

---

# Editorial

## Funcionamento Diferencial de Itens e dos Testes: Teoria de Resposta ao Item ou Equações Estruturais?

DOI: <http://dx.doi.org/10.15689/ap.2019.1802.ed>

Um dos núcleos duros da avaliação psicológica é a possibilidade de comparar pessoas nas suas diferenças, sem o qual a própria psicologia se torna sem sentido. No entanto, essa tarefa esconde algumas armadilhas, entre as quais a justiça (*fairness*), ou melhor, a possibilidade de cometer injustiças nos escores dos examinandos. Por exemplo, um teste de atenção, com estímulos visuais nas cores verde e vermelho, pode subestimar as avaliações de pessoas com daltonismo, pois eventuais erros podem ser atribuídos à dificuldade de reconhecimento de cores, além da capacidade cognitiva-atencional.

Para garantir a justiça nas avaliações, é necessário, entre inúmeros aspectos, que a parametrização dos itens seja equivalente entre os indivíduos, ou que eventuais diferenças dos itens sejam ponderadas nas estimações dos escores dos examinandos. Isso pode ser testado por meio do funcionamento diferencial dos itens (*Differential Item Function* ou DIF), que é uma grande família de modelagens, nas quais os parâmetros dos itens, dos testes ou dos examinandos são comparados entre grupos ou diferentes níveis de uma variável externa. Pode ser avaliado, por exemplo, se a dificuldade dos itens é equivalente para todos os grupos de examinandos.

Desde as primeiras décadas de estudo dos modelos de Teoria de Resposta ao Item (TRI), Frederic Lord (1980) já propôs a ideia de estudar diferenças na parametrização dos itens em função de subgrupos. A ideia era “simples”: considerando que a TRI possibilita a manutenção de uma métrica comum, seria possível estimar os parâmetros dos itens como diferentes entre os grupos e compará-los. Caso as diferenças fossem significativas, concluir-se-ia pelo DIF.

No entanto, para garantir a métrica comum e a identificação do modelo, é necessário fixar alguns parâmetros. Por exemplo, não é possível estimar todos os *thresholds*, discriminação dos itens, médias e variâncias dos grupos simultaneamente como distintos (mesmo por TRI). Aqui, iniciaram-se os problemas: quais parâmetros fixar e quais liberar para serem livremente estimados? Como solução, diversas estratégias metodológicas foram propostas no contexto de Equações Estruturais e de TRI.

Dentro da família dos modelos de Equações Estruturais, a estratégia multigrupos é uma das mais populares. Em linhas gerais, propõem-se um modelo de base (configural) com o maior número possível de parâmetros livres e, em passos progressivos, restringem-se os demais parâmetros como iguais entre os grupos. Para a identificação desse modelo de base, usualmente, fixa-se a média das variáveis latentes de ambos os grupos e uma carga fatorial por fator (ou a variância da variável latente). Ou seja, é necessário fixar alguns parâmetros mesmo para o modelo basal. Outra estratégia é denominada *Multiple Indicators Multiple Causes* (MIMIC). Nesse caso, o viés é testado por meio de efeitos de regressão diretos entre os itens do instrumento e a variável de grupo. Contudo, mais uma vez, não é possível avaliar todas as relações diretas simultaneamente, de modo que uma delas, ao menos, deve ser fixada em 0 (ou qualquer outro valor conhecido). Por fim, é possível, ainda, configurar um modelo de base com o maior número possível de parâmetros livres, e estimar a diferença de parametrização por meio de *bootstrap*. Essa modelagem permite avaliar a invariância considerando todas as possibilidades de itens de referência, sendo tecnicamente equivalente à realização de testes para todas as configurações possíveis, tomando como base diferentes parâmetros fixos (por exemplo, considerando a carga do item 1 fixa na primeira configuração, a carga do item 2 na segunda, e assim por diante). Ainda assim, permanece a necessidade de fixar alguns parâmetros como invariantes entre os grupos (para maiores detalhes sobre essas modelagens no contexto de Equações Estruturais e referências consulte Valentini, Franco, & Iglesias 2018).

Fica claro que, em todas as estratégias descritas, é necessário fixar um conjunto de parâmetros. Isso significa assumir que esses não são enviesados entre os grupos! Ademais, os testes de invariância somente funcionam condicionados à manutenção desse pressuposto. Ou seja, caso o parâmetro fixado não seja, de fato, invariante, os testes dos demais parâmetros ficam comprometidos.

Esse detalhe técnico não é diferente no contexto da TRI. Contudo, as implementações de algoritmos de DIF nos softwares de TRI, normalmente, escolhem automaticamente os parâmetros que serão fixados. A estratégia mais utilizada é configurar um modelo de base restrito (*constrained-baseline*), no qual todos os parâmetros são fixados como iguais entre os grupos e, nos modelos subsequentes, é liberado apenas um parâmetro por vez. Em cada passo, o parâmetro liberado anteriormente volta a ser fixado, e um novo parâmetro é testado como distinto entre os grupos.

---

Então, o modelo com um parâmetro livre é comparado ao de base. Ressalto que, nesse caso, também se assume um pressuposto importante: o teste de DIF para um item assume a invariância dos demais. Outra estratégia em TRI é estimar um modelo de base com o maior número de parâmetros livres (*free-baseline*), à exceção de alguns para a identificação do modelo e da métrica; e, nos passos subsequentes, fixar um parâmetro por vez. Essa estratégia apresenta lógica inversa à primeira. Porém, continua-se com a necessidade de fixar algum parâmetro de algum item como invariante em todos os passos (Drasgow, Nye, Stark, & Chernyshenko, 2018).

Em resumo, em todos os casos, assume-se que algum parâmetro é invariante, tanto por TRI quanto por Equações Estruturais. As estratégias no contexto de Equações Estruturais são levemente mais flexíveis, dando ao pesquisador a possibilidade de decidir qual parâmetro especificamente é mantido fixo no processo. No entanto, essa leve vantagem não se traduz, necessariamente, em maior eficiência do método. Por meio de dados simulados, Stark, Chernyshenko e Drasgow (2006) indicou que o teste de DIF por meio de Equações Estruturais apresenta menos falsos positivos do que por TRI para itens politômicos (principalmente com 5 categorias de respostas ou mais). Porém, a situação se inverte para itens dicotômicos (i.e. TRI apresenta resultados superiores para esse tipo de item).

Portanto, do ponto de vista prático, não parece haver grande diferença. Contudo, ao realizar esse tipo de análise é sempre importante estar ciente que: 1. Algum parâmetro do item será mantido fixo; 2. existe a possibilidade real de ocorrência, principalmente, de falsos positivos, ou seja, indicar DIF, quando, na realidade, o parâmetro é invariante entre os grupos. A sugestão desse editorial é investir em replicação com diferentes amostras e diferentes estratégias de análise.

Prof. Dr. Felipe Valentini  
Editor Associado  
Universidade São Francisco

---

### Referências

- Drasgow, F., Nye, C. D., Stark, S., & Chernyshenko, O. S. (2018). Differential Item and Test Functioning. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 885-899). Manchester, UK: Wiley-Blackwell.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292-1306. doi: 10.1037/0021-9010.91.6.1292
- Valentini, F., Franco, V., & Iglesias, F. (2018). Introdução à análise de invariância: influência de variáveis categóricas e intervalares na parametrização dos itens. In B. F. Damásio & J. C. Borsa (Eds.), *Manual de Desenvolvimento de Instrumentos Psicológicos* (pp. 347-374). São Paulo, SP: Vetor Editora.