
Grau de dificuldade de itens em um teste para 9º ano do ensino fundamental: o caso de uma avaliação externa municipal

Carlos Henrique Delmiro¹

 <https://orcid.org/0000-0001-9055-3909>

Daniel Brandão Menezes²

 <https://orcid.org/0000-0002-5930-7969>

Hermínio Borges Neto³

 <https://orcid.org/0000-0003-4854-6953>

Resumo

O Sistema de Avaliação da Educação de Canindé (SAEC) é uma avaliação que possibilita aos professores e gestores escolares o acesso mais rápido e detalhado de cada turma avaliada. Assim, o objetivo deste estudo é analisar uma característica psicométrica do instrumento SAEC, especialmente no que diz respeito ao grau de dificuldade dos itens abordados no teste de matemática nas turmas de 9.º ano do ensino fundamental, por meio da Teoria Clássica dos Testes (TCT). Este estudo é de natureza quantitativa, com objetivos de pesquisa descritiva e exploratória. Os resultados apontam que dos dez itens avaliados no SAEC, cinco apresentam índice de dificuldade difícil, três; dificuldade moderada; e dois, fáceis, conforme análises estatísticas. A hipótese era de que, para as turmas do 9.º ano no componente de matemática, o SAEC era considerado fácil, e após a análise, verificou-se que esta hipótese não é válida em sua totalidade.

Palavras-chave: SAEC. Teoria Clássica dos Testes. Psicometria Clássica. Grau de dificuldade de itens.

Degree of difficulty of items in a test for the 9th grade of elementary school: the case of a municipal external assessment

Abstract

The Canindé Education Assessment System (SAEC) is an assessment that gives teachers and school managers quicker and more detailed access to each class assessed. Thus, the aim of this study is to analyze a psychometric characteristic of the SAEC instrument, especially regarding the degree of difficulty of the items covered in the mathematics test in 9th grade elementary school classes, by means of the Classical Test Theory (CTT). This is a quantitative study with objectives of descriptive and exploratory research. The results show that out of the 10 items evaluated in the SAEC, 5 were considered difficult, 3 presented moderate difficulty and 2 were considered easy according to the statistical analyses. The hypothesis of this study was that the SAEC was considered easy for 9th graders in mathematics, but after the analysis, it was found that this hypothesis is not entirely valid.

Keywords: SAEC; Classical Test Theory; Classical Psychometrics; Difficulty level of items.

¹ Universidade Federal do Ceará, Fortaleza: delmiro@multimeios.ufc.br.

² Universidade Estadual do Ceará, Fortaleza: brandao.menezes@uece.br.

³ Universidade Federal do Ceará, Fortaleza: herminio@multimeios.ufc.br.

Considerações iniciais

A cidade de Canindé/CE possui um sistema de ensino municipal que atende educação infantil, ensino fundamental e educação de jovens e adultos. Em 2023, a rede possuía pouco mais de 13 200 estudantes matriculados, em 82 instituições de ensino. Desde 2017, a avaliação externa vem sendo um dos parâmetros para políticas educacionais no município, levando-o a criar um sistema próprio de avaliação.

O Sistema de Avaliação da Educação de Canindé (SAEC) é uma avaliação externa promovida pela Secretaria Municipal da Educação de Canindé (SME) que aborda os componentes curriculares língua portuguesa e matemática nos 2.º, 5.º e 9.º anos do ensino fundamental.

De forma legal, o SAEC possui o respaldo da Lei Municipal de Canindé/CE n.º 2286/2015, de 15 de junho de 2015, e no Plano Municipal da Educação (PME), que, em seu item 7.3, indica a aplicação de instrumentos de avaliação com base nos matriculados na rede pública municipal de ensino de Canindé/CE (Canindé, 2015). Nessa linha, a Lei n.º 2532/2021, de 28 de outubro de 2021, instituiu o Projeto Avança Mais Educação, que possui cinco eixos, e um deles é a “inovação e qualidade de excelência pedagógica” (Canindé, 2021).

O SAEC vem da necessidade de contemplar essas duas leis municipais, com o intuito de identificar a visão macro da aprendizagem dos matriculados na rede pública municipal, para que, dessa forma, os professores possam planejar ações pedagógicas baseadas em evidências. São disponibilizados relatórios pedagógicos, apresentando o desempenho individual por estudante. Essa forma de contribuição da avaliação externa municipal é vista em Sousa e Ferrarotto (2019).

O primeiro SAEC foi aplicado nos dias 6, 10 e 11 de março de 2020, com 20 itens (10 de língua portuguesa e 10 de matemática), nas turmas de 2.º, 5.º e 9.º anos, com os testes impressos. Um detalhe dessa aplicação é que não foram utilizados itens inéditos para a avaliação.

Nos dias 4, 5 e 8 de novembro de 2021, houve a segunda edição do SAEC, de forma impressa, preservando as características do SAEC 2020. Para essa avaliação, foram utilizados itens que foram elaborados por professores da SME que participaram da capacitação para elaboradores de itens do SAEC.

O SAEC de 2023, a terceira edição, foi realizado em março, com a intenção de agir como uma avaliação externa diagnóstica. Em todas as edições, os aplicadores do SAEC foram técnicos da SME com o professor da turma.

A lacuna profissional de cada turma da rede pública municipal de ensino está ligada à necessidade de entender quais são os conhecimentos prévios desses estudantes avaliados, seus entraves e potencialidades. É possível que o estudante ingresse na escola (independentemente da série) sem ter sistematizado e/ou assimilado conteúdos formais que ampliem sua visão de mundo. Dessa forma, a escola é uma instituição que oportuniza a ele um aprendizado nos campos cognitivo, social e cultural.

Nas duas primeiras edições do SAEC não houve análise dos itens, de modo a estabelecer algum parâmetro presente na Teoria Clássica dos Testes (TCT) ou na Teoria de Resposta ao Item (TRI). Adolfo (2021) afirma que, para analisar o grau de dificuldade de um item, a TCT é um tipo de análise útil. Por outro lado, Vizzotto (2022a) aponta que o escore bruto não apresenta uma comparação entre indivíduos de forma satisfatória, pois pode ocorrer boas pontuações em uma prova de nível fácil, bem como uma prova com baixas pontuações em uma prova de nível difícil. Dessa forma, tem-se a necessidade de aferir a dificuldade dos itens.

Nessa linha, Soares, Emiliano e Soares (2020) afirmam que TCT e TRI caminham juntas, em que a análise em uma vertente é confirmada pela outra. Além disso, os autores salientam que, ao escolher apenas uma, TCT ou TRI, a análise dos dados será confiável e significativa. Além disso, essas duas teorias apresentam comportamentos semelhantes, quando se referem à análise do grau de dificuldade e discriminação de um item (Vizzotto, 2022b).

Rafi *et al.* (2023) afirmam que a análise de itens além de ser um processo essencial para o desenvolvimento de um teste, ele fornece dados para identificar os êxitos e os entraves dos estudantes, propiciando um panorama da aprendizagem.

A análise dos itens é dividida nas formas qualitativa (não deve conter potencial viés, livre de erros gramaticais e técnicos e medir habilidades que os alunos aprenderam) e quantitativa (utiliza a psicometria, enfatizando o grau de dificuldade e a discriminação do item), segundo Urbina (2004). Com o intuito de realizar uma análise de itens, a TCT concentra-se no grau de dificuldade de um item, discriminação do item, pontuação média do teste, desvio padrão e confiabilidade do teste, seguindo estratégias estatísticas (Rafi *et al.*, 2023).

Uma sugestão apresentada por Shanmugam, Wong e Rajoo (2020) é aplicar testes em um cenário com o qual os estudantes estejam familiarizados e que se tenha um assento para cada um com espaçamento entre eles. Na aplicação do SAEC, cada estudante respondeu ao teste em

sua cadeira, na sala de aula que ele assiste às aulas no ano letivo, ocasionando uma cadeira para cada em um ambiente familiarizado.

O trabalho de Rafi *et al.* (2023) foi em uma escola de uma cidade na Indonésia. Os autores apresentam como lacuna que, por meio da abordagem TCT, devem-se realizar análises do grau de dificuldade de um item, a discriminação de um item, a confiabilidade de um teste e a análise do distrator em pelo menos duas escolas.

Como o SAEC é uma avaliação externa em nível municipal, ainda se faz necessário novos estudos sobre o processo de elaboração, análise e uso dos dados, como apontam Sousa e Ferrarotto (2019).

Considerando que o SAEC foi realizado em 23 escolas públicas municipais e possui o intuito de mensurar as proficiências dos estudantes para superar possíveis entraves e alcançar resultados mais satisfatórios em avaliações externas nacional e estadual, qual foi o grau de dificuldade dos itens abordados no SAEC de 2023? Assim, o presente trabalho objetiva analisar, por meio da TCT, uma característica psicométrica do instrumento SAEC, especialmente no que diz respeito ao grau de dificuldade dos itens abordados no teste de matemática nas turmas de 9.º ano do ensino fundamental.

Uma hipótese para este estudo foi de que os itens abordados no SAEC 2023 nas turmas de 9.º ano eram de grau de dificuldade fácil, pois abrangiam habilidades previstas na Base Nacional Comum Curricular (BNCC) de 6.º, 7.º e 8.º anos do ensino fundamental.

Não obstante, como as avaliações externas nacional (Sistema de Avaliação da Educação Básica – SAEB) e estadual (Sistema Permanente de Avaliação da Educação Básica do Ceará – SPAECE) são parâmetros para políticas públicas (em particular, políticas educacionais), e que o SAEC visa ser uma espécie de pré-resultado nas avaliações em larga escala, espera-se que esse teste organizado pela SME atenda a requisitos mínimos para se ter uma avaliação justa, inclusiva e que cumpra seu objetivo de forma efetiva. Esses requisitos mínimos psicométricos são defendidos por Vizzotto (2022b), tendo em vista o Exame Nacional do Ensino Médio (ENEM).

Considerando o objetivo e a pergunta norteadora, o presente estudo, em um primeiro momento, aborda a fundamentação teórica, envolvendo pesquisas relacionadas ao grau de dificuldade de um item. Em um segundo momento, são apresentados os procedimentos metodológicos seguidos – abordagem quantitativa, descritiva e exploratória (Hartono *et al.*,

2023; Rafi *et al.*, 2023; Vizzoto, 2022a), os procedimentos de coleta de dados e as análises (Pasquali, 2013) – na perspectiva de índice de dificuldade de um item que leve em conta a resposta do estudante dada ao acaso (o chute). No terceiro momento, têm-se as descobertas em relação aos índices de dificuldade de cada item e a consistência interna medida pelo alfa de Cronbach. No quarto momento, são apresentadas as considerações finais, seguidas do registro das referências bibliográficas que foram utilizadas no desenvolvimento da pesquisa.

Grau de dificuldade de um item pela TCT

O conceito de dificuldade pode assumir diferentes significados, a depender do referencial adotado (Vizzotto, 2022b). Uma maneira de atestar a dificuldade que um teste (ou item) possui é por meio da psicometria, área da psicologia que utiliza ferramentas estatísticas para interpretar dados em testes cognitivos com seres humanos.

Uma definição para a psicometria é que esse campo de estudo da psicologia busca explicar o sentido que têm as respostas dos sujeitos em um caderno de itens (Pasquali, 2013). Diante disso, Pasquali (2013) afirma que a TCT se volta para explicar o resultado total desse caderno de itens (o teste). Outro ponto sobre a psicometria é que, tradicionalmente, ela analisa a dificuldade, a discriminação e o chute no teste (Pasquali, 2013).

Para realizar a análise do item, tem-se a análise gráfica e algébrica. A análise algébrica, interesse deste estudo, faz uso de algoritmos estatísticos para avaliar a dificuldade do item, por exemplo (Pasquali, 2013). Com o intuito de ter uma revisão teórica sobre a temática, utilizou-se nesta pesquisa o PRISMA 2020 para a coleta de artigos científicos em bases de dados.

O PRISMA 2020⁴ é uma declaração elaborada para o uso em revisões sistemáticas que avaliem intervenções educacionais, na saúde ou sociais. Ele é uma ferramenta que auxilia o pesquisador a garantir que todas as informações recomendadas sejam contempladas (Page *et al.*, 2022). Apesar de esse estudo não ser uma revisão sistemática de literatura, o uso do PRISMA 2020 tem o intuito de ser transparente quanto à seleção da fundamentação teórica.

⁴ Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Tradução: Itens de relatório preferidos para revisões sistemáticas e meta-análises. *Link*: <http://prisma-statement.org/PRISMAStatement/FlowDiagram>.

Para selecionar trabalhos, seguindo diretrizes do PRISMA 2020, necessita-se de critérios de inclusão e exclusão (Quadro 1). Seguindo a linha de Soto-Cerros, García-González e Pascual-Martín (2023), houve a exclusão de teses, dissertações, monografias, livros e capítulos de livros. Os artigos para este referencial teórico foram limitados àqueles que passaram por um processo de revisão por pares.

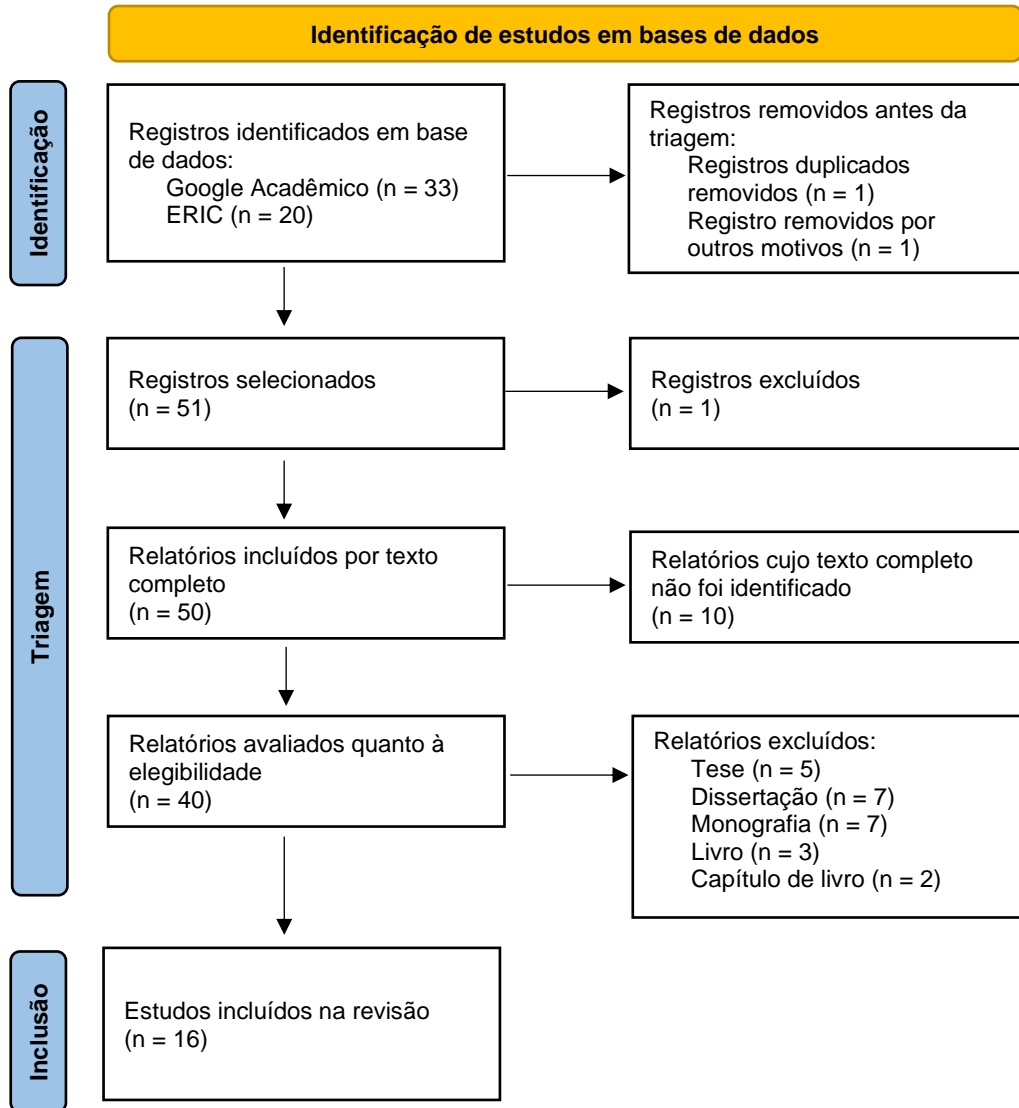
Quadro 1 – Critérios de inclusão e exclusão de documentos de informação

Critérios de inclusão	Critérios de exclusão
<ol style="list-style-type: none"> 1. Estudos que abordam o grau de dificuldade de um item. 2. Estudos publicados em português e em inglês. 	<ol style="list-style-type: none"> 1. Documentos duplicados. 2. Teses, dissertações, monografias e capítulos de livro. 3. Estudos sem relação com o grau de dificuldade de um item.

Fonte: Os autores

A busca do referencial teórico foi realizada em duas bases de dados: Google Acadêmico e ERIC. Na barra de busca em Google Acadêmico foram utilizadas as palavras-chave “dificuldade de um item” e “TCT”, inseridas com o conectivo booleano AND. Para o ERIC, por ser uma plataforma em que predomina o idioma inglês, as palavras-chave foram traduzidas, resultando na escrita na barra de busca por “difficulty level” e “CCT”. Foram consideradas publicações com data desde 2019 (para filtrar os últimos 5 anos). (Quadro 2)

Quadro 2 – Fluxograma PRISMA 2020



Fonte: Adaptado de Page *et al.* (2022)

De posse dos estudos elegidos, houve um estudo dos trabalhos e foi construído um quadro sinótico (Quadro 3) com os resultados obtidos na leitura dos artigos científicos quanto ao ano de publicação, autores, título, objetivo e principais resultados.

Quadro 3 – Investigações que abordam o grau de dificuldade de itens

N	Ano de publicação	Autores	Título	Objetivo	Principais resultados
1	2020	Costa, Lima e Soares	Uma proposta de análise de itens da prova preparatória para o ENADE aplicada aos discentes de engenharia civil do IFMG – <i>campus</i> avançado Piumhi	Analisar dois provões aplicados aos discentes do Curso de Bacharelado em Engenharia Civil do Instituto Federal de Minas Gerais – <i>Campus</i> Avançado Piumhi, e comparar seus resultados.	Os valores de referência, dispostos na literatura, não foram, em sua maioria, atingidos, apontando para a necessidade de uma revisão sistemática no processo de elaboração dessas provas e de um debate sobre os diferentes resultados obtidos por estudantes matriculados em períodos distintos no curso.
2	2021	Pinheiro e Conceição	Do grau de dificuldade para um contínuo de complexidade: análise textual de itens do Enem	Propor a análise do item pela perspectiva da complexidade textual e, dessa forma, contribuir para o debate acerca da sua qualidade e do seu comportamento empírico.	Grande parte dos itens (19 itens, 09 fáceis e 10 difíceis) está no nível 2 de complexidade. Os outros se distribuem pelos demais níveis: 1 item no nível 1 (fácil), 1 item no nível 3 (fácil), 2 itens no nível 4 (1 fácil e 1 difícil), e 1 item no nível 5 (difícil)
3	2020	Soares, Emiliano e Soares	Características psicométricas de uma avaliação de Matemática	Oferecer uma análise estatística, pautada nas duas vertentes da psicometria moderna: a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI), e uma análise pedagógica dos descritores e distratores de alguns itens-chave que compõem uma dessas avaliações, escolhida para um estudo de caso.	Avaliação composta por itens com variados índices de dificuldade e, no geral, com bom poder discriminativo tanto pela análise via TCT quanto pela TRI, cujos resultados foram obtidos através da estimação, pelo método da média <i>a posteriori</i> , de um modelo logístico de dois parâmetros.
4	2020	Sousa e Braga	Teoria clássica dos testes e teoria de resposta ao item em avaliação educacional	Apresentar um breve histórico do desenvolvimento, assim como conceitos e características gerais da TCT e TRI.	A teoria clássica dos testes não foi simplesmente descartada, ela é ainda muito utilizada e diversos contextos, inclusive o da avaliação, principalmente quando não é possível a aplicação da teoria de resposta ao item.

5	2022b	Vizzotto	As “piores” questões de Física do ENEM: Uma análise psicométrica das edições de 2009 a 2019	Analisar a qualidade psicométrica dos itens de Física do ENEM para os anos de 2009 a 2019.	A maioria dos itens foi considerada não adequada para mensurar a proficiência que se propõe medir na prova de Ciências da Natureza.
6	2022a	Vizzotto	Qual é a qualidade psicométrica das questões de Química do ENEM? Uma análise dos itens de 2009 a 2019	Avaliar a qualidade psicométrica das questões de Química do ENEM das edições de 2009 a 2019.	84,52% das questões de Química foram consideradas não adequadas para mensurar aquilo que se propõe a medir
7	2022c	Vizzotto	Análise e classificação das questões de Biologia do ENEM, segundo suas características psicométricas	Aferir a qualidade psicométrica dos itens de Biologia do ENEM das edições de 2009 a 2019.	Mais da metade dos itens não foi considerada boa para aferir a proficiência que a prova de Ciências da Natureza se propõe a medir.
8	2023	Hartono <i>et al</i>	Exploration of Student’ Cognitive Mathematics Ability Diagnostic Instruments: Validity, Reliability, and Item Characteristics	Determinar a validade e a confiabilidade do escore do instrumento, bem como investigar o uso da teoria clássica dos testes para identificar as características dos itens.	Os resultados mostraram que o instrumento de diagnóstico de habilidade matemática desenvolvido apresentou alta validade de conteúdo. No índice de nível de dificuldade, de forma geral, é categorizado como itens moderados.
9	2022	Kiryan, Mashood e Yadav	Development of a Circular Motion Concept Question Item Inventory for Use in Ugandan Science Education	Desenvolver um inventário específico do contexto do Uganda que possa reconhecer as concepções alternativas dos alunos em movimento circular, avaliar os esforços pedagógicos dos professores e medir a mudança na concepção.	A análise TCT determinou que 26,2% dos itens tinham nível de dificuldade razoável e 45,2% tinham nível de dificuldade moderado.
10	2023	Parker <i>et al.</i>	Establishing a physics concept inventory using computer marked free-response questions	Delinear o desenvolvimento de uma versão do Inventário de Conceito de Força que faça uso de questões de resposta livre marcadas automaticamente.	A pesquisa de mecânica alternativa se mostrou confiável por meio da TCT e TRI, além de produzir um inventário adequado para uso prático.

11	2022	Zenger e Bitzenbauer	Exploring German Secondary School Students' Conceptual Knowledge of Density	Relatar o desenvolvimento e a pilotagem de uma versão alemã de um teste conceitual para avaliar o conhecimento conceitual dos alunos sobre densidade.	A análise mostra que o teste de conceito possui boas propriedades psicométricas. Além disso, demonstra permitir um levantamento confiável do conhecimento conceitual dos alunos sobre densidade.
12	2023	Rafi <i>et al.</i>	What might be frequently overlooked is actually still beneficial: Learning from post national-standardized school examination	Descrever as características do teste e seus itens utilizados no exame escolar padronizado nacionalmente, aplicando a teoria clássica dos testes e focando na dificuldade dos itens, na discriminação dos itens, na confiabilidade do teste e na análise de distratores	Tanto os itens de múltipla escolha quanto os de redação contidos na prova apresentavam nível de dificuldade moderado.
13	2021	Subali, Kumaidi e Aminah	The Comparison of Item Test Characteristics Viewed from Classic and Modern Test Theory	Comparar as características dos itens de instrumentos de avaliação do nível de domínio do método científico para alunos do ensino fundamental, conforme analisados pela Teoria Clássica dos Testes (TCT) e pela Teoria de Resposta ao Item (TRI).	Os índices de dificuldade dos itens analisados pelos TCT e pela TRI apresentam padrões semelhantes.
14	2020	Shanmugam, Wong e Rajoo	Examining the quality of english test items using psychometric and linguistic characteristics among grade six pupils	Examinar a qualidade dos itens do teste de inglês usando características psicométricas e linguísticas entre alunos da sexta série	Nenhum dos itens foi considerado muito fácil ou difícil, e todos os itens apresentaram índices de discriminação positivos. O item em expressões idiomáticas era o ideal em termos de dificuldade e discriminação. Descobriu-se que os itens difíceis eram baseados em vocabulário.

15	2019	Dirlik	The Comparison of Item Parameters Estimated from Parametric and Nonparametric Item Response Theory Models in Case of The Violance of Local Independence Assumption	Analisar as diferenças e semelhanças das estimativas de parâmetros de itens do modelo TRI paramétrico e não paramétrico em caso de violação da independência local.	Ficou claro que existem diferenças nos parâmetros estimados a partir das abordagens, mas essas diferenças não são significativas. E os parâmetros de dificuldade dos itens estimados nas diferentes abordagens são alta e negativamente correlacionados entre si, quer a suposição de independência local seja atendida ou não.
16	2022	Hussein <i>et al.</i>	Psychometric Evaluation of Dictations with the Rasch Model	Examinar as propriedades psicométricas dos ditados com o modelo Rasch	A análise de ditados com o modelo Rasch é possível. O teste foi altamente confiável e unidimensional. E o exame do mapa de Wright mostrou que os textos cobrem uma ampla gama de habilidades e direcionam com precisão as habilidades das pessoas.

Fonte: Os autores

Sousa e Braga (2020) afirmam que o parâmetro de dificuldade de um item pode ser observado em avaliações de larga escala (por exemplo, o SAEC), bem como em sala de aula.

Vizzotto (2022a) e Soares, Emiliano e Soares (2020) afirmam que uma das maneiras de verificar a dificuldade de um item é pela quantidade de acertos. Itens com 70% ou mais de acerto são considerados fáceis, de 30% a 70% são moderados e abaixo de 30% são considerados difíceis.

Para esse cálculo, Borgatto e Andrade (2012, p. 148) definem como a “[...] proporção de participantes que responderam o item corretamente”. Nessa linha, Costa, Lima e Soares (2020) modelam a fala de Borgatto e Andrade (2012), definindo a fórmula $D_i = \frac{A}{N}$, em que D_i é o grau de dificuldade, A é a quantidade de respondentes que acertaram o item e N o total de respondentes. Esse tipo de cálculo para o grau de dificuldade foi relatado pelos autores apresentados no Quadro 3. Além disso, Pasquali (2013) afirma que essa fórmula é a forma tradicional de tratamento da dificuldade de um item em um teste.

Por outro lado, na TCT é possível corrigir a resposta dada pelo acaso (o chute) por meio do índice de dificuldade do item, dado pela fórmula $ID = \frac{C - \frac{E}{K-1}}{N}$, em que ID é o índice de dificuldade do item, C é o número de sujeitos que acertaram o item, E é o número de sujeitos que erraram o item, K é o número de alternativas de resposta ao item e N é o número total de sujeitos participantes do teste (Pasquali, 2013).

A TCT não é defasada adiante da TRI, pois elas dialogam e possuem procedimentos complementares (Vizzotto, 2022a). Além disso, para analisar o grau de dificuldade de um item, TCT e TRI apresentam um comportamento semelhante (Soares; Soares; Santos, 2021).

Portanto, considerou-se a TCT como ferramenta para auxiliar na construção da resposta da pergunta de pesquisa. A seção seguinte expõe os itens a serem empregados para analisar a dificuldade, bem como quais os critérios foram utilizados para interpretar os dados.

Procedimentos metodológicos

Esta é uma pesquisa de natureza quantitativa, com objetivo de pesquisa descritivo e exploratório, pois investiga o banco de dados do SAEC e os analisa por meio de técnicas estatísticas (Hartono *et al.*, 2023; Rafi *et al.*, 2023; Vizzoto, 2022a).

Os dados são respostas dos alunos de 9.º ano do ensino fundamental ao instrumento de matemática no SAEC. Foram avaliados 945 estudantes, distribuídos em 23 escolas, sendo que a rede pública municipal de ensino de Canindé/CE, em 2023, possuiu 985 matriculados. Assim, 95,9% dos estudantes de 9.º ano foram avaliados no SAEC. Dessa forma, essa é uma análise pós-exame. Vale ressaltar que todas essas turmas de 9.º ano são em tempo integral.

O teste empregou 10 itens de múltipla escolha com quatro opções (ou seja, alternativas A, B, C e D). Os estudantes tiveram 120 minutos para responder o teste (o de língua portuguesa foi aplicado em conjunto com o de matemática, dentro desse tempo estipulado). Semelhante ao trabalho de Shanmugan, Wong e Rajoo (2020), o teste nesse estudo tinha a pontuação para os itens dicotômica, em que a escolha do gabarito daria uma pontuação, e a escolha de um dos distratores resultaria em nota zero.

A Célula de Avaliação Institucional, da SME, forneceu os itens que foram aplicados nos testes de matemática e língua portuguesa. A definição das habilidades/descriptores avaliados foi de responsabilidade do setor pedagógico da SME.

O teste mediu habilidades cognitivas previstas na Base Nacional Comum Curricular (BNCC) e que são pertinentes nas matrizes de referência de uma avaliação nacional (o SAEB) e de uma avaliação externa estadual (o SPAECE). (Quadro 4)

Quadro 4 – Habilidades avaliadas em cada item

Item	Descritor Matriz de Referência SAEB 2001	Descritor Matriz de Referência SPAECE	Habilidade (BNCC)	Objeto do Conhecimento (BNCC)	Unidade Temática (BNCC)
1	D17	D11	EF07MA10	Números racionais na representação fracionária e na decimal: usos, ordenação e associação com pontos da reta numérica e operações.	Números
2	D21	D13	EF06MA30	Cálculo de probabilidade como a razão entre o número de resultados favoráveis e o total de resultados possíveis em um espaço amostral equiprovável.	Probabilidade e Estatística
3	D28	D17	EF06MA13	Cálculo de porcentagens por meio de estratégias diversas, sem fazer uso da “regra de três”.	Números
4	D29	D18	EF07MA17	Problemas envolvendo grandezas diretamente proporcionais e grandezas inversamente proporcionais.	Álgebra
5	D33	D25	EF07MA18	Equações polinomiais do 1º grau.	Álgebra
6	D02	D52	EF05MA16	Figuras geométricas espaciais: reconhecimento, representações, planificações e características.	Geometria
7	D12	D65	EF04MA20	Medidas de comprimento, massa e capacidade: estimativas, utilização de instrumentos de medida e de unidades de medida convencionais mais usuais.	Grandezas e Medidas
8	D13	D67	EF08MA19	Área de figuras planas.	Grandezas e Medidas

9	D14	D69	EF07MA30	Cálculo de volume de blocos retangulares, utilizando unidades de medida convencionais mais usuais.	Grandezas e Medidas
10	D36	D75	EF06MA32	Leitura e interpretação de tabelas e gráficos (de colunas ou barras simples ou múltiplas) referentes a variáveis categóricas e variáveis numéricas.	Probabilidade e Estatística

Fonte: Os autores

Como o SAEC avaliou língua portuguesa e matemática, os dez primeiros itens avaliaram língua portuguesa; e os dez itens seguintes, habilidades previstas na BNCC na componente curricular matemática, abordando as unidades temáticas. Seguindo as orientações do SAEB, cada item possui quatro alternativas com uma resposta correta (Brasil, 2018).

A primeira análise realizada diz respeito à consistência interna do teste. Como Soares, Emiliano e Soares (2020) indicam, foi utilizado o Alfa de Cronbach. Vizzotto (2022b) salienta que um dos conceitos fundamentais para ser analisado em uma avaliação é quanto à confiabilidade do teste. Esse conceito de consistência interna (a confiabilidade) permite, em hipótese, que o mesmo teste aplicado em um sujeito em tempos diferentes o resultado seria semelhante.

O cálculo do Alfa de Cronbach estabelece uma relação da correlação entre cada item (o quanto cada item contribui no teste) e o teste (Vizzotto, 2022b). Dessa forma, a análise de dados se deu pelo Alfa de Cronbach, para a confiabilidade do teste, e a abordagem TCT, para o grau de dificuldade de cada item. Essas análises foram realizadas no Microsoft Excel, como sugerem Shanmugam, Wong e Rajoo (2020).

Seguindo a maneira de julgar o alfa de Cronbach que Landis e Koch (1977) apresentam, para alfa menor que 0,20, tem-se uma baixa consistência; para valores de 0,21 a 0,40, a consistência é razoável; para valores de 0,41 a 0,60, tem-se a classificação moderado; para valores de 0,61 a 0,80, substancial; e para valores maiores que 0,80, a consistência interna do teste é quase perfeita.

O grau de dificuldade foi calculado pela fórmula $ID = \frac{C - \frac{E}{K-1}}{N}$. A escolha dessa fórmula se deu pelo fato de o sujeito ter a oportunidade de acertar um item pelo acaso, não em função de

sua aptidão na habilidade avaliada (Pasquali, 2013). Se o índice de dificuldade for nulo ou igual a 1, deve-se remover da análise do teste, pois não colabora com informações para a escala de dificuldade (SOARES, 2018). Para interpretar o índice de dificuldade, Condé (2001) sugere que $ID > 0,7$ tem-se o item com índice de dificuldade fácil; se $0,3 \leq ID \leq 0,7$, então o item é de dificuldade moderada; e se $ID < 0,3$, o item é de dificuldade difícil.

Seguindo a ideia de Soares, Emiliano e Soares (2020), a formulação desses itens tem sido de responsabilidade da Célula de Avaliação da SME, a qual possui a pretensão de desenvolver um banco de itens com o grau de dificuldade e para que as futuras edições do SAEC sejam realizadas de forma sistemática.

A seguir, são expostos os resultados psicométricos realizados com os dez itens de matemática abordados no SAEC. Considerando a TCT, apresenta-se a dificuldade do item e sua classificação de acordo com a definição de Condé (2001). Quanto ao alfa de Cronbach, têm-se os resultados do teste e de cada item para entender a consistência interna. Também é descrita a quantidade de marcações em cada alternativa para cada item, gerando o histograma.

Descobertas

A TCT é uma abordagem psicométrica, utilizada para avaliar a qualidade de um teste (Shanmugam; Wong; Rajoo, 2020). Essa abordagem de análise de itens preocupa-se com a estimativa e o controle do teste (Dirlik, 2019). Assume uma relação linear em que a pontuação do teste é igual à pontuação verdadeira do sujeito, somada com o erro na medição do instrumento (Dirlik, 2019).

Cada item do teste foi analisado quanto à sua qualidade, usando o grau de dificuldade. O alfa de Cronbach para o teste resultou em 0,44, classificado em uma consistência interna moderada, segundo Landis e Koch (1977). (Quadro 5)

Quadro 5 – Resultados da análise de itens na visão geral

Item	Alfa de Cronbach	Grau de dificuldade	
1	0,41	0,37	Moderado
2	0,36	0,28	Difícil
3	0,38	0,08	Difícil
4	0,41	0,22	Difícil
5	0,38	0,56	Moderado
6	0,40	0,29	Difícil
7	0,40	0,11	Difícil
8	0,42	0,80	Fácil
9	0,33	0,45	Moderado
10	0,39	0,88	Fácil

Fonte: Os autores

O Quadro 5 revela que o grau de dificuldade dos itens está na faixa de 0,08 a 0,88. Além disso, dos dez itens de múltipla escolha em matemática, havia dois fáceis, três moderados e cinco difíceis. Desses itens, o mais difícil foi o 3, que abordou cálculo de porcentagens. Os itens contidos nesse teste tiveram um nível difícil de grau de dificuldade.

Levando em consideração a TRI, Vizzotto (2022b) e Shanmugam, Wong e Rajoo (2020) aconselham que, para o grau de dificuldade de um item, não haja no teste itens fáceis ou difíceis em sua maioria. Nesse ponto, o SAEC apresentou um teste não satisfatório por abordar itens difíceis em sua maioria. Por outro lado, o item 7, por exemplo, é uma habilidade prevista para o 4.º ano do ensino fundamental, mesmo assim os estudantes tiveram dificuldade em gabaritá-la. Isso demonstra que os estudantes estão com baixo nível cognitivo.

Tratando-se do alfa de Cronbach, os itens se apresentaram na faixa de 0,33 a 0,42, classificados como moderado em relação consistência interna de cada item do teste (Landis; Koch, 1977). Nota-se que para o teste e cada item, o alfa de Cronbach mostrou-se inferior ao mínimo esperado (0,60), semelhante ao ocorrido em Vizzotto (2022a).

Figura 1 – Análise do item 01 de matemática

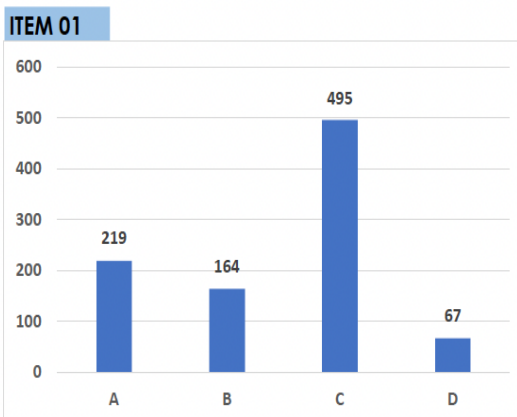
1. (EF07MA10) Observe a reta numérica abaixo. Ela está dividida em segmentos de mesma medida.



Qual ponto dessa reta representa o número

$$-\frac{1}{4}?$$

- (A) X.
- (B) Y.
- (C) Z.
- (D) W.



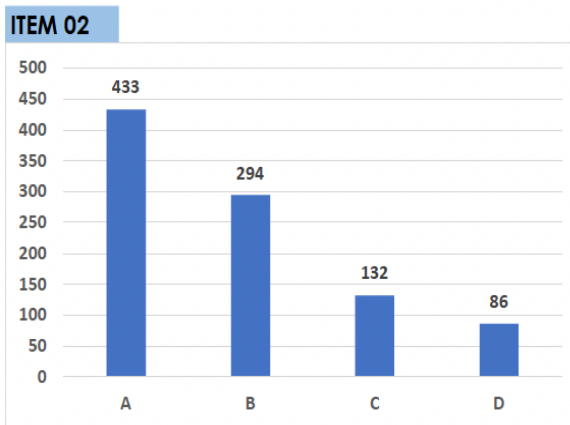
Fonte: Elaborado pelos autores

O item 01 de matemática (Figura 1) versa sobre a localização de números racionais em uma reta numérica. Esse item avalia uma habilidade prevista no 7.º ano do ensino fundamental, segundo a BNCC. De acordo com o Quadro 5, esse item é de dificuldade moderada. Pelo histograma do item, nota-se que a alternativa C foi a mais assinalada (o gabarito).

Figura 2 - Análise do item 02 de matemática

02. (EF06MA30) A final de teatro do Circuito Canindé Arte 2023 poderá acontecer na Escola Carlos Jereissati, na praça Dr. Aramis, na quadra da Infraestrutura ou na quadra da Centro Social Urbano (CSU), a assistência social. Qual a probabilidade da final ser sediada na praça Dr. Aramis?

- (A) $\frac{1}{4}$
- (B) $\frac{1}{2}$
- (C) $\frac{4}{1}$
- (D) $\frac{2}{1}$

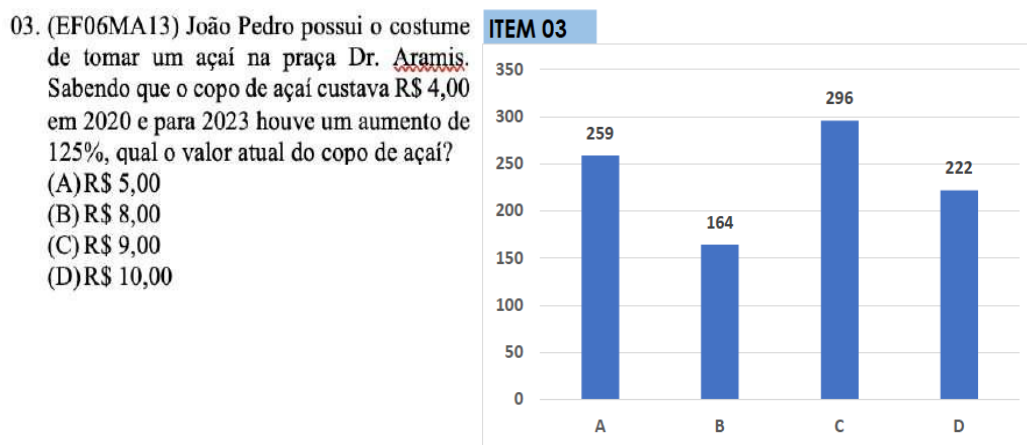


Fonte: Elaborado pelos autores

O item 02 de matemática (Figura 2) aborda o objeto de conhecimento e o cálculo de probabilidade por meio da razão dos eventos favoráveis com o total de eventos. Item considerado difícil, em que 46% dos participantes assinalaram a alternativa correta (alternativa

A). O gabarito foi o item mais escolhido como opção correta pelos participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 6.º ano do ensino fundamental.

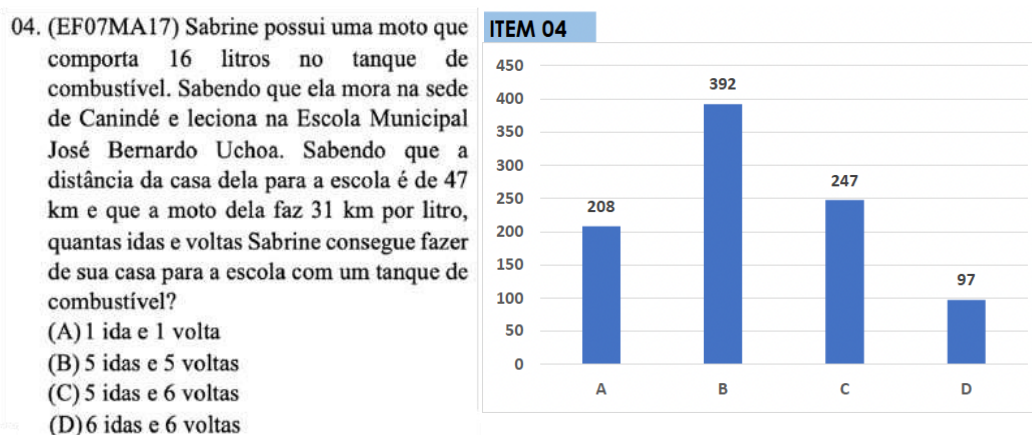
Figura 3 - Análise do item 03 de matemática



Fonte: Elaborado pelos autores

O item 03, assim como o item 02, (Figura 3) apresentou-se com o grau de dificuldade difícil, pela TCT (31,3% dos estudantes assinalaram o gabarito, a alternativa C). Além disso, abordou o conceito de porcentagem. O gabarito foi o item mais escolhido como opção correta pelos participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 6.º ano do ensino fundamental.

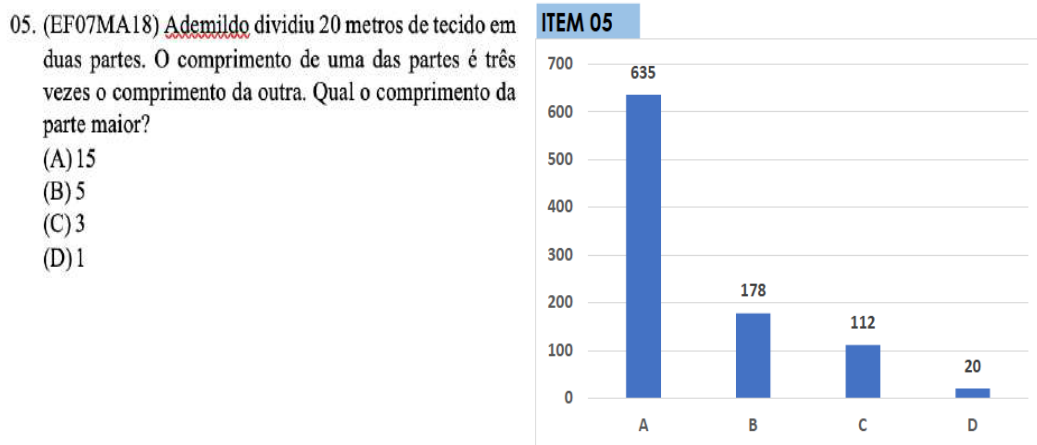
Figura 4 – Análise do item 04 de matemática



Fonte: Elaborado pelos autores

O item 04 de matemática (Figura 4) aborda o objeto de conhecimento que envolve grandezas direta ou inversamente proporcionais. Item considerado difícil, em que 41% dos participantes assinalaram a alternativa correta (alternativa B). O gabarito foi o item mais escolhido como opção correta pelos participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 7.º ano do ensino fundamental.

Figura 5 – Análise do item 05 de matemática

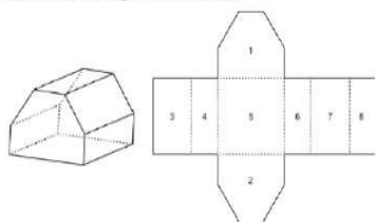


Fonte: Elaborado pelos autores

O item 05 de matemática (Figura 5) aborda o objeto de conhecimento que envolve equações de 1.º grau. Item considerado moderado, em que 67% dos participantes assinalaram a alternativa correta (alternativa A). O gabarito foi o item mais escolhido como opção correta pelos participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 7.º ano do ensino fundamental.

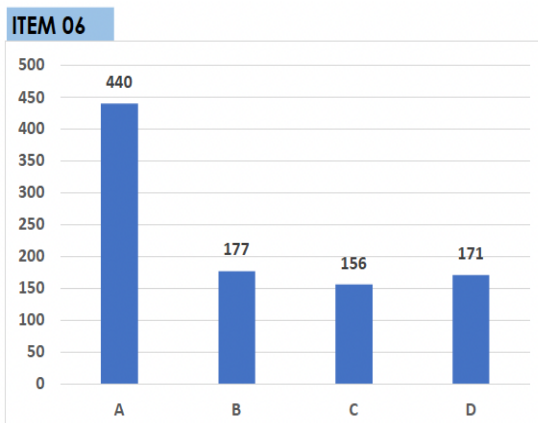
Figura 6 – Análise do item 06 de matemática

06. (D52) Vitor e Talita foram acampar na serra do Salgado. Ele levou a seguinte cabana.



Nessa planificação, os pares de faces paralelas são:

- (A) 1 e 2, 4 e 6, 5 e 8
- (B) 1 e 2, 6 e 8, 7 e 4
- (C) 2 e 3, 4 e 7, 5 e 8
- (D) 3 e 6, 4 e 7, 5 e 8

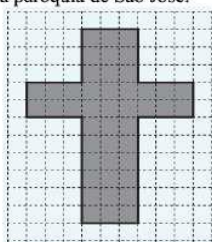


Fonte: Elaborado pelos autores

O item 06 de matemática (Figura 6) aborda o objeto de conhecimento que envolve planificação de figuras geométricas espaciais. Item considerado difícil, em que 46% dos participantes assinalaram a alternativa correta (alternativa A). O gabarito foi o item mais escolhido como opção correta pelos participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 5.º ano do ensino fundamental.

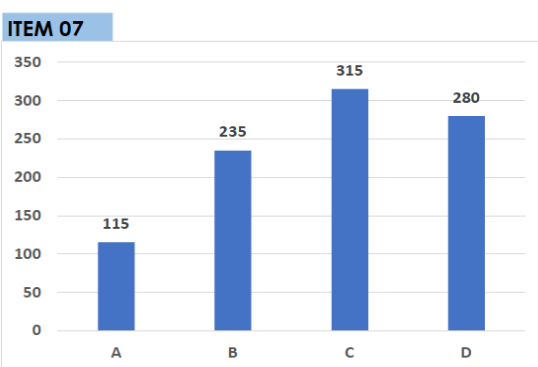
Figura 7 – Análise do item 07 de matemática

07. (D65) John desenhou uma cruz em malha quadriculada para colocar na paróquia de São José.



Sabendo que cada quadrado da malha quadriculada mede 3 m, qual a medida do perímetro da cruz desenhada por John?

- (A) 36 m
- (B) 45 m
- (C) 120 m
- (D) 132 m

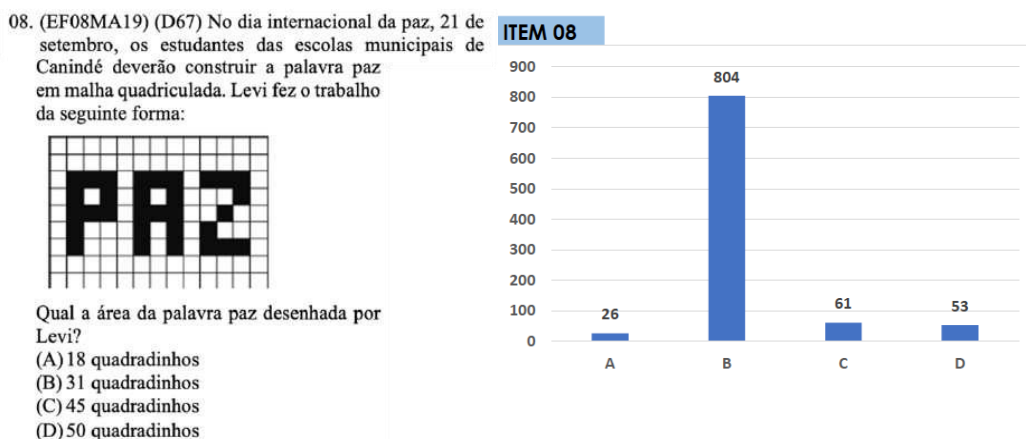


Fonte: Elaborado pelos autores

O item 07 de matemática (Figura 7) aborda o objeto de conhecimento que trabalha unidades de medida. Item considerado difícil, em que 33% dos participantes assinalaram a alternativa correta (alternativa C). O gabarito foi o item mais escolhido como opção correta pelos

participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 4.º ano do ensino fundamental.

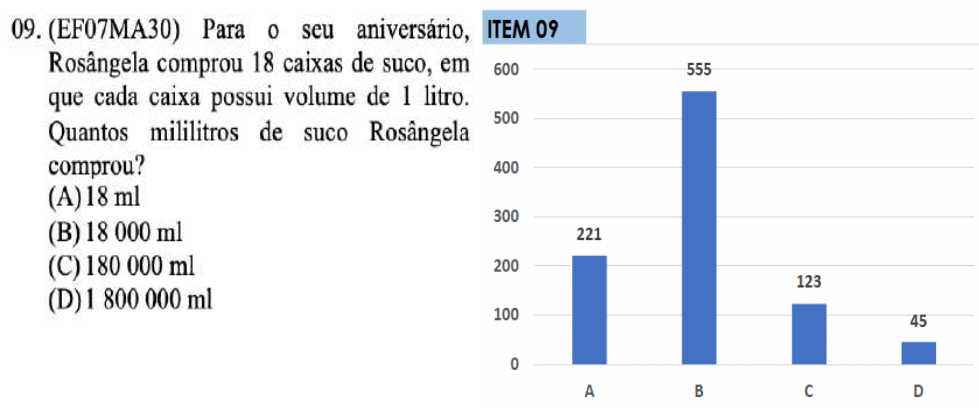
Figura 8 – Análise do item 08 de matemática



Fonte: Elaborado pelos autores

O item 08 de matemática (Figura 8) aborda o objeto de conhecimento área de figuras planas. Item considerado fácil, em que 85% dos participantes assinalaram a alternativa correta (alternativa B). O gabarito foi o item mais escolhido como opção correta pelos participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 8.º ano do ensino fundamental.

Figura 9 – Análise do item 09 de matemática

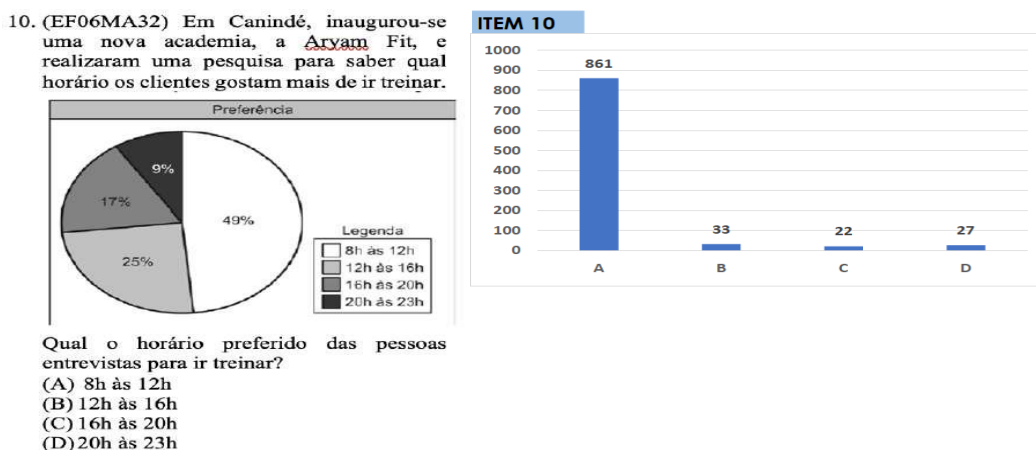


Fonte: elaborado pelos autores.

O item 09 de matemática (Figura 9) aborda o objeto de conhecimento que envolve volume de sólidos geométricos. Item considerado moderado, em que 58% dos participantes

assinaram a alternativa correta (alternativa B). O gabarito foi o item mais escolhido como opção correta pelos participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 7.º ano do ensino fundamental.

Figura 10 – Análise do item 10 de matemática



Fonte: Elaborado pelos autores

O item 10 de matemática (Figura 10) aborda o objeto de conhecimento que envolve leitura e interpretação de gráficos e tabelas. Item considerado fácil, em que 91% dos participantes assinaram a alternativa correta (alternativa A). O gabarito foi o item mais escolhido como opção correta pelos participantes. Pelo Quadro 4, esse item avalia uma habilidade prevista na BNCC para o 6.º ano do ensino fundamental.

Apesar de identificar o grau de dificuldade em cada item, pela abordagem TCT tem-se a limitação de que o mesmo teste aplicado com outros sujeitos poderá ocorrer de ter novos graus de dificuldade para cada item. Essa deficiência da TCT é contornada pela TRI (Shanmugam; Wong; Rajoo, 2020). Por isso, é válida a evolução de análise psicométrica de TCT para a TRI no tocante a avaliações externas, em particular, para o SAEC.

Considerações finais

A pergunta norteadora deste trabalho foi: *qual é o grau de dificuldade dos itens abordados no SAEC de 2023?* O que foi respondido, pois foi abordado como calcular o grau de dificuldade de um item e realizada essa análise pós-teste.

A hipótese adotada para este trabalho foi que *os itens abordados no SAEC 2023 nas turmas de 9.º ano eram de grau de dificuldade fácil, pois abordaram habilidades previstas na BNCC de 6.º, 7.º e 8.º anos do ensino fundamental*. Pelo resultado deste estudo, constata-se que a hipótese foi refutada, visto que 50% dos itens foram classificados estatisticamente com o grau de dificuldade difícil.

O presente trabalho teve como objetivo *analisar uma característica psicométrica do instrumento SAEC, especialmente no que diz respeito ao grau de dificuldade dos itens abordados no teste de matemática nas turmas de 9.º ano do ensino fundamental, por meio da TCT*, o que foi contemplado, visto que cada item abordado no SAEC 2023, para as turmas de 9.º ano, foi analisado estatisticamente, por meio da abordagem TCT, constatando o grau de dificuldade.

Com as duas formas de calcular o grau de dificuldade de um item pela abordagem TCT e pela revisão de literatura, construída seguindo alguns passos do diagrama PRISMA 2020, foi possível elencar pontos sobre o grau de dificuldade de um item para análises estatísticas de um teste.

Este estudo abordou o grau de dificuldade de cada item, empregando aspecto psicométrico. A limitação recai por não levar em consideração a dificuldade do item a critérios relacionados a aspectos do ensino de matemática, o conhecimento prévio dos estudantes, a sua proficiência no período de isolamento social e aulas em formato de ensino emergencial remoto (os estudantes de 9.º ano, em 2023, realizaram os 6.º e 7.º anos de forma *online*).

As lacunas presentes neste trabalho são: (i) para as edições nos anos posteriores, aconselha-se a aplicação de mais itens para se ter um teste mais homogêneo; (ii) abordar a discriminação de um item, com o intuito do teste conseguir distinguir bem os sujeitos que foram satisfatórios no teste dos que não foram; (iii) analisar estatisticamente os distratores apresentados em cada item no SAEC; (iv) empregar uma abordagem de análise Rasch ou a TRI; (v) abordar análises qualitativas, em que apresentem resultados sobre a opinião de professores e/ou estudantes sobre cada item abordado no SAEC.

As contribuições deste trabalho para a psicometria, em particular a análise pós-teste com a abordagem TCT para o cálculo do grau de dificuldade de um item, é que essa maneira de análise é válida para aplicações em sala de aula e em avaliações externas promovidas por secretarias municipais de educação, além de auxiliar na futura construção de um banco de itens, divididos

por grau de dificuldade. Além disso, fornece *insights* sobre o uso da TCT sem a necessidade de *softwares* sofisticados e/ou pagos para se ter essa maneira de análise.

Como produto, este estudo apresenta cinco itens difíceis, três moderados e dois fáceis para compor banco de questões de SMEs e de professores, os quais podem ser aplicados em sala de aula ou em avaliações em larga escala. Além disso, o como analisar o grau de dificuldade de um item pela abordagem TCT.

Para estudos futuros, deve-se considerar uma análise pedagógica dos itens, levando em consideração a natureza qualitativa. Em relação à abordagem quantitativa, realizar a análise de discriminação de cada item. Por outro lado, pode ser realizado um estudo semelhante com a TRI.

Referências

ADOLFO, A. *Ambiente para geração de questionários dinâmicos baseados em teoria da resposta ao item*. 2021. 61 f. TCC (Graduação) – Curso de Engenharia de Computação, Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021. Disponível em: <http://hdl.handle.net/10183/224064>. Acesso em: 29 out. 2023.

BORGATTO, A. F.; ANDRADE, D. F. de. Análise clássica de testes com diferentes graus de dificuldade. *Estudos em Avaliação Educacional*, [S.L.], v. 23, n. 52, p. 146-156, 30 ago. 2012. DOI: <http://dx.doi.org/10.18222/eae235220121934>

BRASIL. Ministério da Educação. Diretoria de Avaliação da Educação Básica. *Sistema de Avaliação da Educação Básica: documentos de referência*. (Versão 1.0). Brasília: Inep, 2018. 200 p. Disponível em: https://download.inep.gov.br/educacao_basica/saeb/2018/documentos/saeb_documentos_d_e_referencia_versao_1.0.pdf. Acesso em: 08 fev. 2023.

CANINDÉ. Secretaria Municipal de Educação. Lei n.º 2286/2015, de 15 de junho de 2015. *Plano Municipal de Educação 2015-2025*. Canindé: SME, 2015.

CANINDÉ. Secretaria Municipal de Educação. *Lei n.º 2532/2021, de 28 de outubro de 2021. Programa Avança Mais Educação*. Canindé: SME, 2021.

CONDÉ, F. N. *Análise empírica de itens*. Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais-DAEB/INEP/MEC, Brasília, 2001.

COSTA, M. C.; LIMA, S. H. O.; SOARES, D. J. M. Uma proposta de análise de itens da prova preparatória para o Enade aplicada aos discentes de engenharia civil do IFMG – *Campus Avançado Piumhi*. *Forscience: Revista Científica do IFMG*, [S.L.], v. 8, n. 1, p. 1-19, 14 maio 2020. DOI: <https://doi.org/10.29069/forscience.2020v8n1.e706>

DIRLIK, E. M. The Comparison of Item Parameters Estimated from Parametric and Nonparametric Item Response Theory Models in Case of The Violance of Local Independence Assumption. *International Journal of Progressive Education*, [S.L.], v. 15, n. 4, p. 229-240, 2 ago. 2019. DOI: <https://doi.org/10.29329/ijpe.2019.203.17>

HARTONO, W. *et al.* Exploration of Diagnostic Testing Instruments: validity, reliability, and item characteristics. *Pegem Journal of Education and Instruction*, [S.L.], v. 13, n. 3, p. 386-394, 1 jan. 2023. DOI: <http://dx.doi.org/10.47750/pegegog.13.03.39>

HUSSEIN, R. A. *et al.* Psychometric Evaluation of Dictations with the Rasch Model. *International Journal of Language Testing*, [S.L.], v. 12, n. 2, p. 118-127, out. 2022. Disponível em: https://www.ijlt.ir/article_157129_52f9c360c0c71818db2c14fc75f7c28f.pdf. Acesso em: 15 nov. 2023.

KIRYA, K. R.; MASHOOD, K. K.; YADAV, L. L. Development of a Circular Motion Concept Inventory for Use in Ugandan Science Education. *Journal Of Turkish Science Education*, [s. l.], v. 20, n. 1, p. 1312-1327, 31 mar. 2023. Disponível em: <https://www.tused.org/index.php/tused/article/view/2563>. Acesso em: 3 nov. 2023.

LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, [S.L.], v. 33, n. 1, p. 159-174, mar. 1977. JSTOR. DOI: <http://dx.doi.org/10.2307/2529310>.

PAGE, M. J. *et al.* A declaração PRISMA 2020: diretriz atualizada para relatar revisões sistemáticas. *Revista Panamericana de Salud Pública*, [S.L.], v. 46, n. 246, p. 1-12, 30 dez. 2022. DOI: <https://doi.org/10.26633%2FRPSP.2022.112>

PARKER, M.A. J. *et al.* Establishing a physics concept inventory using computer marked free-response questions. *European Journal of Science And Mathematics Education*, [S.L.], v. 11, n. 2, p. 360-375, 1 abr. 2023. DOI: <http://dx.doi.org/10.30935/scimath/12680>

PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. 5. ed. Petrópolis: Editora Vozes, 2013.

PINHEIRO, C. L.; CONCEIÇÃO, J. A. Do grau de dificuldade para um contínuo de complexidade: análise textual de itens do Enem. *Revista Horizontes de Linguística Aplicada*, [S.L.], v. 20, n. 1, p. 1-17, 29 jun. 2021. DOI: <http://dx.doi.org/10.26512/rhla.v20i1.36140>

RAFI, I. *et al.* What might be frequently overlooked is actually still beneficial: learning from post national-standardized school examination. *Pedagogical Research*, [S.L.], v. 8, n. 1, p. 1-15, 1 jan. 2023. DOI: <https://doi.org/10.29333/pr/12657>

SHANMUGAM, S. K. S.; WONG, V.; RAJOO, M. Examining the quality of English test items using psychometric and linguistic characteristics among grade six pupils. *Malaysian Journal of Learning and Instruction*, [S.L.], v. 17, p. 63-101, 31 jul. 2020. DOI: <http://dx.doi.org/10.32890/mjli2020.17.2.3>

SOARES, D. J. M. *Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de matemática básica*. 2018. 121 f. Dissertação (Mestrado) – Curso de Programa de Pós-Graduação em Estatística Aplicada e Biometria, Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, 2018. Disponível em: <https://www.locus.ufv.br/handle/123456789/18404>. Acesso em: 06 set. 2023.

SOARES, D. J. M.; EMILIANO, P. C.; SOARES, T. E. A. Características Psicométricas de uma Avaliação de Matemática. *Ensino da Matemática em Debate*, [S.L.], v. 7, n. 3, p. 1-27, 29 dez. 2020. DOI: <https://doi.org/10.23925/2358-4122.2020v7i3p1-27>

SOARES, T. E. A.; SOARES, D. J. M.; SANTOS, W. dos. Medidas de Tendência Central: análise da qualidade das questões do enem de 2016 a 2018. *Jornal Internacional de Estudos em Educação Matemática*, [S.L.], v. 14, n. 1, p. 119-128, 30 abr. 2021. DOI: <https://doi.org/10.17921/2176-5634.2021v14n1p119-128>

SOTO-CERROS, S.; GARCÍA-GONZÁLEZ, M. del S.; PASCUAL-MARTÍN, M. I. Ensayo la relación entre el Dominio Afectivo y el modelo MTSK: una oportunidad de investigación. *Educación Matemática*, [S.L.], v. 35, n. 2, p. 226-246, 1 ago. 2023. DOI: <http://dx.doi.org/10.24844/em3502.09>

SOUSA, L. A. de; BRAGA, A. E. Teoria clássica dos testes e teoria de resposta ao item em avaliação educacional. *Revista de Instrumentos, Modelos e Políticas em Avaliação Educacional*, [S.L.], v. 1, n. 1, p. 1-13, 2 jan. 2020. DOI: <https://doi.org/10.51281/impa.e020002>

SOUSA, M. de; FERRAROTTO, L. Prova São Paulo: pontos de tensão na avaliação externa em larga escala municipal. *Revista Meta: Avaliação*, [S.L.], v. 11, n. 33, p. 611-637, 13 dez. 2019. DOI: <http://dx.doi.org/10.22347/2175-2753v11i33.1892>

SUBALI, B.; KUMAI, K.; AMINAH, N. S. The Comparison of Item Test Characteristics Viewed from Classic and Modern Test Theory. *International Journal of Instruction*, [S.L.], v. 14, n. 1, p. 647-660, 1 jan. 2021. DOI: <http://dx.doi.org/10.29333/iji.2021.14139a>.

URBINA, S. *Essentials of Psychological Testing*. New Jersey: John Wiley & Sons, 2004.

VIZZOTTO, P. A. Qual é a qualidade psicométrica das questões de Química do ENEM? Uma análise dos itens de 2009 a 2019. *Revista de Ensino de Ciências e Matemática*, [S.L.], v. 13, n. 1, p. 1-21, 23 mar. 2022a. DOI: <https://doi.org/10.26843/rencima.v13n1a19>

VIZZOTTO, P. A. As “piores” questões de Física do ENEM: uma análise psicométrica das edições de 2009 a 2019. *Revista Brasileira de Ensino de Física*, [S.L.], v. 44, p. 1-19, 2022b. DOI: <https://doi.org/10.1590/1806-9126-RBEF-2022-0140>

VIZZOTTO, P. Análise e classificação das questões de Biologia do ENEM segundo suas características psicométricas. *Revista de Ensino de Biologia da Sbenbio*, [S.L.], v. 15, n. 1, p. 314-332, 26 jun. 2022c. DOI: <https://doi.org/10.46667/renbio.v15i1.633>

ZENGER, T.; BITZENBAUER, P.; BITZENBAUER, P. Exploring German Secondary School Students' Conceptual Knowledge of Density. *Science Education International*, [S.L.], v. 33, n. 1, p. 86-92, 1 mar. 2022. DOI: <http://dx.doi.org/10.33828/sei.v33.i1.9>

Submetido: 23.11.2023.

Aprovado: 04.03.2024.