



DESVENDANDO HABILIDADES CRÍTICAS DE MATEMÁTICA POR MEIO DA PSICOMETRIA: O USO DA TEORIA CLÁSSICA DOS TESTES E TEORIA DE RESPOSTA AO ITEM EM AVALIAÇÕES EDUCACIONAIS

UNVEILING CRITICAL MATHEMATIC SKILLS THROUGH PSYCHOMETRICS: THE USE OF CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY IN EDUCATIONAL ASSESSMENTS

Bruno Thayguara de Oliveira Ribeiro¹; Claudia Lisete Oliveira Groenwald²,

RESUMO

Este estudo visa explorar a aplicação da psicometria na medição de habilidades críticas de matemática em avaliações de larga escala no Brasil. Adotando uma abordagem metodológica mista, o trabalho combina revisão de literatura, análise documental e técnicas psicométricas. A análise documental incluiu a exploração de políticas públicas e sistemas de avaliação implementados, como o SAEB e o SADEM, permitindo uma compreensão aprofundada das práticas avaliativas no Brasil. Para a análise quantitativa, foram aplicadas a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI) aos dados das avaliações, examinando parâmetro de dificuldade e discriminação dos itens. Exemplos específicos, como a análise do item 22 da prova ADE de 2022, ilustraram a aplicação dessas teorias. Os resultados destacaram a validade e a confiabilidade dos instrumentos de avaliação, com estudos de validade, assegurando que os instrumentos refletissem a diversidade cultural e educacional brasileira. Concluiu-se que o uso de técnicas psicométricas, como a TCT e a TRI, é essencial para melhorar as práticas pedagógicas e formular políticas públicas eficazes. A Divisão de Avaliação e Monitoramento (DAM) desempenhou um papel crucial na elaboração e implementação das avaliações, garantindo a análise estatística e pedagógica dos dados e a divulgação dos resultados para a rede pública municipal de ensino. Este estudo reafirma a importância de adaptar modelos avaliativos às particularidades do contexto educacional brasileiro para alcançar resultados significativos e práticos.

Palavras-chave: Psicometria; Avaliação Educacional; Teoria de Resposta ao Item (TRI); Teoria Clássica dos Testes (TCT); Políticas Públicas Educacionais

ABSTRACT

To explore the application of psychometrics in measuring critical mathematic skills in large-scale assessments in Brazil is the aim of this study. Adopting a mixed methodological approach, the

¹ Mestre em Ensino de Ciências e Matemática pela (ULBRA). Coordenador de Matemática da Divisão de Avaliação e Monitoramento-DAM-(SEMED), Manaus, AM, Brasil. Endereço para correspondência: Rua Otávio Cabral, 27, Petrópolis, Manaus, Amazonas: 69067370. E-mail: thayguara2007@gmail.com
ORCID iD: <https://orcid.org/0000-0002-3156-1797>

² Doutora em Ciências da Educação pela Pontifícia de Salamanca, Espanha. Professora titular da Universidade Luterana do Brasil (ULBRA), Canoas, RS, Brasil. Endereço para correspondência: Rua Protásio Alves, 208, Niterói, Canoas, RS, Brasil. CEP 92120-160. E-mail: claudiag@ulbra.br
ORCID iD: <https://orcid.org/0000-0001-7345-8205>

work combines literature review, documental analysis, and psychometric techniques. The documental analysis included the exploration of public policies and assessment systems implemented, such as SAEB and SADEM, allowing a deep understanding of evaluative practices in Brazil. For the quantitative analysis, the Classical Test Theory (CTT) and the Item Response Theory (IRT) were applied to the assessment data, examining item difficulty and discrimination parameters. Specific examples, such as the analysis of the 22th item from the 2022 ADE exam, illustrated the application of these theories. The results highlighted the validity and reliability of the assessment instruments, with validity studies ensuring that the instruments reflected Brazil's cultural and educational diversity. It was concluded that the use of psychometric techniques, such as CTT Pasquali and IRT, is essential to improve pedagogical practices and formulate effective public policies. The Division of Evaluation and Monitoring (DAM) played a crucial role in the development and implementation of the assessments, ensuring the statistical and pedagogical analysis of the data and the dissemination of results to the municipal public education network. This study reaffirms the importance of adapting evaluative models to the particularities of the Brazilian educational context to achieve significant and practical results.

Keywords: Psychometrics; Educational Assessment; Item Response Theory (IRT); Classical Test Theory (CTT); Educational Public Policies

Introdução

A avaliação educacional tem se configurado como uma faceta essencial da existência humana, sendo uma das diversas formas pelas quais compreendemos e interpretamos o mundo ao nosso redor. De acordo com Luckesi (2018), a capacidade de avaliar é intrínseca aos seres humanos, abrangendo desde a compreensão de fenômenos naturais até a interpretação de acontecimentos sociais e culturais. Esse conhecimento pode se manifestar de várias maneiras, seja por meio do exercício do senso crítico, como nas Ciências e na Filosofia ou da avaliação de valores, que pode ocorrer de forma espontânea a partir das experiências de vida ou de maneira consciente e metodológica em processos avaliativos intencionais. Assim, cada indivíduo toma decisões e realiza ações baseadas no conhecimento adquirido, demonstrando o aspecto prático do agir humano em diversas áreas e atividades criativas.

No contexto brasileiro, a reflexão sobre a avaliação tem sido profunda, dada a vasta literatura existente e as múltiplas perspectivas sobre seu significado e importância no âmbito escolar para a melhoria do processo de ensino e aprendizagem Araújo (2017). A avaliação tem sido objeto de discussão por diferentes segmentos, tanto internos quanto externos à escola, visando redefinir sua função social. Segundo Gatti, Vianna e Davis (2015), é fundamental que a avaliação considere os aspectos específicos do meio educacional em que será implementada, evitando a imposição de modelos que, embora teoricamente bem estruturados, nem sempre se ajustam às diversas particularidades do sistema. A realidade educacional brasileira possui características singulares, sendo mais apropriado adotar ideias gerais que possam orientar e fundamentar o trabalho avaliativo,

conforme desenvolvido por Cronbach (1963). No entanto, a aplicação de modelos importados de outras culturas pode resultar em situações artificiais e sem impacto prático. Além disso, os pesquisadores destacam a carência de pesquisas sobre avaliação em diferentes regiões do país, relacionando os aspectos positivos de um processo avaliativo com as condições dadas às equipes, o que nem sempre é observado em todas as regiões brasileiras.

Diante desse cenário, a utilização da psicometria em avaliações de larga escala se mostra como uma ferramenta vital para a medição de habilidades críticas, especialmente em matemática. A aplicação de teorias como a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI) permite uma análise aprofundada dos instrumentos de avaliação, garantindo a validade e a confiabilidade dos dados obtidos. Este artigo visa explorar a implementação dessas teorias no contexto educacional brasileiro, destacando suas contribuições para a melhoria das práticas pedagógicas e para a formulação de políticas públicas eficazes.

Metodologia

Este estudo adotou uma abordagem metodológica mista, unindo métodos qualitativos e quantitativos para investigar as práticas avaliativas no contexto educacional brasileiro, com ênfase na psicometria aplicada às avaliações de larga escala. A primeira etapa consistiu na revisão de literatura sobre avaliação educacional e psicometria, explorando teorias e práticas históricas e contemporâneas, e na análise documental de relatórios institucionais da Divisão de Avaliação e Monitoramento (DAM) da Secretaria Municipal de Educação de Manaus (SEMED). Essas fontes permitiram compreender o desenvolvimento das avaliações educacionais, como o SAEB e a ADE, e contextualizar a aplicação dessas avaliações no município de Manaus.

A segunda etapa focou na análise quantitativa de dados, utilizando a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI) para examinar o desempenho dos estudantes na avaliação ADE de 2022, especificamente na prova de Matemática da 3ª fase da Educação de Jovens e Adultos (EJA). As análises psicométricas incluíram a avaliação dos parâmetros de dificuldade e discriminação dos itens, além da geração de Curvas Características do Item (CCI), que revelaram a eficácia dos itens em medir habilidades específicas dos alunos. A combinação dessas abordagens metodológicas forneceu uma visão abrangente sobre a validade e confiabilidade dos

instrumentos de avaliação, oferecendo subsídios para melhorias nas práticas pedagógicas e formulação de políticas públicas educacionais.

O conceito de avaliação

De acordo com Luckesi (2018), a avaliação é uma dimensão essencial da experiência humana, pois nos permite compreender e interpretar o mundo ao nosso redor. Todos os seres humanos têm a capacidade de adquirir conhecimento sobre fatos, desde fenômenos naturais até eventos sociais e culturais, manifestando esse saber por meio do senso crítico em disciplinas como Filosofia e Ciências, ou na avaliação de valores. Esses valores podem ser avaliados tanto por reações emocionais do senso comum, que se formam a partir de nossas vivências, quanto por processos conscientes e metodológicos em investigações avaliativas deliberadas. A avaliação, assim, orienta as decisões e ações de cada indivíduo, refletindo o conhecimento adquirido em múltiplas áreas de atividade humana criativa. No contexto educacional brasileiro, como destaca Araújo (2017, p. 26), a questão da avaliação tem sido amplamente discutida, com grande quantidade de literatura sobre o tema, refletindo diferentes perspectivas acerca de seu significado nas escolas e sua importância para a melhoria do ensino e da aprendizagem. Essas reflexões e debates, realizados por diversos segmentos da sociedade, tanto dentro quanto fora das escolas, buscam redefinir o papel social da avaliação, reconhecendo sua relevância para o aprimoramento dos processos educacionais e para a formação integral dos alunos.

De acordo com Gatti, Vianna e Davis (2015), a avaliação deve considerar os aspectos específicos do meio educacional em que será aplicada, evitando a imposição de modelos teóricos que podem não se ajustar às diversas realidades do sistema. No Brasil, é mais adequado adotar ideias gerais para orientar o trabalho de avaliação, como as propostas por Cronbach (1963), válidas no contexto brasileiro, mas não constituem um modelo rígido. A importação de modelos de outras culturas pode resultar em uma situação artificial e sem efeitos práticos. Pesquisadores também destacam a falta de pesquisas sobre avaliação em diferentes regiões do país, apontando que o sucesso de um processo avaliativo está ligado às condições dadas às equipes, o que não é uniformemente observado em todo o Brasil.

A avaliação em Larga escala no Brasil

Conforme aborda Vianna (2003), A partir dos anos 90, a avaliação educacional se tornou uma ferramenta amplamente utilizada no Brasil em diferentes níveis

administrativos, visando abordar problemas educacionais urgentes. Esperava-se que esses processos de avaliação elevassem os padrões de desempenho, especialmente com o auxílio de tecnologias testadas em experiências semelhantes em diversos países, mesmo com culturas diferentes. Essa expectativa era global, com muitos depositando grandes esperanças nos resultados dessas avaliações. No entanto, embora as avaliações identifiquem problemas, elas sozinhas não os resolvem; é necessário considerar e implementar outras abordagens para enfrentar os desafios educacionais.

Conforme Araújo (2017), a prática da avaliação do sistema de ensino se desenvolveu gradualmente, mas só alcançou seu marco legal com a reforma educacional dos anos 90. Nesse período, o governo adotou metodologias de gestão focadas na melhoria dos serviços públicos, influenciado pelo neoliberalismo, que defendia a regulação e o controle dos serviços estatais, incluindo a educação, com base nos resultados apresentados, seguindo a lógica da "gestão de resultados". O objetivo era criar um mecanismo de aferição do desempenho do sistema de ensino, capaz de orientar e reformular políticas públicas para tornar a educação mais eficiente.

Sobre o SAEB analisa Vianna (2003), O Sistema de Avaliação do Ensino Básico (SAEB), proposto pelo Ministério da Educação, é fundamental para a cidadania e base dos níveis educacionais posteriores, devendo receber prioridade governamental. É crucial focar na validade de conteúdo e validade consequencial dos instrumentos de medição de desempenho escolar, pois a validade é específica e não universal. No contexto brasileiro, a validade deve considerar a diversidade social, econômica e cultural do país. Para garantir a precisão dos dados e evitar conclusões inadequadas, os programas de pesquisa sobre o SAEB precisam incluir estudos detalhados sobre a validade, refletindo a realidade cultural diversa do Brasil.

A avaliação em larga escala realizada em Manaus- ADE- Avaliação do Desempenho do Estudante

De acordo com Manaus (2021), a SEMED instituiu a DAM (Divisão de Avaliação e Monitoramento), a DAM é o departamento responsável por criar e implementar suas próprias avaliações, verdadeiramente a partir da web. Desde então, o seu time técnico executa o procedimento de concepção, desenvolvimento, revisão, verificação de itens, diagramação, estruturação de exames e avaliação estatística e educacional das informações. O ponto alto deste esforço é a publicação dos resultados para toda a rede pública municipal de educação.

Assim conforme Ribeiro et. al. (2021), em Manaus, o Sistema de Avaliação do Ensino Básico (SAEB), do Ministério da Educação, é crucial para a cidadania e a base educacional, devendo ser prioridade governamental. É essencial garantir a validade de conteúdo e a validade consequential dos instrumentos de medição, considerando a diversidade social, econômica e cultural do Brasil. Para evitar conclusões inadequadas e assegurar a precisão dos dados, os programas de pesquisa sobre o SAEB devem incluir estudos detalhados sobre validade, refletindo a diversidade cultural do país.

A Psicometria utilizada na ADE

De acordo com Sartes e Formigoni (2013), a Teoria Clássica dos Testes (TCT) oferece um conjunto de conceitos e técnicas utilizados para desenvolver instrumentos de medida e como referência para abordagens psicométricas. Na TCT, a análise dos itens visa selecionar os melhores itens de um conjunto maior, avaliando duas características principais: dificuldade e discriminação. A dificuldade de um item é determinada pela proporção de indivíduos que respondem afirmativamente a itens dicotômicos ou pela proporção de respostas a uma categoria específica, ou pela média das respostas em itens politômicos.

O índice de dificuldade varia entre 0 e 1, no caso do valor ser igual ao zero, isso indica que nenhum indivíduo respondeu ao item, caso seja igual 1, indica que todos responderam ao item, o parâmetro de discriminação na TCT visa diferenciar os grupos de que tiveram uma pontuação mais alta dos que tiveram pontuação baixa, o item é mais discriminativo quanto maior for seu valor, o índice pode variar entre -1 e +1, o que corresponde a diferença entre o índice de dificuldade dos indivíduos com nota alta dos que tiveram nota baixa.

De acordo com Pasquali (2017), no modelo da TCT é preciso distinguir três componentes, o escore bruto ou empírico (T), o escore verdadeiro (V) e o erro (E), além disso, é necessário fazer algumas suposições sobre as relações existentes entre três componentes, então temos:

T = escore bruto ou empírico do sujeito, que é a soma dos pontos obtidos no teste.

V = escore verdadeiro, que seria a magnitude real daquilo que o teste quer medir no sujeito e que seria o próprio T se não houvesse o erro de medida.

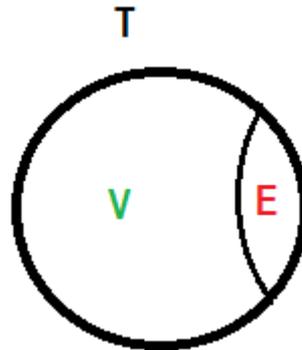
E = o erro cometido nesta medida.

Um dos postulados da Psicometria Clássica é de que o escore bruto do sujeito é a soma do escore verdadeiro e do erro, ou seja:

$$T = V + E$$

O escore empírico é a soma do escore verdadeiro e do erro,
logo $E = T - V$ representado na figura 1.

Figura 1: Componentes do escore T



Fonte: Pasquali (2017).

A Figura 1 mostra a relação entre os elementos do escore empírico, onde se vê que este é a união do escore verdadeiro (V) e do erro (E), ou seja, o escore empírico ou bruto do sujeito (T- resultado do teste) é constituído do escore verdadeiro (V) do sujeito naquilo que o teste pretende medir e o erro (E) de medida, este sempre presente em qualquer operação empírica. De acordo com Pasquali (2017) a grande tarefa da TCT consiste em elaborar estratégias estatísticas para controlar ou avaliar a magnitude do E.

De acordo com Rabelo (2013), pela TRI, o grau de conhecimento dos alunos é obtido por meio das características dos itens, de modo que alunos que acertam um mesmo número de itens de uma prova podem receber notas diferentes em razão de características específicas dos itens acertados. Essas características incluem a discriminação, a dificuldade e a probabilidade de acerto ao acaso. A utilização da TRI para análise de testes de conhecimentos veio para sanar algumas limitações da TCT, principalmente no que diz respeito a discriminação dos itens, fidedignidade dos testes e comparabilidade de desempenho de indivíduos que se submetem a teste diferentes.

Por meio dessas características, o TRI pode medir a proficiência de cada aluno, e construir uma escala a partir dessa unidade de medida. Esta é uma das principais vantagens da TRI sobre a Teoria Clássica dos Testes (TCT). Nesta escala, os alunos que

apresentam testes diferentes podem ser comparados entre si. No entanto, para que essa comparação seja legítima, é necessário fazer equalização³.

De acordo com Rabelo (2013), para descobrir isso, a pergunta inicial que a TRI faz sobre o item é: qual é a probabilidade de um dado indivíduo acertar um item específico? A resposta a essa pergunta depende do nível de aptidão do indivíduo θ e das características do item – dificuldade (b), discriminação (a) e acerto ao acaso (c). Entre os modelos propostos pela TRI, o modelo logístico unidimensional de 3 parâmetros é mais utilizado e representa a probabilidade de um indivíduo j responder corretamente um item i , sendo definido na equação (1).

$$P(X_{ij} = 1 | \theta_j) = c_i + \frac{(1-c_i)}{1 + \exp[-Da_i(\theta_j - b_i)]}, \quad (1)$$

Onde na equação tem-se que:

X_{ij} é a resposta ao item i (igual a 1, se o indivíduo responde corretamente, e igual a 0, caso contrário);

$a_i > 0$ é o parâmetro de discriminação do item i ,

b_i é o parâmetro de posição (ou de dificuldade) do item, medido na mesma escala da habilidade;

$0 < c_i < 1$ é o parâmetro da assíntota inferior do item i , refletindo as chances de um estudante de proficiência muito baixa selecionar a opção de resposta correta;

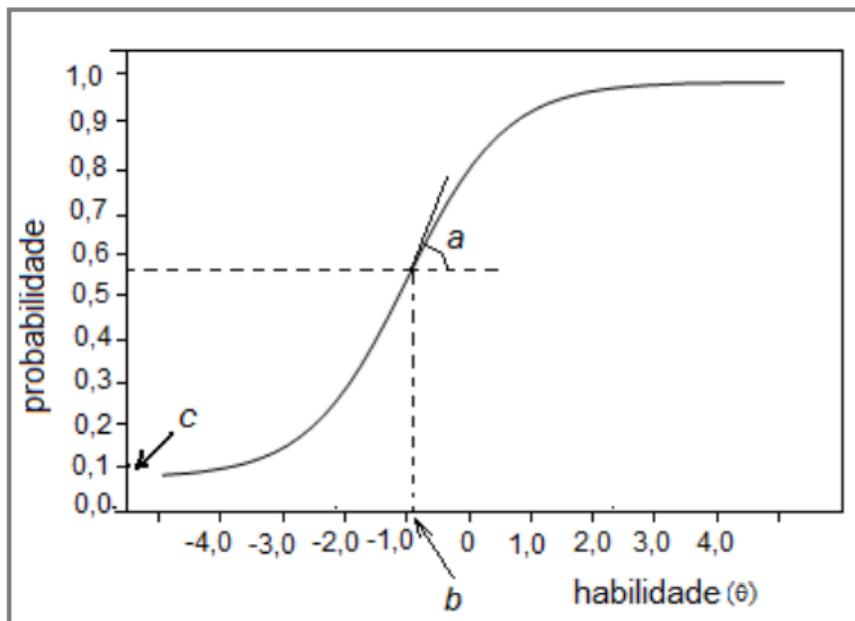
θ_j representa a habilidade ou traço latente do j -ésimo indivíduo;

D é um fator de escala, que é igual a 1 na métrica logística e igual a 1,7 na métrica normal.

Com esses parâmetros é possível construir a Curva Característica do Item representada na (Figura 2).

³ É necessário medir (parâmetros do item e desempenho dos alunos) para que estejam na mesma escala. através dessa análise, que pode avaliar o nível da escola na escala de competência disciplinas e níveis de educação avaliados em termos de habilidades desenvolvidas por alunos. Isso nos permite comparar e acompanhar progresso no sistema educacional ao longo dos anos.

Figura 2: Curva Característica do item (CCI)



Fonte: Rabelo (2013).

O número P é a proporção de respostas corretas ao item i no grupo de indivíduos com habilidade θ_j , podemos inferir que a TRI verifica o comportamento do estudante diante de cada item que ele responde, com base nesse modelo matemático é a curva da figura 2, gráfico em função de $P(\theta)$, que fornece a probabilidade de um estudante com a habilidade θ respondendo corretamente o item teste, a função P assume valores que vão de 0 a 1 e se configura em uma curva na forma de S (sigmoide), na curva, estão indicados os parâmetros a , b , c do item, que serão explicados a seguir. Os valores da habilidade (θ), no eixo horizontal, estão na escala de média igual a zero e desvio-padrão 1. O acerto ao acaso representa as respostas dadas por chute. Isso ocorre principalmente com os itens que são mais difíceis, para os quais indivíduos de baixa aptidão não conhecem a resposta correta, mas arriscam qualquer uma.

Conforme analisa Rabelo (2013), a TRI estima o acerto ao acaso por meio do parâmetro c , que representa a probabilidade de um aluno com baixa habilidade responder corretamente o item. Assim, se não fosse permitido "chutar", c seria igual a zero. A dificuldade é o valor da aptidão (θ) necessário para se obter uma probabilidade de acerto igual a $(1 + c)/2$. Quando não é permitido o "chute", a dificuldade é o valor da habilidade necessário para se obter uma probabilidade de 50% de acertar o item. A dificuldade é o

valor da aptidão necessário para se obter uma probabilidade de acerto igual a $(1 + c)/2$. Quando não é permitido o “chute”, a dificuldade é o valor da habilidade necessário para se obter uma probabilidade de 50% de acertar o item. A dificuldade é o valor da aptidão que é necessário para se obter uma probabilidade de acerto igual a $(1+c) / 2$. Quando não é permitido o chute, a dificuldade é o valor da habilidade necessário para se obter uma probabilidade de 50% de acertar o item.

A análise dos itens da ADE é feita utilizando dois parâmetros a Teoria Clássica de Testes (TCT) e a Teoria de Resposta ao Item (TRI) de acordo com Almeida (1993), em relação à TCT, para que os critérios de precisão e validade de um teste sejam satisfeitos, inicialmente deve se partir para a análise das unidades básicas que compõem o teste, ou seja, os itens. Geralmente, as análises quantitativas incluem a análise da distribuição de respostas nos itens (ou o índice de dificuldade, quando o item é dicotômico), o poder discriminativo, a análise das alternativas, a probabilidade de acerto ao acaso e a validade externa do item.

Segundo Araújo, Andrade e Bortolotti (2009), a Teoria de Resposta ao Item (TRI) oferece modelos matemáticos para traços latentes, representando a relação entre a probabilidade de um indivíduo responder a um item, seu traço latente e as características dos itens na área de estudo. Com base nas respostas de um grupo de indivíduos a um conjunto de itens, a TRI permite estimar os parâmetros dos itens e dos indivíduos em uma escala de medida. Por exemplo, para medir o nível de qualidade de vida, a TRI pode estimar tanto o nível de qualidade de vida do respondente quanto os parâmetros dos itens, criando uma escala de medida para esse construto.

Conforme o Sistema de Relatório de análise dos itens e provas da ADE/DAM (2022), alguns aspectos são considerados durante a análise dos itens, Índice de Dificuldade: Se refere ao nível de complexidade encontrado pelos respondentes para responder o item. O índice de dificuldade indica o nível de proficiência que o aluno deve ter para ter uma probabilidade de 50% de acertar o item, índice de discriminação refere-se à capacidade que o item possui para diferenciar respondentes que desenvolveram a habilidade que está sendo aferida, daqueles que ainda não desenvolveram essa habilidade. Se o parâmetro de discriminação é grande, a inclinação da reta é elevada e a curva sobe rapidamente até um valor alto de probabilidade de acerto, logo, o item possui uma capacidade de discriminar pequenas diferenças no traço latente (Escala de Proficiência).

Na prática, o parâmetro de discriminação varia entre 0 e 3, sendo que geralmente valores menores que 0.65 indica uma baixa discriminação. Probabilidade de Acerto do

Item ao Acaso se refere à probabilidade de acerto ao acaso, ou seja, mede a probabilidade de acerto de uma pessoa com pouco conhecimento do tema. Ele mostra que mesmo que uma pessoa não tenha praticamente domínio algum do que está sendo avaliado, mesmo assim ela ainda tem uma chance de acertar por uma tentativa ao acaso.

A classificação do item varia conforme seu grau de dificuldade. Sendo que o parâmetro de dificuldade conforme a (Tabela 1) foi classificado como:

Tabela 1: Classificação da Dificuldade do item

Classificação	Faixas
Muito Fácil	Menor que -2
Fácil	$[-2;-0,5)$
Médio	$[-0,5;0,5)$
Difícil	$[0,5;2)$
Muito Difícil	Maior que 2

Fonte: Rabelo (2013).

Classificação do item conforme seu grau de discriminação. Sendo que o parâmetro de discriminação foi classificado conforme a (Tabela 2) como:

Tabela 2: Classificação da Discriminação do Item

Classificação	Faixas
Muito Baixo	Menor que 0.30
Baixo	$[0.30;0.65)$
Moderado	$[0.65;1.35)$
Alto	$[1.35;1.70]$
Muito Alto	Maior que 1.70

Fonte: Rabelo (2013).

A quantidade de itens ideais por nível de dificuldade do item foi definida de acordo com a tabela 3:

Tabela 3: Quantitativo/ Classificação dos itens

Quantitativo Ideal de Itens na Avaliação (%esperado)	Índice de Dificuldade do Item (% acerto)	Classificação do Item em relação ao índice de dificuldade
10%	Superior a 0.9	Muito Fácil
20%	De 0.7 a 0.9	Fácil
40%	De 0.3 a 0.7	Médio
20%	De 0.1 a 0.3	Difícil
10%	Até 0.1	Muito Difícil

Fonte: Rabelo (2013).

De acordo com Ribeiro e Souza (2021), desde 2014 a ADE vem sendo analisada pela Teoria Clássica dos Testes, são calculadas as porcentagens de escolha para cada possibilidade de resposta, no entanto, segundo Andrade, Laros e Gouveia (2010, p. 422) ela tem sua limitação “os parâmetros dos itens dependem estritamente da amostra de sujeitos utilizada para estabelecê-los”, isso significa que, no caso da ADE, os testes têm níveis que variam entre muito fácil, fácil, médio, difícil e muito difícil, em que será considerado desempenho que os estudantes apresentam.

Modelo de análise de itens com uso da psicometria na ADE

A análise psicométrica dos itens é um componente fundamental no desenvolvimento e na aplicação de instrumentos de avaliação, especialmente em contextos educacionais. Ao aplicar conceitos e técnicas da psicometria, como a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI), é possível avaliar com precisão a qualidade dos itens em termos de validade, confiabilidade, dificuldade e discriminação. Esse processo assegura que os testes realmente medem o que se propõem a medir e que os resultados são interpretados de forma justa e precisa, levando em consideração as diferenças individuais entre os estudantes. Além disso, a análise psicométrica dos itens permite identificar áreas críticas de aprendizado, guiando tanto as práticas pedagógicas quanto as políticas públicas educacionais. Dessa forma, contribui para o aperfeiçoamento contínuo das avaliações e, conseqüentemente, para a melhoria da qualidade do ensino e da aprendizagem.

Na (Figura3), tem-se o primeiro item desse conjunto a ser analisado, item 22, com o conteúdo de Números e operações, da prova ADE do ano de 2022.

Figura 3: Item 22-1^aADE-3^aFase-EJA-2022

22. Josias e seus amigos foram a uma pizzeria no Jorge Teixeira. Como estava com muita fome, Josias comeu uma parte da pizza que eles pediram.



(Disponível em: <https://tinyurl.com/4vk65l8w>. Acesso em: 10 fev. 2022)

Qual é a fração equivalente ao pedaço de pizza comido por Josias?

(A) $\frac{2}{10}$

(B) $\frac{2}{12}$

(C) $\frac{2}{14}$

(D) $\frac{2}{16}$

(EJAMA4221ADE22SMAO)

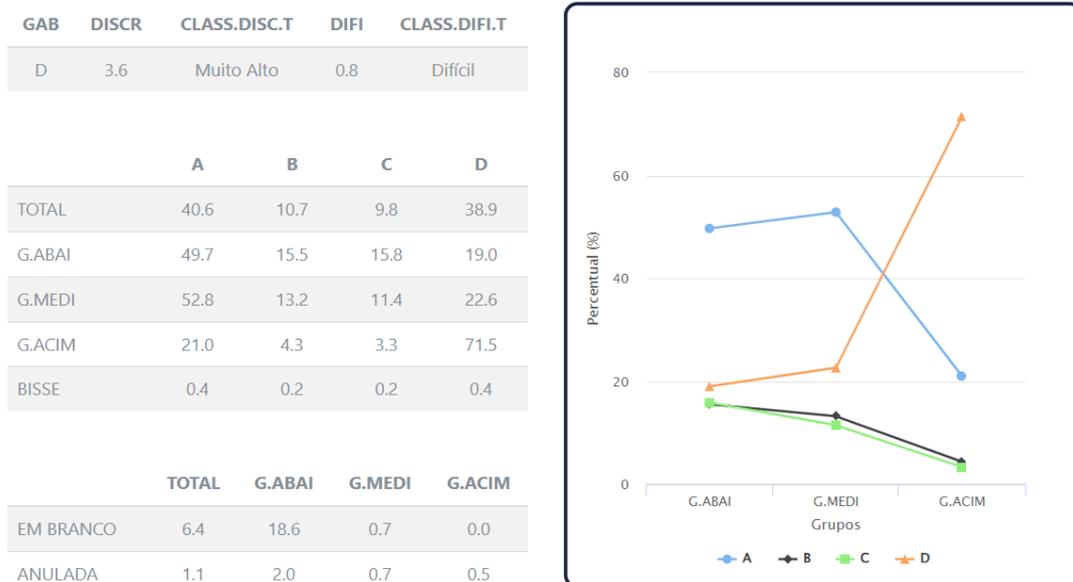
Fonte: DAM/SEMED-Manaus (2022).

O item 22 requer a habilidade de reconhecimento de frações equivalentes, onde se verifica que a pizza foi dividida em 8 pedaços e que 1 foi comido do total da pizza, logo, tem-se:

$$\frac{1}{8} \text{ usando fator de equivalência } 2, \text{ temos } \frac{1 \times 2}{8 \times 2} = \frac{2}{16}$$

Passamos primeiramente para análise das curvas de TCT por grupo de dificuldade na (Figura 4):

Figura 4: Curva do item 22



Fonte: DAM/Semed/Manaus (2022).

Inicialmente verifica-se o item possui um índice alto de discriminação (3.6), analisando a curva do gabarito (D), pela TCT verifica-se que 38,9% do grupo analisado marcou o gabarito, isso se tratando de *score*⁴ geral, para uma análise mais profunda é necessário analisar as curvas em função dos grupos abaixo, médio e acima. Percebe-se que o grupo médio 22,6% marcou a resposta correta, porém 52,8% marcou a curva A, ou seja, esse distrator⁵ chamou mais a atenção que o gabarito, na letra A, a hipótese sugerida é que o estudante aplica duas operações na fração para obtenção da equivalência:

$$\frac{2x2}{8 + 2} = \frac{4}{10}$$

Logo percebe-se que esse grupo possui uma certa noção de como proceder como a equivalência, porém ainda não sabe como operacionalizar as operações no processo, com destaque que esse distrator chamou a atenção 21,0% do grupo acima do proficiente, o restante dos distratores encontram-se em porcentagens aceitáveis do ponto de vista da TCT/TRI, Porém pode-se destacar ainda que 11,2% do grupo médio marcou a letra C, nessa alternativa a hipótese de erro ocorre no fato de fazer a fração inicial com a parte,

⁴ Termo usado para se referir ao total de acertos de uma alternativa.

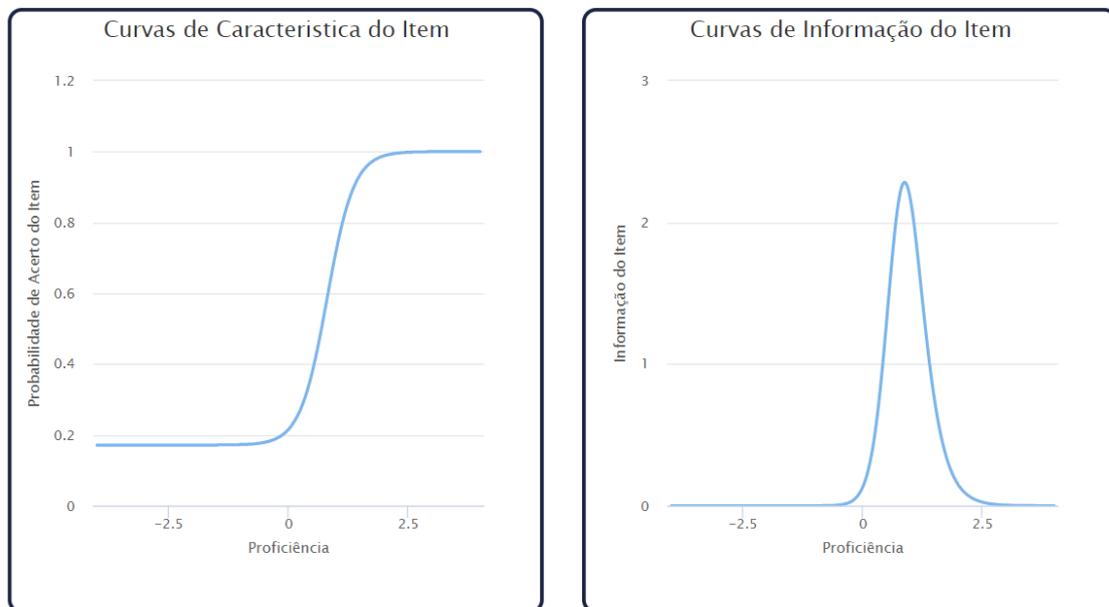
⁵ Distratores são possibilidades de erros plausíveis que podem ocorrer dentro da estrutura de um item de avaliação em larga escala como SAEB, ADE.

que foi retirada e a parte total é feita considerando apenas o número de fatias que sobraram, ou seja:

$$\frac{1}{7(\text{fatias que sobraram})} = \frac{1 \times 2}{7 \times 2} = \frac{2}{14}$$

Nesse caso se percebe que as ideias primárias de fração envolvendo parte-todo não foram consolidadas nesta parte do grupo analisado, servindo como base de reflexão para equipe escolar como um todo, mesmo com esse resultado os outros índices seguem o padrão de normalidade no que tange a psicometria, onde o grupo abaixo marca mais os distratores e o grupo acima marca o gabarito, o próximo passo é a análise das curvas de TRI conforme a (Figura 5):

Figura 5: TCT/TRI do item 22



Fonte: Semed Manaus (2022).

Pela análise da Curva de Característica do Item verifica-se que o traço latente está indicando uma alta discriminação em relação aos grupos acima e abaixo, com ponto de inflexão da curva em 0.6 de probabilidade de acerto do item, incluindo baixo índice de “chute”, isso pode ser verificado pela curva de informação do item que no teste de modo geral a probabilidade de acerto do item foi em sua maioria dos estudantes com alta proficiência, indicando que o grupo com baixa proficiência teve chance de acerto tendendo a 0, ou seja, são características de um item de nível DIFÍCIL, pela análise do item descarta-se possibilidade de erros de elaboração do item, logo essa habilidade pode

ser considerada uma habilidade crítica para a rede municipal de Manaus, podendo ser classificada como uma das habilidades a serem trabalhadas para os professores em formações e encontros.

A análise dos distratores por meio das curvas da Teoria Clássica dos Testes (TCT) e da Teoria de Resposta ao Item (TRI) é uma ferramenta crucial para entender o comportamento dos estudantes em relação a cada item da avaliação. No exemplo do item 22 da ADE, essa análise revelou que, embora 38,9% dos alunos tenham identificado a resposta correta, uma parte significativa dos estudantes do grupo médio (52,8%) foi atraída por um distrator específico. Isso sugere que esses estudantes possuem uma compreensão parcial do conceito de frações equivalentes, mas cometem erros ao operacionalizar as operações necessárias. A curva de discriminação reforça essa observação, mostrando que o item é altamente eficaz em diferenciar os alunos de alta e baixa proficiência. A análise das curvas TRI, por sua vez, confirma que o item é de nível difícil, com uma baixa probabilidade de acerto por chute, evidenciando que os erros estão mais relacionados à falta de domínio da habilidade do que à sorte.

Esses resultados têm uma importância significativa para os professores, pois permitem uma intervenção pedagógica mais direcionada e eficaz. Ao compreender quais distratores estão atraindo mais os alunos, os educadores podem identificar quais conceitos específicos precisam ser reforçados em sala de aula. No caso do item 22, por exemplo, o fato de que muitos estudantes foram atraídos por distratores envolvendo operações incorretas em frações sugere a necessidade de reforçar o ensino das operações de equivalência. Além disso, os dados obtidos pela TCT e TRI auxiliam os professores a reconhecer habilidades críticas que exigem maior atenção, permitindo que eles adaptem suas estratégias de ensino para abordar lacunas de conhecimento de forma mais precisa. Assim, a análise psicométrica dos itens não só melhora a qualidade da avaliação, mas também orienta diretamente a prática pedagógica, promovendo uma aprendizagem mais eficaz e significativa.

Conclusão

Este estudo investigou a aplicação da psicometria na avaliação de habilidades críticas em matemática, focando nas avaliações de larga escala realizadas no Brasil, com destaque para a ADE em Manaus. Ao utilizar uma abordagem metodológica mista, que combinou revisão de literatura, análise documental e técnicas quantitativas, foi possível obter uma compreensão mais profunda das práticas avaliativas no contexto educacional

brasileiro. O uso de teorias psicométricas, como a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI), revelou-se fundamental para garantir a validade e confiabilidade dos instrumentos de avaliação, permitindo que eles reflitam adequadamente a diversidade social, cultural e educacional do país.

A análise detalhada dos itens, exemplificada pelo item 19 da prova ADE de 2021, mostrou como os parâmetros de dificuldade e discriminação podem fornecer informações valiosas para melhorar as práticas pedagógicas. A aplicação da TCT e da TRI foi crucial para identificar as habilidades dominadas pelos alunos e os pontos de dificuldade, orientando intervenções pedagógicas mais eficazes e direcionadas. Assim, conclui-se que o uso de técnicas psicométricas é vital não apenas para aprimorar a qualidade das avaliações educacionais, mas também para apoiar a formulação de políticas públicas eficazes. A atuação da Divisão de Avaliação e Monitoramento (DAM) foi essencial para promover a precisão e o uso estratégico dos dados, reafirmando a necessidade de adaptar os modelos avaliativos às especificidades do contexto brasileiro, a fim de gerar melhorias reais no ensino e na aprendizagem.

Referências

ARAÚJO, Rosemary Farias Rufino de. **Avaliação em Larga Escala**. Manaus: BK Editora, 2017.

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, p. 1000-1008, 2009.

CRONBACH, L.J. Course Improvement through Evaluation. *Teachers College Record*, 64, p.672-83, 1963

DE ANDRADE, Josemberg Moura de; LAROS, Jacob Arie e GOUVEIA, Valdiney Veloso. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. *Aval. psicol.* [online]. 2011, vol.9, n.3, pp. 421-435. ISSN 1677-0471.

GATTI, Bernadete A.; VIANNA, Heraldo Marelim; DAVIS, Cláudia. Problemas e Impasses da Avaliação de Projetos e Sistemas Educacionais: dois casos brasileiros. **A Avaliação da Educação Básica: a experiência brasileira**. -1 ed.-Belo Horizonte, MG: Fino Traço, 2015.

LUCKESI, Cipriano Carlos. *Avaliação da Aprendizagem - Componentes do ato pedagógico*. São Paulo: Cortez, 2011

LUCKESI, Cipriano Carlos. *Avaliação na aprendizagem escolar: estudos e proposições*. São Paulo: Cortez, 2011.

LUCKESI, Cipriano Carlos. Avaliação na educação: questões epistemológicas e práticas. São Paulo: Cortez, 2018.

MANAUS. Secretaria Municipal de Educação de Manaus. **Relatório 1ª e 2ª ADE 2021**. Disponível em :
<<https://drive.google.com/file/d/1UkSxWL1QuzAo1M285XrFqMhP5RZMQh1C/view?usp=sharing>>. Acesso em 10. mar. 2022

PASQUALI, Luiz. **Teoria dos testes na psicologia e na educação**. Editora Vozes Limitada, 2017.

PASQUALI, Luiz. Psicometria – Teoria dos testes na Psicologia e na Educação. 5ed. Petrópolis: Vozes, 2013.

RABELO, M. Avaliação Educacional: Fundamentos, Metodologia e Aplicações no Contexto Brasileiro. Rio de Janeiro: SBM, 2013.

RIBEIRO, Bruno Thayguara De Oliveira et al... **A avaliação na educação de jovens e adultos**. Anais VII CONEDU - Edição Online... Campina Grande: Realize Editora, 2020. Disponível em: <<https://editorarealize.com.br/index.php/artigo/visualizar/67513>>. Acesso em: 11/11/2021 18:06

RUFINO, Rosemary Farias. **Avaliação de desempenho do estudante como ferramenta pedagógica de melhoria do rendimento escolar do ensino fundamental nas escolas de Manaus**. 2024. (Tese de Doutorado em Ciências da Educação) - Universidade Autônoma de Assunção. Assunção, 2024.

SARTES, Laisa Marcorela Andreoli; SOUZA-FORMIGONI, Maria Lúcia Oliveira de. Avanços na teoria clássica dos testes à teoria de resposta ao item. **Psicologia: Reflexão e Crítica**, v. 26, p. 241-250, 2013.

VIANNA, Heraldo Marelim. Avaliações nacionais em larga escala: análises e propostas. **Estudos em avaliação educacional**, n. 27, p. 41-76, 2003.

Recebido em: 09 / 07 / 2024

Aprovado em: 30 / 08 / 2024