

Aplicação do Modelo de Reposta Nominal da TRI a Avaliação Educacional de Larga Escala

André F. Z. Silva^{1†}, Dalton F. Andrade², Adriano F. Borgatto³, Luiz R. Nakamura⁴

¹Mestrando no Programa de Pós-Graduação em Métodos Quantitativos e Gestão em Avaliação (UFSC)

²Professor do Departamento de Informática e Estatística, INE, UFSC

³Professor do Departamento de Informática e Estatística, INE, UFSC

⁴Professor do Departamento de Informática e Estatística, INE, UFSC

Resumo: O sistema de avaliação educacional tem por objetivo organizar informações que viabilizem tomadas de decisões acerca da qualidade da educação. A Avaliação formativa foca em ensino-aprendizagem e pretende, de uma forma rápida, identificar dificuldades dos estudantes. Diante desse contexto, realizou-se um estudo psicométrico utilizando dados de uma avaliação formativa de larga escala do componente curricular de matemática, realizada para estudantes do 4.º ano do ensino fundamental do município de Curitiba-PR. O objetivo deste trabalho é aplicar o Modelo de Resposta Nominal da Teoria da Resposta ao Item aos dados da avaliação. Também foram utilizadas ferramentas da consistência do teste da Teoria Clássica de Testes e uma análise da informação do teste. Ao realizar uma comparação com o Modelo Logístico de 3 Parâmetros, nota-se que o modelo conseguiu capturar alguns problemas nos itens da avaliação não captados pelo Modelo Logístico de 3 Parâmetros da TRI, além de apresentar o crescimento ou decréscimo dos distratores de acordo com a evolução das proficiências dos estudantes.

Palavras-chave: Teoria da Resposta ao Item, Avaliação Educacional; Modelos de Reposta Nominal.

Abstract:

The educational evaluation system aims to organise information that helps the decision making about the education quality. Formative evaluation focuses on teaching-learning and tries to identify students's difficults in a fast way. In this context, a psychometric study was carried out using data from a large-scale mathematics assessment, performed by students of the 4th year of primary school in Curitiba-PR. The main aim of this paper is to apply the nominal response model from item response theory (IRT) to these data. When we compare the three-parameter logistic model, we can note that this model was able to identify problems that were not detected by the three-parameter IRT logistic model. Furthermore, this model presented a growth or decrease of the distractors according to the students' proficiency.

Keywords: Item Response Theory; Large Scale Assessments; Nominal Response Model.

Introdução

A escola, enquanto espaço formal de construção do conhecimento, está sujeita a processos avaliativos, os mais diversos. Avalia-se para diagnosticar, para qualificar e para planejar atividades e estratégias que percebam processos de ensino-aprendizagem, bem como necessidades individuais e coletivas dos estudantes.

O sistema de avaliação educacional tem por objetivo organizar informações que viabilizem reflexões e tomadas de decisão acerca da qualidade da educação.

Segundo FREITAS et.al (2013, p. 10), são três os níveis que compõem a estrutura da avaliação educacional: avaliação de sistema; institucional e da aprendizagem. Embora tenham objetivos diferenciados, estão relacionados e subsidiam a prática pedagógica de todo o contexto educativo.

†Autor correspondente: andrefzs@hotmail.com.

Parecida com a avaliação de aprendizagem, a avaliação formativa foca em ensino-aprendizagem. Pretende, de forma rápida, identificar dificuldades dos estudantes e uma posterior adequação do docente ao trabalho em sala de aula.

A Secretaria Municipal da Educação de Curitiba realiza anualmente a Prova Curitiba. A avaliação é de caráter formativo e apresenta o objetivo de acompanhar sistematicamente o conhecimento dos estudantes nas áreas de língua portuguesa, matemática e produção de texto, permitindo redimensionar ações no trabalho pedagógico nas escolas.

A Rede Municipal de Ensino de Curitiba, nos últimos anos, apresenta um constante crescimento no Índice de Desenvolvimento da Educação Básica (Ideb) – Anos Iniciais, configurando o melhor desempenho do país entre as capitais com mais de um milhão de habitantes. O Ideb mede a qualidade do ensino com base em dados sobre aprovação e desempenho escolar obtidos em provas de língua portuguesa e matemática do Sistema de Avaliação da Educação Básica (Saeb), do Instituto de Pesquisa Anísio Teixeira (Inep).

Diante desse contexto, realizou-se um estudo psicométrico, a partir de dados de uma avaliação formativa de larga escala, de matemática, ofertada a estudantes matriculados no 4.º ano do ensino fundamental do município de Curitiba.

O objetivo deste trabalho será aplicar o modelo de resposta nominal da Teoria da Resposta ao Item aos dados da avaliação.

Material e métodos

O método empregado foi de pesquisa quantitativa, a partir de dados de todos os estudantes matriculados no 4.º ano do ensino fundamental do município de Curitiba e participantes da prova diagnóstica de matemática realizada em 2018. O universo estudado foi constituído por 15.583 estudantes matriculados em 2018 na rede municipal de ensino estudada.

As análises estatísticas deste estudo foram realizadas no software R, utilizando os pacotes MIRT e PSICH.

Para verificar a consistência interna da avaliação utiliza-se o Coeficiente Alpha de Cronbach aferido por Cronbach, L. J. (1951). Esse coeficiente varia de 0 a 1. Quanto mais próximo de 1 maior a consistência do teste. Valores acima de 0,7 são considerados satisfatórios (STREINER, 2003)

O modelo de Resposta nominal é um modelo politômico e foi exposto pela primeira vez por Bock, R. D. (1972) e é definido por:

$$P_{i,k}(\theta_j) = \frac{e^{a_{i,k}^+(\theta_j - b_{i,k}^+)}}{\sum_{h=1}^{m_i} e^{a_{i,h}^+(\theta_j - b_{i,h}^+)}}$$

O modelo, ao contrário do Modelo Logístico de 3 parâmetros, considera todas as categorias de resposta de um item, utilizando toda a informação contida nas respostas dos indivíduos. É utilizado para itens com duas ou mais categorias nominais. Os parâmetros a e b são interpretados em função das categorias de cada item.

Resultados

Perfil dos Estudantes

Observou-se que 15.583 estudantes, o que representa 96% dos estudantes matriculados no 4º ano, provenientes 181 escolas do município estudado realizaram a avaliação diagnóstica de matemática.

Análise Descritiva

Com o auxílio de técnicas de estatística descritiva realizou-se uma análise de distribuição de acertos da avaliação verificando-se o comportamento dos estudantes.

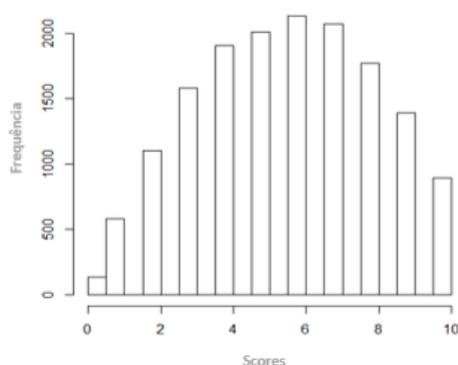


Figura 1 – Histograma – Distribuição de Acertos

A média de acertos foi de 5,658. A distribuição de acertos ficou equilibrada. Não aconteceu uma distribuição maior esperada em torno da média. O percentual de estudantes que obtiveram entre 9 a 10 acertos foi aproximadamente 14%. 136 estudantes não conseguiram acertar nenhuma das 10 questões da avaliação.

Análise pela Teoria Clássica de Testes (TCT)

A figura 2 apresenta a consistência (qualidade) da avaliação estudada pela Teoria Clássica de Testes. O indicador de qualidade global do teste utilizado é o Alfa de Crombach (CRONBACH, 1951). Segundo a literatura é recomendável Alfa superior a 0,70 para qualificar o instrumento como bom. A prova de Matemática apresentou índice igual ao recomendado (0,704) e, no olhar individual, todas as questões contribuem para manter esse índice. Todos itens apresentaram Coeficiente de correlação ponto bisserial positivo para os gabaritos e negativo para os distratores confirmando nenhum problema com os gabaritos. Segundo a literatura coeficientes acima de 0,30 são considerados bons. O índice de discriminação é a diferença entre a proporção de acertos dos participantes com maior escore daqueles com menor escore. Espera-se que a proporção de acertos para um item seja maior no grupo superior em relação ao grupo inferior. Para todos os itens a premissa foi constatada.

Sigmae, Alfenas, v.8, n.2, p. 735-741, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).

18º Simpósio de Estatística Aplicada à Experimentação Agrônoma (SEAGRO).

TABELA 1 – Indicadores e Índices calculados pela TCT

Item	% de Acerto	Grupo		Discriminação	Bisserial
		Grupo 27% Inferiores	Grupo 27% superiores		
I1	75,6%	53,5%	91,9%	0,384	0,533
I2	62,1%	37,6%	89,6%	0,520	0,606
I3	36,6%	16,5%	60,5%	0,440	0,508
I4	49,8%	32,8%	70,6%	0,378	0,410
I5	50,2%	31,1%	72,6%	0,414	0,459
I6	62,8%	43,7%	82,2%	0,385	0,411
I7	65,4%	39,5%	89,5%	0,500	0,541
I8	57,6%	38,4%	77,5%	0,391	0,410
I9	76,8%	59,5%	92,3%	0,328	0,438
I10	28,8%	16,5%	45,5%	0,290	0,382

A figura 2 apresenta a consistência (qualidade) da avaliação estudada pela Teoria Clássica de Testes. O indicador de qualidade global do teste utilizado é o Alfa de Crombach (CRONBACH, 1951). Segundo a literatura é recomendável Alfa superior a 0,70 para qualificar o instrumento como bom. A prova de Matemática apresentou índice igual ao recomendado (0,704) e, no olhar individual, todas as questões contribuem para manter esse índice. Todos itens apresentaram Coeficiente de correlação ponto bisserial positivo para os gabaritos e negativo para os distratores confirmando nenhum problema com os gabaritos. Segundo a literatura coeficientes acima de 0,30 são considerados bons. O índice de discriminação é a diferença entre a proporção de acertos dos participantes com maior escore daqueles com menor escore. Espera-se que a proporção de acertos para um item seja maior no grupo superior em relação ao grupo inferior. Para todos os itens a premissa foi constatada.

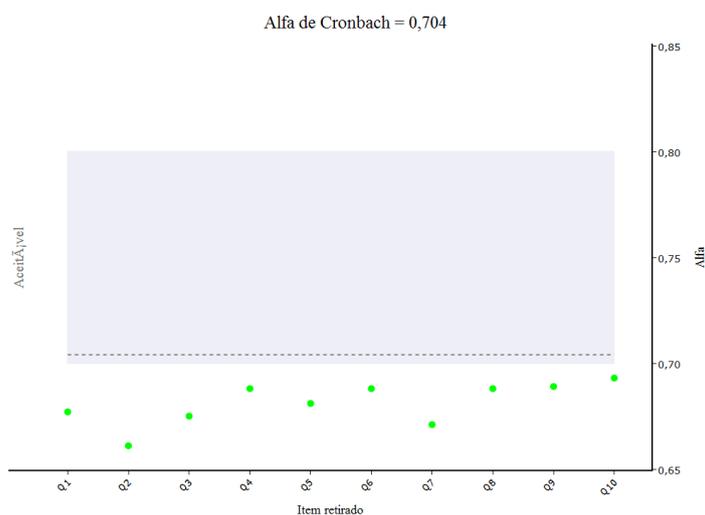


Figura 2 – Consistência da Avaliação

Sigmae, Alfenas, v.8, n.2, p. 735-741, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).
18º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).

Análise pela Teoria da Resposta ao Item (TRI)

Informação do Teste

A figura 3 apresenta a informação do teste. Em outras palavras, mostra onde o teste apresenta maior concentração de itens na escala a ser criada. Toma-se como base o zero (0), o qual significa a proficiência média dos estudantes que realizaram a avaliação. A avaliação de matemática a informação predominante está entre -1,5 a 2,5 apresentando concentração maior de informação à direita, sugerindo uma avaliação com um grau de dificuldade um pouco maior do que a proficiência média dos estudantes.

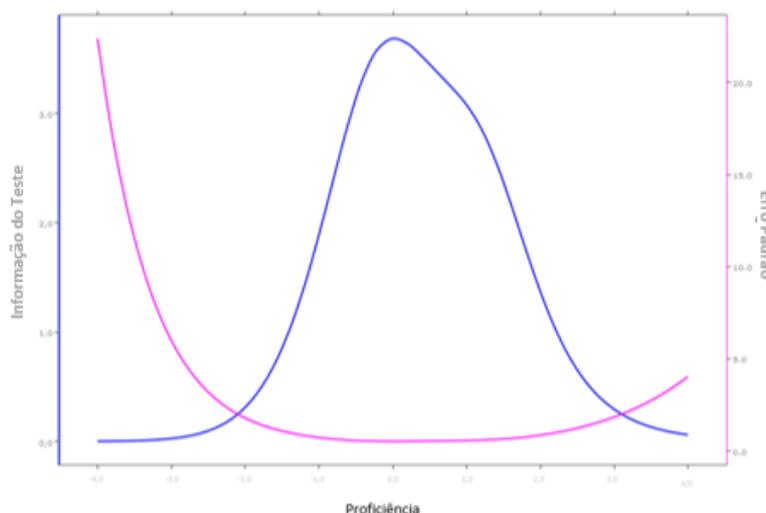


Figura 3 – Informação do Teste

Análise dos Parâmetros dos Itens

Para a validação dos parâmetros e conseqüentemente não eliminação de itens no processo, é necessário que os parâmetros apresentem as seguintes configurações.

Os parâmetros “a” e “b” são interpretados em função das categorias de cada item. Parâmetro “a”: Parâmetro de discriminação ou inclinação do item: Para os parâmetros “a” são esperados valores negativos para as categorias incorretas do item e valor positivo para o suposto gabarito. Quanto maior a proficiência, maior a probabilidade de responder corretamente o item. Apresentar valores positivos sugere que é o gabarito. Parâmetro “b” é definido como o Parâmetro de locação do item sobre a escala contínua de teta na qual as Curvas Características de categorias adjacentes se cruzam (De Ayla, 1993)

Na tabela 2, encontram-se os parâmetros dos itens obtidos a partir do Modelo Nominal da TRI. Em amarelo, estão marcados os maiores valores positivos dos parâmetros dos itens: supostos gabaritos, em amarelo; em negrito, os verdadeiros gabaritos e, em vermelho, os itens que divergiram em relação ao gabarito oficial.

TABELA 2 – Parâmetros dos Itens

	a1	a2	a3	a4	c1	c2	c3	c4
I01	-0,698215518	-0,31897707	-0,186895328	1,204087916	-1,211080134	-0,196590705	-0,705111619	2,112782459
I02	-0,586039019	1,438507484	-0,416089062	-0,436379403	-1,688569903	1,776393358	0,665828952	-0,753652407
I03	-0,45790108	-0,207409506	1,0172593	-0,351948713	-0,805975369	0,944345957	0,552116183	-0,690486771
I04	-0,376016402	-0,047408045	-0,338202448	0,761626895	-0,936953082	0,823477623	-1,009927924	1,123403383
I05	-0,743737142	0,130032944	0,83103269	-0,217328491	-0,277237733	-0,268395654	0,877708124	-0,332074737
I06	0,31614895	-0,312150245	-0,34626221	0,342263505	1,304166462	-0,094058528	-0,343134302	-0,866973633
I07	-0,361166425	1,180173106	-0,766052836	-0,052953845	-1,111732824	1,623173806	-0,593782713	0,082341731
I08	0,511977889	-0,472053608	-0,583104136	0,543179855	1,407074828	-0,45399872	-1,545148834	0,592072726
I09	-0,151292469	0,827029417	-0,575981115	-0,099755832	0,049820865	2,05473844	-0,896256327	-1,208302978
I10	-0,265450587	-0,123141369	0,616143732	-0,227551776	1,009111089	0,036315684	0,361956153	-1,407382927

Fonte: o autor

A figura 4 apresenta as curvas características dos itens calibrados e sem divergência do gabarito oficial, todos os distratores com curvas em queda à medida que as proficiências aumentam e os gabaritos em ascensão medida que aumenta as proficiências.

Comparando as proficiências obtidas através do modelo logístico de 3 parâmetros e o modelo nominal, observa-se uma correlação forte de 0,94341.

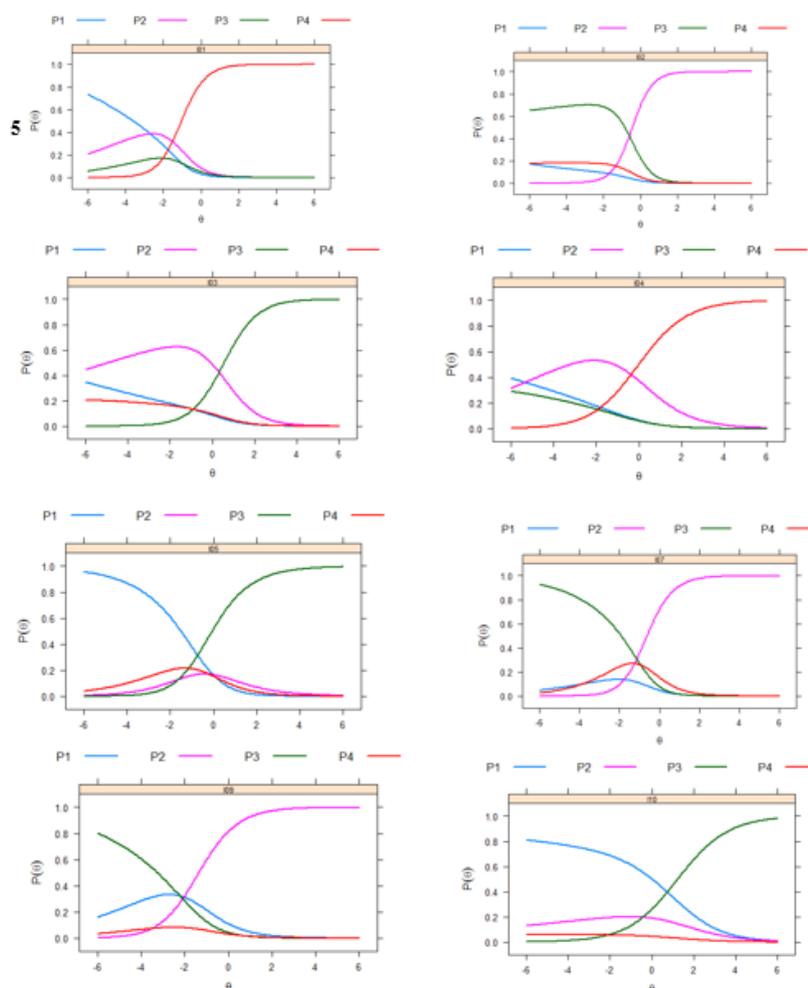


Figura 4 – Itens calibrados e sem divergência do gabarito oficial.

Sigmae, Alfenas, v.8, n.2, p. 735-741, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).
18º Simpósio de Estatística Aplicada à Experimentação Agrônoma (SEAGRO).

A figura 5 apresenta as curvas características para os itens 6 e 8. Ambos os itens apresentam gabarito oficial na letra “A”, porém existe uma tendência de crescimento do distrator “D” por estudantes de maiores proficiências, fazendo com que o distrator possua um parâmetro “a” um pouco mais elevado que o gabarito oficial, algo não constatado na TCT (Teoria Clássica de Testes) e nem pelo Modelo Logístico de 3 parâmetros. Vale ressaltar que as duas questões possuem gabarito correto. Após uma análise com especialistas, constatou-se que os dois itens deram margem ao erro por possuírem distratores muito próximos ao gabarito.

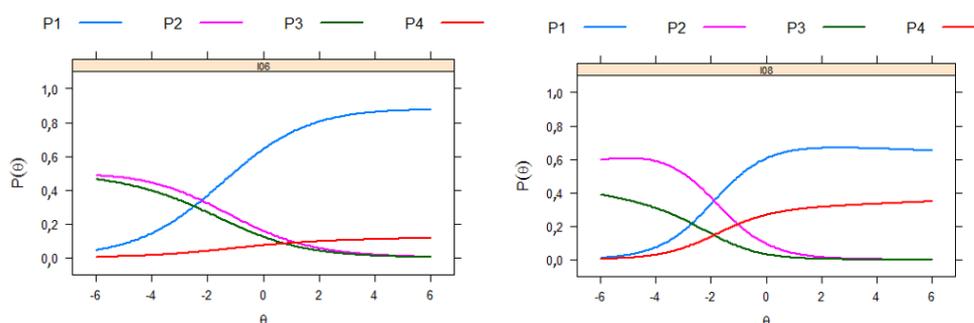


Figura 5 – Itens calibrados e com divergência do gabarito oficial

Conclusões

O Trabalho mostrou resultados da avaliação diagnóstica da Rede Municipal de Ensino de Curitiba, utilizando um modelo da TRI não tão usual em avaliações educacionais. O modelo nominal conseguiu captar alguns problemas nos itens e, conseqüentemente, nos distratores da avaliação não captados pelo modelo logístico de 3 parâmetros (ML3). O ML3 categoriza os itens em acertos e não acertos, ao contrário do modelo nominal, que trabalha com todas as categorias dos itens. Nenhum item foi eliminado por fugir dos pressupostos dos parâmetros. O modelo proposto nominal diverge, em dois dos 10 itens, na identificação do gabarito e da lógica da análise da Teoria Clássica de Testes.

Referências Bibliográficas

- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. 2000. Teoria de Resposta ao Item: conceitos e aplicações. ABE — Associação Brasileira de Estatística, 4^o SINAPE, 2000.
- BLAND, J. M.; ALTMAN, D. G. Cronbach’s alpha. *British Medical Journal*. p. 314-572. 1997.
- FREITAS, L. C.; SORDI, M. R. L.; MALAVASI, M. M. S.; FREITAS, H. C. L. Avaliação educacional: caminhando pela contramão. 5. ed. Petrópolis: Vozes, 2013.
- STREINER, D. L. Being inconsistent about consistency: when coefficient alpha does and doesn’t matter. *Journal of Personality Assessment*. v. 80, p. 217-222. 2003.

Sigmae, Alfenas, v.8, n.2, p. 735-741, 2019.

64^a Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).
18^o Simpósio de Estatística Aplicada à Experimentação Agrônômica (SEAGRO).