

O Funcionamento Diferencial nos Itens de Física do Enem (2021)

Viviana da Cruz Vicente¹

Doutora pelo Programa de Pós-Graduação Interunidades em Ensino de Ciências, Instituto de Física (IFUSP), São Paulo, São Paulo, Brasil
vivianavic@alumni.usp.br

André Machado Rodrigues²

Professor Doutor, pelo Programa de Pós-Graduação Interunidades em Ensino de Ciências, Instituto de Física (IFUSP), São Paulo, São Paulo, Brasil
rodrigues@if.usp.br

Resumo

A pesquisa foi realizada a partir dos microdados, e itens de física, da primeira aplicação de 2021 do Exame Nacional do Ensino Médio (Enem). Os itens foram investigados quanto a existência de representações de seres humanos, gráficos, imagens e fórmulas. Simultaneamente, ocorreu a análise do Funcionamento Diferencial do Item (termo conhecido em inglês pela sigla DIF). Um item com DIF significativa, indica que dois grupos com a mesma proficiência apresentam probabilidades de acerto distintas. A metodologia foi organizada em duas partes: análise de conteúdo e estatística. Sendo que, no que se refere a primeira, foram utilizados os conhecimentos de Bardin (2016) e na segunda o programa R para a obtenção do coeficiente de razão de chances, Mantel-Haenzel (MH) e p-valor. A pesquisa revelou uma tendência de viés em direção aos homens.

Palavras-chave: Enem, Ensino de Física, Gênero.

¹ O autor 1 agradece às pessoas que o ajudaram

² O autor 2 agradece às pessoas que o ajudaram

1 Introdução

Desde a década de 1980, os sistemas de avaliação em larga escala se materializaram em muitos países da América Latina e tinham como objetivo divulgar os níveis de desenvolvimento da qualidade da educação. No Brasil, no ano em questão, a quantidade de literatura educacional que questionava as pedagogias vigentes era ampla. Conforme Silva (1995) e Enguita (1989), a escola incentivava a reprodução das desigualdades sociais afetando principalmente as camadas mais pobres da população. A literatura, dentro do contexto escolar da época, começa a abordar a escola como um espaço em que são colocados em jogo os valores das classes hegemônicas. Em razão do fracasso dos alunos, principalmente das escolas públicas, os processos educativos são questionados e apontados como responsáveis pelo insucesso na promoção de ensino de qualidade.

Na segunda metade dos anos 1990, a avaliação é colocada no centro das políticas educacionais. Na época citada, é observada a valorização dos exames em larga escala por meio de testes que visam classificar e avaliar as unidades de ensino. Na concepção de Freitas (2007, p.1), o incentivo para abordar a avaliação dos sistemas educacionais se baseava no interesse pelo “(..) monitoramento da educação básica e superior, propondo-se não só a aferição da qualidade dos resultados de ambas como a indução da qualificação pretendida para os sistemas e as instituições de ensino”. Na mesma década, com o propósito de avaliar a qualidade, equidade e a eficiência, emerge o Sistema de Avaliação da Educação Básica (SAEB). Posteriormente, a criação do SAEB, surge o Exame Nacional do Ensino Médio (Enem) e o Exame de Desempenho Nacional dos Estudantes (ENADE).

No que se refere ao Enem, criado em 1998 e com forte adesão aos egressos da escola básica, o instrumento foi submetido a uma série de mudanças no decorrer dos mais de 25 anos de história. Entre essas, ocorreram alterações nas funcionalidades do exame, no formato de provas e método de correção, entre outras. Inicialmente, o objetivo principal era avaliar o desempenho de estudantes no final do ensino médio. Após 2009, o Enem tornou-se um instrumento de seleção para acesso ao Ensino Superior e para financiamentos estudantis. A prova que no ano de criação era composta por 63 itens e uma redação, em 2009 passou para 180 e que foram divididos em quatro blocos (Ciências Humanas e suas Tecnologias; Ciências da Natureza suas Tecnologias; Matemática e suas tecnologias; e Linguagens, Códigos e suas Tecnologias). A respeito do método de correção de provas, que utilizava a Teoria Clássica de Testes (TCT), a partir de 2009 empregou a Teoria de Resposta ao Item (TRI) no cálculo das notas. Para Lobato et. al (2024), a TRI é uma mistura de análises envolvendo matemática, estatística e psicometria.

Sobre a psicometria, campo da psicologia que busca produzir e aplicar instrumentos de mensuração, há um método de análise estatística utilizado na verificação de vieses: o Funcionamento Diferencial do Item (termo em inglês Differential item Functioning- DIF). De modo geral, um item apresenta DIF quando a probabilidade de acertos de dois grupos ou mais com o mesmo traço latente é diferente. Assim sendo, um item com DIF fere a o aspecto da confiabilidade e, portanto, interfere na qualidade do exame. Considerando dois grupos de respondentes do Enem, masculino e feminino, o presente trabalho visa identificar se

elementos como representações de seres humanos, gráficos, imagens e fórmulas, ocasionam DIF significativa de gênero e, portanto, influenciam na qualidade do exame.

2 Metodologia

A pesquisa foi desenvolvida a partir dos microdados e do caderno de provas do bloco de Ciências da Natureza, e suas Tecnologias, do Enem (1ª aplicação de 2021). De natureza qualitativa e quantitativa, a metodologia foi composta por duas partes: 1ª) análise de conteúdo e 2ª) investigação estatística. No que se refere a análise de conteúdo, nela foram empregados os conceitos defendidos por Bardin (2016). De acordo com a autora, a análise de conteúdo deve obedecer a três etapas principais: 1ª pré-análise, 2ª exploração do material e 3ª inferências. Na primeira etapa, foi realizada uma leitura flutuante do caderno de provas (com a finalidade de selecionar os itens que eram exclusivamente de física). Na segunda etapa, fase da exploração do material, os itens foram avaliados quanto a presença ou ausência de elementos como: representação de seres humanos, gráficos, imagens e fórmulas. Na terceira etapa, das inferências, foram contabilizadas as frequências para cada um dos elementos descritos na segunda etapa. Por fim, os dados obtidos foram interpretados.

Na parte estatística, viabilizada pelo software R, ocorreu o levantamento dos valores de razão de chances, p-valor e log do coeficiente de Mantel-Haenzel. Sobre a razão de chances, ela quantifica a associação entre duas variáveis categóricas. O p-valor, em inglês denominado *p-value*, mostra a significância da amostra e o coeficiente de Mantel-Haenzel permite classificar qual o tamanho do DIF. Para efeitos de comparação, a habilidade utilizada foi estratificada em cinco níveis. Deste modo, os scores são posicionados na mesma faixa de proficiência.

Por meio do teste de hipótese, ocorreram inferências sobre o modelo probabilístico. Na literatura, dentro das classificações para os testes, são apresentadas dois tipos de hipótese: a nula e a alternativa. A primeira, hipótese nula, é estabelecida como verdadeira. Para valores de $H_0 = 1$ é possível dizer inicialmente que não foi detectada a DIF. A segunda, hipótese alternativa, se contrapõe a hipótese nula. Quando é diagnóstica a hipótese alternativa, significa que foram identificadas diferenças entre os parâmetros de comparação. A hipótese nula, em que não existe relação entre dois fenômenos medidos, para o método MH, pode ser expressa por:

$$H_0 = [C_{mm}/E_{mm}] / [C_{fm}/E_{fm}] = 1 \quad m=1, \dots$$

Onde:

C_{fm} = item correto para o grupo feminino que respondeu m;

C_{mm} = item correto para o grupo masculino que respondeu m;

E_{fm} = item errado para o grupo feminino que respondeu m;

E_{mm} = item errado para o grupo masculino que respondeu m;

C_{tm} = total de respostas corretas considerando os dois grupos.

Apesar dos valores do teste de hipótese expressarem a existência ou não do DIF, é importante determinar a significância dos resultados do teste de hipótese. Os dados do p-valor permitem inferir se realmente a amostra é significativa para que a hipótese nula seja rejeitada ou aceita. Para analisar o p-valor, foi adotada a seguinte escala: $p < 0,05$ rejeita-se a hipótese nula; $p > 0,05$ (ou igual) aceita-se a hipótese nula.

Em relação aos valores da razão de chances, que mede a força da associação entre a exposição e um resultado e em inglês é denominado *odds ratio* (OR), podem ser interpretados da seguinte maneira: $OR = 1$ nenhuma associação; $OR > 1$ o resultado é mais provável e $OR < 1$ associação menos provável. A média ponderada, para a razão de chances de cada estrato, é dada a partir do cálculo do coeficiente de Mantel-Haenzel.

$$\alpha_{MH} = \left[\sum_m C_{mm} E_{fm} / N_{tm} \right] / \left[\sum_m C_{fm} E_{mm} / N_{tm} \right]$$

Para transformar a razão de chances de Mantel-Haenzel (α_{MH}) em logaritmo, foi utilizada a seguinte expressão: $\alpha^*_{MH} = -2,35 \ln(\alpha_{MH})$. A razão de chances α^*_{MH} , segundo Glads et al. (2023), pode ser categorizada em: a) Nenhuma (quando as questões apresentam diferença insignificante entre a razão de chances), b) Moderada (quando as questões apresentam de modo intermediário a significância) e c) Forte (quando as questões apresentam significância estatística e com uma grande diferença). Neste trabalho, adicionalmente as opções acima propostas por Glads et al. (2023) para a análise da razão de chances, foi observada a necessidade e inclusão de uma nova categoria denominada “Tendenciosa” (quando as questões obedeceram ao intervalo $|\alpha^*| < 1$ e $p \leq 0,05$).

3 Resultados e discussão

Por meio da análise do conteúdo, considerando a primeira aplicação do Enem 2021, foram identificados dezesseis itens de física. A partir desses, foram observadas as incidências para as cada uma das seguintes categorias: seres humanos, gráficos, imagens e fórmulas. Simultaneamente, baseando-se dos dados como razão de chances, p-valor e log do coeficiente de Mantel-Haenzel, foi desenvolvida a análise do DIF.

No que se refere aos itens que não apresentaram DIF significativo, estão as questões identificadas no caderno (1ª aplicação de 2021) pelos números (Nº) 92, 94, 102, 105, 109, 115, 120, 125, 128 e 133. À respeito daqueles que manifestaram DIF relevante, os seis itens encontrados foram detalhados e apresentados no Quadro 1.

Enem-2021-1a aplicação-Caderno azul									
Nº	Trechos do conteúdo	α	p-valor	α^*	CATEGORIA α^*_{MH}	S. H.	Gráfico	Imagem	Fórmula
100	No seu estudo sobre a queda dos corpos, Aristóteles afirmava que se abandonarmos corpos leves e pesados de uma mesma altura (...)	2,27	$1,17 \times 10^{-66}$	-1,92	FORTE	S	N	N	N
107	(...) Considerando o	1,7	$4,02 \times 10^{-29}$	-1,24	MODERADO	N	N	S	N

	mesmo automóvel em duas versões, uma delas funcionando a álcool e outra, a gasolina, tem-se os dados apresentados no quadro (...)								
108	TEXTO I No cordel intitulado Senhor dos Anéis, de autoria de Gonçalo Ferreira da Silva (...) Considerando os versos da sextilha e as informações da tabela , a qual planeta o cor-del faz referência?	1,1	$2,74 \times 10^{-2}$	-0,23	TENDENCIOSO	S	N	S	N
126	Um garoto precisa montar um circuito que acenda três lâmpadas de cores diferentes (...). Esse circuito é representado por:	0,86	$5,45 \times 10^{-4}$	0,36	TENDENCIOSO	S	N	S	N
131	O eletrocardiograma é um exame cardíaco que mede a intensidade dos sinais elétricos advindo do coração. De acordo com o eletrocardiograma apresentado , qual foi o número de batimentos cardíacos por minuto desse paciente durante o exame?	1,15	$1,94 \times 10^{-3}$	-0,32	TENDENCIOSO	S	N	S	N
134	O alcoolômetro Gay Lussac é um instrumento destinado a medir o teor de álcool (...). É apresentada parte da tabela de correção de um alcoolômetro (...)	1,34	$4,77 \times 10^{-9}$	-0,69	TENDENCIOSO	S	N	S	N
Observações: A abreviação S. H. indica a representação de seres humanos. Se α^* é negativo, viés masculino. Se α^* é positivo, viés feminino. Se α^* é igual a zero, não existem diferenças entre os grupos.									

Quadro 1: Descrição dos resultados obtidos

Conforme análise das informações do Quadro 1, apesar da investigação considerar apenas as questões de física e uma aplicação, o percentual com DIF significativo foi alto (em torno de 37,5% dos itens). Desses, em relação ao conteúdo, quatro apresentaram simultaneamente algum tipo de imagem e representação de seres humanos. Sobre os dois restantes, um continha apenas representação de seres humanos e o outro somente imagem. Ainda no que diz respeito a análise dos dados, ao julgar o sinal de α^* (Log), cinco itens apontaram viés em direção aos homens (valor negativo de α^*) e um em direção às mulheres.

4 Conclusões e Considerações Finais

A partir do estudo foi possível identificar quais são os elementos, considerando as categorias produzidas para a análise de conteúdo e a investigação estatística, que mais possuem tendência em DIF significativo em itens de física. O

destaque, no que concerne a análise de conteúdo, foi observar a representatividade da categoria imagem e da seres humanos. No que se refere aos itens com DIF, os resultados manifestaram viés em direção aos homens na maior parte. Portanto, o estudo revelou que a qualidade dos itens de física do exame analisado é questionável. Tendo em vistas que, 37,5% dos itens investigados teve DIF significativa. Adicionalmente, ainda a partir dos resultados, é possível sugerir desigualdade de gênero fundamentada a partir da tendência masculina no DIF.

Sobre a análise do DIF foi possível concluir que, parece existir uma relação entre o emprego de imagens, em conjunto com representação de seres humanos, na probabilidade de acertos das pessoas participantes. No entanto, como a amostra é pequena, é recomendada a ampliação. Assim como, sinalizada a necessidade de mais estudos que visem refletir sobre a qualidade das avaliações em larga escala e dos exames.

5 Referências

BARDIN, L. **Análise de Conteúdo**. São Paulo: Edições 70, 2016.

CORDEIRO, L. Sobre a inadequação da metodologia de cálculo das notas do Sisu. **Educação & Sociedade**, v. 35, p. 293-320, 2014.

ENGUIITA, M. F. *A face oculta da escola: educação e trabalho no capitalismo*. Porto Alegre: Artes Médicas, 1989.

FREITAS, D. N. T. de. **A avaliação da educação básica no Brasil**: dimensão normativa, pedagógica e educativa. Campinas: Autores Associados, 2007.

GLADYS, M. J. *et al.* Gender bias in first-year multiple-choice physics examinations. **Physical Review Physics Education Research**, v. 19, n. 2, p. 020109, 2023.

GRIBOSKI, C. M. O Enade como indutor da qualidade da educação superior. **Estudos de Avaliação Educacional**, p. 178-195, 2012.

LOBATO JÚNIOR, J. M. S. *et al.* Uma análise de questões de matemática comercial e financeira do Enem. **Revista Contemporânea**, 4(5), e4545. <https://doi.org/10.56083/RCV4N5-233>.

SILVA, T.T. **Alienígenas na sala de aula**: uma introdução aos estudos culturais em educação. Petrópolis/RJ: Vozes, 1995.