



UM MODELO DE REGRESSÃO LINEAR PARA PREDIZER A DIFICULDADE DE ITENS DE MATEMÁTICA DO ENEM A PARTIR DE SUAS CARACTERÍSTICAS TEXTUAIS E VISUAIS

Maria Eduarda Loschi Xavier¹ - Dejaina Aparecida Teixeira Lima² - Izabelli de Carvalho Oliveira³ - Rafael Santiago do Nascimento⁴ - Daniele Pires Magalhães⁵ - Denilson Junio Marques Soares⁶

RESUMO

Este estudo buscou compreender como a configuração visual e textual dos itens da prova de Matemática do Exame Nacional do Ensino Médio (ENEM) exerce influência sobre sua dificuldade psicométrica, mensurado por meio da Teoria de Resposta ao Item (TRI). Para isso, foram analisados 670 itens aplicados ao longo do período de 2009 a 2023, considerando, em especial, a relação entre duas dimensões estruturais: a extensão textual, medida pelo número de palavras, e a presença de recursos visuais, como figuras, gráficos e tabelas. A investigação utilizou modelos de regressão linear com variáveis padronizadas, permitindo isolar o efeito relativo de cada característica na determinação da dificuldade. Os resultados revelaram que a quantidade de palavras apresentou associação positiva com a dificuldade: quanto mais extenso o enunciado, mais difícil, em média, o item. Por outro lado, a inserção de elementos visuais mostrou-se inversamente relacionada à dificuldade, se configurando como facilitadores do processo cognitivo, provavelmente por condensarem informações de maneira mais clara e acessível. Essas evidências indicam que a apresentação dos itens não deve ser considerada apenas como um aspecto formal ou estético, mas como parte integrante da qualidade da avaliação. O modo como as informações são estruturadas influencia a leitura, a interpretação e a mobilização de estratégias de resolução, afetando diretamente a experiência de quem realiza a prova.

Palavras-chave: ENEM. Matemática e suas Tecnologias. Complexidade Visual. Complexidade Textual.

¹ Estudante do Curso Técnico Integrado em Metalurgia do IFMG Campus Ouro Preto

² Estudante do Curso Técnico Integrado em Metalurgia do IFMG Campus Ouro Preto

³ Estudante do Curso Técnico Integrado em Metalurgia do IFMG Campus Ouro Preto

⁴ Estudante do Curso Técnico Integrado em Metalurgia do IFMG Campus Ouro Preto

⁵ Doutora em Modelagem Computacional, Professora no Instituto Federal de Minas Gerais Campus Ouro Preto

⁶ Doutor em Educação, Professor no Instituto Federal de Minas Gerais Campus Ouro Preto



1 INTRODUÇÃO

A análise da dificuldade dos itens em avaliações de larga escala deve considerar não apenas as competências cognitivas demandadas, mas também os aspectos estruturais e linguísticos que influenciam a forma como a tarefa é processada pelo estudante (PINHEIRO; CONCEIÇÃO, 2021).

No ENEM, a calibração dos itens se apoia na Teoria de Resposta ao Item (TRI), sendo o parâmetro b do modelo utilizado (logístico de três parâmetros) uma medida objetiva da dificuldade do item, expressa na mesma escala da proficiência dos participantes (BRASIL, 2021). Pesquisas recentes têm evidenciado, contudo, que a extensão textual e a presença de recursos visuais, como gráficos e tabelas, podem alterar a experiência de resolução, configurando-se como barreiras ou facilitadores (PINHEIRO; CONCEIÇÃO, 2021; LÓS, 2021; MACHADO; SILVA; ARRUDA, 2023).

Diante disso, justifica-se a necessidade de investigar como esses elementos estruturais interferem na dificuldade psicométrica, uma vez que impactam diretamente a validade, a acessibilidade e a equidade avaliativa. Assim, este trabalho tem como objetivo analisar de que modo a complexidade visual e textual dos itens de Matemática do ENEM influencia sua dificuldade, tomando como referência evidências estatísticas e psicométricas.

2 METODOLOGIA

A pesquisa caracteriza-se por uma abordagem quantitativa, de caráter exploratório e explicativo. A base empírica contemplou todos os itens aplicados entre 2009 e 2023, somando 675 questões (45 por edição, em 15 anos consecutivos) (BRASIL, 2009–2023). Desse total, cinco itens foram excluídos por terem sido anulados e, portanto, não apresentarem parâmetros psicométricos estimados.

Cada item foi classificado segundo quatro variáveis independentes: (i) número de palavras no enunciado, utilizado como indicador da complexidade textual; e três variáveis binárias, que registram a presença (1) ou ausência (0) de figuras, gráficos e tabelas/quadros, representando os elementos de complexidade visual. A variável dependente considerada foi o parâmetro b da TRI.



As análises estatísticas foram realizadas no software R, versão 4.4.3 (R CORE TEAM, 2024), por meio de regressão linear múltipla. Para assegurar a comparabilidade entre os efeitos das variáveis, aplicou-se a padronização por z-score, e os resultados foram reportados em termos de coeficientes padronizados (β).

Complementarmente, foram examinados os intervalos de confiança de 95% e os indicadores globais de ajuste do modelo (R^2 e estatística F), de modo a avaliar a robustez e a significância dos efeitos observados. Os microdados utilizados foram obtidos no portal oficial do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

3 RESULTADOS

O parâmetro de dificuldade apresentou média de 1,92 (DP = 0,99), variando de -1,26 a 6,28. Os enunciados continham, em média, 132,8 palavras (mín. = 26; máx. = 338), e verificou-se a presença de figuras em 30,7% dos itens, gráficos em 14,1% e tabelas em 14,4%. A regressão linear múltipla revelou modelo significativo ($F(4,654) = 12,54$; $p < 0,001$; $R^2 = 0,0713$), com todos os preditores estatisticamente relevantes. O número de palavras apresentou efeito positivo ($\beta = 0,208$), indicando que enunciados mais longos aumentam a dificuldade, enquanto os recursos visuais reduziram o parâmetro b, em especial os gráficos ($\beta = -0,155$).

A Figura 1 apresenta os coeficientes padronizados estimados no modelo, acompanhados de seus respectivos intervalos de confiança de 95%, evidenciando a direção e a magnitude dos efeitos estruturais sobre a dificuldade dos itens.

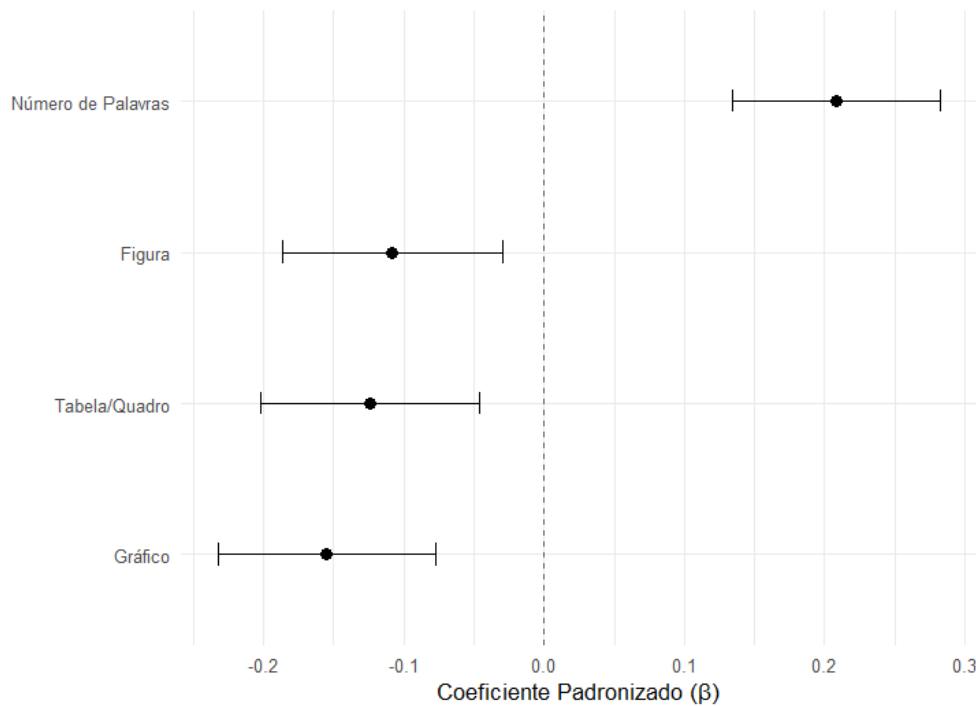


Figura 1- Efeitos Padronizados sobre a Dificuldade dos Itens (Parâmetro b)
Fonte: Elaborado pelos autores (2025)

Os resultados evidenciaram que enunciados mais extensos aumentam a dificuldade psicométrica dos itens, corroborando estudos que relacionam textos longos ao acréscimo da carga cognitiva, por exigirem seleção de informações, retenção na memória de trabalho e articulação textual (LÓS, 2021). Embora a leitura seja parte integrante do construto avaliado pelo ENEM, o uso de vocabulário excessivamente técnico ou enunciados prolixos pode inflar artificialmente a dificuldade, criando barreiras externas à habilidade matemática (HALADYNA; RODRIGUEZ, 2013). Em sentido oposto, a presença de recursos visuais, sobretudo gráficos, reduziu a dificuldade dos itens, conforme teorias que destacam o papel das representações visuais no processamento da aprendizagem (DUVAL, 2003).

Em termos práticos, cada 50 palavras adicionais elevam a dificuldade em 0,22 pontos, enquanto a presença de um gráfico reduz em média 0,44 pontos. Apesar disso, o baixo coeficiente de determinação ($R^2 = 0,0713$) mostra que outros fatores, como o conteúdo matemático e o tipo de habilidade exigida, também influenciam fortemente a dificuldade, reforçando seu caráter multifatorial.



4 CONCLUSÃO

Os resultados obtidos neste estudo confirmam a hipótese de que a complexidade textual dos itens tende a dificultar sua resolução, ao passo que elementos visuais funcionam como facilitadores cognitivos. Reconhece-se, entretanto, que a dificuldade dos itens decorre também de outros fatores, como o conteúdo matemático e às operações cognitivas requeridas, não contemplados no modelo estimado que é multidimensional, o que pode ser verificado por meio do coeficiente de determinação obtido.

Assim, em pesquisas futuras sugere-se ampliar os modelos preditivos, incluindo outras variáveis que podem influenciar a dificuldade dos itens. Também é indicado a realização de pesquisas com abordagens qualitativas para complementar a análise psicométrica, contribuindo para uma interpretação pedagógica do processo.

REFERÊNCIAS

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). **Exame Nacional do Ensino Médio (ENEM)**: documentos técnicos e microdados. Brasília: Inep, 2009–2023.

DUVAL, R. **Semiosis e pensamento humano**: registros semióticos e aprendizagens intelectuais. São Paulo: Editora Livraria da Física, 2003.

HALADYNA, T. M.; RODRIGUEZ, M. C. **Developing and validating test items**. 3. ed. New York: Routledge, 2013.

LÓS, A. F. S. **A visualização matemática na educação básica**: dos percursos teóricos à resolução de problemas do ENEM. 2021. 122 f. Dissertação (Mestrado PROFMAT), Universidade Federal de Alagoas, Maceió, 2021.

MACHADO, P. A. P.; SILVA, F. S.; ARRUDA, T. Z. Análise de conteúdo sobre as imbricações da geometria nas provas do Novo ENEM. **Amazônia**: Revista de Educação em Ciências e Matemáticas, v. 19, n. 42, p. 86–101, 2023.

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**, v. 20, n. 1, p. AG4, 2021.

R CORE TEAM. **R**: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2024.