

**AVALIAÇÃO EDUCACIONAL EM LARGA ESCALA E
ALGUMAS CONSIDERAÇÕES SOBRE A TCT E A TRI**

**LARGE SCALE EDUCATIONAL EVALUATION AND SOME
CONSIDERATIONS ABOUT TCT AND TRI**

José Ailton Rodrigues Soares

Instituto Federal do Tocantins - IFTO, Palmas, TO
ailton.rs@hotmail.com

Aline Ferreira Amorim

Instituto Federal do Pará - IFPA, Cametá, PA
alineamorim19@hotmail.com

Claudionor Renato da Silva

Universidade Federal de Goiás - UFG, Jataí, GO
crenato@uft.edu.br

Resumo: O presente artigo tem como objetivo organizar algumas considerações sobre os principais aspectos da TCT (Teoria Clássica dos Testes) e da TRI (Teoria de Resposta ao Item) como técnicas de avaliação educacional em larga escala na educação básica brasileira. Utiliza-se o método da pesquisa bibliográfica para apresentar os conceitos e aplicações, ainda que breves, destes métodos. Para entender a TCT é preciso conceituar a psicometria, que se fundamenta na teoria da medida em ciências para explicar o sentido que tem as respostas dadas pelos sujeitos a uma série de tarefas e propor técnicas de medida dos processos mentais. Já a TRI, também considerada uma vertente da psicometria, se interessa especificamente por cada um dos itens. Pode-se concluir que a TCT é um método avaliativo menos complexo que TRI, além de necessitar de menor tamanho amostral. Já a TRI trata de método mais detalhista, que refina as análises dos itens e avalia outros aspectos mais subjetivos de uma prova. Contudo, como as duas técnicas são utilizadas juntamente, em estudos na área da matemática, a presente pesquisa pontua como principal crítica a falta de retorno destes resultados em melhoria da qualidade da educação básica brasileira.

Palavras-chave: avaliação em larga escala; teoria clássica dos testes; teoria da resposta ao item.

Abstract: This article aims to organize some considerations about the main aspects of TCT (Classical Theory of Tests) and TRI (Item Response Theory) as techniques of large - scale educational evaluation in Brazilian basic education. The method of bibliographic research is used to present the concepts and applications, albeit brief, of these methods. In order to understand TCT, it is necessary to conceptualize psychometry, which is based on the theory of measurement in sciences to explain the meaning of the answers given by the subjects to a series of tasks and to propose measurement techniques of the mental processes. The TRI, also considered a psychometry strand, is specifically interested in each of the items. It can be concluded that the TCT is an evaluation method less complex than TRI, besides requi-

ring a smaller sample size. The TRI is a more detailed method, which refines the analysis of the items and evaluates other more subjective aspects of a test. However, as the two techniques are used together, in studies in the area of Mathematics, the present research criticizes the lack of return of these results in improving the quality of Brazilian basic education.

Key words: large scale educational evaluation; classical Theory of Tests; item Response Theory.

1 Introdução

Discutir o tema educação e sua cientificidade na grande área das ciências humanas é um processo que não pode ser realizado sem levar em consideração todas as etapas que fazem parte da estrutura do sistema educacional, portanto, das políticas públicas. Enfrenta-se, no Brasil e no mundo, o desafio de universalizar a educação, sendo, a avaliação uma importante parte do processo educativo. As avaliações atuam como recurso que visam quantificar o desenvolvimento psicopedagógico do aluno [5, 11]. Neste sentido, busca-se na educação brasileira a conciliação, de um lado, a qualidade e a excelência e, do outro, a prática de valores que contribuam para a democratização da sociedade [9].

Os resultados obtidos em provas de larga escala, ao longo dos anos, no Brasil, expressos apenas por seus escores brutos ou padronizados, têm sido tradicionalmente utilizados nos processos de avaliação e seleção de indivíduos - principal característica da TCT – Teoria Clássica dos Testes [1]. Mas é importante considerar que mesmo com a chegada da TRI – Teoria de Resposta ao Item – quase não se percebem mudanças no cenário educativo da educação básica que sejam reflexos das avaliações como as do Enem (Exame Nacional do Ensino Médio), Prova Brasil e aquelas avaliações criadas nos Estados da federação [11] que possam traduzir em melhoria do sistema educacional, desde os anos iniciais do ensino fundamental.

Uma discussão importante sobre os resultados das provas e sua avaliação pela TRI, sobretudo as do Enem, é que tais resultados dependem do particular conjunto de questões que compõe a prova e dos indivíduos que a fizeram, ou seja, as análises e interpretações estão sempre associadas à prova como um todo e ao grupo de indivíduos. Assim, a comparação entre indivíduos ou grupos de indivíduos somente é possível quando eles são submetidos às mesmas provas ou, pelo menos, ao que se denomina de provas paralelas, quase sempre difíceis de serem construídas. Desta maneira, fica muito difícil fazer comparações quando diferentes indivíduos fazem provas diferentes e, por outro lado, outra dificuldade tem sido a forma de interpretação dos parâmetros da TRI que realmente identifiquem estas particularidades, tanto das provas, quanto dos contextos em que estas provas são aplicadas, considerando a regionalidade brasileira e sua diversidade [1, 7, 12].

A TRI é a Teoria de Resposta ao Item, é uma avaliação que muda o foco de análise da prova como um todo para a análise de cada item. A TRI é um conjunto de modelos matemáticos em que a probabilidade de resposta a um item é modelada como função da proficiência (habilidade) do aluno (variável latente, não observável) e de parâmetros que expressam certas propriedades dos itens. Quanto maior a proficiência do aluno, maior a probabilidade de ele acertar o item [7, 12].

Considerando esses aspectos, o objetivo do presente artigo é realizar algumas considerações sobre os principais aspectos da TCT e da TRI, seus conceitos e aplicações na avaliação educacional brasileira, particularmente, a educação básica, ainda que, de forma

breve.

Como metodologia adotou-se a pesquisa bibliográfica [13] destacando alguns trabalhos que tiveram como foco a TRI e o tema da avaliação em larga escala. A metodologia proposta permitiu, por meio do levantamento bibliográfico, indicar conclusões, propor pesquisas e clarificar a temática, de forma mais ampla.

2 Desenvolvimento

2.1 Teoria Clássica dos Testes – TCT

Para entender a TCT, é preciso conceituar a psicometria que, segundo Pasquali, (2009) se fundamenta na teoria da medida em ciências para explicar o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas e propor técnicas de medida dos processos mentais.

A teoria clássica dos testes (TCT) é uma das vertentes da psicometria moderna, ela se preocupa em explicar o resultado final total, isto é, a soma das respostas dadas a uma série de itens, expressa no denominado escore total (T). O modelo da TCT foi elaborado por Spearman e detalhado por Gulliksen. O modelo é o seguinte:

$$T = V + E$$

sendo,

T = escore bruto ou empírico do sujeito, que é a soma dos pontos obtidos no teste;

V = escore verdadeiro, que seria a magnitude real daquilo que o teste quer medir no sujeito e que seria o próprio T se não houvesse o erro de medida;

E = o erro cometido nesta medida.

Dessa forma, o escore empírico é a soma do escore verdadeiro e do erro. E sendo assim, a grande tarefa da TCT consiste em elaborar estratégias para controlar ou avaliar a magnitude do erro, que pode ser por defeitos do próprio teste ou estereótipos e vieses do sujeito, ou fatores históricos e ambientais aleatórios [10].

Na TCT são avaliadas duas características dos itens: a dificuldade e a discriminação. O índice de dificuldade é dado pela proporção de indivíduos que respondem afirmativamente ao item e varia entre 0 e 1, sendo que caso seu valor seja igual a 0, significa que nenhum indivíduo respondeu ao item afirmativamente. Ao contrário, se o índice for igual a 1, significa que todos responderam afirmativamente. Já o parâmetro de discriminação visa diferenciar o grupo de indivíduos que tiveram alta pontuação total dos que tiveram baixa pontuação total no teste. O item é mais discriminativo quanto maior for o seu valor [11].

Dois fatores podem influenciar a interpretação do índice de dificuldade: o número de respostas omitidas e a probabilidade de um indivíduo responder corretamente ao item pelo acaso. Sendo assim, se um grande número de indivíduos não responder ao item por falta de tempo, o índice de dificuldade não será avaliado corretamente [11].

Segundo Sartes e Sousa-Formingoni (2013), na TCT se basearam a maioria dos métodos operacionais utilizados para avaliar as duas principais propriedades psicométricas dos instrumentos: a validade e a confiabilidade. A validade de um instrumento pode ser definida como sua capacidade de realmente medir aquilo a que se propõe a medir. Já confiabilidade de um teste se refere à reprodutibilidade da medida, ou seja, o grau de concordância entre múltiplas medidas de um mesmo sujeito inter e intraindivíduos.

Porém, a TCT possui algumas limitações, como as avaliações dos testes somente são válidas se a amostra for representativa, ou se o instrumento for utilizado em outra amostra

com características semelhantes; além disso, testes diferentes com índices de dificuldade e discriminação diferentes geram resultados diferentes para os mesmos indivíduos; se o mesmo construto é medido por dois testes diferentes, os resultados não são expressos na mesma escala, impedindo uma comparação direta; pressupõe-se que a variância dos erros de medida de todos os indivíduos é a mesma, entretanto, deve-se considerar que alguns indivíduos realizam o teste de forma mais consistente do que outros [3, 6, 10].

Assim, conclui-se, de maneira bem ampla, que a TCT busca, em última instância, a interpretação final da resposta dada a determinada questão e, desta forma, tomando como ponto de partida a soma dos itens acertados ou errados, obtém-se “a nota” do sujeito.

Na TCT quanto mais extenso fosse o instrumento para medir um construto, ou seja, a questão, a prova como um todo, melhor seriam suas propriedades de validade. Além disso, um maior número de itens contribuía positivamente no cálculo das fórmulas de confiabilidade.

Entretanto, nas últimas décadas, este conceito tem sido desafiado por uma corrente teórica conhecida como Teoria de Resposta ao Item (TRI). Esta corrente propõe o uso de escalas mais curtas que seriam tão confiáveis quanto as longas, se não mais. Esta era, que se iniciou a partir da década de 1980, é conhecida como a era da psicometria moderna, centrada na TRI [11]. Mas a TCT não foi abandonada de todo ou completamente, já que, para os resultados pela TRI, grande parte dos trabalhos aponta comparações com gráficos e scores da TCT.

2.2 Teoria da Resposta ao Item - TRI

A TRI, segundo Pasquali (2009), assim como a TCT também é uma vertente da psicometria, porém, não está interessada no escore total em um teste, pelo contrário, ela se interessa especificamente por cada um dos itens e quer saber qual é a probabilidade e quais são os fatores que afetam esta probabilidade de cada item individualmente ser acertado ou errado (em testes de aptidão) ou de ser aceito ou rejeitado (em testes de preferência: personalidade, interesses, atitudes).

Em Rabelo (2013) encontramos a fórmula geral da TRI, dada como

$$P(X_{ji} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta_j - b_i)}}$$

em que:

- X_{ji} é a resposta do indivíduo j ao item i (igual a 1, se o indivíduo responde corretamente, e igual a 0, caso contrário);
- $a_i > 0$ é o parâmetro de discriminação do item i ;
- b_i é o parâmetro de posição (ou de dificuldade) do item, medido na mesma escala da habilidade;
- $0 < c_i \leq 1$ é o parâmetro da assíntota inferior do item i , refletindo as chances de um estudante de proficiência muito baixa selecionar a opção de resposta correta;
- θ_j representa a habilidade ou traço latente do j -ésimo indivíduo;
- e representa a função exponencial;

- D é um fator de escala, que é igual a 1 na métrica logística e igual a 1,7 na métrica normal.

Ao contrário da TCT em que o interesse é o de produzir testes de qualidade, a TRI se interessa por produzir tarefas (itens) de qualidade. No final, então, temos ou testes válidos (TCT) ou itens válidos (TRI), itens com os quais se poderão construir tantos testes válidos quantos se quiser ou o número que os itens permitirem. Assim, a riqueza na avaliação psicológica ou educacional, dentro do enfoque da TRI, consiste em se conseguir construir armazéns de itens válidos para avaliar os traços latentes, armazéns estes denominados de bancos de itens para a elaboração de um número sem fim de testes [10].

O modelo da TRI parte da suposição de que existe no indivíduo um traço latente, representado pela letra θ (teta), relativo a uma característica do indivíduo como, por exemplo, a gravidade da dependência de drogas, o nível de depressão, ou o quanto ele sabe sobre matemática. O traço latente pode ser entendido, em outros contextos, como o construto. O nível do traço latente do indivíduo (por exemplo: nível de habilidade em matemática, gravidade da depressão etc.) permite estimar se o indivíduo responderá positivamente ou não a um determinado item, estabelecendo para cada item uma probabilidade de resposta relacionada à presença (ou intensidade) da condição avaliada [4].

O primeiro passo para aplicação da TRI é a estimação dos parâmetros dos itens, que pode ser feita por diferentes modelos estatísticos. Um modelo comum aplicado a itens dicotômicos é o modelo logístico de dois parâmetros, o qual avalia as características de dificuldade e discriminação dos itens, entretanto, diversos outros modelos são propostos.

Pela estimação dos parâmetros é possível construir uma escala que será usada para interpretar o problema que está sendo estudado. Existem vários métodos de estimação. Os mais empregados são o Método da Máxima Verossimilhança e Métodos Bayesianos. A aplicação destes métodos de estimação requer a utilização de ferramentas matemáticas bastante complexas que necessitam de recursos computacionais [2].

O Saeb foi o primeiro a aderir a TRI em 1995, mas, a partir de 2009, o novo Enem também passou a utilizá-la. Por este motivo, a TRI pode ser considerada um importante elemento nas discussões sobre o sistema educacional no Brasil [5].

Segundo Araújo et al (2009), a TRI revolucionou a psicometria ao propor modelos para traços latentes e pela sua suposição de invariância. Pois, para obter a medida de um determinado traço latente deve-se caracterizar a natureza do traço latente a ser medido, construir os itens que devem cobrir todo o traço latente, observar o tipo de resposta que é dado ao item para verificar se os itens têm natureza acumulativa ou de desdobramento e, a partir daí, escolher o modelo da TRI mais adequado que se ajuste aos dados. Em seguida, estimar os parâmetros dos itens e dos respondentes e construir e interpretar a escala do traço latente.

Na TRI há uma “engenharia” específica e complexa na construção de itens para avaliação, que os espaços deste artigo não permitem aprofundar, mas é preciso afirmar que existem vários manuais, estrangeiros e nacionais que explicitam sua aplicabilidade nas avaliações em larga escala [11].

A TRI, em comparação TCT, traz alguns avanços como considerar que o indivíduo possui um traço latente “verdadeiro” específico que não irá variar segundo os itens utilizados, desde que os itens estejam medindo o mesmo traço latente; o cálculo dos parâmetros dos itens independe da amostra de sujeitos utilizada; permite ainda posicionar os itens e os sujeitos em uma mesma escala do traço latente que está sendo avaliado; constitui um modelo que não precisa fazer suposições improváveis, tais como a de que os erros de medida são iguais para todos os indivíduos; não exige que se trabalhe com instrumentos estritamente paralelos

para avaliar a confiabilidade; e ainda permite que as respostas que não se aplicam a alguns entrevistados não sejam consideradas idênticas à resposta nula [6].

3 Conclusões

A TCT é um método avaliativo menos complexo que a TRI, além de necessitar de menor tamanho amostral e talvez por isso não tenha caído em desuso. Já a TRI trata de um método mais detalhista, que refina as análises dos itens e avalia os parâmetros de dificuldade, discriminação e acertos ao acaso, possui grau de acurácia mais apurado que atua como complementos aprofundados de informações necessárias à avaliação aplicada. Porém, o fato de a TRI necessitar de programas mais sofisticados a torna, também, mais dispendiosa e menos disponível ao público, sobretudo à pesquisa de avaliações de larga escala, geralmente, possíveis apenas a órgãos institucionais públicos e, geralmente, só há acesso a provas e resultados pelo TRI, muito tempo após a realização das provas.

Geralmente a TCT atua como parâmetro comparativo para os resultados da TRI [11]. Grande parte dos poucos estudos existentes aponta as duas técnicas em conjunto na avaliação das provas, por exemplo do Enem.

De fato, a aplicação em conjunto deve ser vista como técnica complementar, o que possibilita uma avaliação mais completa. Contudo as provas de avaliação em larga escala ainda não possibilitaram, ao longo desses anos, mudanças no formato das provas e possíveis retornos de melhoria da qualidade no que diz respeito, por exemplo, a melhor preparação dos estudantes para estas provas, sobretudo no ensino médio e o acesso ao ensino superior. Exatamente por isso, como se apontou na Introdução deste artigo, a temática da avaliação em larga escala e o método de construção dos itens e sua avaliação, sobretudo, com a engenharia da TRI são decisivas para se discutir educação, sistema educacional e políticas públicas, ou seja, que resultados destas avaliações promovam a melhoria da educação dos brasileiros e brasileiras.

Referências

- [1] ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. da C. Teoria da Resposta ao Item: Conceitos e Aplicações. 164 p. SINAPE 2000.
- [2] ARAUJO, E. A. C. de; ANDRADE, D. F. de; BORTOLOTTI, S. L. V. Teoria da Resposta ao Item. Revista Escola Enfermagem; v. 43, 1000-1008, 2009.
- [3] EMBRETSON, S. E.; REISE, S. P. Item Response Theory for psychologists. Mahwah, NJ: Lawrence Erlbaum (2000).
- [4] FLETCHER, P. A teoria da resposta ao item: Medidas invariantes do desempenho escolar. Ensaio: Avaliação e Políticas Públicas em Educação, v. 1, n.2, p. 21-28, 1994.
- [5] GUIMARÃES, B. C. Teoria de resposta ao item e aplicação no ensino de geografia. Revista de Estudos e Pesquisas em Ensino de Geografia, Florianópolis, v. 1, n. 2, 2014.
- [6] HAMBLETON, R. K.; SLATER, S. C. Item Response Theory Models and Testing Practices: Current international status and future directions. European Journal of Psychological Assessment, v. 13, n.1, p. 21-28, 1997.

- [7] KLEIN, R. Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). Revista ENSAIO, n. 40, v.11, p.283-296, 2003.
- [8] LORDÊLO, J. A. C.; DAZZANI, M. V. Avaliação educacional: desatando e reatando nós [online]. Salvador: EDUFBA, 2009. 349 p. ISBN 978-85-232-0654-3. Available from Scielo Books.
- [9] PASQUALI, L. Psicometria. Revista Escola Enfermagem, v. 43, p. 992-999, 2009.
- [10] PASQUALI, L.; PRIMI, R. Fundamentos da Teoria de Resposta ao Item – TRI. Avaliação Psicológica, v. 2, n.2, p. 99-110, 2003.
- [11] RABELO, M. Avaliação educacional: fundamentos, metodologia e aplicações no contexto brasileiro. Coleção Profmat. Rio de Janeiro: SBM, 2013.
- [12] SARTES, L. M. A; SOUSA-FORMINGONI, M. L. O. D. Advances in psychometrics: from classical test theory. Psicologia: reflexão e crítica, n.26, v 2, p. 241-250, 2013.
- [13] SAMPIERI, R.H.; COLLADO, C.F.; LUCIO, M.P.B. Metodologia de Pesquisa. Penso, 2006.