

Efeito de posição na dificuldade dos itens do Enem.

Levi Mendes Franklin¹, Leandro Araujo de Sousa², José Airton de F. Pontes Junior³, Nicolino Trompieri Filho⁴

¹ Secretaria de Educação do Ceará (SEDUC-CE). prof.levimf@gmail.com

² Universidade Federal do Ceará (UFC). leandro.sousa@ifce.edu.br

³ Universidade Estadual do Ceará (UECE). japontesjr@gmail.com

⁴ Universidade Federal do Ceará (UFC). trompieri@hotmail.com

Resumo

Este trabalho tem como objetivo analisar o comportamento do parâmetro de dificuldade b de dois grupos de itens do Exame Nacional do Ensino Médio (Enem), em que são dispostos em posições distintas em cada uma das quatro cores de provas (azul, amarela, rosa e branca). Para tanto, estimamos os parâmetros dos itens pela Teoria de Resposta ao Item (TRI) a partir do modelo logístico unidimensional de três parâmetros, mesmo adotado pelo exame. A partir da análise, os itens apresentaram sensível aumento do parâmetro de dificuldade na maioria dos itens. Essa variação não contribui para a uniformidade do exame e pode atribuir proficiência diferentes a candidatos do mesmo nível, contrariando o modelo teórico da TRI.

Palavras-chave: Enem, teoria de resposta ao item, efeito posição.

1. Introdução

A partir de 2009 com a implementação da Teoria de resposta ao Item (TRI) ao Exame Nacional do Ensino Médio (Enem) possibilitou duas finalidades principais; permitir a comparação entre os resultados; permitir a aplicação de vários exames ao longo do ano (BRASIL, 2011). A comparação dos resultados se dá através da definição de uma escala de medida de proficiência estabelecida, na qual possibilita a realização de vários exames ao longo do ano.

É fato que a teoria de resposta ao item obteve maior visibilidade após a sua adoção no Enem, porém, seu uso teve início em 1995 com o Sistema de Avaliação da Educação Básica (Saeb), e posteriormente foi implementado no Encceja, Prova Brasil e no Enem, presente também em vários países, para citar: Estados Unidos da América, França, Holanda, Coréia do Sul e China (BRASIL, 2011).

Conforme Klein (2013) a TRI utiliza um conjunto de modelos estatísticos em que a probabilidade de resposta a um item é modelada como função da proficiência (habilidade) do aluno (variável não observável) e de parâmetros que expressam certas propriedades dos itens, com a propriedade de discriminação do item que apresenta ao aluno de maior a proficiência, maior a probabilidade de acertar o item.

O modelo logístico de 3 parâmetros é definido como

$$P(x = 1 | \theta, a, b, c) = c + \frac{(1 - c)}{1 + \text{EXP}[-Da(\theta - b)]}$$

onde para a resposta correta $x = 1$ e para resposta errada $x = 0$, o valor de a representa o parâmetro de discriminação o valor de a sempre é maior que zero, b representa o parâmetro de dificuldade ou de posição do item e c representa a assíntota inferior do item, na qual representa as chances de um aluno com baixa proficiência acertar a resposta correta (KLEIN, 2013).

Segundo Andrade e Karino (2011), O parâmetro de discriminação diminui a capacidade dos candidatos que possuem proficiências diferentes obterem a mesma probabilidade de acerto dos itens. O parâmetro dificuldade está associado a atribuição de maior valor para o item que possui mais dificuldade, ao fato de candidatos que possuem maior proficiência possuem também maior probabilidade de acerto com isso é fácil perceber a presença de itens de diferentes níveis no Enem. O parâmetro de acerto casual, leva em conta o padrão de resposta do candidato bem como as chances de acerto de um item de dificuldade acima da proficiência apresentada, ocorrendo assim, o fato de que candidatos com o mesmo número de acertos possuam notas diferentes.

O objetivo deste trabalho é analisar o comportamento da dificuldade dos itens quando são apresentados nas diferentes posições dos quatro cadernos de prova do ENEM, segundo Alexandrowicz e Matschinger (2008), vários fatores causam os efeitos de posição dos itens, os quais podem ser identificados como. (1) os de aquecimento, (2) de aprendizado ou prática, (3) fadiga, (4) falta de tempo, o que pode contribuir para a maior ou menor dificuldade do item. Assim, os efeitos de localização de itens distorcem a estimativa das dificuldades do item eles invalidam as conclusões sobre as propriedades psicométricas de um item ou o nível de desempenho de um respondente.

Conforme citam Brennan (1992) e Wainer e Kiely (1987) os efeitos de posição do item, que pertencem ao grupo de efeitos de contexto, indicam que os parâmetros de um item

(por exemplo, dificuldade ou discriminação) em um teste de desempenho variam de acordo com a posição do item na prova (apud ALBANO, 2003). Considerando-se os efeitos da posição do item na sua dificuldade, um item administrado no final geralmente é mais difícil do que o mesmo administrado no início do teste. Se um teste de aproveitamento consistir em vários formulários que podem incluir itens diferentes em uma ordem distinta e que devem ser vinculados a uma escala comum, os efeitos de posição podem violar a suposição de invariância de parâmetro de item nos formulários de teste.

2. Metodologia

Para a realização desta análise, foi feito um recorte nos microdados desse exame, tratando apenas com as avaliações de ciências da natureza aplicadas no estado do Ceará. Além disso, foram retirados da amostra, participantes que alegaram ter algum tipo de necessidade especial.

Dessa forma, no presente trabalho utilizamos uma amostra de 51.585 candidatos que realizaram o Enem 2016. Desses, 13.542 candidatos que responderam a prova azul, 12.683 candidatos que responderam a prova amarela, 12.115 candidatos que responderam a prova rosa e 13.245 candidatos que responderam a prova branca.

Dos itens da prova citada, será destinado para este trabalho dois grupos de questões, grupos esses, que apresentam os mesmos itens dispostos em diferentes posições em cada em cada caderno de prova (azul, amarela, branca e rosa).

Tabela 1- Posição dos Itens nos Cadernos de Prova.

Caderno de prova	Bloco de itens A			Bloco de itens B	
	1	2	3	4	5
	Posição			Posição	
Azul	46	47	48	64	65
Amarela	58	59	60	78	79
Rosa	61	62	63	56	57
Branca	88	89	90	72	73

A análise dos itens foi realizada a partir do modelo logístico unidimensional de três parâmetros, mesmo adotado pelo exame. As estimativas dos parâmetros dos itens foram realizadas pelo *software* R. Para atender ao objetivo deste estudo, comparamos apenas o parâmetro de dificuldade b do mesmo item em diferentes posições.

3. Resultados e Discussão

3.1 Análise dos participantes.

Inicialmente foi avaliada a equivalência dos quatro grupos de participantes correspondentes a cada um dos cadernos de provas, os aspectos analisados são: Média de acertos; Mediana; Moda; Desvio Padrão; Variância; Assimetria; Curtose; Mínimo de Acerto; Máximo de Acerto; Média de Pontos.

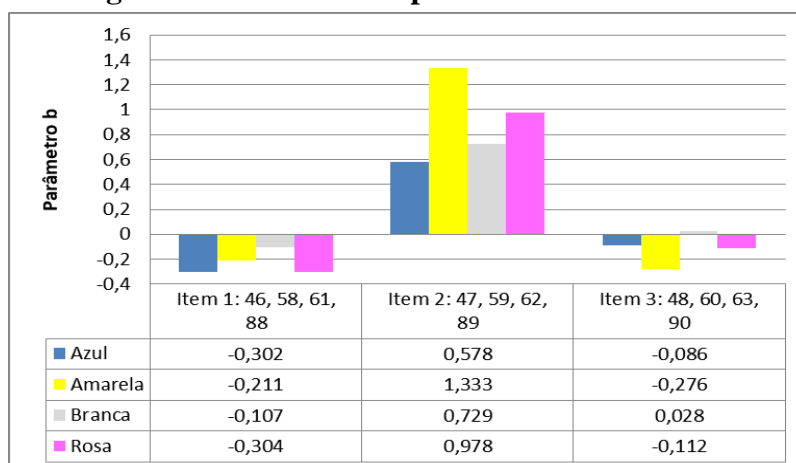
Aspectos\provas	Azul	Amarela	Branca	Rosa	Diferença Máxima
Média de acertos	11.815	11.997	11.864	11.850	133
Mediana	11	11	11	11	0
Moda	11	11	10	10	1
Desvio padrão	4,93	4,96	4,97	4,93	0,03
Variância	24,27	24,64	22,77	24,34	1,87
Assimetria	1,68	1,59	1,49	1,60	0,19
Curtose	4,93	4,44	4,13	4,52	0,81
Mínimo	0	0	0	0	0
Máximo	44	44	43	43	1
Média de pontos	481	485	484	482	4

A tabela II demonstra que os participantes dos quatro cadernos de prova possuem valores muito próximos em todos os aspectos analisados demonstrando assim, a equivalência dos grupos de participantes.

3.1 Análises dos Itens

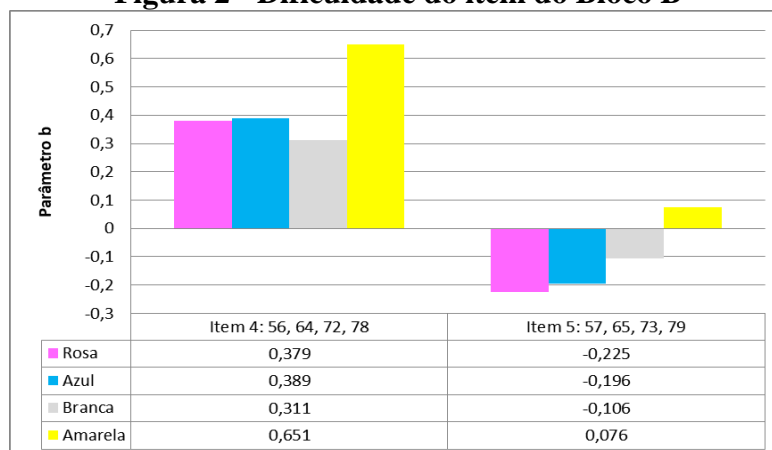
O Grupo 1 de itens apresentam o seguinte gráfico de acerca da dificuldade do item definido pelo parâmetro b para cada item, indicando sua posição em cada uma das 4 cores de provas.

Figura 1 – Parâmetro b para os itens do Bloco A



A figura 1 mostra que na maioria dos casos há um aumento do parâmetro b , o que fica evidenciado no item 1 quando comparamos as provas azul com amarela e branca sendo contrariado apenas na prova rosa. Para o item 2 o parâmetro b aumenta apenas na comparação das provas azul com a prova amarela e quando comparamos as provas branca e rosa. Já para o item 3 existe diminuição de b apenas para quando são analisadas as provas azul e amarela, e branca com rosa.

Figura 2 - Dificuldade do item do Bloco B



Semelhantemente ao que ocorre no grupo 1, no grupo 2 os itens apresentam parâmetros b menores no início dos cadernos de provas para a maioria das situações, para o item 19, o parâmetro b possui pequena variação para as provas rosa, azul e branca, e demonstra maior elevação de b somente na prova amarela aquela posicionada de forma mais avançada. Para o item 20, é mostrada forma mais uniformizada a elevação de b em função da sua posição na prova, com valores negativos para as provas rosa, azul e branca e valor positivo para a prova amarela.

Concordamos com Hohensinn et.al (2008), que cada autor de teste deve estar ciente de que pode haver alguns efeitos de posição de item, devido dois motivos principais. Primeiramente os candidatos iniciam a resolução do exame e ao passar do tempo vão se familiarizando com a prova, ou até mesmo detectam a resolução de problemas estratégias para os itens. Com isso, os efeitos de prática ou aprendizagem poderiam de algum modo, que um item no início de um teste seja mais difícil do que o mesmo item no final do teste. Segundo efeito é a de fadiga, que podem ocorrer, de modo que um item no início de um teste é mais fácil do que o mesmo item administrado no final do teste.

Neste trabalho fica evidenciada que a posição do item influencia na sua dificuldade, podendo ser observada tomando como estimativa o parâmetro b , podemos afirmar de maneira subjetiva, que os itens que apresentam diminuição na dificuldade em relação a sua posição mais avançada no caderno de questões, caderno esse que possui 90 itens daí o fato de apresentarem o efeito da fadiga, e/ou falta de tempo, porém os itens apresentam um aumento na dificuldade em relação a sua posição mais avançada no caderno de questões, pode apresentar o efeito de aprendizado e/ou de aquecimento, estes efeitos influenciam a diminuição da dificuldade do item, haja vista que os itens possuem relações entre si, a organização diferente dos itens pode influenciar a calibração do teste, bem como modificar a proficiência obtida pelo candidato.

4. Conclusões

Concluímos que houve variação considerável nos parâmetros de dificuldade dos itens analisados em função de sua posição nos diferentes cadernos de prova do Enem 2016. A variação do parâmetro de dificuldade dos itens deve ser analisada de maneira totalitária com todos os itens do exame, podendo assim, verificar se há variação na dificuldade total do exame. A variação da dificuldade dos itens para cada cor de prova pode atribuir um valor de proficiência diferente para dois candidatos que possuem a mesma proficiência devido ao fato dos candidatos receberem cadernos de prova de cores diferentes.

Referências

- ALEXANDROWICZ, R.; MATSCHINGER, H. Estimation of item location e effects by means of the generalized logistic regression model: A simulation study and an application. **Psychology Science Quarterly**, 2008.
- ALBANO, A. D. Multilevel modeling of item position e effects. **Journal of Educational Measurement**, 2013.
- ANDRADE, D. F; KARINO C. A; Entenda a Teoria de Resposta ao Item (TRI), Utilizada no ENEM, 2011
- BRASIL, Ministério da Educação, Instituto Nacional de Estudos Pesquisas Educacionais Anísio Teixeira, Nota Técnica. 2011.
- HOHENSINN C; KUBINGER K. D; Reif M; HOLOCHER-ERTL S; KHORRAMDEL L; FREBORT M. Examining item-position effects in large-scale assessment using the Linear Logistic Test Model. **Psychology Science Quarterly**. Volume 50, 2008.
- KLEIN R. Alguns Aspectos da teoria de resposta ao item relativos à estimação das proficiências. Fundação Cesgranrio. 2013.