não simples.
Item M120189H6
<ul style="list-style-type: none"> – É uma prática comum apresentar gráficos de funções parecidas a que é apresentada no suporte do item, bem como as funções identidade da trigonometria. – Distrator da alternativa “D” foi o mais procurado pelos estudantes, apesar dos percentuais de busca das alternativas se distribuírem parcialmente parecidos. – Distratores como o da alternativa “A”, aguçou a atenção de estudantes com boa proficiência, utilizando funções identidade com variações semelhantes a função dada.

Item M090047H6

- É comum utilizar manipulações de dados (geralmente aritméticas) na elaboração de distratores.
- Dentre uma das práticas comuns de construção de distratores, é a de explorar no estudante a possibilidade de confusão entre quadrado e dobro ou quadrado e raiz quadrada.
- Os distratores do item não foram pautados na habilidade avaliada, mas em outras que deveriam estar antecipadamente consolidadas pelo estudante.
- O distrator da alternativa “A” foi bem destacado pelos entrevistados.
- Possivelmente esse modelo de construção possa estar ocasionando confusão nos estudantes (possível indução ao erro).

Item M100093H6

- O item apresentou baixo percentual de acerto.
- Dentre as práticas comuns de elaboração, é recorrente apresentar outras expressões algébricas que se correspondam a algum dos pontos do gráfico apresentado no suporte. Inversões dos coeficientes angular e linear na lei de formação, manipulações com os sinais, o crescimento ou decréscimo da reta, os pontos de intersecção da reta com os eixos também são outras estratégias exploradas.
- O item apresentou três bisseriais positivas, fazendo com que os distratores das alternativas “D” e “E” fossem juntamente com o gabarito (alternativa B), procurados por estudantes de boa proficiência.

Item M1D26I0117

- O item foi aplicado em duas etapas do ensino médio (1EM e 3EM), sendo eliminado somente na aplicação ao 1EM.
- É comum na prática de elaboração, apresentar distratores com variações de sinais, que são alternados entre as alternativas. O mais comum dentre os distratores apresentados, é aquele em que os sinais dos termos independentes dos polinômios dados como alternativas, são iguais aos das raízes dadas no texto do enunciado, correspondendo neste item, ao distrator da alternativa “D”.

– Em ambas as aplicações, ocorreram duas bisseriais positivas, sendo o distrator da alternativa “D”, procurado juntamente com o gabarito (alternativa “E”), por estudantes de boa proficiência.

– Existe a possibilidade do distrator da alternativa “D” estar se tornando um complicador para o item.

Item M040046H6

– O item apresentou apenas uma eliminação ao longo de suas quatro aplicações até 2018.

– O distrator da alternativa “B” se destaca em percentual de procura e bisserial ao longo das aplicações, atraindo atenção de estudantes com boa proficiência.

– Dentre práticas comuns de elaboração observadas, estão a utilização do próprio percentual como valor de desconto ou acréscimo, caso este explorado na alternativa “B”. Associar o percentual como valor final e manter o valor do produto inalterado também foram identificados nesse tipo de produção de itens.

Item M090235H6

– Item aplicado ao 7EF e ao 1EM, sendo eliminado na aplicação ao 1EM.

– Dentre práticas utilizadas na produção de itens deste descritor, a construção de distratores em que são utilizadas a divisão de forma invertida (denominador dividido por numerador) e composições utilizando-se numerador como parte inteira e denominador como parte decimal, e vice-versa, são comumente utilizadas.

– Possivelmente cabe uma discussão acerca da etapa para qual o item foi construído, aparentemente para o 7EF e aplicado no 1EM.

– Possibilidade de distratores construídos a partir de composições com os números da fração dada, semelhante as alternativas “B” e “D” deste item, estarem se tornando complicadores para os estudantes.

Item M050057H6

– Item aplicado e eliminado nas etapas do 7EF e 9EF, sendo mais acertado, próximo de 10% a mais no 7EF.

– Mesmo com bons percentuais de acerto, o item foi eliminado em ambas as aplicações.

– As demais alternativas ficaram parcialmente parecidas em distribuição nos percentuais de busca pelos estudantes.

- Dentre as práticas de elaboração, o mais comum é utilizar Algarismos de outras ordens do número dado, para compor os distratores.
- Existe a possibilidade de o item não ser eficaz na avaliação de ambas as etapas, podendo a matriz de referência estar avaliando uma habilidade já consolidada em etapas anteriores no ensino fundamental.

Fonte: Elaborado pelo autor (2020)

Viu-se também que em nenhuma das análises os especialistas se apoiaram em documentos oficiais, apenas em uma das análises o Entrevistado 1 diz que o item não está mais dentro do padrão de apresentação, algo que deve ter partido de um acordo interno.

A maioria das análises foram realizadas levando em consideração as práticas diárias de trabalho, prevalecendo a experiência profissional. Outro aspecto observado na pesquisa de campo é que, mesmo informando o motivo de eliminação do item, em sua maioria, essa informação não foi relevante.

Dos cinco entrevistados na pesquisa de campo, quatro responderam que norteiam seus trabalhos de elaboração e revisão a partir de acordos internos e do guia de elaboração e revisão do CAEd, embora dois deles tenham salientado que o atual guia se encontra em fase de reformulação. Apenas o Entrevistado 4 alega não utilizar nenhum documento para realizar suas atividades.

Quando interrogados sobre o CAEd promover momentos formativos para o desempenho e aperfeiçoamento na execução de suas tarefas, respondem afirmativamente, que de uma a duas vezes por ano, ocorrem seminários ou quando surge alguma necessidade, se reúnem e discutem assuntos pertinentes ao trabalho. O Entrevistado 5 alega ainda que as reuniões aconteciam semanalmente, mas que no momento são mais pontuais, apenas quando solicitado.

Possivelmente a falta de amparo em documentos oficiais ou encontros mais frequentes, ocasionam um automatismo na realização das tarefas, principalmente quando submetidos a um curto espaço de tempo, não permitindo análises mais aprofundadas.

Em unanimidade, todos os entrevistados concordam que a partir de itens eliminados, é possível obter algum tipo de aprendizagem sobre a elaboração de itens, e diagnósticos sobre a aprendizagem da rede avaliada. Assim, não restam dúvidas de

que algum tipo de trabalho se faz necessário sobre estudo de itens eliminados, com propósito de aprimorar não somente a elaboração, mas a avaliação educacional.

Com a pesquisa de campo, pôde-se observar que a maioria dos colaboradores entrevistados atuam no campo da avaliação em larga escala em torno de quatro anos, sendo apenas uma, a Entrevistada 4, com experiência superior a esse tempo, contabilizando nove anos de atuação na área. No entanto, todos os entrevistados, alegaram não trabalhar anteriormente com avaliações, sendo o CAEd, ao mesmo tempo, um local de labor e formação. Quando questionados se houve algum tipo de formação ou treinamento ao serem admitidos, a Entrevistada 4 é a única que alega não ter sido submetida a nenhum tipo de treinamento, apenas a um monitoramento de outros colaboradores mais antigos. Apesar de todos os demais entrevistados responderem positivamente quanto a esse questionamento (ou seja, houve um treinamento), foi possível detectar, através de seus relatos, uma ausência de padrão de treinamento, tendo o CAEd, oferecido esse treinamento de diferentes modos e intensidades aos seus futuros colaboradores na época.

Este fato talvez traduza a ausência de amparo aos documentos internos durante a realização das análises dos itens, e ainda, o fato responderem que desconhecem alguns tópicos específicos sobre conhecimentos estatísticos, como foi o caso da Entrevistada 2, que desconhece totalmente do que se trata uma eliminação por motivo de ajuste, e a Entrevistada 4 não conhecer as eliminações de itens por DIF.

Com relação aos conhecimentos que são específicos para os colaboradores que atuam na equipe da CME (Coordenação de Medidas Estatísticas), ao longo deste trabalho, é perceptível que, se os colaboradores das demais equipes tiverem ao menos um conhecimento superficial sobre a estatística nos itens, os fatores pedagógicos podem se tornar ainda mais eficazes no processo avaliativo, contribuindo significativamente para o desenvolvimento do trabalho dos elaboradores.

Considerando a experiência que acumulamos no banco de itens, para ser admitido, fui submetido a um processo seletivo, em que houve um treinamento que compunha uma das etapas desse processo de seleção. Porém, o treinamento era específico para elaboração de itens, e mesmo após ser aprovado, e fazer parte da equipe, ao longo dos anos de atuação, percebi a necessidade de obter mais informações e conhecimentos sobre avaliação, e que algumas práticas passam de um colaborador para outro, sem maiores aprofundamentos teóricos ou estudos sobre seus impactos.

Portanto, com a pesquisa de campo, viu-se a necessidade de um aprofundamento da equipe sobre os conhecimentos técnicos estatísticos que envolvem a avaliação. Os conhecimentos pedagógicos são parte da formação de todos, uma vez que em unanimidade são licenciados em Matemática, com passagens e experiências no magistério.

Outro aspecto que ficou perceptível é quanto à possibilidade de implementação de novos incentivos do CAEd para com seus colaboradores na constância em pesquisa, quer sobre itens ou a avaliação como um todo. O Entrevistado 3, ao ser questionado se o CAEd promove eventos formativos para aperfeiçoamento do trabalho, responde que

Sim. Sempre que há a necessidade para alguma nova demanda, em devolutivas das equipes de pesquisa, para adequações de algumas funções e em eventos e congressos voltados para o tema. (ENTREVISTADO 3, Questionário de Pesquisa Acadêmica, 31 jul. 2020).

Com isso, o Entrevistado 3 deixa transparecer o fato de existir uma equipe específica para pesquisas na CIA, o que pode ocasionar o distanciamento daqueles que atuam diretamente com o trabalho, tornando a pesquisa, em termos, parte distante da equipe de elaboração.

Ao serem questionados sobre conhecerem o manual de elaboração e revisão do CAEd, apenas a Entrevistada 4 diz não ter conhecimento. E sobre os manuais de outras instituições, os entrevistados 1 e 2 alegam nunca terem efetuado pesquisas a respeito.

Portanto, até aqui, é possível compreender que se pode pensar em propor um modelo de formação padrão aos colaboradores do CAEd, principalmente no início de suas atividades, uma vez que seus colaboradores não computam experiências anteriores, o que pode afetar sem desvios, a evolução do trabalho. Entende-se, nesse sentido, que o CAEd não funciona somente como um centro de avaliação e políticas públicas, mas também como um centro de formação, levando em consideração que, profissionais com experiência em avaliação ainda somam baixos contingentes no país.

Dessa forma, na seção a seguir, serão estruturadas algumas possibilidades de intervenção, com o propósito de minimizar os efeitos na eliminação de itens,

almejando avanços na pesquisa sobre avaliações em larga escala e na otimização do trabalho realizado pelo CAEd.

4 PLANO DE AÇÃO EDUCACIONAL

O objetivo nessa pesquisa foi identificar possíveis vieses adotados no processo de elaboração de itens da equipe de Matemática na Coordenação de Instrumentos de Avaliação (CIA) do CAEd, que possam estar influenciando na eliminação de itens. Com o objetivo de minimizar a eliminação de itens nas avaliações promovidas pelo CAEd e avançar na discussão acerca da qualidade e produção de itens para o banco. Nesta seção, serão apresentados um conjunto de ações com propósito de minimizar os impactos nesse contexto.

Como visto, nas primeiras seções deste trabalho, desenvolveu-se uma reflexão sobre as bases teóricas que envolvem o conhecimento sobre avaliações em larga escala, no intuito da aproximação de fatos históricos do surgimento da avaliação no Brasil, seus propósitos, a estrutura organizacional do CAEd, a elaboração de itens e as matrizes de referência, com abordagens relacionando-a ao ensino.

Viu-se também a psicometria nos itens, possibilitando uma aproximação dos métodos de correção de testes, a TCT e a TRI, juntamente com os parâmetros estatísticos, galgando conhecimento que pudessem esclarecer como ocorre uma eliminação estatística ou pedagógica de um item em uma avaliação em larga escala.

Com a pesquisa de campo, teve-se a chance de observar que de fato ocorrem algumas práticas que se tornaram comuns no processo de elaboração, conforme apontado no quadro 14 da seção anterior.

Observou-se também que os colaboradores ao serem admitidos no CAEd, não são submetidos ao um treinamento padrão, constatado através dos relatos dos entrevistados e aos diferentes graus de conhecimento sobre aspectos mais aprofundados na avaliação em larga escala.

A falta de incentivo aos colaboradores na continuidade da pesquisa é outro fator que cabe ser salientado, além da falta de amparo nos documentos e acordos internos ao realizarem suas atividades.

Portanto, a partir dos fatores levantados, pretende-se atuar sob dois eixos centrais, que são: Eixo 1 - Estimular a pesquisa com itens no intuito de reduzir as

eliminações e buscar formas de otimizar o trabalho da elaboração; Eixo 2 - Aprimorar os aspectos formativos da equipe.

No quadro 15 a seguir são apontados os principais achados de pesquisa seguidos de possibilidades de intervenção que envolvem esses dois eixos de ações, podendo estes ser acolhidos ou não, pelo Centro de Políticas Públicas, com intuito de aprimorar os aspectos apontados.

Quadro 15 – Achados e propostas de ação

	Principais achados	Propostas de ações
<p>Eixo 1</p> <p>Estimular a pesquisa com itens no intuito de reduzir as eliminações e buscar formas de otimizar o trabalho da elaboração</p>	Existem padrões na construção de distratores	Inicialmente, evitar distratores padronizados, identificando aqueles que mais chamam a atenção de bons estudantes, por meio de análise das bisserias e percentuais de busca por alternativa, verificando se estes são causadores de eliminação do item.
	Alguns distratores não são fundamentados na habilidade avaliada	Evitar o uso de distratores que não se fundamentam em erros provindos da habilidade avaliada.
	Construção de distratores a partir de manipulações com dados do enunciado do item	Evitar ou utilizar somente em último caso, a construção de distratores que utilizam manipulações aritméticas com dados do enunciado, quando estas não são parte da habilidade avaliada do item.
	Existe uma equipe específica para realização de pesquisas sobre itens e avaliação educacional em larga escala	Estimular e aproximar os colaboradores do banco de itens para estudos e realização de pesquisas sobre itens e avaliação educacional em larga escala.
	Não identificamos estudos especificamente voltados aos itens eliminados como parte do fluxo de trabalho	Semestral ou anualmente, efetuar o levantamento do quantitativo de itens eliminados por projeto, buscando identificar padrões, fatores e itens que não estão sendo eficazes na avaliação.
	Não identificamos estudos das matrizes referência no fluxo de	Alguns itens podem não estar sendo eficazes na avaliação,

	trabalho, considerando itens que não possuem bom desempenho nas avaliações	despertando a necessidade de análise dos descritores da matriz de referência, verificando se ainda são compatíveis a etapa avaliada ou mesmo necessários.
Eixo 2 Aprimorar os aspectos formativos da equipe	Análises e correções de itens realizadas a partir das experiências profissionais	Estimular a ênfase nos documentos oficiais que norteiam a elaboração e a revisão, incluindo o de outras instituições para estudos.
	A formação inicial dos colaboradores não é padrão	Criar um modelo padronizado de formação de colaboradores, que reúna fatores pedagógicos e psicométricos da avaliação em larga escala.

Fonte: Elaborado pelo autor (2020)

A luz desses dois eixos centrais, nas próximas seções serão detalhadas ações que serão especificadas em subdivisões. Como auxílio nesse tópico, será utilizada a ferramenta 5w2h com a função de definir metodologicamente ações que serão apresentadas e propostas ao CAEd, no modelo e formato apresentado no apêndice B deste trabalho.

4.1 EIXO 1 – ESTIMULAR A PESQUISA COM ITENS NO INTUITO DE REDUZIR AS ELIMINAÇÕES E OTIMIZAR O TRABALHO DA ELABORAÇÃO

Para estimular a pesquisa com itens, inicialmente pode-se adotar experimentalmente as propostas desse trabalho, investigando itens eliminados pertencentes a cada um dos eixos temáticos da Matemática. Assim, levantar os quantitativos, e buscar similaridades e padrões nos itens, pode revelar respostas não apenas sobre a eliminação, mas sobre a qualidade do item e da avaliação como um todo.

4.1.1 – Levantamento do quantitativo de itens eliminados por áreas temáticas e produção de pesquisa

Para verificar a qualidade dos itens a partir de itens eliminados, será necessário o levantamento do quantitativo de itens utilizados nos principais projetos de atuação do CAEd. Assim, é possível identificar qual das áreas temáticas apresenta maior

quantitativo de eliminações e realizar um estudo a partir dos descritores mais recorrentes, ou a partir de itens com baixos percentuais de acertos, como foi o caso dessa pesquisa.

Quadro 16: Proposta de ação I

O que será feito?	Levantamento do quantitativo de itens eliminados pertencentes a áreas temáticas da Matemática
Por que será feito?	Identificar padrões nas eliminações de itens, buscando fatores que possam contribuir para eliminação e afetar a qualidade do item, além de identificar itens que já não são eficazes na avaliação.
Onde?	CIA
Quando?	O tempo deve ser ajustado em razão da demanda de trabalho, podendo ser definido para ser realizado durante seis meses, duas vezes ao ano. Dessa forma é possível realizar duas pesquisas por ano, em duas áreas temáticas da Matemática. Esse estudo também pode ser adaptado e realizado anualmente.
Quem?	Colaboradores da equipe de Matemática da CIA.
Como?	Organizando estudos individuais por colaborador, promovendo encontros semanais para reunir informações. Essas reuniões podem durar em torno de 2 horas cada. Ao final de cada semestre/ano, será produzido um artigo ou produto educacional por toda equipe, divulgando os resultados.
Custo	A princípio não gera custos extras na realização dessas tarefas, uma vez que fará parte da rotina de trabalho.

Fonte: Elaborado pelo autor (2020)

O que propomos com esta ação é de realizar um levantamento de itens eliminados e verificar a recorrência ou padrões de eliminação. Assim, seria possível analisar itens ou distratores que não estão sendo eficazes para a avaliação, assim mantendo os itens sempre com boa qualidade. Esse levantamento possibilita a reflexão sobre o uso de distratores padrões, estimulando a pesquisa com itens dentro da CIA, proporcionando aos colaboradores, uma reflexão contínua sobre as práticas de trabalho e a pesquisa sobre avaliação.

O fortalecimento da pesquisa se daria diante de um trabalho, texto ou produto educacional a ser construído coletivamente pela equipe a cada semestre, a partir dos resultados obtidos com o levantamento realizado. Nesse levantamento, podem surtir efeitos sobre a construção e eficácia dos instrumentos de avaliação e a funcionalidade da Matriz de Referência.

Nesta ação, pode-se ainda incluir alguns aspectos que envolvem a eliminação de itens por motivos que rodeiam a Matriz de Referência. Como visto na pesquisa de campo, existe a possibilidade de alguns itens se tornarem ineficazes na avaliação ao

decorrer dos anos. Isso nos possibilita analisar e estipular um período para renovação/atualização das matrizes de referência.

Entende-se que, quanto mais se avalia uma determinada rede, seu ensino e suas consolidações, podem evoluir ou até mesmo regredir, fazendo com que algumas habilidades possam se tornar relevantes ou não, em determinadas etapas, com foi o caso do **Item M050057H6** dessa pesquisa, eliminado por DIF.

4.1.2 Possibilitar Caminhos Para Produção De Novos Itens e Realizar Pré-Testes

Para possibilitar caminhos para produção de novos itens, será preciso levantar informações a partir das pesquisas realizadas nas ações propostas em 4.1.1. Identificadas recorrências de eliminação e padrões de elaboração que possam estar contribuindo para eliminação, será necessário construir novos itens e realizar novos pré-testes. Os novos itens podem partir de ideias formuladas pelos próprios elaboradores através de novas propostas. Contudo, é interessante observar o comportamento de estudantes, ou mesmo, na forma como as redes avaliadas promovem o ensino. Dessa forma, a realização de seminários investigativos pode proporcionar estudos, que inspirem a produção de novos itens com novos distratores. Esses seminários podem ser amparados através de referenciais teóricos que envolvam a produção de significados na aprendizagem da Matemática, ou mesmo produzi-los durante a realização deste seminário, através de cenários previamente observados pelos professores participantes.

A produção de significados acontece a partir do Modelo dos Campos Semânticos (MCS), desenvolvido pelo educador matemático Rômulo Campos Lins. Do ponto de vista de Lins (2012), os campos semânticos servem para articular a produção de conhecimentos, significados e objetos, oferecendo formas de produção de leituras suficientemente finas de processos de produção de significados (LINS, 2012).

Lopes (2013) aponta que o MCS

[...] é um modelo epistemológico que afirma que o conhecimento é dado pela crença-afirmação e justificação. Não basta que o sujeito creia e afirme sobre uma dada enunciação, é preciso que ele justifique o que foi afirmado. Ou seja, a crença-afirmação é aquilo que o sujeito enuncia algo que acredita e a justificação é o que o sujeito entende como aquilo que ele está autorizado a dizer. (LOPES, 2013, p. 2).

Portanto, entende-se que este método possa fundamentar inúmeros estudos e possibilidades, por exemplo, na elaboração de itens, principalmente na criação de distratores, permitindo uma coleta apurada de dados, tornando-os teoricamente fundamentados nos significados produzidos em sala de aula.

Essa investigação pode se dar também em oficinas de elaboração de itens, sejam elas presenciais ou virtuais, em que professores das redes participantes busquem por exemplos de respostas, resoluções de atividades em sala, garimpando produções dos próprios alunos, que servirão de referencial para os elaboradores levantarem novas propostas de itens. Ao final das oficinas, os professores participantes podem ser certificados como forma de incentivo na participação da pesquisa.

Quadro 17: Proposta de ação II

O que será feito?	Produção de novos itens
Por que será feito?	Para obter novos itens, com novas estruturas, conforme informações levantadas nas pesquisas realizadas em 4.1.1
Onde?	Pode ser realizado na CIA, oficinas de elaboração de itens ou em ambientes virtuais, em forma de oficinas e seminários presenciais ou vídeo conferências
Quando?	Após cada artigo ou produto educacional produzido em 4.1.1. Duas vezes por ano ou anualmente.
Quem?	Colaboradores da CIA, com ênfase aos membros das equipes de elaboração.
Como?	Na realização de seminários ou oficinas, propor aos cursistas o levantamento de significados produzidos por seus alunos, através de exercícios resolvidos, provas ou quaisquer atividades do cotidiano escolar.
Custo	Nos custos podem envolver a disponibilização e deslocamentos de colaboradores para ministrar tais oficinas ou seminários ao optarem por meios presenciais. Em caso de propor em ambientes virtuais, disponibilização de recursos digitais otimizando em custo quanto aos deslocamentos.

Fonte: Elaborado pelo autor (2020)

É importante ressaltar que se está dando ênfase quanto a construção do item nesta ação, e o quanto um estudo aprofundado pode revelar novos caminhos para aprimorarmos o trabalho. Um distrator bem colocado contribui positivamente para a

avaliação de uma habilidade, mas, ao mesmo tempo, um outro que apresente comportamento negativo, pode evidenciar ou até mesmo denunciar uma prática equivocada no processo de ensino adotado na rede avaliada, e não apenas representar uma falha de elaboração. Entende-se que pode existir algo muito maior do que um simples erro de elaboração, e que deve ser interpretado e pesquisado de forma criteriosa por especialistas em avaliação.

Por isso a importância em propor maior incentivo a pesquisa com itens eliminados, o que destaca ainda mais a importância das avaliações em larga escala, no sentido do monitoramento, tomada de decisões e ações futuras.

Concluindo a ação, após a realização das oficinas investigativas, junta-se toda produção de significados produzida, e inicia-se um processo de elaboração de novos itens fundamentados nessa pesquisa realizada. Em seguida, monta-se o teste e realiza-se os pré-testes para interpretação final dos resultados com todos os dados estatísticos gerados.

4.2 EIXO 2 – APRIMORAR OS ASPECTOS FORMATIVOS DA EQUIPE

Com o propósito de otimizar a performance dos futuros elaboradores, e ainda, situá-los diante do cenário avaliativo na educação, a formação do elaborador torna-se fundamental, proporcionando a produção de itens de boa qualidade pedagógica e estatística. Nesse quesito, poderia ser desenvolvido no, CAEd, um modelo de capacitação padrão, que atinja os principais aspectos que envolvem a avaliação em larga escala, como seus propósitos, a construção de instrumentos, a montagem de testes e os conhecimentos estatísticos fundamentais para sua compreensão. Para isso, são propostas as seguintes ações:

Quadro 18: Proposta de ação III

O que será feito?	Curso de formação de elaboradores/colaboradores da CIA
Por que será feito?	Aperfeiçoamento dos colaboradores que iniciarão suas atividades na CIA
Onde?	CAEd
Quando?	Sempre que houver novas contratações de colaboradores para a CIA
Quem?	Será necessário montar equipes de formadores com colaboradores experientes da CIA
Como?	Por meio de um seminário com duração de três semanas. 1º Semana: Ministrando conteúdos que apresentem os propósitos da avaliação educacional no Brasil, seu início e alguns exemplos de projetos de avaliação no Brasil.

	2ª Semana: Elaboração de itens e montagem de testes. 3ª Semana: A TCT e a TRI; A estatística nos itens.
Custo	Os custos envolverão a disponibilização de colaboradores experientes, selecionados para ministrar a formação.

Fonte: Elaborado pelo autor (2020)

O propósito deste curso de formação, pode se relacionar ao fato da necessidade de formar novas ou acrescentar novos colaboradores, e ainda, oferecê-lo na forma de um curso de capacitação com certificação ou habilitação. Com isto, seria possível capacitar um profissional para atuar com mais clareza no universo da avaliação em larga escala, principalmente com a produção de instrumentos.

A princípio seria um curso de formação com duração de aproximadamente 3 semanas, em que seriam ministrados em cada semana, um tema sobre avaliação em larga escala. Na primeira semana, seria desenvolvido um estudo sobre os propósitos da avaliação no Brasil, fatos históricos e o conhecimento de alguns projetos de avaliação nacional, por exemplo, o SAEB.

Na semana seguinte, o tema estaria voltado para a produção de itens. Como elaborar um item, o funcionamento da revisão de itens, expondo que para se tornar um revisor é necessário mais experiência, deixando claro que a revisão aconteça seguindo as instruções dos manuais e, dos critérios específicos da disciplina, ou seja, que conteúdo da habilidade esteja correto. É importante que a formação ofereça condições do participante assimilar como de fato deve ser a estrutura de um item em uma avaliação educacional.

Por último, seriam explicitados os métodos de correção de testes e a formas de leitura de um item a partir de seus dados estatísticos. A TCT e a TRI seriam abordadas de maneira mais compreensível, como proposto no início deste trabalho, de maneira que, seja possível entender que para ser um bom elaborador não se torna necessário ser um bom estatístico, mas que ambos os profissionais saibam se comunicar com clareza e entendimento com suas áreas de atuação. É importante desenvolver nessa formação, a capacidade do indivíduo ler e interpretar o que os dados estatísticos revelam sobre o comportamento pedagógico do item na avaliação, para assim, possibilitar o cursista efetuar as devidas interpretações e pesquisas.

O intuito desse modelo de capacitação, é servir ou inspirar uma nova proposta de formação com formato padrão, de maneira que possa atender não exclusivamente apenas novos elaboradores para a equipe de Matemática, estendendo também, aos

futuros colaboradores de outras áreas do conhecimento, como ciências humanas, língua portuguesa ou ciências da natureza. Com este modelo, o CAEd, com todo prestígio que o situa dentro do cenário educacional brasileiro, pode utilizá-lo não somente com propósitos de formar novas equipes internas, possibilitando estendê-lo ao público externo, que deseja mergulhar mais profundamente no campo da avaliação educacional em larga escala, capacitando e/ou habilitando/certificando profissionais a exercer atividades voltadas a avaliação, em qualquer outra instituição ou rede que desenvolva projetos neste universo.

Dessa forma, com essas propostas de intervenção, espera-se de fato, poder contribuir e promover avanços significativos no processo de elaboração de itens, visando a redução de itens eliminados, e de certa maneira, avançar no debate e pesquisa acerca da avaliação em larga escala, possibilitando ainda, contribuir para formação de novos talentos neste ramo ainda tão contrariado e avesso aos olhos de muitos profissionais da educação.

5 CONSIDERAÇÕES FINAIS

Identificar possíveis vieses adotados no processo de elaboração de itens da equipe de Matemática na CIA do CAEd, que possam estar influenciando na eliminação de itens, foi o objetivo principal deste trabalho. No entanto, para que se possa alcançar tal objetivo, teve-se a oportunidade, até aqui, de se aprofundar um pouco mais estudos sobre o universo da avaliação educacional em larga escala.

Teve-se ainda, como questão de pesquisa, a busca por uma forma de como se poderia intervir no fluxo de trabalho da elaboração para que ocorram cada vez menos eliminações de itens. Dessa forma, mergulhou-se em aspectos históricos sobre as políticas de avaliação educacional no Brasil, trazendo um breve histórico de seu início através do SAEB e o estado de Minas Gerais com o PROEB, que tem forte destaque neste trabalho.

Como o estudo girou em torno da produção de itens e avaliações realizadas pelo CAEd, viu-se também como é a sua estrutura organizacional, além das grandes contribuições e atuações dessa instituição no cenário da educação pública, demonstrando sua grande importância no país.

Conheceu-se um pouco da dinâmica que é realizada para a construção de um banco de itens, a partir dos processos de elaboração de itens e testes, trazendo destaques para os elementos que compõem um item e seu processo de validação, inspirado em estudos e documentos oficiais do CAEd e do INEP.

As matrizes de referência também fizeram parte desse estudo. Foram trazidos como foco a matriz de Matemática do PROEB, com destaque para a área de Números e Operações, Álgebra e Funções. Realizou-se um debate em torno do uso indevido das matrizes de referência como currículos no processo de ensino, além de alguns aspectos relevantes sobre a importância do ensino e avaliação de Números e da Álgebra na Matemática.

Nesse estudo, realizou-se uma análise de nove itens eliminados de Matemática da área de Números e Operações, Álgebra e Funções do PROEB, aplicados entre os anos de 2014 a 2018. Com ajuda de especialistas da equipe de Matemática da CIA do CAEd, e com intuito de verificar se existem fatores ou práticas adotadas pela equipe de elaboração, que possam estar contribuindo para eliminação de itens, realizou-se em conjunto a análise dos itens uma pesquisa de campo, aplicado através de um questionário de pesquisa acadêmica.

Para a seleção desses nove itens, realizou-se o levantamento de todos os itens aplicados e utilizados nas avaliações do PROEB de 2014 a 2018, por meio de uma análise documental. Selecionados os itens, partiu-se para o estudo da psicometria nos itens, aprofundando o estudo sobre os métodos de correção clássica e de resposta ao item, acompanhados dos parâmetros estatísticos.

Tentou-se amenizar a complexidade que é interpretar os dados estatísticos de um item, trazendo uma abordagem menos técnica na tentativa de atingir grande parte dos profissionais da educação, além da busca em compreender o que significa um item eliminado em um teste, seguido de suas possíveis causas de eliminação.

Com a análise dos itens, foi possível identificar alguns padrões adotados no processo de elaboração de itens, que chamaram a atenção para uma discussão sobre novas possibilidades de pesquisa sobre itens e avaliação. Com a pesquisa de campo, foi possível identificar alguns fatores que nos levaram a propor novos modelos para a formação de futuros elaboradores das equipes da CIA, que oportunizam o aprofundamento dos conhecimentos estatísticos que envolvem os itens em avaliações educacionais.

Sendo assim, finalizou-se este trabalho propondo ações que serão apresentadas ao CAEd, ficando a critério da instituição, acatá-las ou não, mas que possuem o intuito de amenizar as implicações que giram em torno da eliminação de itens. Lembrando que as etapas finais dessa pesquisa foram realizadas em um período conturbado e atípico do ano de 2020, meio a pandemia de Covid-19, que em todos os aspectos, impossibilitou o contato direto com os atores envolvidos, impedindo talvez um aproveitamento mais inventivo e estreito com todos. Espera-se ainda que por meio deste trabalho, se possa inspirar novos estudos e pesquisas acerca da educação Matemática e da avaliação educacional em larga escala no Brasil, viabilizando formas de se conceberem rumos cada vez mais propícios ao desenvolvimento de uma educação pública e de qualidade em no país.

REFERÊNCIAS

- _____. **Promoção da Autonomia Escolar em Minas Gerais**. In: Cadernos Educação Básica, Série Inovações, Volume 2. Brasília, 1993. Disponível em: <<http://www.dominiopublico.gov.br/download/texto/me002334.pdf>>, Acesso em: 10 nov. 2019.
- AMERICAN EDUCATIONAL RESEARCH ASSOCIATION et al. **Standards for educational and psychological testing**. Amer Educational Research Assn, 2014.
- ARAUJO, Eutalia Aparecida Candido de; ANDRADE, Dalton Francisco de; BORTOLOTTI, Silvana Ligia Vincenzi. Teoria da resposta ao item. **Revista da Escola de Enfermagem da USP**, v. 43, n. SPE, p. 1000-1008, 2009.
- BEKMAN, Roberto M. Aplicação dos blocos incompletos balanceados na Teoria de Resposta ao Item. **Estudos em avaliação educacional**, n. 24, p. 119-136, 2001.
- BOLLELA, Valdes Roberto; BORGES, Marcos de Carvalho; TRONCON, Luiz Ernesto de Almeida. Avaliação Somativa de Habilidades Cognitivas: Experiência Envolvendo Boas Práticas para a Elaboração de Testes de Múltipla Escolha e a Composição de Exames. **Revista Brasileira de Educação Médica**, v. 42, n. 4, p. 74-85, 2018.
- BONAMINO, Alícia; FRANCO, Creso. Avaliação e política educacional: o processo de institucionalização do SAEB. **Cadernos de pesquisa**, n. 108, p. 101-132, 1999.
- BONAMINO, Alicia; SOUSA, Sandra Zákia. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. **Educação e Pesquisa**, v. 38, n. 2, p. 373-388, 2012.
- BRASIL. Base Nacional Comum Curricular. Brasília: MEC, 2017. Disponível em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=79601-anexo-texto-bncc-reexportado-pdf-2&category_slug=dezembro-2017-pdf&Itemid=30192>. Acesso em: 19 de março de 2020.
- BROOKE, Nigel (Org.). **Marcos históricos na reforma da educação**. Belo Horizonte: Fino Traço, 2012.
- BROOKE, Nigel; ALVES, Maria Teresa Gonzaga; DE OLIVEIRA, Lina Kátia Mesquita (Ed.). **A avaliação da educação básica: a experiência brasileira**. Fino Traço Editora, 2015.
- CAEd. **Guia de Elaboração de Itens**. Juiz de Fora: CAEd/UFJF, 2008.
- CAMPOS, Fernanda Cristina dos Santos. **Elaboração da prova do ENADE no modelo do banco nacional de itens**. Dissertação (mestrado profissional) - Universidade Federal de Juiz de Fora, Faculdade de Educação/CAEd. Programa de Pós-Graduação em Gestão e Avaliação da Educação Pública, 2013. 89 p.

COSTA, Helio Furtado. **PRODUÇÃO E REVISÃO DE ITENS DE MÚLTIPLA ESCOLHA**. 2014. 88 f. Dissertação (Mestrado) - Curso de Profmat – Mestrado Profissional em Matemática em Rede Nacional, Matemática, Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro, 2014. Disponível em: <https://impa.br/wp-content/uploads/2016/12/helio_f_costa.pdf>. Acesso em: 10 maio 2019.

COUTO, Gleiber; PRIMI, Ricardo. Teoria de resposta ao item (TRI): conceitos elementares dos modelos para itens dicotômicos. **Boletim de Psicologia**, v. 61, n. 134, p. 1-15, 2011.

DE ANDRADE, Dalton Francisco; TAVARES, Heliton Ribeiro; DA CUNHA VALLE, Raquel. Teoria da Resposta ao Item: conceitos e aplicações. **ABE, São Paulo**, 2000.

DE ANDRADE, Dalton Francisco; TAVARES, Heliton Ribeiro; DA CUNHA VALLE, Raquel. Teoria da Resposta ao Item: conceitos e aplicações. **ABE, Sao Paulo**, 2000.

DE ANDRADE, Josemberg Moura; LAROS, Jacob Arie; GOUVEIA, Valdiney Veloso. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. **Avaliação Psicológica**, v. 9, n. 3, p. 421-435, 2010.

DE OLIVEIRA HARDEN, Alessandra Ramos. Construindo Testes: Como Elaborar e Validar Itens de Múltipla Escolha. **Examen: Política, Gestão e Avaliação da Educação**, v. 1, n. 1, p. 7-7, 2017.

EDUCATIONAL TESTING SERVICE. ETS guidelines for fairness review of assessments. 2009.

FLETCHER, Philip R. A Teoria da Resposta ao Item: medidas invariantes do desempenho escolar. **Ensaio**, v. 1, n. 2, p. 21-27, 1994.

FLETCHER, Philip Ralph. Da teoria clássica dos testes para os modelos de resposta ao item. **Rio de Janeiro: Escola Nacional de Ciências Estatísticas**, 2010.

FONTANIVE, Nilma Santos; ELLIOT, Ligia Gomes; KLEIN, Ruben. Os desafios da apresentação dos resultados da avaliação de sistemas escolares a diferentes públicos. **REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 5, n. 2e, p. 262-273, 2007.

GARCÍA, Ma Isabel Barbero. Desarrollos recientes de los modelos psicométricos de la teoría de respuesta a los ítems. **Psicothema**, v. 11, n. 1, p. 195-210, 1999.

GATTI, Bernardete A. Avaliação de sistemas educacionais no Brasil. **Sísifo**, n. 9, p. 7-18/EN 7-18, 2016.

GATTI, Bernardete Andelina. Avaliação de sistemas educacionais no Brasil. **Sísifo . Revista de Ciências da Educação**, São Paulo, v. 1, n. 9, p.7-18, 2009. Quadrimestral. Disponível em:

<https://www.researchgate.net/profile/Bernardete_Gatti/publication/28320450_Avaliao_de_sistemas_educacionais_no_Brasil/links/56c7103708ae8cf82903ae8d/Avaliao-de-sistemas-educacionais-no-Brasil.pdf>. Acesso em: 12 out. 2019.

GOMES, Leonardo da Silva. A TEORIA DE RESPOSTA AO ITEM NA AVALIAÇÃO EM LARGA ESCALA: um estudo sobre o Exame Nacional de Acesso do Mestrado Profissional em Matemática em Rede Nacional – PROFMAT 2012. 2014. 78 f. Dissertação (Mestrado) - Curso de Mestrado Profissional em Matemática, Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro, 2014.

HALADYNA, Thomas M. **Developing and validating multiple-choice test items**. Routledge, 2004.

HALADYNA, Thomas M.; DOWNING, Steven M. **Handbook of test development**. Routledge, 2011.

INEP. **Guia de Elaboração e Revisão de Itens**: volume 1. Brasília: Daeb, 2010.

KLEIN, Ruben. Alguns aspectos da teoria de resposta ao item relativos à estimação das proficiências. **Ensaio: avaliação e políticas públicas em educação**, v. 21, n. 78, p. 35-56, 2013.

KLEIN, Ruben. Utilização da teoria de resposta ao item no Sistema Nacional de Avaliação da Educação Básica (Saeb). **Revista Meta: Avaliação**, v. 1, n. 2, p. 125-140, 2009.

LINS, Rômulo Campos. Modelo dos Campos Semânticos e Educação Matemática: 20 anos de história. **Org. Claudia Laus Angelo [et al.]**. São Paulo: Midiograf, 2012.

LOPES, K. T. **Uma investigação sobre o ensino de porcentagem no 6º ano do Ensino Fundamental**. 161 .f . Dissertação (Mestrado em Educação Matemática) – Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

MESSICK, Samuel. Validity. **ETS Research Report Series**, v. 1987, n. 2, p. i-208, 1987.

MISLEVY, Robert J.; ALMOND, Russell G.; LUKAS, Janice F. A brief introduction to evidence-centered design. **ETS Research Report Series**, v. 2003, n. 1, p. i-29, 2003.

MORAES, TATIANE GONÇALVES. **SISTEMA DE AVALIAÇÃO DO ESTADO DE GOIÁS (SAEGO): INTERPRETAÇÃO ESTATÍSTICA E PEDAGÓGICA DOS ITENS DE MATEMÁTICA**. Dissertação (mestrado profissional) - Universidade Federal de Juiz de Fora, Faculdade de Educação/CAEd. Programa de Pós Graduação em Gestão e Avaliação da Educação Pública. P. 161. 2017.

PASQUALI, Luiz. Validade dos testes. **Revista Examen**, Brasília, v. 1, n. 1, p.14-48, 2017. Semestral.

PASQUALI, Luiz; PRIMI, Ricardo. Fundamentos da teoria da resposta ao item: TRI. **Avaliação Psicológica: Interamerican Journal of Psychological Assessment**, v. 2, n. 2, p. 99-110, 2003.

PAULA, PATRÍCIA DE LIMA. **A INFLUÊNCIA DA ELABORAÇÃO DE ITENS NA EFICÁCIA DAS AVALIAÇÕES EM LARGA ESCALA**. Dissertação (mestrado profissional) - Universidade Federal de Juiz de Fora, Faculdade de Educação/CAEd. Programa de Pós Graduação em Gestão e Avaliação da Educação Pública. P. 100. 2018.

PEREIRA, Vinicius Ribeiro. **Métodos Alternativos no Critério Brasil para Construção de Indicadores Sócio-Econômico: Teoria da Resposta ao Item**. 2004. 103 f. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Elétrica, Departamento de Engenharia Elétrica do Centro Técnico Científico da Puc-Rio, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

PONTE, João Pedro da. Números e Álgebra no currículo escolar. **XIV EIEM- Encontro de Investigação em Educação Matemática**, p. 5-27, 2006.

PONTES, Luís Antônio Fajardo. Avaliação educacional em grande escala: a experiência brasileira. **DAVID, Marcus Vinícius et al. Avaliação e Indicadores Educacionais e Políticas Públicas e escola. Coleção Gestão e Avaliação da Educação Pública**, v. 2, p. 105-123, 2012.

QUADROS, Fernanda Maciel de. **Comparação da Teoria Clássica dos Testes com a Teoria de Resposta ao Item aplicado ao subteste semelhanças da WASI**. 2010. 18 f. TCC (Graduação) - Curso de Estatística, Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010. Disponível em: <<https://lume.ufrgs.br/handle/10183/24863>>. Acesso em: 05 jul. 2019.

RABELO, M. Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro. 1ª edição. **DRQ Gráfica e Editora-SBM-Sociedade Brasileira de Matemática. Rio de Janeiro**, 2013.

RIBEIRO, Alessandro Jacques; CURY, Helena Noronha. **Álgebra para a formação do professor: explorando os conceitos de equação e de função**. Autêntica, 2018.

SARTES, Laisa Marcórela Andreoli; SOUZA-FORMIGONI, Maria Lucia Oliveira de. Avanços na psicometria: da Teoria Clássica dos Testes à Teoria de Resposta ao Item. **Psicologia: Reflexão e Crítica**, [s.l.], v. 26, n. 2, p.241-250, 2013. FapUNIFESP (SciELO). Disponível em: <<http://dx.doi.org/10.1590/s0102-79722013000200004>>.

SCULLY, Darina. Constructing multiple-choice items to measure higher-order thinking. **Practical Assessment, Research, and Evaluation**, v. 22, n. 1, p. 4, 2017.

SOARES, Carlos Renato. **Sistemas de avaliações em larga escala na perspectiva histórico-cultural: o caso do Sistema Mineiro de Avaliação da**

Educação Pública – SIMAVE. 2011. 204f. Dissertação (Mestrado Profissional em Educação Matemática) – Universidade Federal de Juiz de Fora, Juiz de Fora, 2011.

SOARES, Tufi Machado; DE MENEZES GENOVEZ, Silene Felizardo; GALVÃO, Ailton Fonseca. Análise do Comportamento Diferencial dos Itens de Geografia: estudo da 4a série avaliada no Proeb/Simave 2001. **Estudos em Avaliação Educacional**, v. 16, n. 32, p. 81-110, 2005.

TENÓRIO, André; COSTA, Zélia de Souza Santos; TENÓRIO, Thaís. Resolução de exercícios e problemas de função polinomial do 1º grau com e sem o GeoGebra. **Revista do Instituto GeoGebra Internacional de São Paulo. ISSN 2237-9657**, v. 3, n. 2, p. 104-119, 2014.

TOCANTINS, Andreia Cristina Teixeira. **Avaliações em Larga Escala em Ciências Humanas no Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz De Fora (CAEd/UFJF): A Construção de um Guia de Orientação do Trabalho de Analistas e Auxiliares de Instrumentos de Avaliação em Questão**. Dissertação (mestrado profissional) - Universidade Federal de Juiz de Fora, Faculdade de Educação/CAEd. Programa de Pós-Graduação em Gestão e Avaliação da Educação Pública. P.183. 2018.

TOFFOLI, Sônia Ferreira Lopes et al. Avaliação com itens abertos: validade, confiabilidade, comparabilidade e justiça. **Educação e Pesquisa**, v. 42, n. 2, p. 343-358, 2016.

TOFFOLI, Sônia Ferreira Lopes; ANDRADE, Dalton Francisco de; BORNIA, Antônio Cezar; QUEVEDO-CAMARGO, Gladys. Avaliação com itens abertos: validade, confiabilidade, comparabilidade e justiça. **Educação e Pesquisa**, [s.l.], v. 42, n. 2, p.343-358, jun. 2016. FapUNIFESP (SciELO). Disponível em: <<http://dx.doi.org/10.1590/s1517-9702201606135887>>.

VIANNA, Heraldo Marelim. Avaliação Educacional: uma perspectiva histórica. **Estudos em Avaliação Educacional**, n. 12, p. 7-24, 1995.

YIN, Robert K. **Pesquisa qualitativa do início ao fim**. Penso Editora, 2016.

YIN, Robert K. **Estudo de Caso: Planejamento e métodos**. Bookman editora, 2015.

APÊNDICE A – Questionário de pesquisa acadêmica

	O(A) Sr.(a) concorda que o material coletado possa ser utilizado em outros projetos do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora, sendo assegurado que sua identidade será tratada com padrões profissionais de sigilo, atendendo a legislação brasileira, utilizando as informações somente para os fins acadêmicos e científicos? Caso sua manifestação seja positiva, esta autorização poderá ser retirada a qualquer momento sem qualquer prejuízo.	Declaro que concordo em participar da pesquisa e que me foi dada a oportunidade de ler e esclarecer as minhas dúvidas.	Qual é a sua graduação? Ex.: Licenciatura em História	Possui alguma especialização a nível lato-sensu? Ex.: Educação - Filósofica - Concluída	Possui especializações a nível stricto-sensu? Ex.: Mestrado Acadêmico em Estatística - Concluído
Entrevistado 1 (E1)	Sim	Concordo	Licenciatura em Matemática	Não.	Mestrado Acadêmico em Educação
Entrevistado 2 (E2)	Sim	Concordo	Licenciatura Plena em Matemática	Métodos Estatísticos Computacionais	Mestrado Profissional em Educação Matemática
Entrevistado 3 (E3)	Sim	Concordo	Licenciatura Plena em Matemática	Não	Não
Entrevistado 4 (E4)	Sim	Concordo	Licenciatura Plena em Matemática	Não	Mestrado em Gestão e Avaliação da Educação Pública
Entrevistado 5 (E5)	Sim	Concordo	Licenciatura Plena em Matemática	Não	Não

	Qual a sua função no CAEd?	A quanto tempo você atua no CAEd? Ex.: 9 anos	Antes de atuar no CAEd, já havia trabalhado com avaliação educacional em larga escala? Caso afirmativo, relate brevemente suas experiências.	Já atuou em outros departamentos do CAEd?	Ao ser admitido no CAEd, houve algum tipo de treinamento oferecido pela instituição? Comente brevemente como foi o início de suas atividades.
(E1)	Analista de Instrumentos de Avaliação III	3 anos	Não, embora na época cursava mestrado e pesquisava sobre as avaliações externas no Estado de Minas Gerais.	Não	Sim, com alguns membros mais experientes da equipe de Matemática.
(E2)	Analista de instrumento de avaliação	4 anos	Não	Não	Sim. Duas semanas de palestra e elaboração de itens.
(E3)	Trabalho na equipe de modelagem de testes, escolhendo itens, montando avaliações, preparando materiais para boletins entre outros.	4 anos	Não	Sim	Sim, iniciei meu trabalho no CAEd na elaboração de itens. Antes de ser admitida participei de uma oficina de itens, com duração de 4 dias, em que a última etapa avaliativa era a entrega de um item.
(E4)	Analista de instrumentos de avaliação	9 anos	Não	Não	Não houve treinamento especificamente apenas uma espécie de monitoramento por parte de analistas mais antigos.
(E5)	Analista de Instrumentos de Avaliação e Pesquisa	4 anos	Não	Sim	Sim. Durante a última etapa do processo seletivo que participei, houve um treinamento para elaboração de itens. Após ter sido aprovada, na primeira semana houve um treinamento, de elaboração e algumas palestras sobre TRI. Ao longo dos anos, alguns pesquisadores e professores que trabalham com projetos específicos, voltado para avaliação, compartilharam suas experiências. Em 2018, participei de um treinamento em Luxemburgo, para manipular uma plataforma de avaliação digital.

	O CAEd promove momentos formativos para seu desempenho e aperfeiçoamento na execução de suas tarefas? Caso afirmativo, qual a frequência e como são essas atividades?	Já atuou como professor? Se sim, diga qual/quais disciplina(s) e em quais segmentos lecionou?	Quais documentos norteiam sua elaboração/revisão de itens no CAEd?	Tem conhecimento sobre o manual de elaboração e revisão de itens do CAEd?	Já procurou pesquisar sobre os manuais de outras instituições?
(E1)	Sim.	Sim. Já atuei por 3 anos como professor no ensino fundamental II, ensino médio e em cursos preparatórios para o Vestibular/ENEM.	Guia de elaboração, acordos internos firmados e orientações já existentes na equipe.	Sim	Não
(E2)	Sim. Seminários, uma a duas vezes ao ano.	Sim. Matemática e Geometria. Educação Básica e EJA	Guia de Elaboração de itens	Sim	Não
(E3)	Sim. Sempre que há a necessidade para alguma nova demanda, em devolutivas das equipes de pesquisa, para adequações de algumas funções e em eventos e congressos voltados para o tema. Esses momentos que já eram	Minhas únicas atuações em sala de aula se deram durante a graduação, em duas bolsas de treinamento profissional realizadas no colégio de aplicação João XXIII no 5º e 6º ano do ensino fundamental com duração de 1 ano aproximadamente cada uma, porém também houveram dias em que fui solicitada para substituir a professora em salas do 8º ano do fundamental.	Minhas elaborações e revisões sempre se pautaram no guia de elaboração e revisão do CAEd, documento este que atualmente está sendo revisado. Apesar de que existem padrões de elaboração e revisão para além deste guia.	Sim	Sim
(E4)	Nós mesmos, dentro da equipe preparamos, quando possível, encontros que acabam sendo de formação ao objetivar discussões acerca de itens ou classes que necessitem ser revistos. Tentamos fazer uma vez por semana, mas na maioria das vezes o trabalho não torna possível.	Atuei e atuo como professora no Ensino Fundamental e Médio lecionando Matemática.	Atualmente nenhum. Temos um em construção, pois um antigo que temos está muito ultrapassado.	Não	Sim
(E5)	Sim. As atividades de elaboração e revisão de itens eram discutidas em reuniões coletivas, sempre com a participação de pelo menos um professor da universidade. Durante aproximadamente 18 meses elas ocorriam semanalmente, agora são mais pontuais, sempre quando solicitamos a discussão de um tema específico.	Sim. Atualmente atuo como professora de Matemática da rede estadual de educação de Minas Gerais.	Atualmente, o manual de elaboração e revisão e a BNCC.	Sim	Sim

O que você entende por um item eliminado?	Cite um exemplo, pedagógico ou estatístico que pode eliminar um item em um teste ou pré-teste?	Acredita que através de itens eliminados seja possível obter algum tipo de aprendizado sobre a elaboração de itens ou diagnóstico de aprendizagem?	O que você entende por uma eliminação de item por motivo de ajuste?
(E1) Item que por algum motivo técnico, pedagógico ou estatístico não cumpriu com o seu papel de avaliar em um teste em larga escala.	Alternativas que chamem muito a atenção do estudante (pegadinhas), duas bisseriais positivas (o que pode indicar gabarito errado), parâmetro C muito alto, entre outros.	Sim	Alguma alternativa chamando a atenção, alguma pegadinha etc.
(E2) Um item que não tenha se comportado bem quanto às estatísticas ou com erro pedagógico.	Gabarito sinalizado incorretamente, erro conceitual.	Sim	Não sei o que isso significa.
(E3) É um item que não atendeu de alguma forma (pedagógica ou estatisticamente) ao seu propósito avaliativo.	Um exemplo de motivo de eliminação para um item pode ser, por exemplo, a ausência de um gabarito.	Sim	Acredito que a pergunta poderia ser melhor elaborada, mas imagino que se trate da eliminação estatística de adequação ao grupo avaliado.
(E4) Um item que não se comporta como esperado no teste sendo considerado inabilitado na avaliação da habilidade para a qual foi proposto.	Um distrator que por algum motivo atrai mais os estudantes do que o gabarito do item pode fazer ele ser eliminado assim como, estatisticamente, um item que não discrimina bem os estudantes que sabem ou não a habilidade.	Sim	Talvez que o item possa ser ajustado a partir do motivo que o fez ser eliminado para que possa estar mais adequado para a avaliação da habilidade para a qual foi proposto. Não necessitando ser completamente descartado.
(E5) Acredito que um item eliminado seja um item que de alguma forma não atenda à expectativa do teste. Isso pode acontecer tanto estatisticamente quando pedagogicamente.	Pedagogicamente, um item pode ser eliminado por conta da linguagem utilizada (não apropriada para a etapa), por conta de distratores mal construídos, um suporte que não ofereça informações ou que ofereça informações em excesso. Estatisticamente, algum dos parâmetros dos itens pode levar a eliminação deste. Como discriminação ou dificuldade.	Sim	É uma eliminação estatística dos itens, quando o item não se comporta com o esperado.

O que você entende por uma eliminação de item por motivo da bisserial?	O que você conhece sobre uma eliminação de item por comportamento diferencial (DIF)?		
(E1)	Duas bisseriais positivas (o que pode indicar gabarito errado).	São itens em que as probabilidades de acertos diferem entre distintos grupos. Por exemplo, um item tem um parâmetro de dificuldade de 280 no Sul do país, e para um grupo similar a esse na região norte o mesmo item tem dificuldade 600.	<p>Nesta seção, serão apresentados 9 itens para serem analisados, verificando os fatores que possivelmente tenham influenciado sua eliminação no PROEB entre os anos de 2014 a 2018. Lembre-se da confidencialidade sobre as imagens e quaisquer outros dados dos itens.</p> <p>Em cada um dos itens, responda, as seguintes questões:</p> <p>(Obs.: Identifique suas respostas no campo abaixo de cada um dos quadros)</p> <p>1º Questão: A partir do histórico de utilização do item no PROEB e seus respectivos dados estatísticos, qual/quais motivo(s) podem ter ocasionado a eliminação do item?</p> <p>2º Questão: Sobre os distratores, eles são coerentes? Você reconhece alguma prática comum de elaboração na construção desses distratores?</p> <p>3º Questão: Observando os dados estatísticos, como você avalia o comportamento do item no teste?</p>
(E2)	Significa o estudo do comportamento das marcações das alternativas em relação aos estudantes que tiveram alto índice de acerto no teste.	São itens que apresentam grau de dificuldade diferente em relação ao público a que foi submetido.	
(E3)	A bisserial indica a tendência de resposta dos estudantes, quando essa tendência, por exemplo, revela que uma resposta errada está se comportando como gabarito e que os bons estudantes (que possuem os melhores padrões de resposta) estão marcando-a, é um motivo de eliminação. Podendo indicar, por exemplo, a inadequação daquele item para o grupo ao qual foi destinado.	É uma eliminação do item por apresentar comportamentos diferentes quando utilizado para diferentes grupos.	
(E4)	Indica que possivelmente os estudantes que demonstram estarem além da habilidade estão errando o item, ou seja, possivelmente tem um distrator que atrai esses estudantes, influenciando o erro.	Confesso que desconheço esse motivo.	
(E5)	Significa que o item não tem uma boa discriminação e assim, não podemos concluir que um acerto desse item pelo indivíduo irá garantir um bom resultado no teste.	Acredito que essa eliminação ocorre quando o item se comporta de forma diferente quando aplicado em grupos de indivíduos diferentes, com a mesma proficiência. Esse comportamento é a probabilidade de acerto.	

Item 1	
(E1)	Neste item temos duas bisseriais positivas, ou seja, bons respondentes deram como resposta as alternativas D e.
(E2)	<p>1º Questão: Muitos alunos que obtiveram bons resultados no teste, se confundiram e marcaram a alternativa E como resposta.</p> <p>2º Questão: Acho os distratores coerentes. Geralmente colocam-se expressões em que pelo menos um dos pontos do gráfico faça sentido.</p> <p>3º Questão: Não avaliou claramente, visto que bons alunos marcaram a E. Esse tipo de item não é fácil.</p>
(E3)	<p>1ª – Uma possibilidade para a eliminação deste item pode estar relacionada a bisserial positiva da alternativa E, mostrando que "bons" estudantes estão sendo atraídos para ela.</p> <p>2ª – Uma das formas que os estudantes resolvem esse tipo de item é pela substituição de pontos do gráfico nas alternativas, o que já indica uma análise das mesmas, o que é inadequado na concepção de item. Percebe-se também uma tentativa na construção dos distratores A e B de identificar os estudantes que não conhecem a curva do gráfico e a lei de formação de uma função exponencial, e que foram atraídos ao encontrar uma equivalência na substituição de algum ponto.</p> <p>Acredito que os "bons" estudantes que foram atraídos para o distrator E, conhecem o gráfico e a forma da função exponencial, porém podem indicar uma defasagem na operação de potenciação, também muito comum, entendendo que 3 elevado a 0 é igual a 0.</p> <p>Já vi em muitos itens esse tipo de construção dos distratores, porém o mais indicado na construção dos distratores é que se baseiem em erros pautados na construção da habilidade avaliada, e não em erros que teoricamente já deveriam ter sido superados em etapas anteriores.</p> <p>3ª – Apesar do item, em partes, não avaliar de fato o que se propõe, ele traz uma riqueza em sua análise, podendo inclusive denunciar defasagens anteriores que não foram superadas.</p>
(E4)	Outra alternativa além do gabarito com bisserial positiva. O distrator E foi mais marcado que o gabarito. Apesar de coerentes, o distrator da alternativa C destoou um pouco dos demais por conter a base fracionária e o da alternativa E chamou atenção pelos dados do gráfico, uma vez que essa habilidade, pela minha experiência, muitas vezes não é plenamente consolidada pelos alunos.
(E5)	<p>1) Quanto ao histórico, percebo que o percentual de acerto do item praticamente se manteve constante e baixo (~17%), mesmo sendo aplicado em etapas distintas. Isso é um indicativo que o item possa ter algum problema ou que de fato os estudantes não conhecem ou não estão sabendo associar o gráfico de funções exponenciais à representação algébrica, uma vez que os estudantes do 1EM e do 3EM não estão acertando.</p> <p>2) Analisando os distratores, não percebo incoerência na construção, pelo contrário, o estudante que reconhece o gráfico de uma função exponencial crescente, poderia inclusive eliminar 3 das 5 alternativas, e isso não acontece quando percebemos os percentuais de marcação do gabarito e da alternativa E. De fato, a alternativa E atraiu mais respondentes, talvez por ser a única exponencial crescente que apresenta os valores assinalados no gráfico. É uma prática comum na elaboração.</p> <p>3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa E também é positiva e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelo distrator E.</p>

Item 2	
(E1)	Assim como no item anterior, aqui temos duas bisseriais positivas, ou seja, bons respondentes deram como resposta as alternativas A e B. Mas compreendo neste caso que tal prática ainda é comum, visto que tais alternativas são similares, uma corresponde a um gráfico crescente e outro decrescente.
(E2)	<p>1º Questão: Os alunos ficaram em dúvidas entre as alternativas A e B.</p> <p>2º Questão: Sim. Geralmente, tenta-se apresentar gráficos de funções mais comuns aos alunos.</p> <p>3º Questão: Foi pouco acertado. Além disso, a maioria marcou a opção B e os demais ficaram igualmente divididos entre A, C, D e E. Não é um item muito fácil, tem pré-requisitos que o estudante precisaria dominar.</p>
(E3)	<p>1ª – Podemos seguir nesse item uma linha de análise bastante semelhante ao item anterior. Uma possibilidade para a eliminação deste item pode estar relacionada a bisserial positiva da alternativa B, mostrando que "bons" estudantes estão sendo atraídos para ela.</p> <p>2ª – Não é possível visualizar a lei de formação apresentada no item, mas imagino que o erro da alternativa B se baseie em algum erro de sinal, considerando a reflexão do gabarito em relação ao eixo das ordenadas.</p> <p>Também já vi em muitos itens esse tipo de construção dos distratores, porém o mais indicado na construção dos distratores é que se baseiem em erros pautados na construção da habilidade avaliada.</p> <p>3ª – O comportamento do item nas duas etapas avaliadas mostra um padrão semelhante, mostrando que as defasagens que levaram os "bons" estudantes do 1EM a considerar como correta a alternativa B permaneceram nos estudantes do 3EM.</p>
(E4)	Novamente temos duas bisseriais positivas. O distrator da alternativa B atraiu mais da metade dos respondentes. Julgo os distratores plausíveis e acredito que essa eliminação tem mais a ver com questões pedagógicas do que técnicas. Demonstra que os estudantes, quando se trata de função exponencial, tentam buscar no gráfico os dados que levam à solução.
(E5)	<p>1) Quanto ao histórico, percebo que o percentual de acerto do item praticamente se manteve constante e baixo (~12%), mesmo sendo aplicado em etapas distintas. Observe que é esperado que pelo menos no 3EM esse percentual seja maior do que foi no 1EM.</p> <p>2) Analisando os distratores, não percebo uma incoerência na construção. O que me chama a atenção é o fato do contradomínio da função ser os reais não negativos e mesmo assim cerca de 40% (total) dos respondentes marcaram uma das 3 alternativas que isso não acontece (C, D ou E). A alternativa B chamou a atenção dos respondentes, mais do que o gabarito, talvez por apresentar um expoente fracionário os estudantes têm associado ao gráfico decrescente.</p> <p>3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa B também é positiva e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelo distrator B.</p>

Item 3	
(E1)	Mesmo comentário dos itens anteriores, possivelmente os itens foram eliminados pelo critério das bisseriais.
(E2)	<p>1º Questão: Os alunos não tiveram muito critério, marcaram, quase de maneira igual uma alternativa.</p> <p>2º Questão: Os distratores são coerentes. Geralmente, procura-se pelas funções parecidas a que é apresentada no suporte.</p> <p>3º Questão: Item pouco acertado. Os alunos menos marcaram o gabarito. Mas, de modo geral, foi distribuída a marcação entre as alternativas, sem muito critério, ou considerando apenas um.</p>
(E3)	<p>1ª – Da mesma forma que nos itens anteriores esse também apresenta duas alternativas com bisseriais positivas, mostrando que o distrator A chamou a atenção dos "bons" estudantes.</p> <p>2ª – Muito complicado ficar tentando descobrir em que erro foi baseado um distrator, mas claramente o distrator da letra A chamou a atenção de muitos "bons" estudantes, o que não deveria acontecer.</p> <p>3ª – O item apresentou um comportamento inadequado.</p>
(E4)	Os alunos se distribuíram entre as alternativas. Novamente existem duas bisseriais positivas e um distrator chamando mais atenção do que o gabarito. Acredito mais uma vez que as razões sejam mais técnicas do que pedagógicas uma vez que os estudantes escolhem marcar a alternativa que lembra a função seno padrão, demonstrando que talvez não saibam lidar com as composições das funções trigonométricas. Julgo que os distratores estão de acordo com a habilidade avaliada.
(E5)	<p>1) Em linhas gerais, percebo que o percentual de acerto do item é baixo (~13%), a bisserial é menor do que 0,3 e que houve uma distribuição "homogênea" na marcação das alternativas, e isso pode ter ocasionado a eliminação do item.</p> <p>2) Analisando os distratores, todos são compostos por funções trigonométricas que de fato podem causar dúvida, talvez por isso apresentem percentual de marcação semelhante. Uma prática comum, por exemplo, é abordar a função sem suas variações, como $f(x) = \sin(x) + 1$ apresentada na alternativa A.</p> <p>3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa A também é positiva e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelo distrator A.</p>

Item 4	
(E1)	Este item apresenta duas bisseriais positivas. Além disso trata-se de um item de equação do segundo grau, em que o estudante deve modelar uma equação para solucioná-lo, o que não é algo simples. Pelo que vejo dos itens eliminados, é comum tal habilidade estar presente, sobretudo nesses casos que exige do estudante modelar algebricamente o problema.
(E2)	1º Questão: O enunciado ser um pouco extenso, com informações que exigiam a algebrização por parte do aluno. 2º Questão: Os distratores são coerentes. O estudante apenas manipular os dados do problema ou confundir quadrado com raiz, e ainda, quadrado com o dobro ou metade. 3º Questão: Os alunos, possivelmente, não montaram ou não souberam montar a equação e resolvê-la. Marcaram mais as alternativas que se referiam à metade, que era uma das informações do problema, e um dos dados informados.
(E3)	1ª – Da mesma forma que nos itens anteriores esse também apresenta duas alternativas com bisseriais positivas, porém nesse caso a maior bisserial não é a do gabarito, ou seja, essa alternativa se comportou como gabarito. 2ª – A escolha dos "bons" estudantes pelo distrator A pode indicar uma escolha massiva da primeira alternativa, ou então uma grande defasagem na habilidade avaliada, visto que o estudante que marca como correta a metade da distância total percorrida mostra um desentendimento do problema sugerido, se apegando a apenas algumas informações apresentadas. O que costuma ser comum na construção de distratores em itens que também avaliam a resolução de problemas. 3ª – O item apresentou um comportamento inadequado.
(E4)	Nesse caso houve um distrator que chamou atenção de mais da metade dos estudantes, se tornando o segundo com bisserial positiva esse distrator atraiu estudantes que não equacionaram o problema. Apenas dividiram a distância por dois. Acredito que esse item também foi condenado pela dificuldade na modelagem. Creio que o distrator seja plausível e que o problema não seja muito técnico uma vez que o gabarito Não foi nem o segundo mais marcado. Caso ele não estivesse lá, parece que os estudantes optariam pelo 364.
(E5)	1) Acredito que esse item tem um baixo percentual de acerto pelo fato de abordar uma modelagem de equação de 2º grau, e por exigir do respondente uma boa manipulação algébrica, já que envolve raiz quadrada de 2916. A eliminação desse item deve estar associada a uma baixa discriminação, uma vez que a bisserial está muito baixa, quase nula. 2) Analisando os distratores, percebo que eles poderiam explorar outras possibilidades de erro, que envolvessem erro na modelagem ou até mesmo na interpretação da situação-problema, não apenas uma mera manipulação dos dados do problema. 3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa A também é positiva e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelo distrator A. E isso ao me ver é um problema, até mesmo pedagógico, uma vez que o, possivelmente, os respondentes apenas efetuaram a divisão de 364 por 2.

Item 5	
(E1)	Note que tal item apresenta um baixo percentual de acerto, além disso duas alternativas apresentam bisseriais iguais e positivas, o que indica que bons respondentes se sentiram atraídos pela alternativa A.
(E2)	<p>1º Questão: O desconhecimento em relação à habilidade de determinar expressão algébrica de uma função.</p> <p>2º Questão: Sim são coerentes. Geralmente, apresenta-se expressões algébricas com os dados do gráfico e que tenham algum ponto que satisfaça ao gráfico.</p> <p>3º Questão: A alternativa que eles menos marcaram foi o gabarito. Optaram mais pelas alternativas que tinham algum número diretamente relacionado aos que apareceram no gráfico.</p>
(E3)	<p>1ª – O item apresenta 3 bisseriais positivas, sendo uma delas maior que a do gabarito.</p> <p>2ª – O distrator mais marcado e comumente utilizado é aquele em que o estudante utiliza os pontos de interseção da função com os eixos, como coeficientes, salvo algumas alterações de sinal. O outro distrator que apresentou uma bisserial positiva se pautou em um erro de sinal, considerado para a análise de crescimento/decrescimento da função, também muito utilizado.</p> <p>3ª – O item apresentou comportamento inadequado.</p>
(E4)	Considero o item tecnicamente bom. Os distratores estão de acordo porém o gabarito foi o menos marcado, gerando três bisseriais positivas. Acredito que, mais uma vez, o item entregue que os estudantes procuram a resolução nos dados do item, demonstrando que, de fato estão cometendo erros dentro da habilidade. Uma pequena minoria está a par do procedimento para a obtenção da resposta. Em termos de técnica de elaboração, não consigo ver como esse item poderia ser diferente do apresentado.
(E5)	<p>1) Esse item apresentou outras duas bisseriais positivas além do gabarito. Sendo assim, ou o item tem de fato um problema pedagógico ou os estudantes desconhecem de fato a habilidade.</p> <p>2) Analisando os distratores, percebo que eles exploram manipulações de sinais, mas me chamou a atenção o distrator D ter sido muito marcado, cerca de 36%. Isso revela claramente que muitos estudantes não sabem sequer associar o gráfico de uma função do 1º grau decrescente ao sinal do coeficiente angular. Acho que é comum abordar essas manipulações de sinais e considerar os coeficientes como sendo a abscissa e a ordenada dos pontos de interseção com os eixos.</p> <p>3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, as bisseriais das alternativas D e E também foram positivas e a bisserial do gabarito não é maior que 0,3. Isso quer dizer que indivíduos com boa nota no teste se "atraíram" pelos distratores D e E.</p>

Item 6	
(E1)	Era om para analisarmos este item conhecer a ancoragem do item em sua primeira aplicação, pois este pode ser um caso em que o item tenha sido eliminado pelo seu comportamento diferencial (DIF), visto que as bisseriais entre as duas aplicações se diferem um pouco. As alternativas D e E, que apresentaram bisseriais positivas atraíram bons respondentes, e o distrator D é o mais comum de utilizarmos nestes itens, usados até hoje em nossas elaborações, quando o estudante confunde os sinais das raízes na forma fatorada de um polinômio.
(E2)	1º Questão: Desconhecimento da habilidade envolvida. 2º Questão: Sim, eles são coerentes. Alternar entre os sinais das raízes do polinômio. 3º Questão: Um item relativamente fácil, mas que os estudantes marcaram sem nenhum critério as alternativas.
(E3)	1ª – No 1º quadro a discrepância entre as bisseriais positivas é menor que no segundo, o que mostra que no 1EM o distrator se comportou como gabarito e no 3EM o distrator apenas chamou mais atenção que os demais distratores. 2ª – O distrator que gerou essas discrepâncias de bisserial mostra que o estudante que o marcou entende que na forma fatorada a raiz do polinômio é o valor que acompanha x. Distrator também muito comum. 3ª – O item apresentou além do comportamento inadequado da bisserial no primeiro quadro, também apresentou um comportamento diferente em grupos diferentes, o que também não deveria acontecer.
(E4)	Nesse item, quando eliminado, parece que a bisserial da alternativa D superou a do gabarito. As estatísticas de quando não foi eliminado demonstram que a habilidade torna o item de alto nível na escala, ou seja, muito difícil para os alunos. A maioria desconsidera o zero como raiz e outros apesar de considerar o zero como raiz, troca o sinal dos fatores. Não vejo problema técnicos com os distratores. Trata-se de uma habilidade bem direta...
(E5)	1) Esse item apresentou no 1EM outra bisserial positiva além do gabarito e igual a 0,120. Sendo assim, ou o item tem de fato um problema pedagógico ou os estudantes desconhecem de fato a habilidade. 2) Analisando os distratores, percebo que eles exploram manipulações de sinais, relacionados às raízes do polinômio. O distrator D chamou a atenção dos respondentes por apresentar apenas os valores do enunciado. A presença desse distrator e a seu percentual de marcação mostram que os estudantes não conhecem essa "tarefa"/"habilidade". É usual a presença desse distrator na elaboração. 3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa D também foi positiva e a bisserial do gabarito, embora não tenha sido maior que 0,3, foi próximo e portanto "aceitável".

Item 7	
(E1)	Analisando os parâmetros A, B e C do item nota-se que eles são próximos, então tal item não deve ter sido eliminado pelo critério do DIF. Um fator que pode ter sido o da eliminação deste item corresponde ao fato de em algumas aplicações duas alternativas apresentarem bisseriais positivas, o que indica que alguma alternativa possa ter chamado mais a atenção do que outras. Neste caso trata-se da alternativa B, que é muito comum de ser utilizada quando o aluno desconhece o significado do símbolo "%".
(E2)	1º Questão: Os estudantes não reconhecerem o conceito de porcentagem. 2º Questão: Sim, eles são coerentes. Apresentar alternativas em que o estudante confunde desconto com acréscimo e porcentagem com o valor absoluto. 3º Questão: Um item relativamente fácil, mas que mostrou que os estudantes, a maioria deles, procurou pela resposta do desconto absoluto, não reconhecendo o conceito de porcentagem.
(E3)	1ª – O item apresenta comportamento semelhante em mesma etapa, porém diferente em etapas diferentes. No 5EF vemos uma maior atração dos estudantes "bons" para a alternativa B. 2ª – Esse comportamento diferencial, evidencia um distrator também muito usual, aquele em que o estudante soma ou subtrai, de acordo com o contexto, a porcentagem ao invés de fazer o cálculo da mesma. 3ª – O item apresentou além do comportamento inadequado da bisserial no 5EF, também apresentou um comportamento diferente em grupos diferentes, o que também não deveria acontecer.
(E4)	Vejo que mais da metade dos estudantes escolheu a alternativa B demonstrando não estar a par do cálculo das porcentagens mas não consegui identificar o porquê o item teria sido eliminado.
(E5)	1) Não compreendi a eliminação do item, analisando apenas o histórico apresentado, sem os parâmetros pela TRI, uma vez que ele se comportou de forma semelhante em 2016 (9EF). Talvez fosse necessário apresentar os escores dos diferentes grupos que compõe o teste. 2) Analisando os distratores, distrator B chamou a minha atenção pois mais de 50% dos respondentes consideram que um desconto de 10% é de R\$ 10,00. Os distratores são de naturezas diferentes e plausíveis. Talvez, para compor, e para o próprio gabarito não chamar a atenção, o que não aconteceu, apresentaria outra alternativa com centavos. É usual a presença desses distratores na elaboração. 3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso pra garantir uma boa discriminação dos indivíduos respondentes. Nesse caso, a bisserial da alternativa B também foi positiva e a bisserial do gabarito, embora não tenha sido maior que 0,3, foi próximo e portanto "aceitável" (9EF de 2016).

Item 8	
(E1)	Esta é uma habilidade muito comum de se ver itens sendo eliminados, ainda mais quando se tem uma alternativa em que o numerador da fração corresponde à parte inteira do número e o denominador corresponde à parte decimal. Vale destacar que tal formato que o item foi utilizado já não corresponde aos padrões atuais de elaboração, quando o enunciado está junto ao comando.
(E2)	<p>1º Questão: Os estudantes marcaram mais as alternativas em que aparecia um dos números do enunciado.</p> <p>2º Questão: Sim, são coerentes. Apresentar a composição do numerador com o denominador e a divisão do denominador pelo numerador.</p> <p>3º Questão: Item relativamente fácil, mas que se apresentou de forma difícil, visto que os estudantes apenas procuram as alternativas onde havia um dos números do problema. Não reconhecendo conceito de fração e número decimal.</p>
(E3)	<p>1ª – O item apresenta comportamento diferente entre as etapas, sendo o distrator D mais atrativo para o 7EF.</p> <p>2ª – Esse comportamento diferencial, evidencia um distrator também muito usual, aquele em que o estudante utiliza o numerador da fração como parte inteira e o denominador como parte decimal, sendo bem mais atrativo para a etapa 7EF.</p> <p>3ª – O item apresentou nesse caso apenas um comportamento diferente entre as etapas avaliadas, o que também é inadequado.</p>
(E4)	Ao meu ver o item estaria de acordo com as técnicas e não consigo identificar um motivo pelo qual teria sido eliminado.
(E5)	<p>1) Não compreendi a eliminação do item, analisando apenas o histórico apresentado, sem os parâmetros pela TRI. Apenas com a análise clássica não é possível fazer tal inferência.</p> <p>2) Analisando os distratores, percebo que eles estão abordando o que é usual na técnica de elaboração de itens e no 7EF, me chamou a atenção o percentual de marcação da alternativa D, cerca de 44% dos respondentes apenas acrescentaram a vírgula entre os números de numerador e denominador.</p> <p>3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas. Isso de fato ocorreu e, embora a alternativa D tenha sido muito marcada, ela não foi marcada por respondentes que tiveram uma boa "nota" (escore) no teste.</p>

Item 9	
(E1)	Analisando tal item não consigo identificar possíveis motivos para tal item ser eliminado, pois apresentam bisseriais próximas nas duas aplicações.
(E2)	<p>1º Questão: O fato de ele ter sido mais acertado em uma etapa do que em relação à outra.</p> <p>2º Questão: Sim, eles são coerentes. Apresentar os algarismos das outras ordens do número.</p> <p>3º Questão: Esse item foi mais bem acertado no 5EF do que no 9EF, quando, teoricamente, não pode haver esse comportamento diferencial. Além disso, os estudantes do 9EF, depois do gabarito, marcaram qualquer alternativa, sem critério. Enquanto no 5EF eles ficaram mais em dúvida entre B e D.</p>
(E3)	<p>1ª – O item apresenta comportamento diferenciado em cada etapa.</p> <p>2ª – Os distratores apresentados são comumente utilizados, pautados nas diferentes ordens do número.</p> <p>3ª – O item apresentou um comportamento diferente entre as etapas avaliadas, o que é inadequado.</p>
(E4)	Única coisa que penso é que, como a maioria dos estudantes (mais de 50%) marcaram a alternativa correta, o item pode ter se tornando ineficaz para diferenciar os alunos que sabem daqueles que não sabem. Tornou-se um item que não diferencia bem os estudantes.
(E5)	<p>1) Não compreendi a eliminação do item, analisando apenas o histórico apresentado, sem os parâmetros pela TRI. Suponho que esse item tenha sido tão fácil para a etapa que ele não discriminou os respondentes.</p> <p>2) Analisando os distratores, percebo que eles estão abordando o que é usual na técnica de elaboração de itens e o que é possível (os outros algarismos que compõe o número em questão).</p> <p>3) O esperado é que o item apresente uma bisserial positiva (bisserial do gabarito) e maior que 0,3 e, que as demais sejam negativas e isso de fato ocorreu. Suponho que esse item tenha sido tão fácil para a etapa que ele não discriminou os respondentes.</p>

Apêndice B – Plano de ação educacional proposto

Ao Centro de Políticas Públicas e Avaliação da Educação

Por meio deste, apresenta-se esta proposta contendo um conjunto de ações que visam minimizar a eliminação de itens devido a práticas comuns adotadas pelos elaboradores da equipe Matemática da CIA, além de uma sugestão de capacitação inicial que poderá ser oferecida a novos colaboradores, quando necessário.

A partir dos fatores levantados nessa pesquisa, propõe-se atuar sob dois eixos centrais, que são: Eixo 1 - Estimular a pesquisa com itens no intuito de reduzir as eliminações e buscar formas de otimizar o trabalho da elaboração; Eixo 2 - Aprimorar os aspectos formativos da equipe;

A luz desses dois eixos centrais, detalhamos ações com o auxílio da ferramenta 5w2h, que possibilita-nos definir metodologicamente, as ações que trouxemos como proposta.

Eixo 1 – Estimular a pesquisa com itens no intuito de reduzir as eliminações e otimizar o trabalho da elaboração

Para estimular a pesquisa com itens, inicialmente pode-se adotar experimentalmente as propostas desse trabalho, investigando itens eliminados pertencentes a cada um dos eixos temáticos da Matemática. Assim, levantar os quantitativos, e buscar similaridades e padrões nos itens, pode nos revelar respostas não apenas sobre a eliminação, mas sobre a qualidade do item e da avaliação como um todo.

Levantamento do quantitativo de itens eliminados por áreas temáticas e produção de pesquisa

Para verificar a qualidade dos itens a partir de itens eliminados, será necessário o levantamento do quantitativo de itens utilizados nos principais projetos de atuação do CAEd. Assim, é possível identificar qual das áreas temáticas apresenta maior quantitativo de eliminações e realizar um estudo a partir dos descritores mais

recorrentes, ou a partir de itens com baixos percentuais de acertos, como foi o caso dessa pesquisa.

O que será feito?	Levantamento do quantitativo de itens eliminados pertencentes a áreas temáticas da Matemática
Por que será feito?	Identificar padrões nas eliminações de itens, buscando fatores que possam contribuir para eliminação e afetar a qualidade do item, além de identificar itens que já não são eficazes na avaliação.
Onde?	CIA
Quando?	O tempo deve ser ajustado em razão da demanda de trabalho, podendo ser definido para ser realizado durante seis meses, duas vezes ao ano. Dessa forma é possível realizar duas pesquisas por ano, em duas áreas temáticas da Matemática. Esse estudo também pode ser adaptado e realizado anualmente.
Quem?	Colaboradores da equipe de Matemática da CIA.
Como?	Organizando estudos individuais por colaborador, promovendo encontros semanais para reunir informações. Essas reuniões podem durar em torno de 2 horas cada. Ao final de cada semestre/ano, será produzido um artigo ou produto educacional por toda equipe, divulgando os resultados.
Custo	A princípio não gera custos extras na realização dessas tarefas, uma vez que fará parte da rotina de trabalho.

Nesta ação, pode-se incluir ainda alguns aspectos que envolvem a eliminação de itens por motivos que rodeiam a matriz de referência. Como visto na pesquisa de campo, existe a possibilidade de alguns itens se tornarem ineficazes na avaliação ao decorrer dos anos. Isso nos possibilita analisar e estipular um período para renovação/atualização das matrizes de referência.

Entende-se que, quanto mais avaliamos uma determinada rede, seu ensino e suas consolidações, podem evoluir ou até mesmo regredir, fazendo com que algumas habilidades possam se tornar relevantes ou não, em determinadas etapas, como foi o caso do **item 9 - Item M050057H6** dessa pesquisa, eliminado por DIF.

Possibilitar caminhos para produção de novos itens e realizar pré-testes

Para possibilitar caminhos para produção de novos itens, será preciso levantar informações a partir das pesquisas realizadas nas ações propostas em 4.1.1. Identificadas recorrências de eliminação e padrões de elaboração que possam estar

contribuindo para eliminação, será necessário construir novos itens e realizar novos pré-testes. Os novos itens podem partir de ideias formuladas pelos próprios elaboradores através de novas propostas. Contudo, é interessante observar o comportamento de estudantes, ou mesmo, na forma como as redes avaliadas promovem o ensino. Dessa forma, a realização de seminários investigativos pode proporcionar estudos, que inspirem a produção de novos itens com novos distratores. Esses seminários podem ser amparados através de referenciais teóricos que envolvam a produção de significados na aprendizagem da Matemática, ou mesmo produzi-los durante a realização deste seminário, através de cenários previamente observados pelos professores participantes.

A produção de significados acontece a partir do método dos campos semânticos (MCS), desenvolvido pelo educador matemático Rômulo Campos Lins. Do ponto de vista de Lins (2012), os campos semânticos servem para articular a produção de conhecimentos, significados e objetos, oferecendo formas de produção de leituras suficientemente finas de processos de produção de significados (LINS, 2012) .

Lopes (2013) aponta que o MCS

[...]é um modelo epistemológico que afirma que o conhecimento é dado pela crença-afirmação e justificação. Não basta que o sujeito creia e afirme sobre uma dada enunciação, é preciso que ele justifique o que foi afirmado. Ou seja, a crença-afirmação é aquilo que o sujeito enuncia algo que acredita e a justificação é o que o sujeito entende como aquilo que ele está autorizado a dizer. (LOPES, 2013, p. 56).

Portanto, entende-se que este método possa fundamentar inúmeros estudos e possibilidades, por exemplo, na elaboração de itens, principalmente na criação de distratores, permitindo uma coleta apurada de dados, tornando-os teoricamente fundamentados nos significados produzidos em sala de aula.

Essa investigação pode se dar também em oficinas de elaboração de itens, sejam elas presenciais ou virtuais, em que professores das redes participantes busquem por exemplos de respostas, resoluções de atividades de seus próprios alunos, garimpando produções dos próprios alunos, que servirão de referencial para os elaboradores levantarem novas propostas de itens. Ao final das oficinas, os professores participantes podem ser certificados como forma de incentivo na participação da pesquisa.

O que será feito?	Produção de novos itens
Por que será feito?	Para obter novos itens, com novas estruturas, conforme informações levantadas nas pesquisas realizadas em 4.1.1
Onde?	Pode ser realizado na CIA, oficinas de elaboração de itens ou em ambientes virtuais, em forma de oficinas e seminários presenciais ou vídeo conferências
Quando?	Após cada artigo ou produto educacional produzido em 4.1.1. Duas vezes por ano ou anualmente.
Quem?	Colaboradores da CIA, com ênfase aos membros das equipes de elaboração.
Como?	Na realização de seminários ou oficinas, propor aos cursistas o levantamento de significados produzidos por seus alunos, através de exercícios resolvidos, provas ou quaisquer atividades do cotidiano escolar.
Custo	Nos custos podem envolver a disponibilização e deslocamentos de colaboradores para ministrar tais oficinas ou seminários ao optarem por meios presenciais. Em caso de propor em ambientes virtuais, disponibilização de recursos digitais otimizando em custo quanto aos deslocamentos.

Após a realização das oficinas, junta-se toda produção de itens obtida, monta-se o teste e realiza-se os pré-testes.

Eixo 2 – Aprimorar os aspectos formativos da equipe

Com o propósito de otimizar a performance dos futuros elaboradores, e ainda, situá-los diante do cenário avaliativo na educação, a formação do elaborador se torna fundamental, proporcionando a produção de itens de boa qualidade pedagógica e estatística. Nesse quesito, poderia ser desenvolvido no CAEd, um modelo de capacitação padrão, que atinja os principais aspectos que envolvem a avaliação em larga escala, como seus propósitos, a construção de instrumentos, a montagem de testes e os conhecimentos estatísticos fundamentais para sua compreensão. Para isso, propomos as seguintes ações.

O que será feito?	Curso de formação de elaboradores/colaboradores da CIA
-------------------	--

Por que será feito?	Aperfeiçoamento dos colaboradores que iniciarão suas atividades na CIA
Onde?	CAEd
Quando?	Sempre que houver novas contratações de colaboradores para a CIA
Quem?	Será necessário montar equipes de formadores com colaboradores experientes da CIA
Como?	Por meio de um seminário com duração de três semanas. 1º Semana: Ministrando conteúdos que apresentem os propósitos da avaliação educacional no Brasil, seu início e alguns exemplos de projetos de avaliação no Brasil. 2º Semana: Elaboração de itens e montagem de testes. 3º Semana: A TCT e a TRI; A estatística nos itens.
Custo	Os custos envolverão a disponibilização de colaboradores experientes, selecionados para ministrar a formação.

O intuito desse modelo de capacitação, é servir ou inspirar uma nova proposta de formação com formato padrão, de maneira que possa atender não exclusivamente apenas novos elaboradores para a equipe de Matemática, estendendo também, aos futuros colaboradores de outras áreas do conhecimento, como Ciências Humanas, Língua Portuguesa ou Ciências da Natureza. Com este modelo, o CAEd, com todo prestígio que o situa dentro do cenário educacional brasileiro, pode utilizá-lo não somente com propósitos de formar novas equipes internas, possibilitando estendê-lo ao público externo, que deseja mergulhar mais profundamente no campo da avaliação educacional em larga escala, capacitando e/ou habilitando/certificando profissionais a exercer atividades voltadas a avaliação, em qualquer outra instituição ou rede que desenvolva projetos neste universo.

Dessa forma, com essas propostas de intervenção, esperamos de fato, poder contribuir e promover avanços significativos no processo de elaboração de itens, visando a redução de itens eliminados, e de certa maneira, avançarmos no debate e pesquisa acerca da avaliação em larga escala, possibilitando ainda, contribuir para formação de novos talentos neste ramo ainda tão contrariado e avesso aos olhos de muitos profissionais da educação.