

AVALIAÇÕES DE SISTEMA: ESTIMAÇÃO DE PROFICIÊNCIAS POR MEIO DO USO DE MODELAGENS MATEMÁTICA BASEADA NA TEORIA DE RESPOSTA AO ITEM

Wagner Bandeira Andriola
Universidade Federal do Ceará
w_andriola@yahoo.com

Resumo

O presente texto tem o intuito de descrever como são estimadas as proficiências dos alunos (qualidade das aprendizagens) nas avaliações de sistema educacionais, a partir de modelo matemático baseado na Teoria de Resposta ao Item (TRI). A TRI visa solucionar uma grave limitação da sua antecessora, a Teoria Clássica dos Testes (TCT), que foi muito utilizada em sistemáticas nacionais, estaduais e municipais, com o intuito de avaliar a qualidade do aprendizado do alunado através de testes de rendimento (ANDRIOLA, 2004). Ressalta-se que a Teoria de Resposta ao Item (TRI) possui duas características essenciais. A primeira é a possibilidade de comparação de resultados de sujeitos distintos ao longo de períodos temporais. Com efeito, pode-se comparar resultados de anos diferentes e até mesmo de estados ou países distintos. A segunda é o fato de poder colocar em uma mesma escala o item e a habilidade do aluno. Nessa perspectiva, a utilização da TRI produz informações sobre as características dos itens dos testes, tal como seu grau de dificuldade, a capacidade que os itens têm de discriminar diferentes grupos de estudantes e a probabilidade de acerto ao acaso. Pretende-se, assim, apresentar o modelo matemático da TRI e discorrer sobre as premissas: i) Os alunos têm distintos níveis de aprendizado que podem ser identificados; ii) as diferentes habilidades componentes de cada nível de proficiência podem ser avaliadas de forma precisa; iii) a magnitude das diferenças de proficiências entre os alunos pode ser estimada; iv) a média das proficiências dos alunos pode ser calculada e, assim, determinar-se as diferenças entre escolas; v) os fatores internos associados às escolas de elevadas médias podem ser identificados e, assim, se podem expandi-los às demais; vi) é possível agrupar as médias das escolas e, assim, ter indícios da qualidade do próprio Sistema Municipal de Educação ao longo de certo período temporal (uso de séries históricas); vii) é possível introduzir ações pedagógicas de relevo, de modo a proporcionar mudanças substantivas nos níveis de proficiências dos alunos.

Palavras-chave: Avaliação educacional; Teoria de Resposta ao Item (TRI); Modelagem Matemática.

1. Introdução

A partir da década de 1990, a avaliação de sistemas escolares passou a ocupar posição central nas políticas públicas de educação, sendo recomendada e promovida por agências internacionais, pelo Ministério da Educação e por Secretarias de Educação de numerosos estados brasileiros, como elemento privilegiado para a realização das expectativas

de promoção da melhoria da qualidade do ensino básico e superior. Os diversos níveis e modalidades de ensino - da educação básica à pós-graduação - têm sido objeto de avaliação por parte do poder público sob o pressuposto de que a avaliação pode "produzir" um ensino de melhor qualidade.

As possíveis contribuições da avaliação na promoção da melhoria do ensino têm sido discutidas com base em dois tipos de argumentos: a. O primeiro considera a lógica interna dos processos de avaliação, examina seus princípios e procedimentos, procura explicitar os enfoques e critérios adotados, tomando-os como referência para a discussão das potencialidades e dos limites das práticas adotadas. As possibilidades de interferência da avaliação na melhoria da qualidade de ensino estariam presentes na própria lógica intrínseca ao seu delineamento; b. O segundo considera também a utilização dos resultados na condução das políticas educacionais, examinando a validade, a relevância e a oportunidade das iniciativas adotadas como consequência do conhecimento obtido mediante o processo de avaliação.

No Brasil, os debates a propósito das iniciativas do governo federal e de governos estaduais tendem a focalizar os princípios e os delineamentos dos programas realizados, procurando explicitar suas potencialidades ou deficiências e/ou riscos inerentes à sua implementação, explorando, por vezes, concepções de qualidade de ensino implícitas nos programas de diferentes administrações. São ainda escassas as investigações que se propõem a explorar o uso dos resultados nos processos decisórios pelas diversas instâncias do sistema educacional, o que torna oportunas as pesquisas dessa natureza. Sua institucionalização, tanto em âmbito nacional quanto nas unidades federadas, tem implicações nas políticas públicas de educação envolvendo as instituições, o currículo, os profissionais e os estudantes, e evidencia uma redefinição do Estado em relação ao seu papel na área.

No Brasil, a partir de 1995, a implementação de processos avaliativos intensificou-se e a visão de melhoria da qualidade do ensino e do papel do Estado na educação, intrínseca à avaliação em larga escala, assume crescentemente maior importância no debate educacional. Nesse período, agências internacionais passam a estimular e influenciar de modo mais direto o delineamento das propostas de avaliação de sistemas, demonstrando claro interesse sobre a eficácia dos investimentos externos na educação. O Sistema de Avaliação da Educação Básica - Saeb -, de nível nacional, implementado a partir de 1990, passa, por exemplo, a ser parcialmente financiado pelo Banco Mundial e, conseqüentemente, sofre transformações de acordo com a mudança de seus agentes, influências de concepções, pesquisas e técnicas avaliativas internacionais. A consolidação do Saeb estimulou ainda as propostas de avaliação

no âmbito de governos estaduais, iniciativas que passaram a ter centralidade nas políticas em curso. Nessa direção, a adoção de procedimentos estatísticos avançados começou a ser algo perseguido pelos sistemas de avaliação que estavam sendo criados. Um desses procedimentos é conhecido pelo nome de Teoria da resposta ao Item (TRI).

2. Breve revisão histórica da TRI

Como sempre ocorre no âmbito das Ciências, os modelos ou teorias não surgem de um momento para outro, da noite para o dia, nem tampouco de forma linear (ANDRIOLA, 1998). A Teoria de Resposta ao Item (TRI) não é, como se poderia ingenuamente acreditar, um novo enfoque psicométrico (ANDRIOLA, 2001). Suas raízes foram estabelecidas nos anos 1930 pelo psicólogo Louis Leon Thurstone (1887-1956). Não obstante, foi somente com o avanço dos sistemas de informação e dos programas de computação que a nova teoria foi, de fato, posta em uso, de modo massivo, pelos sistemas educacionais mundo afora (ANDRIOLA, 2009). Para ilustrar o exposto, cumpre destacar que, no âmbito brasileiro, a TRI vem sendo usada, desde 1990, pelo Sistema Nacional de Avaliação da Educação Básica (SAEB) implementado pelo Ministério de Educação (MEC) desde 1998 (ANDRIOLA & BARRETO, 1997).

A TRI visa solucionar uma grave limitação da sua antecessora, a Teoria Clássica dos Testes (TCT), modelo este que foi muito utilizado em sistemáticas nacionais, estaduais e municipais com o intuito de avaliar a qualidade do aprendizado do alunado através de avaliações em larga escala empregando-se testes de rendimento (ANDRIOLA, 2004). Tal limitação se refere ao fato de que, *na TCT, o grupo de alunos selecionado com base nos seus desempenhos depende, quase exclusivamente, do teste utilizado* (ANDRIOLA, LEITE & MAIA, 2013).

De modo a ilustrar a limitação referida, vamos supor que um grupo de 300 alunos concorresse a 100 vagas e, para tal, se submetesse a um Teste 1. Ao final do certame, haveria 100 selecionados com base no uso do Teste 1. Na linha das suposições, imaginemos que o certame tivesse sido anulado por causas estranhas ao certame (omitamos a hipótese de problemas na elaboração do teste) e que, novamente, os 300 alunos devessem ser submetidos a uma nova seleção, através do uso de um Teste 2. Após a aplicação desta, nada asseguraria que os 100 alunos selecionados no primeiro momento, em que foi empregado o Teste 1, fossem os mesmos selecionados no momento posterior, em que foi usado o Teste 2, embora este medisse os mesmos conteúdos do seu antecessor, o Teste 1. *Esse caso fictício exemplifica*

como o instrumento de medida (teste) utilizado pode comprometer a seleção justa dos melhores alunos (ANDRIOLA, 2003). É como se o comprimento de uma mesa, por exemplo, dependesse do instrumento (fita métrica) utilizado para a sua medida. Ao usar a fita métrica A obteríamos uma medida de comprimento A, enquanto ao usarmos a fita métrica B obteríamos uma medida de comprimento B. Como explicar as diferenças destas duas medidas hipotéticas, se o comprimento da mesa é invariável? Com os dois exemplos, pode-se compreender como a estimativa do aprendizado dos alunos é contaminada pelo instrumento de medida utilizado (o Teste), deixando de garantir, assim, a seleção dos alunos de maior mérito (ANDRIOLA, 2008).

Para solucionar esse grave problema a TRI fundamenta-se em dois princípios: *convergência e separabilidade*. O princípio da *convergência* fundamenta-se na idéia de que o instrumento utilizado para medir a aprendizagem discente deve maximizar a probabilidade de escolher os candidatos com maior grau de aprendizagem, independentemente de estes alunos serem submetidos a outros instrumentos de medida em outras situações de seleção. Em síntese: *os resultados devem convergir para a escolha dos melhores candidatos (ANDRIOLA, 2005).*

Já o segundo princípio, o da *separabilidade*, está associado à possibilidade de *distinguir os indivíduos conforme o grau de aprendizagem destes, de modo a garantir a meritocracia e, assim, contribuir para a efetividade da inclusão e da justiça sociais*. Por outro lado, a TRI permite comparar o progresso do aluno ao longo da sua formação acadêmica, bem como estimar o valor agregado pelo curso e/ou pela instituição, contribuindo, com isso, para estudos com vistas ao aprimoramento da gestão pedagógica, bem como da formação acadêmica vislumbrada (ANDRIOLA, 2006; ANDRIOLA 2011).

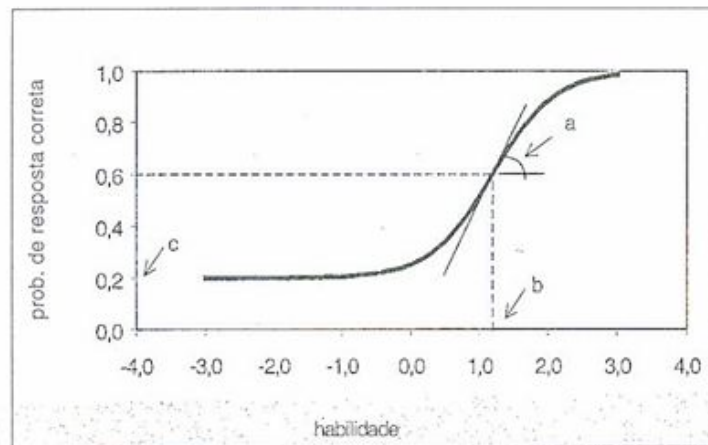
3. Bases Matemáticas da TRI

A TRI propõe um modelo matemático que formaliza a relação entre elementos essenciais da situação na qual uma pessoa responde a um problema (item de uma prova de rendimento, por exemplo). Nessa situação, quanto maior a habilidade da pessoa no fator requerido pelo problema maior será a probabilidade de que ela responda corretamente (MUÑIZ & HAMBLETON, 1992; LORD, 1980).

Conforme Muñiz (1997), a TRI representa matematicamente essa situação por meio da Curva Característica do Item (CCI), que indica a probabilidade de acerto em função da habilidade das pessoas que o respondem e da dificuldade do problema (dependendo do

modelo, podem ser incluídas outras características como a discriminação e a probabilidade de acertos ao acaso). O modelo de um parâmetro inicialmente criado por Georg Rasch em 1980 caracteriza o item somente pela sua dificuldade (parâmetro a).

Segundo GOLDSTEIN e WOOD (1989), em 1968 Allan Birnbaum apresentou o modelo de dois parâmetros logísticos, caracterizando a qualidade do item a partir da dificuldade (parâmetro a) e do poder de discriminação (parâmetro b). Atualmente, o modelo logístico de três parâmetros é o mais utilizado mundialmente, posto que se baseia em três componentes do item: a dificuldade, a discriminação e a probabilidade de ser acertado ao caso (parâmetro c). A partir da estimação destes três parâmetros é possível representar a qualidade do item através da Curva Característica do Item (CCI), conforme a figura abaixo.



Na figura acima:

- ✓ a representa a dificuldade do item;
- ✓ b representa o poder de discriminação do item;
- ✓ c representa a probabilidade de acerto ao acaso do item;
- ✓ a *habilidade* é o construto latente (representada pela letra grega θ).

Como destacou Pasquali (1997), a CCI é a característica particular de cada item, é sua carteira de identidade, uma vez que compartilha uma forma geral muito parecida com um S.

Para finalizar, convém destacar, por oportuno, que a prática da avaliação psicológica e educacional sofreu profundas modificações com o surgimento dos modelos de TRI. Desde então, foram propostos distintos modelos unidimensionais para dados dicotômicos; desenvolveram-se variados procedimentos para a estimação dos parâmetros dos itens e da magnitude dos sujeitos na variável latente θ – estimação de máxima verossimilhança condicional e estimação de máxima verossimilhança conjunta; formularam-se distintas provas para verificar o grau de ajuste do modelo aos dados empíricos – qui-quadrado e análise dos resíduos. Nesse novo contexto, o foco das sofisticadas análises estatísticas desviou-se do teste

(que representa o macro) para o item (que representa o micro). A lógica para tal câmbio é a mesma usada pela Física Quântica: o micro (nano)¹ exerce influência sobre o macro. Justifica-se, assim, a acentuada ênfase no uso dos modelos de TRI, visto que se prestam à análise acurada do micro - os itens (ANDRIOLA, 2009).

4. O uso da TRI para estimar proficiências

Após a elaboração dos itens, em conformidade com as especificações pedagógicas adotadas (p. ex.: matrizes de referência), procede-se à parametrização destes. A parametrização consiste na determinação dos seguintes aspectos inerentes ao item: a) Dificuldade; b) Discriminação; c) Probabilidade de acerto ao acaso.

Com estes parâmetros determinados pode-se, agora, estimar o grau de aprendizagem (ou de proficiência) dos alunos. A partir disto, pode-se avaliar a eficácia do ensino, agrupando os alunos em turmas, em escolas e em sistemas. Os três parâmetros (dificuldade, discriminação e probabilidade de acerto ao acaso) são estimados através do seguinte modelo matemático.

$$P(\theta) = c_i + (1 + c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

- ✓ $P(\theta)$ é a probabilidade de acerto do item i dada determinada magnitude de θ ;
- ✓ θ é a variável latente medida pelo item i ;
- ✓ a_i é o índice de discriminação do item i ;
- ✓ b_i é o índice de dificuldade do item i ;
- ✓ c_i é a probabilidade de que o item i seja acertado ao acaso;
- ✓ e é o valor (2,72) na base dos logaritmos neperianos;
- ✓ D é uma constante de valor 1,7

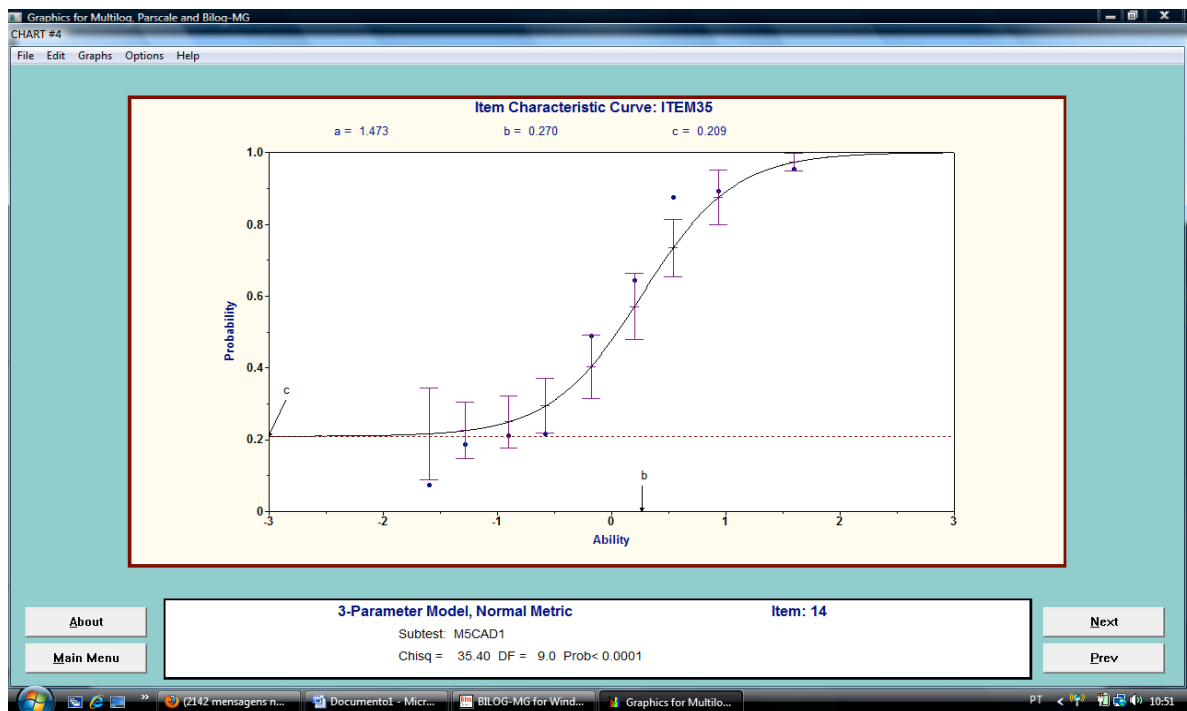
Os passos para a estimação da proficiência serão descritos a seguir

¹ O prefixo nano indica uma unidade de medida derivada, igual a 10^{-9} vezes a primeira (Cf. Dicionário Novo Aurélio, Século XXI). No campo da Física, o termo quark foi cunhado pelo físico norte-americano Murray Gell-Man e refere-se a um dos constituintes fundamentais da matéria. No seu livro, *El quark y el Jaguar* (4ª edição, Madrid: Ed. Metatemas, 2003), há uma citação que sintetiza muito bem a associação entre o micro e o macro, o simples e o complexo: (...) *los quarks son las partículas elementales que constituyen el núcleo atómico (...). El Jaguar representa la complejidad del mundo que nos rodea, especialmente tal como se manifiesta en los sistemas complejos adaptativos (...). La imagen del quark e del Jaguar transmite perfectamente mi idea de lo simple y lo complejo: de un lado las leyes físicas subyacentes de la materia y el universo, y del otro, el rico entramado del mundo que percibimos directamente y del que formamos parte* (p. 29).

1. Inicialmente se procede à estimação dos três parâmetros de cada item (dificuldade, discriminação e probabilidade de acerto ao acaso).
2. Para tal, uma primeira equação estima os parâmetros dos itens, independentemente da magnitude dos alunos na variável latente (θ), usando, para tal, o número total de respostas corretas.
3. Posteriormente, sobre as bases de uma segunda equação, pode-se estimar a magnitude dos alunos (as proficiências) na variável latente (θ) conhecendo-se tão-somente os parâmetros dos itens, já estimados pela primeira equação.
4. Finalmente, uma terceira equação permite o contraste dos resultados da primeira equação com os da segunda equação (Método de Máxima Verossimilhança), verificando, assim, sua adequação.
5. Desse modo, pode-se dizer que houve, quase concomitantemente, a estimação:
 - ✓ Dos parâmetros dos itens (dificuldade, discriminação e probabilidade de acerto ao acaso);
 - ✓ Da proficiência de cada aluno;
 - ✓ Da plausibilidade dos dois conjuntos de valores (dos parâmetros dos itens e da proficiência gerada).

Após a parametrização, o modelo gera a CCI (curva característica do item) para cada item, conforme apresenta-se na figura 1.

Figura 1: Curva característica do Item



Com os valores do *parâmetro b* (a dificuldade) se pode organizar os itens de modo crescente. Por exemplo:

Q30 (Cad. 2): $b = -0,025$ (Fácil);

Q39 (Cad. 6): $b = 2,421$ (Muito Difícil);

Q23 (Cad. 4) $b = 0,319$ (Intermediária);

Q23 (Cad. 3): $b = 0,341$ (Intermediária).

6. Estes valores explicam algo inusitado, à primeira vista: *como alunos com mesma quantidade de acertos podem ter proficiências distintas*. Por fim, a seguir serão descritas as considerações finais dessa pesquisa.

5 Considerações Finais

O texto discorreu acerca das potencialidades dos Sistemas de Avaliação existentes no Brasil, em distintos níveis da administração pública: municipal, estadual e federal. Desde os anos 1990, com a criação do Sistema de Avaliação da Educação Básica (SAEB), foram introduzidos, no Brasil, procedimentos para averiguar a qualidade dos sistemas educacionais, a partir da estimação das proficiências dos alunos em testes de rendimento. Sob esta ótica, introduziram-se, também, inovações metodológicas nos procedimentos de análises dos dados resultantes do desempenho dos alunos nas referidas provas. A mais destacada inovação foi efetivada com a adoção da Teoria de Resposta ao ITEM (TRI), dado que esta possibilita comparar resultados de sujeitos distintos ao longo de períodos temporais. Igualmente, podem ser comparados resultados de outras entidades (escolas e sistemas educacionais) obtidos em anos diferentes, bem como de municípios, estados e países distintos. A segunda é o fato de poder colocar em uma mesma escala o item e a proficiência dos alunos.

Para concluir, há que se destacar que o uso sistemático da TRI proporcionou, dentre outros avanços, os seguintes:

- i) A identificação de distintos níveis de aprendizado do alunado (proficiências);
- ii) A identificação das diferentes habilidades componentes de cada nível de proficiência;
- iii) A estimação da magnitude das diferenças de proficiências entre alunos;
- iv) A determinação da média das proficiências dos alunos que podem ser agrupados por escola, possibilitando a observação de possíveis diferenças entre escolas;
- v) A identificação de fatores internos associados às escolas de elevadas médias de proficiências;

- vi) A identificação do nível de qualidade do próprio Sistema Municipal de Educação ao longo de certo período temporal (uso de séries históricas), a partir da observação do comportamento das médias de proficiências dos alunos agrupados em escolas;
- vii) A introdução de ações pedagógicas de relevo, de modo a proporcionar mudanças substantivas nos níveis de proficiências dos alunos.

7. Referências

ANDRIOLA, W. B. Utilização da Teoria da Resposta ao Item (TRI) para a organização de um banco de itens destinados a avaliação do raciocínio verbal. *Psicologia: Reflexão e Crítica, Rio Grande do Sul*, v. 11, n.2, p. 295-308, 1998.

ANDRIOLA, W. B. Descrição dos principais métodos para detectar o Funcionamento Diferencial dos Itens (DIF). *Psicologia: Reflexão e Crítica, Rio Grande do Sul*, v. 14, n.3, p. 643-652, 2001.

ANDRIOLA, W. B. Descripción de los principales métodos para detectar el funcionamiento diferencial de ítem (DIF): un análisis crítico. *Bordón: Revista de Pedagogia, Madrid*, v. 2, n.55, p. 7-19, 2003.

ANDRIOLA, W. B. Avaliação da aprendizagem: conceitos, dados, problemas e perspectivas. *Psicologia Argumento, Curitiba*, v. 22, n.39, p. 45-55, 2004.

ANDRIOLA, W. B. Breve retrospectiva histórica acerca dos estudos que abordam o viés dos itens de testes psicológicos. *Psicologia Argumento, Curitiba*, v. 23, p. 35-44, 2005.

ANDRIOLA, W. B. Estudo sobre o viés de itens em testes de rendimento: uma retrospectiva. *Estudos em Avaliação Educacional, São Paulo*, v. 17, p. 115-134, 2006.

ANDRIOLA, W. B. Uso da Teoria de Resposta ao Item (TRI) para analisar a equidade do processo de avaliação do aprendizado discente. *Revista Iberoamericana de Evaluación Educativa, Santiago de Chile*, v. 1, p. 171-189, 2008.

ANDRIOLA, W. B. Psicometria moderna: características e tendências. *Estudos em Avaliação Educacional, São Paulo*, v. 20, p. 319-340, 2009.

ANDRIOLA, W. B. Doze motivos favoráveis à adoção do Exame Nacional do Ensino Médio (ENEM) pelas Instituições Federais de Ensino Superior (IFES). *Ensaio: Avaliação e Políticas Públicas, Rio de Janeiro*, v. 19, p. 107-126, 2011.

ANDRIOLA, W. B.; BARRETO, J. A. Análise métrica de instrumento de medida da aprendizagem através da Teoria da Resposta ao Item-TRI. *Ensaio: Avaliação e Políticas Públicas, Rio de Janeiro*, v.5, n.5, p. 59-74, 1997.

ANDRIOLA, W. B.; LEITE, R. H.; MAIA, J. L. Análise métrica de questões componentes de testes de rendimento: mecanismo de feedback para aprimorar sua elaboração. *Foro Educacional, Santiago de Chile*, v. 21, p. 13-31, 2013.

GOLDSTEIN, H.; WOOD, R. Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology, London*, v. 42, p. 139-167, 1989.

LORD, F. M. *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates, 1980.

MUÑIZ, J. *Introducción a la teoría de respuesta a los ítems*. Madrid: Ediciones Pirâmide, 1997.

MUÑIZ, J.; HAMBLETON, R. K. Medio siglo de Teoría de Respuestas a los Ítems. *Anuario de Psicología, Madrid*, n. 52, p. 41-66, 1992.

PASQUALI, L. *Psicometria: teoria e aplicações*. Brasília: UnB, 1997.

