

PSICOMETRIA: LIMITAÇÕES E AVANÇOS

Cristina Régia Barreto Moreira¹

RESUMO

A Psicometria é o conjunto de técnicas capaz de medir processos mentais e promover a interpretação destes. Suas principais ferramentas são a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI). A avaliação educacional é um tema polêmico na ação docente, mas também nos sistemas de ensino. Assim, é evidente a importância de realização de estudos nessa área e a utilização de testes psicométricos na Educação, uma vez que a aprendizagem é um traço latente. A TCT, amplamente incorporada à avaliação educacional, apresenta várias limitações. Os avanços oriundos da TRI são uma resposta às restrições da TCT, uma vez que com a TRI torna-se possível comparar desempenho de populações que respondem a testes distintos, desde que se preservem alguns itens comuns; é também possível comparar indivíduos da mesma população que respondem a testes completamente diferentes. Consequentemente, muitas são as aplicações exequíveis com uso da TRI, inclusive em Educação.

Palavras-chave: Psicometria. TRI. Avaliação Educacional.

INTRODUÇÃO

A Psicometria é uma área da Psicologia que trata de atribuir significados às respostas de um indivíduo a uma série de tarefas ou testes, através de processos mentais. Há duas vertentes para a Psicometria: a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI). Nesta última, os testes recebem tratamento criterioso, uma vez que os traços latentes são características não observáveis diretamente, tais como inteligência e ansiedade.

De fato, a TRI se apresenta como um avanço da Psicometria em relação à TCT. As vantagens da TRI sobre a TCT são importantes, por isso a TRI tem sido largamente utilizada nas avaliações educacionais em seus diferentes contextos, assim como em outros campos da atividade humana e demais ciências.

Alguns exemplos de uso dos modelos unidimensionais da TRI são os utilizados pelo *Programme for International Student Assessment* (PISA) e pelo exame de proficiência em língua inglesa (TOEFL). No Brasil, a partir da aplicação da TRI no Exame Nacional do Ensino Médio (ENEM), tem crescido o interesse de diversos pesquisadores e instituições no uso dessa ferramenta, principalmente em avaliações de larga escala, para averiguar com eficiência as proficiências dos respondentes.

Segundo Rabelo (2013), as limitações da TCT estão relacionadas a:

¹ Mestre em Matemática (UFERSA/SBM). E-mail: cristina.moreira@ifrn.edu.br

[...] discriminação dos itens, fidedignidade dos testes e comparabilidade de desempenho de indivíduos que se submeteram a testes diferentes. As estatísticas clássicas são restritas a determinado momento, a determinado teste e ao grupo que respondeu à prova, dificultando o acompanhamento pedagógico ao longo dos anos. Além disso, na TCT, os parâmetros dos itens de um teste dependem diretamente da amostra de indivíduos para os quais eles foram calculados. Desse modo, a dificuldade de um item vai variar de pesquisa para pesquisa, em função da amostra de respondentes, isto é, o parâmetro é dependente dos sujeitos (RABELO, 2013, p. 45).

No entanto, apesar dessas restrições, a TRI não busca substituir a TCT, mas complementar e aprofundar os métodos de análise das respostas a testes aplicados a um número grande de respondentes. No momento, a TRI utiliza medidas da TCT para considerar os resultados do grupo de modo a qualificar os itens do teste e melhor interpretar o resultado de cada indivíduo avaliado.

Diante da importância das inúmeras aplicações da Psicometria, esta pesquisa objetiva introduzir os principais conceitos, limitações e avanços da TCT e da TRI, enfatizando as premissas dos modelos matemáticos que as sustentam e algumas aplicações na avaliação educacional.

No cenário contemporâneo, a pesquisa bibliográfica realiza-se sob a perspectiva dos trabalhos publicados por Rabelo (2013), Andrade *et al.* (2000), Ferreira (2009), Pasquali e Primi (2003), dentre outros.

As hipóteses são de que a Psicometria oferece um campo produtivo, ainda em pleno desenvolvimento, capaz de ampliar efetivamente o conhecimento sobre a aprendizagem educacional, assim como de outros construtos. A cada dia, com o avanço tecnológico e matemático, novos algoritmos e softwares vão resolvendo problemas de restrições dessas teorias e permitindo melhorar as estimativas de traços latentes.

PSICOMETRIA: MODELOS MATEMÁTICOS

Com o objetivo de medir traços latentes (θ), ou seja, características de indivíduos não mensuráveis diretamente, a Psicometria propõe dois modelos matemáticos: a TCT e a TRI. As referências para as medidas mudam em cada teoria: a TCT utiliza o escore do teste como um todo; e a TRI concentra-se no item, como bem diz seu nome. No entanto, ambas realizam a análise através de parâmetros estimados e apresentam uma avaliação do instrumento de medida.

Teoria Clássica dos Testes

Durante muitos anos, a TCT foi a principal ferramenta de análise estatística nas avaliações de traços latentes e, desde então até os dias atuais, é a teoria dominante em se tratando de aplicação

na avaliação educativa. Por conseguinte, a TCT não foi suplantada pela TRI e nem deve ser abolida. Como será visto, ela é bastante útil. Além disso, ampara as teorias recentes na obtenção de análises mais qualitativas.

Ora, todo modelo matemático necessita de pressupostos que delimitam sua aplicação. Segundo Ferreira (2009) as premissas para a TCT são as abaixo descritas.

[...] No modelo clássico, dois construtos são introduzidos: o escore verdadeiro e o erro de medida. O escore verdadeiro para um indivíduo pode ser definido como um valor esperado dos seus escores obtidos em vários testes. O erro de medida pode ser definido como a diferença entre o escore verdadeiro e o observado. O modelo clássico supõe que: (1) os erros de medida são aleatórios com média zero e não correlacionados entre si e com os escores verdadeiros e (2) os escores verdadeiros, os observados e os erros de medida são linearmente relacionados. (FERREIRA, 2009, p. 8).

Disso resulta que o modelo matemático da TCT pode ser assim expresso (FERREIRA, 2009, p. 8):

$$x = t + \epsilon,$$

em que x , t e ϵ são, respectivamente, o escore observado, o escore verdadeiro e o erro de medida.

Ou seja, os pressupostos acima nos levam a:

1. $E(\epsilon) = 0$;
2. $\rho(t, \epsilon) = 0$;
3. $\rho(\epsilon_1, \epsilon_2) = 0$,

em que ϵ_1 e ϵ_2 são os erros de medida em duas aplicações de um teste.

Limitações da TCT

Ferreira (2009) resume as limitações teóricas graves da Psicometria Clássica em quatro pontos principais.

- a) Os parâmetros clássicos dos itens (dificuldade e discriminação) dependem diretamente da amostra de sujeitos utilizada para estabelecê-los. É o que Pasquali e Primi (2003) chamam de *subject-dependent*, isto é, o parâmetro varia de pesquisa para pesquisa em função da amostra de sujeitos.
- b) A avaliação das aptidões dos respondentes depende do teste utilizado. Nesse caso tem-se *test-dependent*, ou seja, a resposta fica ligada ao instrumento utilizado, de tal forma que o objeto medido afeta diretamente o instrumento utilizado (PASQUALI; PRIMI, 2003).

- c) A definição do conceito de fidedignidade na TCT apresenta dificuldades, uma vez que se baseia na correlação entre escores obtidos de um teste como oposto do erro de medida. O erro de medida é considerado o mesmo para todos os itens e examinados. Isso é muito improvável, pois seria admitir que indivíduos com baixa proficiência mobilizam erros iguais aos de proficiências mais altas.
- d) A informação gerada do item é resultado das análises globais do teste. Não se sabe o que cada item individualmente provoca no respondente. Nota-se uma incongruência: por que analisar itens que presumidamente são assumidos como adequados? Se são ou não adequados, como saber, já que a análise é feita em função do escore total, o qual cada item influencia? O problema é avaliar a adequação de um item utilizando-o na análise e considerando que os demais estejam adequados, quando, na verdade, nada se sabe sobre isso.

Por tudo isso é que se tornou pungente o surgimento de uma nova ferramenta de avaliação psicométrica para a superação dos fatos levantados. Mas apesar desses problemas serem conhecidos desde os anos 1930, a sua solução só ocorreu de forma efetiva nos anos 1980. A demora está relacionada à complexidade dos algoritmos matemáticos propostos pela TRI para obter a independência do instrumento de medida em relação ao objeto de medida.

A aplicação de tais algoritmos foi viabilizada somente após o avanço computacional e a disponibilidade de softwares específicos, sendo o primeiro deles o BICAL em 1979, seguidos pelo LOGIST e pelo BILOG em 1984 (PASQUALI; PRIMI, 2003).

Avanços com a Teoria de Resposta ao Item

Também conhecida como Teoria do Traço Latente, a TRI fundamenta-se na relação entre variáveis observáveis e variáveis hipotéticas para, fixando as aptidões, estimar os parâmetros dos itens respondidos pelos indivíduos.

Desse modo, são introduzidos dois postulados na nova teoria: (1) o traço latente θ é a causa e o desempenho do sujeito no item do teste. É o efeito, ou seja, comportamento é função de θ ; (2) a relação entre desempenho e traço latente ou aptidão de um indivíduo pode ser descrita por uma equação matemática monotônica crescente, isto é, à medida que θ aumenta, também aumenta a probabilidade de acerto ao item (PASQUALI; PRIMI, 2003).

Deve-se enfatizar que a TRI substitui a TCT em parte, principalmente a que se refere à análise dos itens do teste. Os avanços oriundos da TRI são uma resposta às restrições da Teoria Clássica. Pasquali e Primi (2003) dividem estas vantagens em cinco avanços.

- [...] a) o cálculo do nível de aptidão do sujeito independe da amostra de itens utilizados: diz-se que a habilidade do sujeito é independente do teste (*not test-dependent*). [...];
- b) o cálculo dos parâmetros dos itens (dificuldade e discriminação) independe da amostra de sujeitos utilizada: diz-se que os parâmetros são independentes dos sujeitos (*not group-dependent*). [...];
- c) a TRI permite emparelhar itens com a aptidão do sujeito. Isto quer dizer que se avalia a aptidão de um sujeito, utilizando itens com dificuldade tal que se situam em torno do tamanho da aptidão do sujeito, sendo, assim, possível utilizar itens mais fáceis para sujeitos com habilidades inferiores e itens mais difíceis para sujeitos mais aptos, produzindo escores comparáveis em ambos os casos. [...];
- d) a TRI constitui um modelo que não precisa fazer suposições que aparentam ser improváveis, tais como os erros de medida serem iguais para todos os testandos;
- e) a TRI não necessita trabalhar com testes estritamente paralelos como exige a psicometria clássica (PASQUALI; PRIMI, 2003, p.102-103).

Assim, torna-se possível comparar desempenho de populações que respondem a testes distintos, desde que se preservem alguns itens comuns ou, ainda, comparar indivíduos da mesma população que respondem a testes completamente diferentes. Isso amplia as aplicações da TRI porque possibilita a comparação longitudinal de populações.

Por outro lado, existem dois pré-requisitos postulados para a TRI: a unidimensionalidade e a independência local. Sobre a unidimensionalidade sabe-se, desde os anos 1930, que todo desempenho humano na execução de determinadas tarefas exige um conjunto de traços latentes, chamados de n dimensões (PASQUALI; PRIMI, 2003). Assim, o comportamento pode ser expresso como vetor de vários tetras. Isso ocorre simplesmente porque é evidente que todo comportamento humano é multideterminado, ou seja, as motivações são sempre variadas e simultâneas. No entanto, para a TRI a suposição é de que há uma única aptidão dominante responsável pelo desempenho no teste.

Agora, sobre a independência local supõe-se que mantidas constantes as aptidões que afetam o teste, menos o teta dominante, as respostas dos sujeitos a quaisquer dois itens são estatisticamente independentes (PASQUALI; PRIMI, 2003). A importância desse postulado é assumir que o desempenho no teste como o todo é o produto das probabilidades de acerto a cada item individualmente. Não significa que os itens não possam estar correlacionados, mas que um não interfere na probabilidade de acerto ao outro.

O que se deseja com essas suposições é apresentar um modelo matemático capaz de estimar a aptidão de um indivíduo realizando análises das respostas dadas por ele a uma série de itens. Isto é, a TRI fornece a probabilidade que um determinado sujeito tem de acertar a um item específico, conhecidos seus parâmetros e a proficiência dos respondentes.

Assim, quando um indivíduo responde um conjunto de itens, este produz um padrão de respostas com uma sequência de 1 (quando acerta) e 0 (quando erra). A partir dessas respostas é

gerada uma matriz formada pelo número de linhas igual ao número de respondentes e pelo número de colunas igual à quantidade de itens no teste. Essa tabela é conhecida como matriz de padrão de resposta dos indivíduos. Em avaliações de larga escala, como, por exemplo, o ENEM, essas matrizes são gigantescas, com milhões de linhas e muitas colunas. Esse é mais um motivo, além da complexidade dos algoritmos dos modelos da TRI, para a necessidade de recursos computacionais poderosos na manipulação desses dados.

TRI: breve retrospectiva histórica

A TRI teve início na década de 50 com modelos que consideravam somente uma aptidão em um único grupo e com respostas dicotômicas sim ou não, zero ou um. Mais precisamente foi a partir dos trabalhos de Lord, em 1952, que a TRI toma corpo através do primeiro modelo unidimensional de dois parâmetros. Concomitante e independentemente, em 1960, Rasch apresenta o modelo unidimensional de um parâmetro, descrito, em seguida, como modelo logístico por Wright. Após algumas aplicações, Lord responde a necessidade de incluir um terceiro parâmetro que considerasse o problema do acerto casual, surgindo o modelo de três parâmetros (ANDRADE *et al*, 2000).

Em 1968, Birnbaum propõe a substituição da função ogiva normal pela função logística, matematicamente mais adequada por não envolver integração e explicitar os parâmetros do item e a habilidade. O modelo de resposta gradual foi introduzido por Samejima em 1969. No entanto, os modelos de mais de duas categorias de respostas só foram implementados em 1972 por Bock, em 1978 por Andrich, em 1982 por Masters e em 1992 por Muraki (ANDRADE *et al*, 2000).

A partir de 1997, Bock e Zimowski introduziram os modelos logísticos de 1, 2 e 3 parâmetros para duas populações de respondentes, portanto, bem recentemente. O método da máxima verossimilhança marginal foi proposto por Bock e Zimowski, em 1970, para solucionar problemas de estimação simultânea de parâmetros e aptidões. Em 1981, Bock e Aitkin modificam esse método introduzindo o algoritmo EM de Dempster, Laird e Rubin (ANDRADE *et al*, 2000).

Já os métodos Bayesianos foram pensados para resolver problemas relacionados à estimação dos parâmetros dos itens respondidos corretamente ou incorretamente por todos os indivíduos e, ainda, os relacionados com a estimação das proficiências de indivíduos que acertaram ou que erraram todos os itens do teste.

MODELO LOGÍSTICO DE TRÊS PARÂMETROS

A função logística tem substituído o modelo da ogiva normal, em parte devido à simplificação do algoritmo, dado que não se usam integrais, além de apresentar melhor ajuste dos

dados. A preferência também se explica por que o método da máxima verossimilhança é matematicamente mais fácil de produzir do que as dos mínimos quadrados (PASQUALI; PRIMI, 2003).

A função logística aplicada a TRI é expressa da seguinte forma (PASQUALI; PRIMI, 2003, p. 106):

$$p(\theta) = \frac{1}{1 + e^{-a(\theta-b)}}$$

em que

- e é a constante de Euler, equivalente a 2,7182818...;
- b é o parâmetro de dificuldade do item;
- a é a discriminação do item;
- θ é a aptidão, traço latente, do indivíduo;
- $a(\theta - b)$ é o desvio logístico também chamado de logit.

Por outro lado, um modelo muito utilizado na TRI, inclusive pelo SAEB e o ENEM, é o logístico de três parâmetros (ML3 ou 3LP) que considera a dificuldade (b), a discriminação (a) e a probabilidade de acerto ao acaso (c) do item. O favoritismo do modelo ML3 está relacionado ao melhor ajuste desse modelo a itens politômicos com uma única opção correta que, em vista disso, podem ser dicotomizados nas categorias certo (1) ou errado (0).

Ora, o ML3 modela a probabilidade de um sujeito j de aptidão θ acertar um item i , cujas características são os parâmetros a , b e c , ou seja, discriminação, dificuldade e acerto casual, respectivamente.

Veja essa função expressa conforme Andrade *et al.* (2000, p. 9):

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)'}}$$

com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, n$, em que:

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i ;

θ_j representa a habilidade (traço latente) do j -ésimo indivíduo;

$P(U_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j com habilidade θ_j responder corretamente o item i e é chamada de Função de Resposta do Item – FRI;

- b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da habilidade;
- a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional à inclinação da Curva Característica do Item – CCI no ponto b_i ;
- c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item i (muitas vezes referido como a probabilidade de acerto casual);
- D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça resultados semelhantes ao da função ogiva normal (ANDRADE *et al.*, 2000, p. 9-10).

Esse modelo prevê que sujeitos com maior habilidade têm maior probabilidade de acertar o item e que esta não é uma relação linear. Para melhor compreensão dos elementos envolvidos nesse algoritmo e sua interpretação sugere-se a leitura de Rabelo (2013).

APLICAÇÕES NA EDUCAÇÃO

Pode-se dizer que a TRI está “em moda”, visto beneficiar pesquisas em todos os campos do conhecimento, incluindo no desenvolvimento de modelos próprios para testar desempenhos e reduzir a complexidade de seus métodos.

Na educação, a TRI tem sido frequentemente utilizada.

Alguns exemplos de avaliações em larga escala que utilizam a TRI são o teste TOEFL (*Test of English as a Foreign Language*) e o teste GRE (*Graduate Record Examination*), aplicados via testagem adaptativa por computador (*computerized adaptive testing - CAT*) pelo *Educational Testing Service – ETS* (ANDRADE *et al.*, 2010, p. 423).

Nesse contexto, influenciado por pesquisas em Psicometria e sistemas de avaliação em larga escala internacionais, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) implanta, a partir de 1995, as técnicas da Teoria de Resposta ao Item reestruturando o Sistema de Avaliação da Educação Básica (SAEB) que, desde 1990, levantava dados sobre o ensino brasileiro, mas que não permitia a comparabilidade desses dados ao longo dos anos.

O pioneirismo do SAEB e a implementação da Prova Brasil, com o uso da TRI, possibilitou a construção de uma série histórica do desempenho dos estudantes em todas as regiões do país. A partir de então, outras avaliações nacionais, tais como o Exame Nacional para Certificação de Competência de Jovens e Adultos (ENCCEJA) e o Exame Nacional do Ensino Médio (ENEM) passaram a utilizar a TRI.

Em se tratando de ENEM, sua adesão a TRI ocorre em 2009, depois de decorridos 11 anos de sua implementação. Segundo Rabelo (2013),

[...] isso permitirá o acompanhamento longitudinal do desempenho nas quatro áreas avaliadas, constituindo excelente indicador para a implantação de políticas públicas de melhoria da qualidade de ensino na educação básica (RABELO, 2013, p. 50).

Portanto, é possível, a partir dessas estimativas, realizar a interpretação da escala de proficiência, gerando relatórios para as escolas com intuito de orientar a ação pedagógica das mesmas e a implantação de políticas públicas pertinentes.

No entanto, deve-se lembrar de que dados educacionais dificilmente podem ser considerados unidimensionais. Por isso, espera-se que nos próximos anos modelos como os de Hambleton e McDonald sejam aplicados fortemente, uma vez que dados politômicos e multidimensionais têm maior ocorrência no âmbito educacional (ANDRIOLA, 2009).

LIMITAÇÕES DA TRI

Apesar de pouco tempo, o desenvolvimento da TRI é notável. No entanto, ainda será necessário superar algumas limitações dessa teoria. Para isso, o principal cuidado diz respeito ao ajuste dos modelos para a obtenção de itens e habilidades invariantes, ou seja, a escolha de modelos e pressupostos adequados que garantam a validade e comparabilidade dos resultados. (ANDRADE *et al.*, 2010).

Além disso, como o objetivo é descobrir um teste que seja a representação legítima do construto, isso exige uma estreita relação entre a Psicometria e a Psicologia Cognitiva, ou seja, a confiabilidade do construto de um teste pode ser vista de diferentes ângulos: análise da representação comportamental do construto, análise por hipótese, a curva de informação do item. (PASQUALI, 2009).

Andrade *et al.* (2000) traz um resumo das vantagens e desvantagens dos métodos da máxima verossimilhança marginal (MVM) e Bayesiano, tanto para a estimação dos parâmetros dos itens, quanto à estimação das habilidades. Segundo esse autor, as limitações da MVM relacionadas à estimação dos parâmetros dos itens são:

- (1) não está definido para itens com acerto total ou erro total;
- (2) é bastante trabalhoso computacionalmente;
- (3) necessidade do estabelecimento de uma distribuição para a média populacional;

- (4) apresenta problemas na estimação do parâmetro de acerto casual em alguns casos;
- (5) deve ser usado somente com um número suficientemente grande de respondentes. (ANDRADE *et al.*, 2000, p. 76).

Para a estimação das habilidades, o MVM não está definido para alguns padrões de respostas.

Já o método Bayesiano para estimação dos parâmetros dos itens apresenta as seguintes restrições: dar mais trabalho computacionalmente que o MVM e necessita de distribuições a priori para os parâmetros dos itens (ANDRADE *et al.*, 2000).

Por fim, o método Bayesiano para a estimação das habilidades tem a desvantagem de: produzir estimadores viciados; exigir cálculos mais complexos que o MVM e necessitar de uma distribuição a priori para θ .

CONSIDERAÇÕES FINAIS

A Psicometria Clássica na abordagem de avaliações psicológicas e educacionais apresenta como principais limitações: a dependência entre os parâmetros dos itens do teste e a amostra de respondentes; a dependência entre o teste como um todo e a aptidão dos sujeitos examinados; a fidedignidade da teoria, uma vez que assume a premissa de que o erro de medida é o mesmo para todos os respondentes.

Nesse contexto, surge a Psicometria Moderna para suplantando as limitações da TCT. Atualmente, existem vários modelos matemáticos para a TRI que se diferenciam quanto à função e à quantidade de parâmetros, o número de dimensões estudadas (unidimensionais ou multidimensionais), para traço latente cumulativo ou não cumulativo, quanto ao tipo de item (dicotômico ou politômico) e quanto ao número de populações envolvidas.

Os modelos de estimação de parâmetros dos itens e de estimação das habilidades propostos dessa nova teoria trouxeram os seguintes avanços: a medida da proficiência do sujeito independe do teste; os parâmetros dos itens são independentes dos respondentes; possibilita emparelhar itens em torno da aptidão dos examinados; supõe que a probabilidade de acerto não é a mesma para todos os sujeitos, portanto, indivíduos com maior proficiência têm maiores chances de acerto. Assim, a TRI fornece a probabilidade que um determinado sujeito tem de acertar a um item específico, conhecidos seus parâmetros e a proficiência dos respondentes.

Em vista disso, torna-se possível comparar avaliações de populações ao longo do tempo e aplicar testes distintos para os examinados, uma vez que os itens são invariantes e não interferem

na aptidão medida. Como a aprendizagem ou cognição é um traço latente, muitos são os exemplos de uso da TRI na educação no Brasil e no mundo.

Apesar desses avanços, os modelos da TRI precisam superar algumas restrições. As principais delas são: a complexidade computacional; estimação de acertos totais e erros totais no teste; estabelecimento de distribuição a priori para média, parâmetros dos itens ou para o traço latente; problemas de estimação de alguns acertos casuais; e aplicação apenas a um grande número de respondentes.

Como recomendação para pesquisas futuras, a partir dos estudos aqui revelados, sugerimos investigar: os modelos matemáticos de estimação dos parâmetros dos itens e das habilidades pela TRI; os recursos computacionais e sua evolução; a aplicação da TRI na educação; a engenharia de elaboração de itens; e a construção e interpretação das escalas de proficiência.

REFERÊNCIAS

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. Introdução à teoria da resposta ao ítem: conceitos e aplicações. In: **SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA**, 14, Caxambu, MG, 2000. Disponível em: <http://www.ufpa.br/heliton/arquivos/LivroTRI.pdf>. Acesso em: 16/02/2020.

ANDRADE, J. M.; LAROS, J. A.; GOUVEIA, V. V. O uso da Teoria de Resposta ao Item em avaliações educacionais: diretrizes para pesquisadores. **Avaliação Psicológica** [online]. Revista da Universidade São Francisco, São Paulo, v. 9, n. 3, p. 421-435. 2010. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300009. Acesso em: 10/02/2020.

ANDRIOLA, W. B. Psicometria Moderna: características e tendências. **Estudos em Avaliação Educacional**. Periódico da Fundação Carlos Chagas – FCC, São Paulo, v.20, n. 43, maio/ago., p. 319-340. 2009. Disponível em: www.fcc.org.br/pesquisa/publicacoes/ae/arquivos/1496/1496.pdf. Acesso em: 10/01/2020.

FERREIRA, F. F. G. **Escala de Proficiência para o ENEM: utilizando teoria da resposta ao ítem**. 105 p. 2009. Dissertação (Mestrado Programa de Pós-graduação em Matemática e Estatística) - UFPA, Belém, PA, 2009. Disponível em: <http://www.ppgme.ufpa.br/doc/diss/fialhoguedes.pdf>. Acesso em: 04/03/2019.

PASQUALI, L. Psicometria. **Revista da Escola de Enfermagem da USP**, n. 43, p. 992-999, 2009. Disponível em: http://www.scielo.br/scielo.php?pid=S0080-62342009000500002&script=sci_arttext. Acesso em: 17/02/2020.

PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item: TRI. **Avaliação Psicológica**, v. 2, n. 2, p. 99 – 110, 2003. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712003000200002&lng=pt&nrm=iso. Acesso em: 06/04/2019.

RABELO, Mauro. **Avaliação educacional**: fundamentos, metodologia e aplicações no contexto brasileiro. Rio de Janeiro: SBM, 2013. 258 p.