

# DIFERENÇAS NA DIFICULDADE DOS ITENS DE MATEMÁTICA DO ENEM: UMA ANÁLISE POR COMPETÊNCIA AVALIATIVA

<https://doi.org/10.5902/2318133896143>

Denilson Junio Marques Soares<sup>1</sup>  
 Izabelli de Carvalho Oliveira<sup>2</sup>  
 Maria Eduarda Loschi Xavier<sup>3</sup>  
 Rafael Santiago do Nascimento<sup>4</sup>  
 Dejaina Aparecida Teixeira Lima<sup>5</sup>  
 Daniele Pires Magalhães<sup>6</sup>

## Resumo

Neste artigo, apresenta-se uma análise da dificuldade dos itens de Matemática do Exame Nacional do Ensino Médio, entre 2009 e 2023, em função das competências da matriz de referência, com base na Teoria de Resposta ao Item. Foram considerados 661 itens válidos, analisados por meio de estatísticas descritivas, Análise de Variância e teste de Tukey. Os resultados indicaram diferenças significativas entre as competências, com efeito de pequena magnitude e diferenças concentradas em poucos pares. A competência 7, relacionada à Estatística e Probabilidade, apresentou médias de dificuldade mais elevadas em comparação com algumas competências. Os achados indicam a existência de assimetrias pontuais na distribuição da dificuldade dos itens entre as competências avaliativas, sugerindo que a organização da matriz não se traduz, necessariamente, em equivalência plena de dificuldade entre seus diferentes domínios.

Palavras-chave: Enem; matemática; teoria de resposta ao item; dificuldade dos itens; avaliação por competências.

<sup>1</sup> Instituto Federal de Minas Gerais, Ouro Preto, Minas Gerais, Brasil. E-mail: [denilson.marques@ifmg.edu.br](mailto:denilson.marques@ifmg.edu.br). Orcid: <https://orcid.org/0000-0003-3075-3532>.

<sup>2</sup> Instituto Federal de Minas Gerais, Ouro Preto, Minas Gerais, Brasil. E-mail: [izabelli.coliveira07@gmail.com](mailto:izabelli.coliveira07@gmail.com). Orcid: <https://orcid.org/0009-0005-2277-5522>.

<sup>3</sup> Instituto Federal de Minas Gerais, Ouro Preto, Minas Gerais, Brasil. E-mail: [melxgalo@gmail.com](mailto:melxgalo@gmail.com). Orcid: <https://orcid.org/0009-0009-7054-4867>.

<sup>4</sup> Instituto Federal de Minas Gerais, Ouro Preto, Minas Gerais, Brasil. E-mail: [rafaelsantiago1304@gmail.com](mailto:rafaelsantiago1304@gmail.com). Orcid: <https://orcid.org/0009-0005-0289-5260>.

<sup>5</sup> Instituto Federal de Minas Gerais, Ouro Preto, Minas Gerais, Brasil. E-mail: [dejainaaparecida@gmail.com](mailto:dejainaaparecida@gmail.com). Orcid: <https://orcid.org/0009-0007-9815-7071>.

<sup>6</sup> Instituto Federal de Minas Gerais, Ouro Preto, Minas Gerais, Brasil. E-mail: [daniele.magalhaes@ifmg.edu.br](mailto:daniele.magalhaes@ifmg.edu.br). Orcid: <https://orcid.org/0009-0004-7083-2539>.

Crerios de autoria: os autores, coletivamente, realizaram a concepção, criação e consolidação do artigo.

Recebido em 9 de abril de 2026. Aceito em 31 de maio de 2026.



## DIFFERENCES IN THE DIFFICULTY OF ENEM MATHEMATICS ITEMS: AN ANALYSIS BY COMPETENCY

### Abstract

This study analyzes the difficulty of Mathematics items from the Brazilian National High School Examination (Enem) between 2009 and 2023, according to the competencies defined in the assessment framework, based on Item Response Theory. A total of 661 valid items were analyzed using descriptive statistics, Analysis of Variance, and Tukey's test. The results indicated statistically significant differences among competencies, although with a small effect size and differences concentrated in a limited number of pairwise comparisons. Competency 7, related to Statistics and Probability, presented higher average difficulty levels compared to some other competencies. The findings reveal localized asymmetries in the distribution of item difficulty across assessment competencies, suggesting that the organization of the framework does not necessarily translate into full equivalence of difficulty among its different domains.

Keywords: Enem; mathematics; item response theory; item difficulty; competency-based assessment.

### Introdução

A expansão das avaliações em larga escala tem sido acompanhada por um crescente interesse em compreender seus efeitos sobre os sistemas educacionais, especialmente no que se refere à produção de evidências para formulação de políticas, organização curricular e monitoramento do desempenho escolar (Verger; Novelli; Altinyelken, 2012; Lingard; Sellar; Lewis, 2017). Nesse contexto, tais avaliações passam a ser analisadas não apenas pelos resultados que produzem, mas também pelos modos como são elaboradas e pelas implicações decorrentes de suas decisões de medida (Verger; Normand, 2015).

No Brasil, o Exame Nacional do Ensino Médio, especialmente após a reformulação de 2009, passou a ser utilizado como principal mecanismo de acesso ao ensino superior, constituindo-se como porta de entrada para instituições públicas, por meio do Sistema de Seleção Unificada, e para instituições privadas, por meio de programas como o Programa Universidade para Todos e o Fundo de Financiamento Estudantil. Mais recentemente, o Ministério da Educação anunciou a ampliação das atribuições do exame, que passa a integrar o Sistema de Avaliação da Educação Básica, assumindo também a função de avaliar a qualidade da educação básica e produzir indicadores educacionais, além de manter seu papel na certificação da conclusão do ensino médio (Inep, 2026).

Esse conjunto de funções, que articula seleção, certificação e avaliação sistêmica, amplia o escopo de uso dos resultados do exame, na medida em que passam a subsidiar tanto decisões individuais quanto o monitoramento de políticas educacionais. Nesse contexto, torna-se pertinente examinar o funcionamento da prova em sua configuração atual, especialmente no que se refere à relação entre seus referenciais curriculares e sua operacionalização.

O Enem é composto por quatro áreas do conhecimento – Linguagens, Códigos e suas Tecnologias; Ciências Humanas e suas Tecnologias; Ciências da Natureza e suas Tecnologias; e Matemática e suas Tecnologias – além de uma redação. Cada área é

estruturada a partir de uma matriz de referência organizada em competências e habilidades, que orientam a elaboração dos itens e buscam operacionalizar a avaliação de conhecimentos em situações contextualizadas.

Especificamente, a prova de Matemática e suas Tecnologias do Enem está estruturada a partir de uma matriz composta por sete competências e trinta habilidades, que orientam a elaboração dos itens e delimitam os conteúdos e processos cognitivos a serem avaliados. Essa organização estabelece, no plano normativo, uma distribuição dos conteúdos e das demandas cognitivas entre diferentes domínios do conhecimento matemático (Brasil, 2009). No entanto, a existência dessa distribuição não assegura, por si só, que os itens associados a cada competência apresentem níveis equivalentes de exigência, nem que essa exigência se distribua de forma homogênea ao longo da prova.

Em avaliação educacional, a interpretação dos resultados de um teste não depende apenas da definição prévia das competências e habilidades que compõem sua matriz, mas também de evidências de que essas dimensões estão representadas de forma consistente no conjunto dos itens. Entre tais evidências, a literatura destaca a importância da estrutura interna do instrumento, isto é, da análise de como suas diferentes partes se comportam empiricamente e contribuem para a medida pretendida (Messick, 1989; Aera; APA; NCME, 2014).

Em exames utilizados para seleção, essa questão assume relevância adicional, pois eventuais assimetrias na dificuldade dos itens entre competências podem afetar a comparabilidade interna da prova e influenciar a interpretação dos desempenhos obtidos (Soares; Soares; Santos, 2021). Nesses casos, diferenças na exigência associada a determinadas dimensões podem elevar o nível de proficiência requerido para que os estudantes demonstrem domínio de certos conteúdos, com possíveis implicações para a equidade do exame e para o uso de seus resultados em processos decisórios (Messick, 1989).

Apesar dessa possibilidade analítica, os estudos sobre o Enem têm se concentrado em análises de desempenho, fatores associados e impactos de políticas educacionais, sendo menos frequentes as investigações voltadas à estrutura interna da prova e à distribuição das características dos itens em função das competências avaliadas (Soares; Soares; Santos, 2021). Essa lacuna limita a compreensão sobre a relação entre os objetivos expressos na matriz e sua materialização nos itens, particularmente no que se refere à exigência demandada dos estudantes.

Diante desse quadro, tem-se a seguinte questão: as competências avaliativas da área de Matemática do Enem apresentam níveis equivalentes de dificuldade, conforme sugerido pela organização da matriz, ou há variações sistemáticas entre elas? Assim, este estudo teve como objetivo analisar a distribuição da dificuldade dos itens de Matemática do Enem, no período de 2009 a 2023, segundo as sete competências avaliativas da matriz da área, buscando identificar diferenças estatisticamente significativas entre elas e discutir suas implicações para a avaliação educacional.

### Referencial teórico

A análise da dificuldade dos itens em avaliações organizadas por competências exige a articulação entre fundamentos psicométricos e a forma como os construtos são operacionalizados nos instrumentos de medida. Nesse sentido, esta seção apresenta,

Regae: Rev. Gest. Aval. Educ.	Santa Maria	v. 15	n. 24	e96143	2026
-------------------------------	-------------	-------	-------	--------	------

inicialmente, os elementos centrais da Teoria de Resposta ao Item – TRI –, que sustentam a análise da dificuldade dos itens e, em seguida, discute a organização da prova de Matemática do Enem a partir de competências, de modo a situar o problema investigado.

A TRI constitui um conjunto de modelos probabilísticos voltados à análise da relação entre a probabilidade de resposta a um item e o nível de proficiência do respondente num traço latente. Diferente da Teoria Clássica dos Testes, que se baseia em escores totais e apresenta forte dependência da amostra e do conjunto de itens, a TRI permite estimar parâmetros dos itens e habilidades dos indivíduos numa mesma escala, favorecendo a comparabilidade dos resultados entre diferentes aplicações de um teste (Soares, 2018).

Uma das propriedades centrais da TRI é a invariância da medida, segundo a qual as estimativas de habilidade são independentes do conjunto específico de itens utilizados, assim como os parâmetros dos itens são independentes da amostra de respondentes. Essa característica constitui um dos principais fundamentos para o uso da TRI em avaliações educacionais em larga escala, pois permite a comparação de desempenhos ao longo do tempo e entre diferentes versões de prova (Pasquali, 2020).

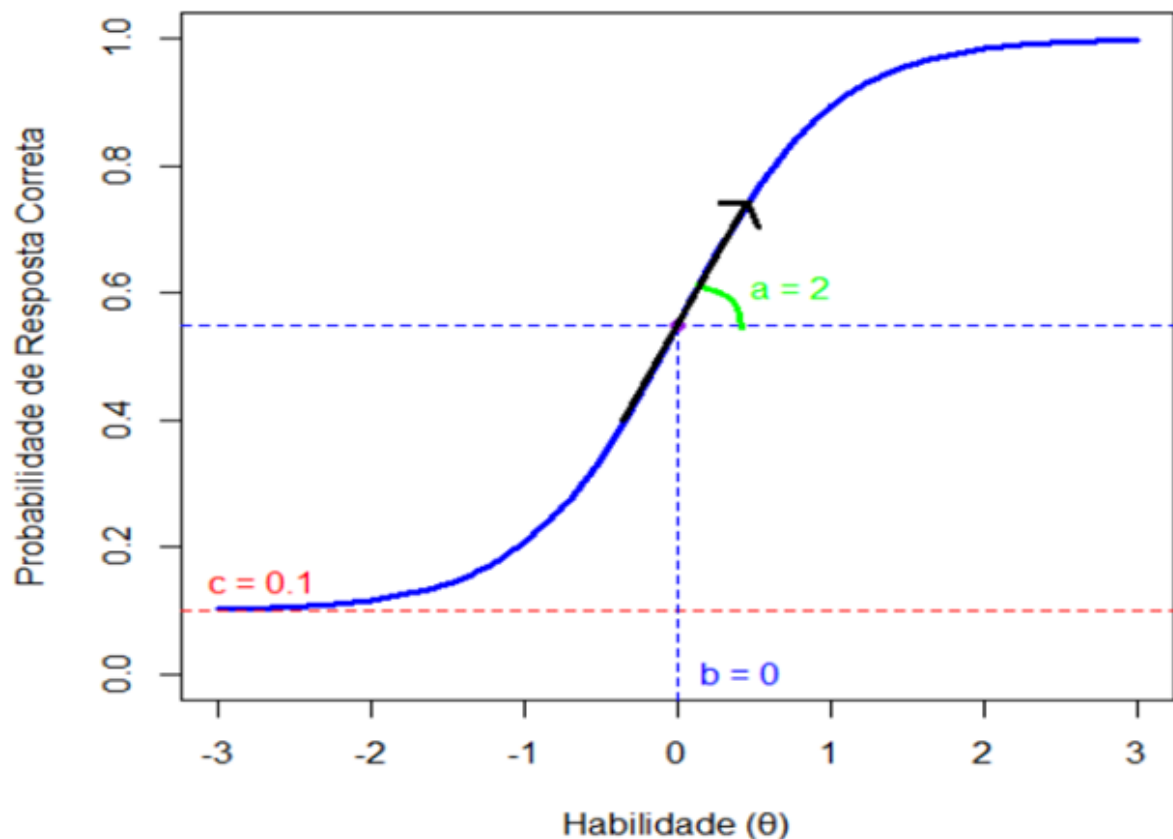
A aplicação da TRI baseia-se em dois pressupostos fundamentais: a unidimensionalidade e a independência local. A unidimensionalidade pressupõe que o conjunto de itens mede um único traço latente subjacente, representado pela variável  $\theta$ , enquanto a independência local estabelece que, dado o nível de habilidade do indivíduo, as respostas aos itens são estatisticamente independentes entre si (Rabelo, 2013).

Os modelos mais utilizados são os modelos logísticos unidimensionais, dentre os quais se destacam os modelos de um, dois e três parâmetros. No contexto do Enem, utiliza-se o modelo logístico de três parâmetros – ML3P –, no qual cada item é descrito pelos parâmetros de discriminação ( $a$ ), dificuldade ( $b$ ) e acerto ao acaso ( $c$ ) (Brasil, 2021).

O parâmetro  $a$  indica a capacidade do item de discriminar indivíduos com diferentes níveis de proficiência. Itens com valores mais elevados apresentam maior inclinação na Curva Característica do Item (CCI), sendo mais sensíveis às variações de habilidade dos respondentes. O parâmetro  $b$  representa a posição do item na escala de proficiência e indica o nível de habilidade em que a probabilidade de resposta correta atinge o ponto médio da Curva Característica do Item entre o limite inferior e superior da curva. Em geral, seus valores situam-se numa escala padronizada entre  $-3$  e  $+3$ , na qual valores mais elevados estão associados a itens que requerem maiores níveis de proficiência para serem respondidos corretamente. Já o parâmetro  $c$  expressa a probabilidade de acerto ao acaso, refletindo o comportamento do item em níveis baixos de proficiência (Soares; Soares; Santos, 2021).

A relação entre esses parâmetros pode ser visualizada por meio da CCI, que descreve a probabilidade de resposta correta em função da habilidade. Nessa representação, o parâmetro  $b$  desloca a curva ao longo da escala, o parâmetro  $a$  define sua inclinação e o parâmetro  $c$  estabelece seu limite inferior. A figura 1 ilustra essa estrutura.

Figura 1 –  
Representação de uma curva característica do item.



Fonte: autores, 2026.

Embora os três parâmetros desempenhem papéis complementares no funcionamento dos itens, o presente estudo concentra-se no parâmetro de dificuldade ( $b$ ), por ser ele o mais diretamente associado à exigência cognitiva requerida dos estudantes. A comparação desse parâmetro entre diferentes conjuntos de itens permite examinar como a dificuldade se distribui ao longo da prova e identificar possíveis variações sistemáticas entre as dimensões definidas na matriz avaliativa.

### Avaliação por competências na prova de Matemática do Enem

A prova de Matemática do Enem é estruturada a partir de uma matriz de referência que expressa uma concepção de avaliação por competências, orientada à mobilização de conhecimentos em situações contextualizadas. Na área de Matemática e suas Tecnologias, essa matriz organiza-se em sete competências, apresentadas no quadro 1.

Quadro 1 –  
Competências avaliativas da área de Matemática no Enem.

Competência de área	Descrição
Competência 1	Construir significados para os números naturais, inteiros, racionais e reais.
Competência 2	Utilizar o conhecimento geométrico para realizar a leitura e a representação da realidade e agir sobre ela.

Competência 3	Construir noções de grandezas e medidas para a compreensão da realidade e a solução de problemas do cotidiano.
Competência 4	Construir noções de variação de grandezas para a compreensão da realidade e a solução de problemas do cotidiano.
Competência 5	Modelar e resolver problemas que envolvem variáveis socioeconômicas ou técnico-científicas, usando representações algébricas.
Competência 6	Interpretar informações de natureza científica e social obtidas da leitura de gráficos e tabelas, realizando previsão de tendência, extrapolação, interpolação e interpretação.
Competência 7	Compreender o caráter aleatório e não-determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas numa distribuição estatística.

Fonte: Brasil, 2009.

A adoção de uma estrutura baseada em competências está associada à proposta de avaliar a capacidade dos estudantes de articular conceitos e procedimentos na resolução de situações que exigem interpretação, análise e tomada de decisão. Essa orientação encontra respaldo em referenciais curriculares mais amplos, como a Base Nacional Comum Curricular, que enfatiza o desenvolvimento de competências e habilidades como eixo estruturante da aprendizagem na educação básica (Brasil, 2018). Nesse contexto, os itens são elaborados de modo a integrar conteúdos matemáticos a situações diversas, demandando o estabelecimento de relações entre variáveis e a interpretação de diferentes formas de representação.

Do ponto de vista da elaboração do teste, a definição prévia de competências e habilidades funciona como um princípio organizador dos itens, orientando sua distribuição ao longo da prova. Essa distribuição estabelece, implicitamente, a expectativa de cobertura dos diferentes domínios do conhecimento matemático, bem como de contemplação de distintos níveis de complexidade. No entanto, essa organização é estabelecida em nível normativo, não implicando, necessariamente, que os itens associados a diferentes competências apresentem comportamentos equivalentes do ponto de vista psicométrico (Both, 2001; Gatti, 2005; Alves et al., 2020).

Nesse contexto, a organização dos itens segundo competências também se relaciona à expectativa de distribuição equilibrada das exigências ao longo da prova. Ainda que não se estabeleça, a priori, uma equivalência estrita entre os agrupamentos, pressupõe-se que a avaliação contemple diferentes níveis de dificuldade, de modo a cobrir uma faixa ampla da escala de proficiência. Entretanto, a concentração de itens mais exigentes ou mais acessíveis em determinadas competências pode alterar esse equilíbrio, afetando a forma como a prova distribui suas demandas cognitivas.

## Metodologia

Este estudo adota uma abordagem quantitativa, de natureza descritiva e inferencial, com o objetivo de analisar a dificuldade dos itens de Matemática do Enem em função das competências avaliativas da matriz de referência.

Os dados utilizados foram obtidos a partir dos microdados do Enem, disponíveis no site do Inep no momento da realização deste estudo<sup>7</sup>. Foram considerados os itens da área de Matemática aplicados entre os anos de 2009 e 2023, período posterior à reformulação do exame, quando passou a adotar a TRI como base para a estimação das proficiências.

Foram selecionados apenas os itens considerados válidos para análise, isto é, aqueles com parâmetros psicométricos estimados e divulgados nos microdados. Itens com parâmetros ausentes, inconsistentes ou não estimados foram excluídos. Ao final do processo de filtragem, o conjunto analisado foi composto por 661 itens.

As informações relativas à competência e à habilidade de cada item foram obtidas diretamente dos microdados do Enem, não sendo necessária classificação adicional. Essa informação permitiu agrupar os itens em sete conjuntos, correspondentes às competências avaliativas da área de Matemática, que constituem a unidade de análise comparativa deste estudo.

A variável principal analisada foi o parâmetro de dificuldade ( $b$ ) do modelo TRI, que representa o nível de proficiência associado ao ponto da curva em que a probabilidade de acerto atinge um valor intermediário, sendo interpretado como um indicador da exigência cognitiva da questão.

Cabe destacar que os valores do parâmetro  $b$  foram utilizados conforme estimados e divulgados pelo Inep, não sendo realizada reestimação dos parâmetros no presente estudo e que esses valores foram tratados como variável contínua, permitindo a comparação entre itens pertencentes a diferentes competências.

A análise dos dados foi realizada em duas etapas. Inicialmente, foram calculadas estatísticas descritivas do parâmetro  $b$  para cada competência, incluindo média, desvio padrão e valores mínimo e máximo, com o objetivo de caracterizar a distribuição da dificuldade dos itens.

Na sequência, foi aplicada a Análise de Variância de um fator, tendo como variável dependente o parâmetro de dificuldade ( $b$ ) e como fator independente a competência associada ao item, com o objetivo de comparar as médias entre os grupos. Os pressupostos do modelo foram verificados previamente, incluindo a normalidade dos resíduos e a homogeneidade das variâncias. A normalidade foi avaliada por inspeção gráfica e testes estatísticos, enquanto a homogeneidade das variâncias foi verificada por meio do teste de Levene. Diante da rejeição da hipótese nula, foram realizadas comparações múltiplas por meio do teste de Tukey, para identificar diferenças entre pares de competências. Além da significância estatística, foi estimado o tamanho do efeito por meio do coeficiente eta quadrado ( $\eta^2$ ), permitindo avaliar a magnitude das diferenças observadas.

Todas as análises foram realizadas no software R, adotando-se nível de significância de 5% ( $\alpha = 0,05$ ).

<sup>7</sup> Os microdados do Enem podem ser acessados em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>.

## Resultados

Esta seção apresenta os resultados das análises realizadas sobre a distribuição do parâmetro  $b$  entre as competências avaliativas da prova de Matemática do Enem. Inicialmente, são apresentadas as estatísticas descritivas, seguidas dos testes de pressupostos e das análises inferenciais.

### Estatísticas descritivas do parâmetro de dificuldade

A tabela 1 apresenta as estatísticas descritivas do parâmetro de dificuldade ( $b$ ) por competência, incluindo medidas de tendência central, dispersão e amplitude.

Tabela 1 –  
Estatísticas descritivas do parâmetro de dificuldade ( $b$ ) por competência.

Competência	n	Média	DP	Mediana	Mín	Máx
1	118	1,86	0,99	1,82	-0,72	5,61
2	95	1,75	1,10	1,93	-0,97	6,28
3	102	1,98	0,75	1,94	-0,05	4,78
4	82	1,78	0,96	1,87	-0,42	4,49
5	107	2,12	1,05	2,05	-1,26	5,35
6	76	1,67	1,00	1,79	-0,42	4,28
7	90	2,21	0,98	2,13	-0,24	4,65

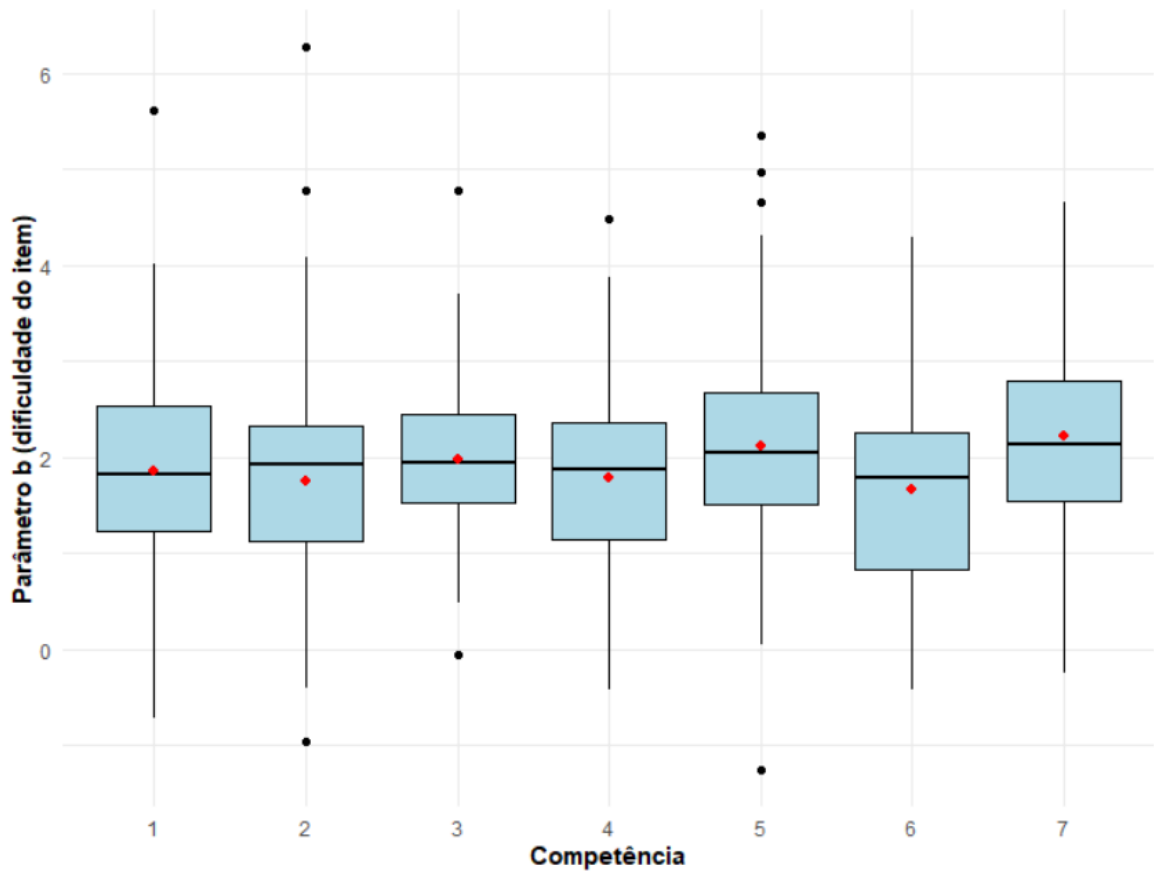
Fonte: autores, 2026.

Os resultados mostram que a distribuição dos itens entre as competências não é homogênea. A competência 1 reúne o maior número de itens analisados ( $n = 118$ ), enquanto a competência 6 apresenta o menor quantitativo ( $n = 76$ ). As demais competências apresentam frequências intermediárias, variando entre 82 e 107 itens.

Em relação ao parâmetro  $b$ , observa-se variação nas médias absolutas entre as competências. A competência 7 apresenta o maior valor médio ( $M = 2,21$ ), seguida da competência 5 ( $M = 2,12$ ), enquanto a competência 6 apresenta o menor valor médio ( $M = 1,67$ ). As demais competências situam-se em níveis intermediários, com médias entre 1,75 e 1,98.

As diferenças na dispersão e na amplitude dos valores podem ser observadas na figura 2, evidenciando variações na distribuição do parâmetro  $b$  entre as competências, bem como a presença de valores extremos em alguns agrupamentos.

Figura 2 –  
Boxplots das variações nos parâmetros de dificuldade ( $b$ ) por competências.



Fonte: autores, 2026.

De modo geral, os resultados descritivos indicam que, embora haja variação na distribuição do parâmetro  $b$  entre as competências, os valores encontram-se concentrados em intervalos relativamente próximos ao longo da escala, com sobreposição considerável entre os agrupamentos.

#### Análise de variância e comparações múltiplas

A aplicação da Anova foi precedida pela verificação dos pressupostos de normalidade e homogeneidade das variâncias. O teste de Shapiro-Wilk indicou que a distribuição do parâmetro  $b$  apresentou comportamento compatível com a normalidade na maioria das competências, com exceção da competência 2 ( $p < 0,001$ ). O teste de Levene não indicou violação da homogeneidade das variâncias entre os grupos ( $F(6, 654) = 1,65$ ;  $p = 0,1296$ ). A Tabela 2 apresenta os resultados Anova.

Tabela 2 –  
Resultados da Anova.

Fonte	gl	Soma de quadrados	Quadrado médio	F	p
Competência	6	21,4	3,562	3,71	0,0012
Resíduo	654	628,0	0,960		

Fonte: autores, 2026.

Os resultados indicam efeito estatisticamente significativo da competência sobre o parâmetro de dificuldade ( $F(6, 654) = 3,71$ ;  $p = 0,0012$ ), evidenciando que as médias do parâmetro  $b$  não são equivalentes entre os agrupamentos. O tamanho do efeito foi pequeno ( $\eta^2 = 0,03$ ;  $\omega^2 = 0,02$ ), indicando que a proporção da variabilidade explicada pela competência é reduzida.

Diante da significância observada, foram realizadas comparações múltiplas por meio do teste de Tukey, com o objetivo de identificar quais competências diferem entre si. A Tabela 3 apresenta apenas os pares de competências para os quais foram observadas diferenças estatisticamente significativas no teste de Tukey

Tabela 3 –  
Pares de competências com diferenças estatisticamente significativas segundo o teste de Tukey.

Comparação	Diferença	IC 95%	p
2 – 7	-0,462	[-0,891; -0,033]	0,025
5 – 6	0,447	[0,009; 0,885]	0,042
6 – 7	-0,544	[-0,999; -0,088]	0,008

Fonte: autores, 2026.

Os resultados indicam que as diferenças estatisticamente significativas concentram-se num número restrito de pares de competências. Observa-se que a competência 7 apresenta valores médios de dificuldade superiores aos da competência 2 e da competência 6, enquanto a competência 5 apresenta valores superiores aos da competência 6.

Para os demais pares de competências, não foram observadas diferenças estatisticamente significativas ao nível de 5%. De modo geral, os resultados indicam a presença de variações entre as competências na dificuldade dos itens, embora com magnitude reduzida, conforme evidenciado pelo tamanho do efeito observado.

## Discussão

Os resultados deste estudo indicam que a dificuldade dos itens de Matemática do Enem varia entre as competências avaliativas, mas essa variação ocorre de forma limitada. A significância estatística observada na Anova mostra que a competência à qual o item pertence está associada a diferenças no parâmetro  $b$ , porém o tamanho do efeito foi pequeno e as comparações múltiplas revelaram diferenças significativas em apenas três pares de competências.

Esse resultado permite evitar duas leituras simplificadoras: de um lado, a ideia de que todas as competências seriam empiricamente equivalentes; de outro, a interpretação de que haveria um desequilíbrio amplo e generalizado na prova. A análise dos dados sugere um quadro intermediário, em que a matriz apresenta assimetrias localizadas, mas não uma ruptura global de coerência.

Cabe destacar que em avaliações organizadas por competências, a matriz de referência define, em nível normativo, os domínios que devem compor a prova. No entanto, como argumenta Messick (1989), a validade não se esgota na definição do construto, exigindo evidências de que a interpretação dos escores é sustentada pela forma como esse construto é efetivamente operacionalizado. Nessa direção, a estrutura interna do teste constitui uma fonte importante de evidência, pois permite examinar se os diferentes agrupamentos previstos na matriz se comportam de maneira coerente com a medida pretendida.

Os resultados encontrados neste estudo não indicam ausência de coerência, mas mostram que a simples distribuição dos itens por competências não garante equivalência empírica na exigência psicométrica. Em outras palavras, a organização formal da prova não se traduz automaticamente em homogeneidade de dificuldade entre seus diferentes domínios. Ao mesmo tempo, o fato de a competência explicar apenas uma proporção reduzida da variabilidade do parâmetro  $b$  indica que a dificuldade dos itens não é determinada de forma predominante pela competência à qual pertencem.

Isso sugere que outros fatores, não examinados diretamente neste estudo, podem exercer papel importante na configuração da dificuldade dos itens, como a formulação dos enunciados, o grau de contextualização, a complexidade linguística, o tipo de representação mobilizada, a presença de gráficos ou tabelas e o número de operações cognitivas exigidas. Assim, a competência aparece como um fator relevante, mas insuficiente, para explicar a variabilidade observada.

Esse achado é consistente com uma leitura mais complexa da medida educacional, segundo a qual o desempenho em testes de larga escala resulta da combinação entre conteúdo, forma do item, condições de escolarização e repertórios previamente elaborados pelos estudantes.

Há, contudo, um resultado que merece maior atenção: os valores médios do parâmetro  $b$  foram elevados em todas as competências. Tomando como referência os intervalos propostos por Rabelo (2013), segundo os quais valores de  $b$  iguais ou superiores a 1,28 correspondem à faixa de itens muito difíceis, observa-se que todas as médias por competência situam-se acima desse ponto.

Ainda que essa classificação tenha sido originalmente formulada para itens individuais e não para médias agregadas, ela oferece um parâmetro útil de interpretação, pois sugere um padrão geral de elevada exigência psicométrica da prova de Matemática. Esse aspecto aprofunda a leitura dos resultados: não se trata apenas de discutir qual competência é relativamente mais difícil, mas de reconhecer que, em média, todas as competências se concentram numa faixa elevada da escala de dificuldade.

Nesse sentido, o problema da coerência interna aqui investigado, para além da comparação entre agrupamentos, sinaliza para a localização geral dos itens ao longo da escala de proficiência. Se a maior parte dos itens se concentra em níveis elevados de  $b$ , a prova tende a ser mais informativa para estudantes com proficiências mais altas, o que

pode reduzir sua sensibilidade em faixas inferiores da escala (Rabelo, 2013). Essa questão é particularmente importante em avaliações de larga escala com funções seletivas, nas quais a distribuição dos itens ao longo da escala não é apenas uma questão técnica, mas também um aspecto com implicações substantivas para a interpretação dos desempenhos.

No plano das comparações específicas, a competência 7 apresentou a maior média absoluta de dificuldade e diferenças estatisticamente significativas em relação às competências 2 e 6. Esse resultado sinaliza que a competência 7 ocupa, no conjunto analisado, uma posição relativamente mais exigente, embora não permita sustentar que ela seja, de forma generalizada e inequívoca, a mais difícil da prova em todos os contrastes possíveis.

A interpretação mais adequada é que essa competência tende a concentrar itens de maior dificuldade média, sobretudo quando comparada a agrupamentos específicos. Essa nuance é importante de ser destacada, porque preserva a fidelidade aos resultados empíricos e evita extrapolações indevidas. Do ponto de vista substantivo, essa tendência é plausível, visto que a competência 7 está relacionada à compreensão do caráter aleatório e não determinístico dos fenômenos e ao uso de instrumentos estatísticos e probabilísticos.

A literatura especializada tem apontado que Estatística e Probabilidade, embora previstas nos documentos curriculares, ainda ocupam posição instável no currículo praticado, frequentemente subordinadas a conteúdos algébricos e aritméticos mais tradicionais. Frei, Rosa e Biazzi (2023), por exemplo, destacam fragilidades na formação de professores para o ensino desses conteúdos, enquanto Magalhães (2022) chama atenção para dificuldades persistentes na aprendizagem estatística, mesmo quando se reconhece sua importância social e educacional.

Sob essa perspectiva, o comportamento da competência 7 pode refletir não apenas uma característica intrínseca dos itens, mas também uma história curricular marcada por menor consolidação pedagógica desses conteúdos.

Ainda assim, é preciso evitar leituras causalistas fortes. Os dados deste estudo não permitem afirmar que a maior dificuldade média da competência 7 decorra exclusivamente das fragilidades do ensino de Estatística e Probabilidade. O que eles mostram é que os itens classificados nessa competência tendem, em média, a ocupar posições mais altas na escala de dificuldade.

A explicação dessa tendência exige investigações adicionais que articulem análise psicométrica, análise de conteúdo e estudos sobre currículo e prática pedagógica. Em especial, seria importante examinar se esses itens se distinguem por maior complexidade textual, por uso mais intenso de linguagem técnica, por exigirem múltiplas etapas de raciocínio ou por mobilizarem formas específicas de inferência e tomada de decisão sob incerteza, o que pode ser feito em estudos futuros.

Por outro lado, a competência 6 apresentou a menor média de dificuldade e diferenças significativas em relação às competências 5 e 7. Esse resultado também merece atenção, pois sugere que os itens relacionados à leitura e interpretação de gráficos e tabelas tendem a exigir menor nível de proficiência.

Uma hipótese plausível é que esses itens sejam mais acessíveis por mobilizarem habilidades de leitura de informações visuais e comparação de dados, frequentemente associadas a estratégias mais imediatas de interpretação. Isso não significa, evidentemente, que tais itens sejam simples em sentido absoluto, mas indica que ocupam posição relativamente menos exigente no conjunto analisado.

Esse contraste entre competências 6 e 7 é particularmente interessante, porque ambas se aproximam do campo do tratamento da informação, embora mobilizem operações cognitivas distintas: a competência 6 tende a concentrar tarefas de leitura e interpretação, enquanto a competência 7 envolve maior aproximação com raciocínio probabilístico e estatístico.

Outro aspecto relevante é que a maior parte das comparações entre competências não foi significativa. Esse resultado é tão importante quanto as diferenças encontradas, porque reforça a ideia de que a prova preserva um padrão geral de relativa proximidade entre seus agrupamentos. Assim, em termos de estrutura da avaliação, isso sugere que as competências não configuram blocos radicalmente distintos de exigência.

### **Considerações finais**

Este estudo teve como objetivo analisar a dificuldade dos itens de Matemática do Enem, no período de 2009 a 2023, com base no parâmetro  $b$  da TRI, comparando as sete competências avaliativas que compõem a matriz de referência da área. A análise evidenciou que a dificuldade dos itens varia entre as competências, com diferenças estatisticamente significativas, porém de pequena magnitude.

De modo geral, os resultados indicam que a organização da prova por competências não se traduz em equivalência empírica plena na dificuldade dos itens, ainda que as assimetrias observadas sejam pontuais e não generalizadas. Observou-se, adicionalmente, um padrão de elevada dificuldade média em todas as competências, o que desloca a análise do contraste entre agrupamentos para a configuração global da prova ao longo da escala de proficiência.

Nesse contexto, a competência 7 destacou-se por apresentar valores médios mais elevados e diferenças significativas em relação a parte das demais competências, especialmente as competências 2 e 6, configurando-se como um ponto de atenção relativo dentro da matriz avaliativa.

As implicações desses resultados tornam-se particularmente relevantes quando se considera o uso atual e projetado do Enem. Além de seu papel consolidado no acesso ao ensino superior, o exame tem sido progressivamente associado à produção de indicadores educacionais e ao monitoramento da qualidade da educação básica (Inep, 2026). À medida que se ampliam suas funções, torna-se ainda mais necessário investigar a consistência interna de sua estrutura e a forma como suas dimensões se distribuem empiricamente, de modo a assegurar interpretações adequadas dos desempenhos.

Este estudo apresenta algumas limitações que devem ser consideradas. A análise não considerou características estruturais dos itens, como extensão dos enunciados, presença de recursos visuais ou nível de contextualização, que podem influenciar a dificuldade. Também não foram analisadas possíveis variações ao longo do tempo ou entre diferentes edições do exame, o que limita a compreensão de dinâmicas temporais.

Diante disso, pesquisas futuras podem aprofundar a análise da prova de Matemática do Enem a partir de abordagens mais abrangentes, incorporando características dos itens e variáveis contextuais. Investigações que articulem a análise quantitativa com estudos de conteúdo e com o campo do currículo também podem contribuir para compreender de forma mais ampla as relações entre a matriz avaliativa, a prática pedagógica e o desempenho dos estudantes. Por fim, análises longitudinais e estudos sobre invariância de medida podem oferecer evidências adicionais sobre a estabilidade e a comparabilidade das estimativas de proficiência ao longo do tempo.

Em síntese, os resultados reforçam a importância de monitorar empiricamente a relação entre a estrutura normativa da matriz de referência e o funcionamento psicométrico do exame, contribuindo para uma compreensão mais qualificada do papel das avaliações em larga escala no contexto educacional brasileiro.

### Agradecimentos

Os autores agradecem ao Instituto Federal de Minas Gerais, à Fundação de Amparo à Pesquisa do Estado de Minas Gerais e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio à realização desta pesquisa, desenvolvida com a participação de estudantes bolsistas de Iniciação Científica Júnior vinculados ao edital n. 373/2024 do IFMG.

### Referências

AERA; APA; NCME. *Standards for educational and psychological testing*. Washington, American Educational Research Association, 2014.

ALVES, Ana Cristina dos Santos; MARQUES, Cíntia Bueno; AMARAL, Lisandra; CENTENARO, Luciano; BERTOLLO, Maria Rita; CAPITANI, Renato; PIASSINI, Rui; SHEILA CARDOSO, Shirley. Avaliação por habilidades e competências. *Caderno Marista de Educação*, Porto Alegre, v. 9, n. 1, 2020, p. e39574, 2020.

BOTH, Ivo José. Mais do que avaliar por competência cabe valorizar a capacidade criadora e empreendedora. *Revista Diálogo Educacional*, Curitiba, v. 2, n. 4, 2001, p. 199-234.

BRASIL. *Base nacional comum curricular*. Brasília: MEC, 2018. Disponível em: <http://basenacionalcomum.mec.gov.br/>. Acesso em: 9 abr. 2026.

BRASIL. *Matriz de referência do Enem*. Brasília: Inep, 2009. Disponível em: [https://download.inep.gov.br/download/enem/matriz\\_referencia.pdf](https://download.inep.gov.br/download/enem/matriz_referencia.pdf). Acesso em: 17 abr. 2025.

BRASIL. *Microdados do Exame Nacional do Ensino Médio (Enem)*. Brasília: Inep, 2009-2023. Disponível em: <https://www.gov.br/inep/ptbr/acesso-a-informacao/dados-abertos/microdados/enem>. Acesso em: 17 abr. 2025.

BRASIL. *Entenda a sua nota no Enem: guia do participante*. Brasília: Inep, 2021. Disponível em: [https://download.inep.gov.br/publicacoes/institucionais/avaliacoes\\_e\\_exames\\_da\\_educacao\\_basica/entenda\\_a\\_sua\\_nota\\_no\\_enem\\_guia\\_do\\_participante.pdf](https://download.inep.gov.br/publicacoes/institucionais/avaliacoes_e_exames_da_educacao_basica/entenda_a_sua_nota_no_enem_guia_do_participante.pdf). Acesso em: 17 abr. 2025.

FREI, Fernando; ROSA, Jussara Santos; BIAZI, Ângelo Henrique. Professores de Matemática estão preparados para o ensino de Estatística e Probabilidade? *Revista Internacional de Pesquisa em Educação Matemática*, Brasília, v. 13, n. 2, 2023, p. 1-17.

GATTI, Bernardete Angelina. Habilidades cognitivas y competencias sociales. *Enunciación*, Bogotá, v. 10, n. 1, 2005, p. 123-132.

INEP. *Enem será utilizado para avaliar educação brasileira*. Brasília, DF, 30 mar. 2026. Disponível em: <https://www.gov.br/inep/pt-br/centrais-de-conteudo/noticias/enem/enem-sera-utilizado-para-avaliar-educacao-brasileira>. Acesso em: 9 abr. 2026.

LINGARD, Bob; SELLAR, Sam; LEWIS, Steven. *Accountabilities in schools and school systems*. Oxford: Oxford University Press, 2017.

MAGALHÃES, Marcos Nascimento. Duas propostas para a aprendizagem em Estatística. *Revista Baiana de Educação Matemática*, Juazeiro, v. 3, n. 1, 2022, p. 1-16.

MESSICK, Samuel. Validity. In: LINN, Robert L. (org.). *Educational measurement*. New York: Macmillan, 1989.

PASQUALI, Luiz. *TRI teoria de resposta ao item: teoria, procedimentos e aplicações*. Curitiba: Appris, 2020.

R CORE TEAM. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2024. Disponível em: <https://www.rproject.org/>. Acesso em: 17 mar. 2026.

RABELO, Mauro. *Avaliação educacional: fundamentos, metodologia e aplicações no contexto brasileiro*. Rio de Janeiro: SBM, 2013.

SOARES, Denilson Junio Marques. *Teoria clássica dos testes e teoria de resposta ao item aplicadas em uma avaliação de matemática básica*. Viçosa: Universidade Federal de Viçosa, 2018.

SOARES, Denilson Junio Marques; SOARES, Talita Emidio Andrade; SANTOS, Wagner dos. Análise da qualidade psicométrica da prova de matemática do Exame Nacional do Ensino Médio brasileiro de 2018. *Actualidades Investigativas en Educación*, San José, v. 21, n. 1, 2021, p. 1-28.

VERGER, Antoni; NORMAND, Romuald. New public management and education: theoretical and conceptual elements for the study of a global education reform model. *Educação e Sociedade*, Campinas, v. 36, n. 132, 2015, p. 599-622.

VERGER, Antoni; NOVELLI, Mario; ALTINYELKEN, Hülya K. *Global education policy and international development: new agendas, issues and policies*. London: Bloomsbury, 2012.