
Análise do Exame Nacional do Ensino Médio via Teoria Clássica dos Testes e Teoria de Resposta ao Item

Leandro Araujo de Sousa¹, Levi Mendes Franklin², José Airton de Freitas Pontes Junior³,
Nicolino Trompieri Filho⁴

¹ Universidade Federal do Ceará (UFC). leandro.sousa@ifce.edu.br

² Secretaria de Educação do Ceará (SEDUC-CE). prof.levimf@gmail.com

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

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

Resumo

Esta pesquisa objetivou comparar os resultados de matemática do Enem 2016 via Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI). Para tanto, aplicou-se um teste de correlação entre as notas dadas pela TCT e TRI a partir de uma amostra de 500 participantes do exame. As estimativas dos escores em matemática via TCT e TRI apresentaram uma correlação alta, no entanto, o método de análise pode ter causado influência na classificação dos participantes do exame. Concluímos que os resultados da prova de matemática do Enem 2016 dados pelos dois modelos podem ser comparáveis.

Palavras-chave: avaliação em larga escala, Enem, psicometria

1. Introdução

As avaliações educacionais em larga escala de abrangência nacional surgiram no Brasil no início da década de 1990, entre essas o Exame Nacional do Ensino Médio (Enem). A priori seus resultados eram analisados, tanto a nível de teste/item como de sujeito, a partir da Teoria Clássica dos Testes (TCT), embora com algumas limitações, principalmente em relação a comparabilidade dos resultados (ANDRADE; LAROS; GOUVEIA, 2010).

No entanto, com o avanço tecnológico na área computacional alcançado no final do século 20, pôde-se implementar modelos matemáticos mais sofisticados, conhecidos hoje como Teoria de Resposta ao Item (TRI). Esse modelo propôs superar determinados problemas da TCT, tais como instabilidade dos parâmetros dos itens e problemas de comparabilidade dos resultados entre grupos de sujeitos ou entre instrumentos de avaliações com itens diferentes (KLEIN, 2013; MUÑIZ, 2010; SARTES; SOUSA-FORMIGONI, 2013; VALLE, 2000).

Para os autores, a TRI supera essas fragilidades por possibilitar parâmetros estáveis dos itens independente do particular grupo de sujeitos avaliados, mesmo que com instrumentos distintos, desde de que satisfeitos alguns pressupostos.

Desse modo, os modelos de TRI para itens dicotômicos passaram a ser utilizados para análise dos resultados das avaliações e exames educacionais em larga escala no Brasil, a exemplo do Sistema de Avaliação da Educação Básica (Saeb) e do Enem a nível nacional (INEP, 2012).

No entanto, algumas pesquisas têm comparado os resultados de avaliações pela TCT e TRI (ADEDOYIN; ADEDOYIN, 2013; AWOPEJU; AFOLABI, 2016; KOHLI; KORAN; HENN, 2014; PROGAR; SOČAN, 2008), apontando indícios empíricos de que os resultados pela TCT são comparáveis ao da TRI. Além disso, há questionamentos em relação aos pressupostos do modelo no sentido de dificuldade de satisfazê-los (TAVARES, 2013).

Dessa forma, partimos do seguinte questionamento: As notas dos candidatos do Enem quando calculadas via TCT e TRI são comparáveis? Diante disso, esta pesquisa teve o objetivo de comparar os resultados de matemática do Enem 2016 via Teoria Clássica dos Testes e Teoria de Resposta ao Item.

2. Metodologia

O estudo foi realizado com uma amostra aleatória de 500 sujeitos, pois esse tamanho amostral já se mostra suficiente para obter estimativas estáveis da TRI (NUNES; PRIMI, 2005; VALENTINI; HAUCK FILHO, 2013). Foram selecionados os participantes que responderam todas as questões dentre os que realizaram o exame no ano de 2016. Utilizou-se os itens da prova matemática de cor azul. Essas escolhas deram-se de forma aleatória. Os dados foram extraídos dos microdados do exame disponibilizados pelo Inep.

A partir de 2009 a prova do Enem passou a ser composta por 180 questões de múltipla escolha com cinco alternativas divididas em quatro áreas de conhecimento, Linguagens e Códigos, Matemática, Ciências Humanas e Ciências da Natureza, dessa forma, 45 item para cada área. Além disso, os candidatos também produzem uma redação. Com essa mudança, as notas passaram a ser calculadas por área de conhecimento via TRI, modelo de três parâmetros.

Para atender ao objetivo desta pesquisa, foram estimados os escore dos sujeitos pela TCT, que se deu pela soma dos escores nos itens dicotomizados em 0 para errado e 1 para

certo, com escore máximo de 45. Também foi estimado os parâmetros dos sujeitos pela TRI a partir do modelo logístico de três parâmetros (adotado no exame), dado a seguir:

$$P(u_{ji} = 1 | \Theta_j, a_i, b_i, c_i) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\Theta_j - b_i)}}$$

em que,

$P(u_{ji} = 1 | \Theta_j, a_i, b_i, c_i)$ = probabilidade de resposta correta do participante j ao item i ($u_{ji} = 1$),

em função do parâmetro de proficiência Θ_j e dos parâmetros do item a_i , b_i e c_i .

a = parâmetro de discriminação.

b = parâmetro de dificuldade.

c = parâmetro de acerto casual.

e = base logarítmica de valor 2,72.

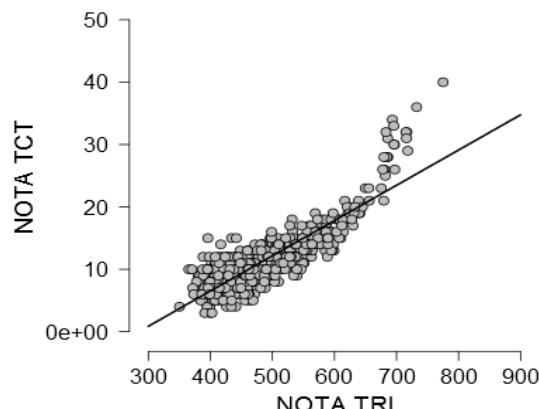
Θ = proficiência do participante.

As habilidades obtidas foram convertidas para a escala de 0 a 1000, com média 500 e desvio padrão 100. Para fins de comparação, utilizou-se a análise de correlação de Pearson. Além disso, foi analisado as diferenças nas classificações dos 20 melhores sujeitos com notas calculadas pela TCT e TRI. Todas as análises foram realizadas através do *software R*.

3. Resultados e Discussão

Na análise de correlação realizada entre as notas calculadas pela TRI (3 parâmetros) e a soma total dos escores nos itens, obteve-se valor de correlação alto ($r = 0.85$, $p < 0,001$). Neste caso, um percentual alto de acerto influenciou para uma nota alta em TRI. A correlação pode ser visualizada no Gráfico 1.

Gráfico 1 – Correlação entre notas de matemática obtidas pela TCT e TRI, Enem 2016.



Fonte: Da pesquisa.

Resultados semelhantes foram encontrados em outros estudos (ADEDOYIN; ADEDOYIN, 2013; GÜLER; UYANIK; TEKER, 2014; PROGAR; SOČAN, 2008). Entretanto, correlações altas foram identificadas entre as estimativas pela TCT e os modelos de um e dois parâmetros de TRI, em casos de estimativas de três parâmetros em que os itens tenham uma probabilidade alta de acerto casual, há diferenças significativas quando comparadas com a TCT (GÜLER; UYANIK; TEKER, 2014). De modo similar, outros estudos (ADEGOKE, 2013; COSTA; FERRÃO, 2015) compararam os parâmetros de dificuldade e discriminação dos itens pela TCT e TRI, encontrando forte relação. Nesse sentido, esses estudos têm questionado a superioridade teórica da TRI, enquanto outros têm advogado pela complementaridade entre os dois modelos.

A Tabela 1 apresenta os 20 melhores participantes com nota pela TRI, bem como a posição dos mesmos sujeitos considerando a nota pela TCT. A partir disso, houve uma mudança na classificação dos participantes. Ao considerar as notas para fins de seleção, como é o caso do Enem hoje, o método de correção da prova pode interferir na seleção dos participantes que vão ou não ingressar nos cursos em Instituições de Ensino Superior. Se tomarmos como exemplo, o participante 19 não seria selecionado, caso a nota fosse dada pela TCT, por outro lado, o participante 21 seria selecionado.

Tabela 1 – Classificação dos participantes pela TCT e TRI.

Posição	TRI		TCT	
	Sujeito	Nota	Sujeito	Nota
1	1	775,12	1	40
2	2	732,44	2	36
3	3	718,16	11	34
4	4	716,64	10	33
5	5	716,00	4	32
6	6	715,18	6	32
7	7	697,15	15	32
8	8	696,55	5	31
9	9	695,94	13	31
10	10	695,51	8	30
11	11	693,30	9	30
12	12	686,11	3	29
13	13	685,79	12	28
14	14	684,13	14	28
15	15	683,07	18	28
16	16	681,91	7	26

17	17	680,59	17	26
18	18	679,75	20	26
19	19	679,41	16	25
20	20	677,90	21	23

Fonte: Da pesquisa.

Nesse sentido, embora as estimativas pela TCT e TRI apresentem correlações altas, indicando comparabilidade entre as duas estimativas, em caso particular de classificação dos participantes, há diferenças consideráveis. Em termos de avaliação de redes de ensino, as repercussões do método podem não ser tão perceptíveis, já que o foco está no sistema como um todo, não em cada sujeito.

Entretanto, teoricamente a TRI proporciona vantagens de comparabilidade e invariância dos parâmetros, tarefa muito difícil pela TCT. No entanto, a suposição de unidimensionalidade deve ser satisfeita para que a propriedade da invariância possa se manifestar para que o modelo apresente vantagens em relação a TCT (CONDE; LAROS, 2007). Por outro lado, evocando o princípio da parcimônia, caso não satisfeitas as condições necessárias, o modelo clássico seria preferível. É importante destacar que o modelo de resposta ao item proporcionou avanços importantes para as medidas educacionais, como a composição de banco de itens na mesma escala da habilidade dos sujeitos, permitindo a elaboração de provas distintas para medir a mesma dimensão e possibilitando um melhor uso pedagógico dos resultados das avaliações.

4. Conclusões

Concluímos que há comparabilidade entre as notas dos participantes em matemática dadas pela TCT e TRI no Enem, todavia, o método influenciou na classificação dos participantes. Consideramos importante, dessa forma, estudos mais profundos que analisem este exame, por exemplo, que comparem a variabilidade e a comparabilidade das estimativas pela TCT e TRI entre e dentro de diferentes tamanhos amostrais. Além disso, que comparem os resultados entre os anos. Também é importante a análise da dimensionalidade dos itens e o seu efeito na estimativa dos parâmetros dos itens e dos sujeitos pelo modelo de resposta ao item.

Agradecimentos

Agradecemos ao Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) pelo apoio e financiamento da pesquisa.

Referências

ADEDYIN, O. O.; ADEDYIN, J. A. Assessing the comparability between classical test theory (CTT) and item response theory (IRT) models in estimating test item parameters. **Herald Journal of Education and General Studies**, v. 2, n. 3, p. 107–114, 2013.

ADEGOKE, B. A. Comparison of item statistics of physics achievement test using classical test and item response theory frameworks. **Journal of Education and Practice**, v. 4, n. 22, p. 87–96, 2013.

ANDRADE, J. M. DE; LAROS, J. A.; GOUVEIA, V. V. O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores. **Avaliação Psicológica**, v. 9, n. 3, p. 421–435, 2010.

AWOPEJU, O. A.; AFOLABI, E. R. I. Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination. **European Scientific Journal**, v. 12, n. 28, p. 263–284, 2016.

CONDE, F. N.; LAROS, J. A. Unidimensionalidade e a propriedade de invariância das estimativas da habilidade pela TRI. **Avaliação Psicológica**, v. 6, n. 2, p. 205–215, 2007.

COSTA, P.; FERRÃO, M. E. On the complementarity of classical test theory and item response models: item difficulty estimates and computerized adaptive testing. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 23, n. 88, p. 593–610, 2015.

GÜLER, N.; UYANIK, G. K.; TEKER, G. T. Comparison of classical test theory and item response theory in terms of item parameters. **European Journal of Research on Education**, v. 2, n. 1, p. 1–6, 2014.

INEP. Nota técnica – assunto: teoria de resposta ao item. . 2012.

KLEIN, R. Alguns aspectos da Teoria de Resposta ao Item relativos à estimação das proficiências. **Ensaio: Aval. Pol. Públ. Educ.**, v. 21, n. 78, p. 35–56, 2013.

KOHLI, N.; KORAN, J.; HENN, L. Relationships among classical test theory and item response theory frameworks via factor analytic models. **Educational and Psychological Measurement**, v. 75, n. 3, p. 389–405, 2014.

MUÑIZ, J. Las teorías de los tests: Teoria clasica y teoria de respuesta a los tems. **Papeles del Psicólogo**, v. 31, n. 1, p. 57–66, 2010.

NUNES, C. H. S. DA S.; PRIMI, R. Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item. **Avaliação psicológica**, v. 4, n. 2, p. 141–153, 2005.

PROGAR, Š.; SOČAN, G. An empirical comparison of item response theory and classical test theory item/person statistics. **Horizons of Psychology**, v. 17, n. 3, p. 5–24, 2008.

SARTES, L. M. A.; SOUSA-FORMIGONI, M. L. O. DE. Avanços na Psicométria: aa teoria clássica dos testes à teoria de resposta ao item. **Psicologia: Reflexão e Crítica**, v. 26, n. 2, p. 241–250, 2013.

TAVARES, C. Z. Teoria da resposta ao item: uma análise crítica dos pressupostos epistemológicos. **Estudos em Avaliação Educacional**, v. 24, n. 54, p. 56–76, 2013.

VALENTINI, F.; HAUCK FILHO, N. Estabilidade da estimação dos parâmetros dos itens em diferentes tamanhos amostrais. **III CONBRATRI - Congresso Brasileiro de Teoria da Resposta ao Item**, n. 1, p. 43–48, 2013.

VALLE, R. DA C. Teoria de resposta ao item. **Estudos em Avaliação educacional**, v. 21, p. 07–92, 2000.