

APLICAÇÕES DIFERENCIADAS DA TEORIA DE RESPOSTA AO ITEM

Jacob Arie Laros¹

A quantidade de pesquisadores no Brasil que usam a Teoria de Resposta ao Item (TRI) está aumentando. Além disso, também têm aumentado as aplicações diferenciadas da TRI. Nesta mesa redonda as seguintes aplicações serão discutidas: (1) a revisão simultânea de um teste em vários países, (2) avaliação da qualidade psicométrica de um teste (3) equalização de dois testes aplicados em anos diferentes com o intuito de comparar os resultados.

A NECESSIDADE DE ANÁLISE EXPLORATÓRIA DE DADOS ANTES DA UTILIZAÇÃO DA TEORIA DE RESPOSTA AO ITEM.

Josemberg Moura de Andrade²

Este estudo tem como objetivo verificar as implicações da análise exploratória de dados, mais especificamente, análise gráfica dos itens, sobre a estimação dos parâmetros de itens, a partir da Teoria de Resposta ao Item (TRI). A análise exploratória de dados consiste em um conjunto de procedimentos estatísticos realizado com o intuito de revelar o que está contido em um banco de dados. Ainda, através desse tipo de análise é possível verificar a adequação dos dados aos pressupostos requeridos pelas análises estatísticas. Nesse estudo, considerou-se a análise gráfica de itens como técnica exploratória. Tal técnica, utilizada especialmente em testes de aptidão, avalia a qualidade psicométrica dos itens e permite analisar os itens de um teste em função do escore total, verificando se, com o aumento do escore total, também aumenta a probabilidade de resposta correta a um dado item e, inversamente, diminui a probabilidade de resposta errada. Para atingir o objetivo do estudo, foram realizados dois conjuntos de análises: (1) análise gráfica dos itens da prova de Língua Portuguesa do SAEB 2001 para a 3ª série do ensino médio, seguida da estimação dos parâmetros dos itens através da TRI, usando o modelo logístico de 3 parâmetros; e (2) estimação dos parâmetros dos itens sem a análise gráfica prévia dos itens. Foi considerada uma amostra de 36.263 alunos que responderam a prova de Língua Portuguesa da 3ª série do ensino médio do SAEB 2001. Para as análises de dados, foram utilizados os programas AGI-1.1 para a análise gráfica, bem como o BILOG-3.0 para a estimação dos parâmetros dos itens. Em relação aos resultados do primeiro conjunto de análises, a análise gráfica identificou dois itens como problemáticos. Esses itens foram excluídos para a estimação dos parâmetros dos itens a partir da TRI e os resultados foram observados. Em seguida, os parâmetros dos itens foram estimados sem a exclusão dos itens problemáticos. Foram verificadas diferenças em relação ao parâmetro b (dificuldade do item). A média desse parâmetro sem os dois itens identificados como problemáticos foi de 0,12, desvio-padrão de 0,83, com valor mínimo de -1,94 e máximo de 2,20. Por outro lado, a média do parâmetro b com a inclusão dos itens problemáticos foi de 0,56, desvio-padrão de 4,45, com valor mínimo -1,94 e máximo de 53,75. Este último índice é incoerente, já que na prática os valores do parâmetro b variam de -3 a +3. Em relação aos parâmetros a (discriminação) e c

¹ Coordenador. Instituto de Psicologia - Universidade de Brasília. jalaros@unb.br.

² Universidade de Brasília.

(resposta certa dada ao acaso) não foram verificadas diferenças na estimação com itens problemáticos e sem itens problemáticos. Ainda, verificou-se o ajuste dos modelos através da diferenças dos qui-quadrados. O modelo com itens excluídos a partir da análise gráfica prévia mostrou-se mais adequado aos dados. Conclui-se que (1) a análise exploratória de dados deve ser utilizada antes do uso da TRI e (2) a análise gráfica é uma técnica exploratória eficaz para identificar itens problemáticos capazes de distorcer os resultados. Palavras-chave: Análise Exploratória, Análise Gráfica de Itens, Teoria de Resposta ao Item.

O USO DA TEORIA DE RESPOSTA AO ITEM NA REVISÃO DE UM TESTE DE INTELIGÊNCIA NÃO-VERBAL.

Jacob Arie Laros

A última versão do teste não-verbal de inteligência SON (Snijders-Oomen Não-verbal) construído na Holanda em 1990 para a faixa etária de 5½ a 17 anos está em uso por há 15 anos. Sabe-se que os testes precisam ser revisados depois de 15 a 20 anos para incluir normas atuais e mudar o conteúdo dos itens que não são mais adequados. Atualmente, o teste SON-R 5½-17 está em processo de revisão. O novo teste consistirá de 5 subtestes: Categorias, Situações, Analogias, Mosaicos, e Padrões. O presente estudo concentrou-se na pré-testagem dos itens do subteste Categorias. Um aspecto inovador no desenvolvimento dos itens do novo teste é que a sua adequação ao contexto brasileiro será verificada durante o processo de desenvolvimento dos itens e não - como de costume com outros instrumentos psicológicos - depois da validação e normatização. Assim, é pertinente esperar que para o estabelecimento das normas brasileiras do teste, adaptações adicionais para este contexto não precisem ser realizadas. O objetivo do presente estudo foi verificar o grau de concordância entre os valores dos parâmetros psicométricos dos itens estimados com a Teoria de Resposta ao Item (TRI), obtidos com uma amostra de 418 crianças, adolescentes e adultos do Brasil, com os valores obtidos com a TRI baseada em uma amostra de 1.545 crianças, adolescentes e adultos da Indonésia, África e Holanda. O conjunto de 39 itens do subteste Categorias foi aplicado coletivamente. Os dados foram analisados com os programas SPSS 11.5 e BILOG-MG.3. Os parâmetros psicométricos dos itens foram estimados utilizando o modelo logístico da TRI de 3 parâmetros. O parâmetro b (dificuldade do item) variou entre -3,31 e 2,15 com média de 0,52 e desvio-padrão de 1,43. O parâmetro a (discriminação do item) variou entre 0,33 e 1,86 com um valor médio de 0,99 e desvio-padrão de 0,42. Um item do conjunto de 39 itens mostrou um índice de discriminação muito baixo (0,33) e será excluído ou modificado na segunda pré-testagem dos itens. A correlação entre os valores obtidos para o parâmetro b baseado na amostra Brasileira e os valores obtidos baseados na amostra da Indonésia, África e Holanda foi de 0,90. A correlação para o parâmetro a foi de 0,77. Os resultados do atual estudo mostraram um alto grau de concordância entre os países comparados em relação ao parâmetro b, porém, um grau moderado de concordância para o parâmetro a. Estudos futuros precisam esclarecer porque alguns itens do subteste Categorias discriminam melhor no Brasil enquanto outros discriminam melhor na Indonésia, Holanda e África. Palavras-chave: SAEB, Equalização de Testes, TRI.

ANÁLISE DA QUALIDADE PSICOMÉTRICA DE TESTES EDUCACIONAIS UTILIZANDO A TEORIA DE RESPOSTA AO ITEM

Anamara Ferreira Ribeiro³

O objetivo do presente estudo foi avaliar a qualidade psicométrica do teste de Matemática da 4ª série do Ensino Fundamental aplicado pelo Sistema Nacional de Avaliação da Educação Básica (SAEB) em 2001. Considerando que a qualidade de um teste está diretamente relacionada com a qualidade de seus itens, este estudo se concentrou na análise dos itens que compõem o teste em questão. Outra razão, para focalizar o presente estudo na análise dos itens, é o fato de que praticamente todo o tipo de análise estatística que utiliza o escore total de um teste depende da qualidade dos itens. Os itens do teste de Matemática foram avaliados no que se refere às suas propriedades psicométricas, tanto pela Teoria Clássica dos Testes (TCT) como pela Teoria de Resposta ao Item (TRI) utilizando os programas SPSS 11.5 e BILOG-3. Um objetivo específico do presente estudo foi verificar a concordância entre a TCT e a TRI, por meio da correlação entre seus índices de discriminação e dificuldade. A amostra do presente estudo foi composta por 57.220 alunos dos 27 estados brasileiros, os quais responderam ao teste de Matemática. A primeira análise realizada foi a de unidimensionalidade por meio da análise fatorial full-information e também por meio dos índices complementares de unidimensionalidade. Os resultados indicaram que o teste de Matemática pode ser considerado unidimensional. Contudo, cinco itens que não contribuíam para a avaliação do fator único foram excluídos, ficando 164 itens para as análises psicométricas posteriores. Com base nos parâmetros de discriminação e validade dos itens, 135 itens (82%) apresentaram uma boa qualidade e 29 itens (18%) apresentaram uma qualidade razoável. Os resultados mostraram que o teste de Matemática é um teste difícil, com itens discriminativos e válidos. Observou-se um alto grau de concordância entre os parâmetros de dificuldade da TCT e da TRI ($r=-0,96$) e uma correlação baixa entre os parâmetros de discriminação de ambas as teorias ($r = 0,42$). Constatou-se uma melhoria na qualidade do teste de Matemática de 2001 em relação aos testes do SAEB de anos anteriores. Recomenda-se a melhoria dos 29 itens com qualidade razoável e a inclusão de mais itens fáceis no teste de Matemática. Palavras-chave: Teoria de Resposta ao Item (TRI), Análise Psicométrica de Itens, SAEB.

COMPARAÇÃO DE QUATRO MÉTODOS DE EQUALIZAÇÃO DE TESTES BASEADOS NA TEORIA DE RESPOSTA AO ITEM.

Guilherme Coelho Rabello⁴

Avaliação educacional em larga escala demanda com grande intensidade a utilização de técnicas estatísticas para comparação de resultados das aplicações ao longo do tempo, as ditas técnicas de equalização. Particularmente, técnicas baseadas na Teoria de Resposta ao Item (TRI) têm sido aplicadas freqüentemente. Os métodos de equalização estão sendo utilizados regularmente no SAEB, um sistema de avaliação educacional em larga escala levado a efeito pelo Ministério da Educação (MEC) e pelo Instituto Nacional de Estudos e

³ Universidade de Brasília.

⁴ Ministério de Desenvolvimento Social e Combate à Fome.

Pesquisas Educacionais Anísio Teixeira (INEP). Muitos estudos têm focado na comparação de métodos de equalização baseados na TRI, descrevendo e aplicando os mesmos a bases de dados reais. Um interessante estudo foi aplicado a avaliação do SAEB, comparando resultados do SAEB 1995 e SAEB 1997 da 4ª série do Ensino Fundamental na disciplina de Matemática, com o intuito de transformar a escala dos dados de 1997 para a escala dos dados de 1995, tornando assim, os escores dos alunos nos dois anos comparáveis. Neste estudo, foram comparados quatro métodos de equalização baseados na Teoria de Resposta ao Item. Estes métodos são descritos na literatura como: (a) método de Stocking & Lord; (b) método de Haebara; (c) método da Média/Média; (d) método da Média/Desvio. Os dois primeiros métodos são baseados nas Curvas Características dos Itens (CCI), e os dois últimos, têm em comum o fato de usarem estatísticas de resumo das estimativas dos itens comuns das duas aplicações a serem comparadas (médias e desvios-padrão) para realizar a equalização em estatísticas de resumo das estimativas dos parâmetros dos itens. Os métodos foram comparados por meio da análise dos coeficientes de equalização. A raiz dos erros quadráticos médios das diferenças entre os escores transformados foram calculadas para todos os pares de métodos, tornando possível avaliar a congruência entre os escores transformados. Os resultados analisados correspondem a uma amostra de 11.886 estudantes da 4ª Série do Ensino Fundamental, em 1995, e 23.535 estudantes, em 1997. Os testes de Matemática contêm, respectivamente, 130 e 143 itens, sendo aproximadamente 20 deles, comuns. Os itens de Matemática foram arranjados em 13 blocos, e então distribuídos em 26 cadernos de teste (formas), de acordo com o delineamento de Blocos Incompletos Balanceados (BIB) com distribuição em espiral. O Desenho “BIB” permite que cada estudante responda a apenas uma forma de teste, variando de 30 a 33 itens cada. Com este estudo foi possível mostrar que, antes da aplicação de uma equalização, é necessário levar a efeito uma análise cuidadosa dos pressupostos subjacentes aos modelos e técnicas, bem como avaliar a congruência entre os métodos. A título de exemplo, não é sempre possível usar todos os itens comuns no processo de equalização, porque podem existir itens que não estão satisfazendo os pressupostos da TRI. Quando tais itens são excluídos do processo de equalização, os resultados mudaram para todos os métodos. Este estudo mostrou também que os resultados de equalização baseados nos métodos que utilizam as CCI e no método Média/Desvio foram similares entre si e diferentes do método da Média/Média.

Palavras-chave: SAEB, Equalização de Testes, TRI.