



Figueiredo Damásio, Bruno
Contribuições da Análise Fatorial Confirmatória Multigrupo (AFCMG) na avaliação de
invariância de instrumentos psicométricos
Psico-USF, vol. 18, núm. 2, agosto, 2013, pp. 211-220
Universidade São Francisco
São Paulo, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=401036095005>

Contribuições da Análise Fatorial Confirmatória Multigrupo (AFCMG) na avaliação de invariância de instrumentos psicométricos

Bruno Figueiredo Damásio – Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

Resumo

A Análise Fatorial Confirmatória Multigrupo (AFCMG) é uma técnica da modelagem de equações estruturais que avalia em que medida a configuração e os parâmetros de determinado instrumento psicométrico são invariantes (equivalentes) para diferentes grupos. Tal técnica tem se apresentado como um importante recurso no desenvolvimento, no uso, na avaliação e no refinamento de instrumentos psicométricos. Entretanto, no Brasil há enorme escassez de publicações sobre o tema. Assim, o presente estudo discute a AFCMG, apresentando suas potencialidades. Ao longo do artigo três grandes tópicos são abordados: os diferentes modelos a serem testados, a invariância completa *versus* a invariância parcial, e os métodos de avaliação de invariância dos modelos. Espera-se que esta leitura auxilie pesquisadores na compreensão, na interpretação e no uso da AFCMG.

Palavras-chave: Análise Fatorial Confirmatória Multigrupo, Invariância de medida, Equivalência de medida, Psicometria, Modelagem de equações estruturais.

Contributions of the Multigroup Confirmatory Factor Analysis in the invariance evaluation of psychometric tests

Abstract

The Multigroup Confirmatory Factor Analysis (MGCFA) is a structural equation modeling technique which evaluates to what extent the configuration and the parameters of a psychometric instrument are invariant (equivalent) in several groups. This technique is being presented as an important resource in the development, use, evaluation and refinement of psychometric instruments. However, in Brazil, there is a significant lack of publications considering the topic. Thus, the present study discusses the MGCFA technique, presenting its potentialities. Throughout the article, three major topics are covered: the different models to be tested, the issue of complete and partial measurement invariance, and the measurement invariance evaluation. It is expected that this article assist researchers in the comprehension, interpretation and use of the MGCFA.

Keywords: Multigroup Confirmatory Factor Analysis, Measurement invariance, Measurement equivalence, Psychometrics, Structural equation modeling.

Contribuciones del Análisis Factorial Confirmatorio Multigrupo (AFCMG) en la evaluación de invariancia de instrumentos psicométricos

Resumen

El Análisis Factorial Confirmatorio Multigrupo (AFCMG) es una técnica de modelaje de ecuaciones estructurales que evalúa hasta qué punto la configuración y los parámetros de determinado instrumento psicométrico son invariantes (equivalentes) para diferentes grupos. Esta técnica ha sido presentada como un importante recurso en el desarrollo, uso, evaluación y refinamiento de los instrumentos psicométricos. Sin embargo, en Brasil hay una gran escasez de publicaciones sobre el tema. Así, este estudio discute el AFCMG, presentando su potencial. A lo largo del artículo, tres grandes temas son presentados: los diferentes modelos a ser evaluados, la invariancia completa versus la invariancia parcial, y la evaluación de invariancia de los modelos. Se espera que esta lectura contribuya a los investigadores en la comprensión, la interpretación y el uso de la AFCMG.

Palabras-clave: Análisis Factorial Confirmatorio Multigrupo, Invariância de medida, Equivalência de medida, Psicometría, Modelado de ecuaciones estructurales.

A análise fatorial (AF) é, para a Psicologia, uma das principais ferramentas no desenvolvimento, na avaliação, no refinamento e no uso de testes psicométricos. A AF refere-se a um conjunto de técnicas estatísticas que tem por objetivo definir a estrutura subjacente em uma matriz de dados (Hair, Anderson, Tatham & Black, 2005), avaliando de que maneira um determinado número de itens pode ser agrupado em um número menor de variáveis latentes (fatores) que expliquem as suas inter-relações (Jöreskog, 2007).

As técnicas de análises fatoriais podem ser divididas em duas grandes vertentes: análise fatorial exploratória (AFE) e análise fatorial confirmatória

(AFC). A primeira tem por objetivo investigar como um conjunto de itens se agrupa, sendo, portanto, uma técnica exploratória, visto que o pesquisador não estipula, *a priori*, a estrutura dos dados. Já a AFC é entendida como uma técnica confirmatória, uma vez que o pesquisador precisa predeterminar a estrutura a ser avaliada. Refere-se a um tipo de modelagem de equações estruturais (MEE) voltada, especificamente, a modelos de mensuração, ou seja, modelos que avaliam a relação entre variáveis observadas e variáveis latentes (Millsap & Meredith, 2007). Diferente das AFEs, ao conduzir uma AFC, o pesquisador deve apresentar uma estrutura fatorial já delineada, a qual expõe claramente o número de fatores e os itens específicos

correspondentes a cada fator. Nesse sentido, a AFC é um método teórico e/ou empiricamente baseado (Brown, 2006). Em muitos casos, ambas as técnicas, AFE e AFC, podem ser utilizadas de maneira complementar, com vistas a avaliar a plausibilidade de determinada estrutura fatorial (Schmitt, 2011).

Desde o seu desenvolvimento, o uso da AFC tem crescido de maneira exponencial. Isso porque tal técnica oferece diversos recursos de análise de adequação da estrutura fatorial não existentes na AFE (Brown, 2006). Dentre outras, uma das vantagens da AFC é a possibilidade de avaliar a invariância da estrutura e dos parâmetros de determinado instrumento em diversos grupos, simultaneamente. Esta técnica, denominada Análise Fatorial Confirmatória Multigrupo (AFCMG), vem sendo amplamente empregada em estudos de validação de instrumentos psicométricos e em estudos de comparação de grupos (Sass, 2011). Apesar de ser uma técnica amplamente utilizada em estudos internacionais, no Brasil há uma enorme escassez de publicações que tenham considerado esse tipo de análise. Assim, o objetivo deste artigo é o de apresentar a técnica da AFCMG e discutir suas potencialidades na avaliação de invariância de instrumentos psicométricos.

Análise Fatorial Confirmatória Multigrupo (AFCMG)

Conforme mencionado, a AFCMG é uma técnica que avalia em que medida a configuração e os parâmetros de determinado instrumento psicométrico são invariantes, ou equivalentes, para diferentes grupos de pessoas. Na literatura internacional, os termos invariância de medida (*measurement invariance*, em inglês) e equivalência de medida (*measurement equivalence*, em inglês) são utilizados indistintamente. Tais termos são sinônimos e não apresentam nenhuma distinção, seja psicométrica, seja interpretativa. Para fins didáticos, a primeira forma (invariância de medida) será adotada ao longo deste artigo, visto que se refere à nomenclatura tradicional, utilizada no início do desenvolvimento da técnica (Jöreskog, 1971; Meredith, 1964).

A avaliação do pressuposto de invariância de estruturas e de parâmetros, por meio da AFCMG, responde a algumas questões relevantes, como por exemplo: 1) a estrutura fatorial de determinado instrumento é igual entre diferentes grupos (mesmos itens avaliando o mesmo construto)?; 2) os itens que compõem determinado fator apresentam a mesma importância (mesma carga fatorial) para diferentes grupos?; 3) os escores de um grupo podem ser comparados com os escores de outros grupos ou existem vieses de resposta que impossibilitam essa comparação?; 4) os itens apresentam erros de medida

semelhantes para os diferentes grupos?; 5) o nível de variância do(s) fator(es) difere entre os grupos?; 6) a covariância entre os fatores é a mesma entre os grupos?

Avaliar a invariância de medida é um aspecto crucial para uma série de conclusões no desenvolvimento e no uso de instrumentos psicométricos (Cheung & Rensvold, 2002). A não ser que seja rigorosamente testada, os pesquisadores não podem afirmar que a configuração e os parâmetros de determinado instrumento são semelhantes em diferentes populações (Vandenberg & Lance, 2000). Assim, a avaliação da invariância de medida, por meio da AFCMG, tem sido utilizada em diferentes contextos, tais como, validação de instrumentos em estudos transculturais (e.g., Murayama, Zhou & Nesbit, 2009), validação de instrumentos para diferentes subgrupos de uma mesma amostra (e.g., jovens e adultos; Damásio, Borsa & da Silva, 2011), estabilidade da medida em estudos longitudinais (Spurk, Abele & Volmer, 2011); avaliação sobre a invariância do instrumento para diferentes métodos de coleta de dados (e.g., coleta virtual *versus* coleta presencial, Davidov & Depner, 2009), etc. Além disso, a avaliação de invariância de medida é um pré-requisito para qualquer estudo que tenha por objetivo avaliar diferenças entre grupos (e.g., Damásio e cols., 2011). Isso porque, se o instrumento estiver enviesado para um ou outro grupo, quaisquer diferenças encontradas entre os grupos podem estar relacionadas não às diferenças reais nos níveis de traço latente dos sujeitos, mas sim a parâmetros não equivalentes do instrumento utilizado (Nimon & Reio Jr., 2011; Sass, 2011).

A condução de uma AFCMG envolve a avaliação de diversos parâmetros. Estes são subdivididos em duas grandes categorias: 1) invariância de medida (*measurement invariance*), que engloba a invariância configural, invariância escalar e invariância residual, e 2) invariância estrutural ou heterogeneidade populacional (*structural invariance* ou *populational heterogeneity*), que engloba a invariância da (co)variância da(s) variável(is) latente(s) e a invariância da média da variável latente. A ordem dos parâmetros avaliados segue uma hierarquia de complexidade, de maneira que um modelo mais complexo só é avaliado caso o modelo prévio tenha apresentado invariância (Brown, 2006; Byrne, 2010). A Tabela 1 apresenta os parâmetros que podem ser avaliados em uma AFCMG. Cada um deles será discutido separadamente. Entretanto, é importante salientar que nem todos os parâmetros precisam ser necessariamente testados. A escolha sobre quais parâmetros avaliar dependerá dos objetivos da pesquisa (Schmitt & Kuljanin, 2008).

Tabela 1. Modelos de invariância e seus respectivos parâmetros avaliados

Modelo	Nome(s) do modelo	Parâmetros avaliados
Invariância de medida	Modelo 1	Invariância configural; Igualdade de forma; Modelo irrestrito
	Modelo 2	Invariância métrica; Igualdade de cargas fatoriais
	Modelo 3	Invariância escalar; Igualdade de interceptos
Heterogeneidade populacional	Modelo 4	Invariância residual dos itens; Invariância factorial estrita
	Modelo 5	Invariância estrutural; Igualdade de (co)variância da(s) variável(is) latente(s)
	Modelo 6	Invariância da média latente; Igualdade da média latente

Modelo 1: Teste de Invariância Configural

O primeiro modelo a ser testado em uma AFCMG é denominado invariância configural (*configural invariance*), também chamado de modelo irrestrito (*unconstrained model*) ou igualdade de forma (*equal form*). Este modelo avalia em que medida a estrutura de um determinado instrumento é plausível para todos os grupos que estão sendo analisados. Considerando que a AFC é, consequentemente, a AFCMG é um tipo de método confirmatório, do qual o pesquisador deve estipular, *a priori*, a estrutura fatorial, este modelo avalia se é possível afirmar que a estrutura proposta – mesmos itens sendo explicados pela(s) mesma(s) variável(is) latente(s) – se mantém para os diferentes grupos avaliados. Neste modelo, nenhum parâmetro de medida é estipulado para ser invariante entre os grupos (por isso, é chamado de modelo irrestrito).

Caso o pressuposto de invariância configural não seja acatado, isso significa que os itens carregam em diferentes fatores nos diferentes grupos, ou que os grupos produzem diferentes números de fatores, ou, ainda, ambas as hipóteses (Cheung & Rensvold, 1999). Nesses casos, testes subsequentes de invariância de medida não devem ser conduzidos, tampouco estudos de comparação de grupos devem ser realizados, visto que a estrutura inicial do instrumento não pode ser considerada equivalente para os diferentes grupos.

Modelo 2: Teste de Invariância Métrica

O segundo modelo a ser testado em uma AFCMG refere-se à invariância métrica (*metric invariance*), também chamado de igualdade de cargas fatoriais (*equal factor loadings*). Neste modelo, o

pesquisador investiga em que medida os pesos de regressão dos itens (cargas fatoriais) são equivalentes para os grupos avaliados.

No teste de invariância métrica estipula-se que as cargas fatoriais de todos os itens do instrumento devam ser estatisticamente iguais para todos os grupos analisados (Cheung & Rensvold, 1999). Se este pressuposto é aceito, isso significa que os itens apresentam relações semelhantes com o construto subjacente para todos os grupos (Brown, 2006). Em outras palavras, é possível afirmar que os itens apresentam a mesma importância para o construto, independente do grupo (Byrne, 2010), ou ainda, que os sujeitos de diferentes grupos respondem aos itens de maneira semelhante, não havendo vieses de resposta a um ou mais itens (Sass, 2011).

Nos casos em que o pressuposto da invariância métrica não pode ser acatado, isso significa que um ou mais itens do instrumento está sendo respondido de maneira enviesada por um ou outro grupo. Em consequência, inferências sobre as diferenças entre os grupos podem estar equivocadas. Qualquer teste de comparação de grupos conduzido (e.g., teste T, ANOVA, MANOVA, Mann-Whitney, Kruskal-Wallis) poderá estar enviesado, visto que os escores dos sujeitos, ou seus postos, para os testes não-paramétricos, estarão influenciados por respostas diferenciadas para um ou outro grupo. Assim, é importante que o pesquisador avalie quais e quantos são os itens que apresentam funcionamento diferencial, para que sejam adotados métodos adequados de equiparação dos grupos (Byrne, Shavelson & Muthén, 1989; Sass, 2011). Tais procedimentos de avaliação e

equiparação serão abordados adiante, no tópico “Invariância completa *versus* Invariância parcial”.

Apesar de o pressuposto da invariância métrica apresentar indicações sobre vieses de resposta aos itens, este não é suficiente em garantir a comparação entre os grupos. Para tanto, é necessário que seja satisfeita, também, o pressuposto de invariância escalar.

Modelo 3: Teste de Invariância Escalar

O teste de invariância escalar, ou invariância de interceptos, é outro aspecto necessário para assegurar, adequadamente, a comparação de escores entre grupos. Garante que os escores obtidos estão totalmente relacionados com o nível de traço latente dos sujeitos, independente do seu grupo (Milfont & Fischer, 2010). Tal teste é realizado ao estipular que os interceptos dos itens são equivalentes para os diferentes grupos.

Para compreender o teste de invariância escalar é necessário entender o conceito de intercepto. Para isso, é interessante se remeter a algumas equações de regressão simples, visto que as análises de equações estruturais são, basicamente, equações de regressão.

A equação de regressão linear simples postula que: $y = \alpha + \beta x + \epsilon$, onde: y se refere ao escore do sujeito na variável dependente (ou seja, o escore no fator); α se refere a uma constante, denominada intercepto que é o valor de y quando $x=0$. O intercepto também pode ser definido como o valor inicial do indivíduo na variável y ; β refere-se à outra constante, que significa a taxa de crescimento em y para cada valor de x , sendo representado graficamente pelo declive da linha de regressão; por fim, ϵ se refere ao resíduo da equação, ou seja, os erros da medição, que seria uma parcela do valor de y a qual não pode ser explicada pelas variáveis presentes no modelo da equação.

Quando o objetivo é comparar as médias de diferentes grupos, a equação é assumida da seguinte maneira: $\mu y = \alpha + \beta \mu x$, onde: μy se refere à média da variável y , ou seja, a média no fator; μx , se refere à média dos escores na variável x . Nesses casos, o resíduo (ϵ) não é considerado, pois se assume que sua média é igual à zero (Byrne, 2010; Meredith, 1993).

Para se assumir que as médias dos diferentes grupos são equivalentes ($\mu y_1 = \mu y_2 = \dots = \mu y_n$) é preciso assumir, também, que o restante da equação seja invariante, ou seja: $(\alpha_1 + \beta \mu x_1) = (\alpha_2 + \beta \mu x_2) = \dots = (\alpha_n + \beta \mu x_n)$, assim, tanto a média das respostas ao item para diferentes grupos ($\mu x_1, \mu x_2, \dots, \mu x_n$, avaliado pelo pressuposto da invariância métrica) quanto o intercepto do item para os diferentes grupos ($\alpha_1, \alpha_2, \dots, \alpha_n$, avaliado pelo pressuposto da invariância escalar) necessitam ser equivalentes para que seja possível

assumir que $\mu y_1 = \mu y_2 = \dots = \mu y_n$. Em termos literais, testes de diferenças de médias só devem ser conduzidos quando se estabelece a invariância configural, métrica e escalar (Brown, 2006; Byrne, 2010; Sass, 2011). Na ausência desses pressupostos, quaisquer diferenças encontradas entre os grupos podem estar relacionadas não a diferenças reais nos níveis do traço latente dos sujeitos, mas à não-equivalência nos parâmetros do instrumento (Chen, 2008).

Modelo 4: Teste de Invariância Residual dos Itens

O teste de invariância residual dos itens avalia em que medida os erros (resíduos) dos itens são iguais para os diferentes grupos. O resíduo dos itens se refere à porção do item que não está associada à variância da variável latente. Assim, avaliar a invariância deste parâmetro refere-se a observar, em diferentes grupos, em que medida o item do instrumento está associado ao construto latente assumindo o mesmo grau de erro (Cheung & Rensvold, 2002).

Não há um consenso sobre a necessidade de avaliar a invariância residual dos itens (Wu, Li, & Zumbo, 2007). A avaliação deste parâmetro é, em geral, considerada como um procedimento opcional, visto que ele é extremamente restritivo e dificilmente alcançado nas pesquisas em Psicologia (Brown, 2006; Byrne, 2010; Harrington, 2009; Vandenberg & Lance, 2000). Diversos argumentos endossam esta perspectiva. Por exemplo, é possível argumentar que diferenças residuais dos itens entre grupos referem-se mais a diferenças na confiabilidade das variáveis observadas do que propriamente uma violação da invariância de medida do construto (Lubke & Dolan, 2003; Wu e cols., 2007). Também pode ser considerado que os sujeitos de diferentes grupos respondem de maneira inconsistente ao instrumento pela não familiaridade com a escala e com o método de pontuação (Mullen, 1995), ou, ainda, que dificuldades em relação ao vocabulário, idioma, expressões ou diferenças culturais não-específicas podem produzir resíduos diferenciados entre os grupos (Malpass, 1977). Em síntese, para os que defendem a não-necessidade de avaliação de invariância residual dos itens, argumenta-se que as diferenças residuais encontradas devem-se, em geral, a variações aleatórias, não relacionadas diretamente com o construto.

Entretanto, outras perspectivas postulam que a invariância residual dos itens deve ser estabelecida para que a medida seja considerada, de fato, invariante (DeShon, 2004). Esse autor argumenta que, ao contrário do que é amplamente aceito, a variância de erro dos itens, em geral, não é um processo aleatório, mas sim, reflete uma variância sistemática que deve ser adequadamente avaliada. Nesta perspectiva, os resíduos

dos itens são considerados erros de medida oriundos de diversas variáveis externas não mensuradas, que interferem nas respostas dos sujeitos (Cronbach, 1947). Essas variáveis podem ser aleatórias, conforme explicitado anteriormente, mas também podem ser variáveis sistemáticas, como por exemplo, itens que deveriam estar carregando em outro fator não explicitado no modelo, gerando resíduos e respostas não acuradas (DeShon, 2004). Independentemente de o erro ser aleatório ou sistemático, a medida perde confiabilidade com o aumento dos resíduos (Cronbach, 1947; DeShon, 2004). Entretanto, erros sistemáticos tendem a ser mais prejudiciais do que os aleatórios.

Os erros sistemáticos, em geral, tendem a se correlacionar fortemente entre si, ou tendem a se correlacionar com outros fatores do modelo testado. Ambas as características dão indícios de que a configuração do modelo poderia ser mais bem estipulada (Wu e cols., 2007). Assim, quando o pressuposto da invariância residual não pode ser estabelecido, é importante avaliar em que medida os resíduos são aleatórios ou sistemáticos. Para melhor compreender esse procedimento de avaliação, sugere-se a leitura de DeShon (2004) e de Wu e cols. (2007). Dependendo dos resultados obtidos, pode-se concluir sobre a confiabilidade dos itens do instrumento para os diferentes grupos. A implicação do não-estabelecimento de invariância residual dos itens dependerá do tipo de resíduo (aleatório ou sistemático), bem como dos objetivos do estudo e da finalidade da aplicação do instrumento (DeShon, 2004).

Modelo 5: Teste de Invariância Estrutural

O teste da invariância estrutural ou invariância da (co)variância da(s) variável(is) latente(s) se refere a um modelo de invariância cuja interpretação dependerá da configuração do modelo a ser testado (Byrne e cols., 1989). Em síntese, a estrutura fatorial de um instrumento pode ser unidimensional, quando o instrumento é composto por um único fator, ou multidimensional, quando o instrumento é composto por dois ou mais fatores.

Quando o instrumento é unidimensional, o teste de invariância estrutural avalia em que medida o nível de variância da variável latente é o mesmo para os diferentes grupos. Em outras palavras, avalia em que medida o nível de variabilidade do construto é equivalente para os sujeitos de diferentes grupos (Brown, 2006). Nos casos em que o instrumento é multidimensional, o teste de invariância estrutural avalia em que medida os níveis das correlações (covariâncias) das variáveis latentes são semelhantes para os diferentes grupos. Assim, é investigado em que medida o construto subjacente ao instrumento

apresenta uma mesma estrutura empírica para os diferentes grupos (Byrne, 2010).

Como pode ser percebido, a interpretação de cada um dos modelos é distinta, trazendo, portanto, diferentes consequências. Em geral, é compreendido que a avaliação da invariância da covariância das variáveis latentes apresenta implicações mais significativas quando comparada à invariância da variância da variável latente (Brown, 2006; Byrne, 2010). Quando o pressuposto da invariância estrutural não é alcançado, entende-se que os grupos estudados apresentam diferentes níveis de variabilidade no traço latente investigado (nos casos de instrumentos unidimensionais) ou os níveis de correlação dos fatores não são os mesmos para os diferentes grupos (nos casos de instrumentos multidimensionais).

É importante salientar que o teste da invariância estrutural não avalia a variabilidade nos parâmetros psicométricos do instrumento. Quaisquer diferenças encontradas aqui se referem à homogeneidade ou heterogeneidade populacional (ver Tabela 1), não estando relacionadas a falhas ou vieses da medida (Brown, 2006). Desse modo, estabelecer a invariância da variância e covariância das variáveis latentes não é um pré-requisito para avaliar diferenças de médias entre os grupos (Schmitt & Kuljanin, 2008).

Modelo 6: Teste de Invariância da Média Latente

O teste de invariância da média latente avalia se os sujeitos de diferentes grupos apresentam igualdade de médias nos escores dos fatores. Este teste de invariância é similar aos testes paramétricos de diferença de média (e.g., teste T, ANOVA), no qual é avaliada a hipótese nula de que as médias dos sujeitos de diferentes grupos sejam iguais (Vandenberg & Lance, 2000). Quando a hipótese nula é rejeitada (ou seja, quando os grupos apresentam diferenças de média), é necessário conduzir testes subsequentes para avaliar as diferenças específicas entre os grupos, especialmente quando se trata de uma AFCMG com mais de dois grupos. Assim como os testes *post hoc* das análises tradicionais, os níveis de α precisam ser ajustados com vistas a prevenir a inflação das taxas de erro Tipo I (Vandenberg & Lance, 2000).

A condução do teste de invariância da média latente por meio da AFCMG apresenta vantagens em relação aos testes multivariados comuns (e.g., teste T, ANOVA). Isso porque os testes tradicionais raramente levam em consideração os pressupostos de invariância, não avaliando se os grupos são, de fato, equivalentes (Vandenberg & Lance, 2000). Em contrapartida, para a AFCMG, a condução do teste de invariância da média latente só deve ser realizada quando os pressupostos de invariância configural, métrica e escalar já foram

satisfeitos. Assim, quaisquer similaridades ou diferenças obtidas mediante o teste de invariância da média latente são diferenças reais entre os grupos, visto que quaisquer problemas na medida já foram descartados anteriormente.

Invariância completa versus Invariância parcial

Durante a condução de uma AFCMG, o pesquisador pode se deparar tanto com uma invariância de medida completa (IMC) quanto com uma invariância de medida parcial (IMP, Byrne e cols., 1989). A IMC ocorre quando todos os parâmetros do instrumento (configuração, cargas fatoriais, interceptos, e resíduos dos itens) são invariáveis para todos os grupos avaliados. Já a IMP pode ocorrer em duas situações: 1) quando apenas alguns parâmetros do instrumento são invariantes para todos os grupos, ou 2) quando todos os parâmetros do instrumento são invariantes para alguns, mas não todos os grupos (Milfont & Fischer, 2010; Vandenberg & Lance, 2000). Por exemplo, em um instrumento de dez itens, é possível que o pesquisador encontre que as cargas fatoriais de oito itens estão operando de maneira equivalente entre os grupos, enquanto dois itens específicos estão operando de maneira não-similar. Nesse caso, denomina-se uma invariância métrica parcial. O mesmo pode acontecer para os interceptos (invariância escalar parcial), para os resíduos (invariância residual parcial), e para a (co)variância das variáveis latentes (invariância estrutural parcial).

O pressuposto da IMP não compromete totalmente a comparação entre os grupos (Byrne, 2010; Byrne e cols., 1989). Diversas opções podem ser acatadas pelo pesquisador para lidar com a IMP, tais como: 1) excluir os itens não-equivalentes e usar apenas os itens invariantes nas análises estatísticas; 2) utilizar modelos de invariância parcial (ver Byrne e cols., 1989; Millsap & Kwok, 2004), liberando (*unconstraining*) os parâmetros não-equivalentes para que eles assumam valores individuais para cada um dos grupos, deixando fixos apenas os parâmetros que se mostraram invariantes; ou 3) utilizar todos os itens, sem a liberação de nenhum parâmetro, mesmo sabendo que alguns deles não são equivalentes (Sass, 2011).

A primeira opção pode ser acatada, principalmente, quando se trata de um instrumento longo, cuja remoção dos itens não irá afetar o construto mensurado, tampouco as propriedades psicométricas do instrumento. Entretanto, quaisquer modificações adotadas devem ser teoricamente baseadas (Nimon & Reio Jr., 2011). A segunda ou a terceira opção deve ser acatada apenas quando um pequeno número de itens não seja equivalente. A lógica

por trás dessa afirmação é a de que se os parâmetros não-equivalentes são a minoria, a média final dos grupos não será drasticamente impactada e não afetará os resultados (Sass, 2011). Exemplos práticos da avaliação de invariância parcial podem ser encontrados em Tuokko e cols. (2010) e Rossi, Elkli e Simonsen (2010).

Nos casos em que a IMP é muito severa, ou seja, quando diversos itens, ou outros parâmetros, não são equivalentes entre os grupos, o pesquisador pode ou analisar os escores separadamente e não comparar os grupos ou, ainda, decidir não utilizar o instrumento (Sass, 2011). É importante salientar, porém, que a comparação ou não-comparação de grupos em casos de IMP não é consensual na literatura, havendo, atualmente, um amplo debate sobre qual o melhor procedimento a ser adotado quando o instrumento não se mostra equivalente entre os grupos. Diversos pesquisadores (e.g., Millsap & Kwok, 2004; Sass, 2011) têm salientado a importância de se estabelecer critérios mais claros e precisos, que auxiliem os pesquisadores nas decisões a serem tomadas ao se deparar com IMP.

Identificando a (in)variância entre os parâmetros

Conforme explicitado anteriormente, os modelos apresentados em uma AFCMG são, em geral, avaliados de maneira hierárquica. Desse modo, um modelo é comparado a um modelo prévio menos restritivo (Schmitt & Kuljanin, 2008). Na medida em que o pesquisador estabelece a invariância de um parâmetro, este pode “dar um passo adiante” e avaliar a invariância do modelo seguinte. Entretanto, como se avalia a invariância dos modelos?

O início da avaliação de uma AFCMG dá-se a partir da avaliação da invariância configural (Modelo 1). A invariância configural é o modelo inicial, que postula sobre a plausibilidade da configuração da medida para os grupos de interesse (i.e., número de fatores e itens por fator). Por ser o modelo inicial, este não pode ser comparado a nenhum modelo prévio. Dessa maneira, apenas este modelo deve ser avaliado com os índices de adequação de ajuste padrões à AFC (e.g., razão qui-quadrado por graus de liberdade, $\chi^2/g.l.$; índice de ajuste comparativo, CFI; índice de Tucker-Lewis, TLI; índice de ajuste não-normalizado, NNFI; raiz do erro quadrático médio de aproximação, RMSEA; raiz quadrada média residual padronizada, SRMR, etc.)¹. A

¹ Alguns artigos de revisão podem ser consultados para auxiliar os pesquisadores na escolha de índices de ajuste adequados (Hooper, Coughlan, & Mullen, 2008; Marsh, Balla, & McDonald, 1988), embora os índices acima mencionados sejam amplamente aceitos (Schreiber, Stage, King, Nora, & Barlow, 2006).

plausibilidade desse modelo, indicada pela adequação dos índices de ajuste, sugere que a configuração proposta ao instrumento é plausível para todos os grupos de interesse.

Para todos os modelos subsequentes (modelo 2, invariância métrica; modelo 3, invariância escalar; modelo 4, invariância residual dos itens; modelo 5, invariância estrutural; e modelo 6, invariância da média latente), o pesquisador deve avaliar em que medida a imposição de igualdade de parâmetros para os diferentes grupos reflete em uma piora nos índices de ajuste, quando comparado ao modelo anterior (Vandenberg & Lance, 2000). Entretanto, apenas alguns índices de adequação de ajuste servem a este propósito (Meade, Johnson & Braddy, 2008).

A forma mais tradicionalmente utilizada para avaliar a invariância entre os modelos refere-se ao teste de diferença do qui-quadrado ($\Delta\chi^2$). O valor do $\Delta\chi^2$ refere-se ao valor do χ^2 do modelo avaliado diminuído do valor do χ^2 do modelo prévio. Diferença estatisticamente significativa ($\Delta\chi^2 \leq 0,05$) entre os valores do χ^2 de um modelo, em comparação com o modelo anterior, sugere que o pressuposto de invariância não foi aceito e que o parâmetro avaliado opera diferentemente entre um ou mais grupos (Brown, 2006; Steiger, Shapiro & Browne, 1985). A interpretação sobre diferenças significativas nos valores do $\Delta\chi^2$ devem considerar os valores de referência em relação à diferença nos graus de liberdade (Δgl) do modelo testado em relação ao modelo prévio.

Apesar de o teste $\Delta\chi^2$ ser amplamente utilizado como ferramenta para avaliar invariância dos modelos, alguns pesquisadores têm demonstrado que as diferenças nos valores do χ^2 são sensíveis a pequenas diferenças nas matrizes de covariância entre os grupos, sendo também influenciados pelo tamanho amostral (Brannick, 1995; Cheung & Rensvold, 2002; Hu & Bentler, 1993; Kelloway, 1995). Assim, sugere-se que, quando utilizado, o teste $\Delta\chi^2$ seja interpretado em conjunto com outros indicadores de invariância.

A utilidade de outros índices de adequação de ajuste em avaliar a invariância de parâmetros e estruturas na AFCMG, foi inicialmente testada por Cheung e Rensvold (2002). Em um estudo de simulação Monte-Carlo, esses autores avaliaram o desempenho de mais de 20 índices de adequação de ajuste em mais de 768.000 modelos de AFC simulados. Os autores concluíram que apenas o CFI (Bentler, 1990), o Gamma-hat (Steiger, 1989) e o índice de ajuste não-centralizado, NCI (McDonald, 1989) foram suscetíveis à identificação de invariância de parâmetros, não sendo influenciados pela complexidade do modelo, tampouco pelo tamanho amostral. Para se assumir

invariância de medida, o modelo testado não deve apresentar diferenças no CFI ($\Delta CFI > 0,01$), diferenças no Gamma-Hat ($\Delta \text{Gamma-hat} > 0,001$), e diferenças no McDonald's NCI ($\Delta \text{McDonald's NCI} > 0,02$), quando comparado ao modelo prévio (Cheung & Rensvold, 2002). Tais diretrizes têm sido aceitas e utilizadas em diversos estudos (e.g., Damásio, Pacico, Poletto & Koller, 2012; Meade e cols., 2008; Rossi e cols., 2010).

Em casos de não-equivalência da medida, quando o pressuposto da invariância não é estabelecido, o pesquisador deve identificar quais itens, ou parâmetros, não estão operando de maneira similar. Byrne (2010) apresenta estratégias interessantes na busca desses parâmetros. Em síntese, devem-se conduzir análises individuais (item a item, ou parâmetro a parâmetro) para o modelo avaliado. Por exemplo, se o pressuposto da invariância métrica não foi estabelecido (ou seja, a carga fatorial de um ou mais itens não é equivalente para um ou mais grupos), o pesquisador pode conduzir análises de invariância individual para os itens do instrumento e avaliar, a partir dos testes de invariância de medida ($\Delta\chi^2$, ΔCFI , $\Delta \text{Gamma-hat}$, e/ou $\Delta \text{McDonald's NCI}$), quais itens estão degradando o ajuste do modelo. É importante salientar que, em alguns programas estatísticos (e.g., EQS, Bentler, 2004), a análise de invariância dos itens ou parâmetros é feita de maneira coletiva, e não individual, o que facilita a inspeção e a identificação de possíveis problemas.

Outro aspecto importante é que, por vezes, os indicadores de invariância são contraditórios entre si (Byrne, 2010). Por exemplo, o teste $\Delta\chi^2$ aponta para não-invariância de parâmetros entre os grupos, enquanto o ΔCFI indica homogeneidade entre os grupos. Em casos de inconsistência, Byrne (2010) sugere que o pesquisador deve tomar uma decisão arbitrária em aceitar ou rejeitar o pressuposto de invariância do modelo. Para minimizar o subjetivismo desse procedimento, aqui se sugere que os pesquisadores levem em consideração os quatro indicadores apresentados (χ^2 , CFI, Gamma-hat e McDonald's NCI) e pondere seus resultados, considerando que o χ^2 é um indicador sensível e, por vezes, enviesado para detectar invariância de parâmetros (Hu & Bentler, 1993; Sass, 2011).

Considerações finais

O objetivo deste estudo foi o de apresentar alguns dos principais aspectos da AFCMG e discutir suas potencialidades na avaliação de invariância de instrumentos psicométricos. A AFCMG tem sido compreendida, atualmente, como uma das principais ferramentas que possibilitam a avaliação da

configuração e dos parâmetros de determinado instrumento para diferentes grupos simultaneamente. Suas ferramentas podem auxiliar os pesquisadores a conduzirem adequadamente tanto estudos de comparação de grupos, ao assegurar que testes de diferenças de médias não estão enviesados, quanto estudos de validação de instrumentos, ao possibilitar a avaliação da invariância da medida para diferentes grupos e contextos.

Ao longo deste estudo, três grandes tópicos da área da AFCMG foram abordados, a saber: os modelos a serem testados; a invariância completa *versus* invariância parcial; e a avaliação de invariância dos modelos. Em relação ao primeiro tópico, cabe ressaltar que nem todos os modelos apresentados precisam, necessariamente, ser testados para que o pressuposto de invariância seja aceito. A invariância configural (modelo 1), a invariância métrica (modelo 2) e a invariância escalar (modelo 3) são os modelos que têm recebido maior atenção no âmbito das pesquisas empíricas. Isso porque, em geral, esses são considerados os parâmetros que, de fato, asseguram a invariância da medida, garantindo a possibilidade ou não de comparação entre grupos, seja em estudos transculturais ou em estudos de comparação de subgrupos de uma mesma amostra.

Em relação à invariância residual, não há consenso na literatura sobre a necessidade de sua avaliação e as implicações da não-invariância desse pressuposto. Neste estudo, são apresentadas perspectivas favoráveis e contrárias à necessidade de avaliação desse parâmetro. Leituras mais aprofundadas sobre a temática deverão auxiliar os pesquisadores a compreenderem as situações teóricas e empíricas em que a invariância residual dos itens apresenta maior ou menor importância.

Referente aos modelos subsequentes (invariância da covariância das variáveis latentes, e invariância da média latente), estes são compreendidos como invariâncias estruturais, e estão relacionados a diferenças entre os grupos (heterogeneidade populacional) e não aos aspectos métricos do instrumento. A invariância de medida é, portanto, necessária para a avaliação da invariância estrutural e da invariância da média latente. Ou seja, o instrumento precisa estar adequadamente calibrado para todos os grupos de interesse para que estes possam ser comparáveis entre si.

Em relação à invariância completa *versus* invariância parcial, este é um tópico controverso na literatura internacional, visto que, atualmente, ainda não há um ponto de corte definido sobre quantos parâmetros e sob quais circunstâncias esses parâmetros devem ser invariantes entre os grupos para que as

comparações entre os grupos sejam aceitáveis. As pesquisas sobre este tópico ainda precisam avançar consideravelmente para que os pesquisadores possam tomar decisões pautadas em diretrizes claras e amplamente aceitas.

Por fim, no que se refere à avaliação de invariância de estrutura e dos parâmetros, quatro índices de ajuste (χ^2 ; CFI; Gamma-hat; McDonald's NCI) suscetíveis ao uso na AFCMG foram apresentados. Apesar de, até este momento, não haver total consenso sobre quais e quantos índices de ajuste avaliar, é possível observar na literatura que os índices CFI, Gamma-hat e McDonald's NCI têm recebido crescente credibilidade no uso das AFCMG, desde a publicação de Cheung e Rensvold (2002). Assim, em consonância com estudos internacionais (e.g., Meade e cols., 2008) sugere-se que, ao avaliar a invariância de estrutura e dos parâmetros de um instrumento, todos os índices apresentados sejam utilizados, com vistas a aceitar ou refutar a hipótese de invariância com maior nível de confiabilidade.

Por fim, um último aspecto precisa ser ressaltado. Para que os resultados da AFCMG sejam confiáveis, é preciso que os pesquisadores utilizem métodos de estimação adequados para os seus dados. Como se sabe, no âmbito da Psicologia, a maioria dos instrumentos psicométricos apresenta itens com distribuição não-normal. Isso exige a utilização de métodos de estimação que levem em consideração essa característica, como por exemplo, os métodos de distribuição assintótica livre (*asymptotic distribution free*, ADF) ou o método da máxima verossimilhança robusto (*Robust Maximum Likelihood*; Satorra & Bentler, 2001). Em casos de itens categóricos (dicotômicos ou polítômicos), métodos que utilizem matrizes de correlações tetracóricas e/ou policóricas, como o *Weighted Least Squares Mean and Variance-Adjusted* (WLSMV; Muthén & Muthén, 1999), são, também, requeridos. Apenas utilizando métodos de estimação condizentes com as características dos itens do instrumento é que estimativas precisas poderão ser obtidas.

Referências

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.

Bentler, P. M. (2004). *EQS 6 Structural Equations Program manual*. Encino, CA: Multivariate Software, Inc.

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, 16(3), 201-213.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Nova Iorque: The Guilford Press.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: basic concepts, applications, and programming* (2nd ed.). Nova Iorque: Routledge, Taylor & Francis.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95(5), 1005-1018.

Cheung, G. W. & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new model. *Journal of Management*, 25(1), 1-27.

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.

Cronbach, L. J. (1947). "Test reliability": Its meaning and determination. *Psychometrika*, 12(1), 1-16.

Damásio, B. F., Borsa, J. C. & da Silva, J. P. (2011). 14-Item Resilience Scale (RS-14): psychometric properties of the Brazilian version. *Journal of Nursing Measurement*, 19(3), 131-145.

Damásio, B. F., Pacico, J. C., Poletto, M. & Koller, S. H. (2012). Refinement and psychometric properties of the eight-item Brazilian Positive and Negative Affective Schedule for Children (PANAS-C8). *Journal of Happiness Studies*. doi:10.1007/s10902-012-9383-x

Davidov, E. & Depner, F. (2009). Testing for measurement equivalence of human values across online and paper-and-pencil surveys. *Quality & Quantity*, 45(2), 375-390.

DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science*, 46(1), 137-149.

Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (2005). *Análise multivariada de dados*. A. S. Sant'Anna & A. C. Neto (Trans). Porto Alegre: Bookman.

Harrington, D. (2009). *Confirmatory factor analysis*. Nova Iorque: Oxford University Press.

Hooper, D., Coughlan, J. & Mullen, M. R. (2008). Structural equation modeling: guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.

Hu, L. & Bentler, P. M. (1993). Evaluating model fit. Em R. H. Hoyle (Ed.) *Structural equation modeling: concepts, issues, and applications* (pp. 16-99). Newbury Park, CA: Sage.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409-426.

Jöreskog, K. G. (2007). Factor analysis and its extensions. Em R. Cudeck & R. C. MacCallum (Eds.). *Factor analysis at 100: historical development and future directions* (pp. 47-78). Mahwah, NJ: Lawrence Erlbaum.

Kelloway, E. K. (1995). Structural equation modeling in perspective. *Journal of Organizational Behavior*, 16(3), 215-224.

Lubke, G. H. & Dolan, C. V. (2003). Can unequal residual variance across groups mask difference in residual means in the common factor model? *Structural Equation Modeling*, 10(2), 175-192.

Malpass, R. S. (1977). Theory and method in cross-cultural psychology. *American Psychologist*, 32(12), 1069-1079.

Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size. *Psychological Bulletin*, 103(3), 391-410.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification* 6(1), 97-103.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93(3), 568-592.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29(2), 177-185.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.

Milfont, T. L. & Fischer, R. (2010). Testing measurement invariance across groups: applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111-121.

Millsap, R. E. & Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. Em R. Cudeck & R. C. MacCallum (Eds.). *Factor analysis at 100: historical development and future directions* (pp. 131-152). Mahwah, NJ: Lawrence Erlbaum.

Millsap, R. E. & Kwok, O. M. (2004). Evaluating the impact of partial measurement invariance on selection in two populations. *Psychological Methods*, 9, 93-115.

Mullen, M. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 26(3), 573-596.

Murayama, K., Zhou, M. & Nesbit, J. C. (2009). A cross-cultural examination of the psychometric properties of responses to the Achievement Goal Questionnaire. *Educational and Psychological Measurement, 69*(2), 266-286.

Muthén, L. K. & Muthén, B. (1999). *Mplus user's guide*. Los Angeles: Muthén & Muthén.

Nimon, K. & Reio Jr, T. G., (2011). Measurement invariance: a foundational principle for quantitative theory building. *Human Resource Development Review, 10*(2), 198-214.

Rossi, G., Elkliit, A., Simonsen, E. (2010). Empirical evidence for a four factor framework of personality disorder organization: multigroup confirmatory factor analysis of the Millon Clinical Multiaxial Inventory-III Personality Disorder Scales across Belgian and Danish data samples. *Journal of Personality Disorders, 24*(1), 128-150.

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*(4), 347-363.

Satorra, A. & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29*(4), 304-321.

Schmitt, N. & Kuljanin, G. (2008). Measurement invariance: review of practice and implications. *Human Resource Management Review, 18*(4), 210-222.

Schreiber, J. B., Stage, F. K., King, J., Nora, A. & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research, 99*(6), 324-337.

Spurk, D., Abele, A. E. & Volmer, J. (2011). The Career Satisfaction Scale: longitudinal measurement invariance and latent growth analysis. *Journal of Occupational and Organizational Psychology, 84*(2), 315-326.

Steiger, J. H. (1989). EzPATH: Causal modeling. Evanston, IL: SYSTAT.

Steiger, J. H., Shapiro, A. & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50*(3), 253-264.

Tuokko, H. A., Chou, P. H. B., Bowden, S. C., Simard, M., Ska, B. & Crossley, M. (2010). Partial measurement equivalence of French and English versions of the Canadian Study of Health and Aging neuropsychological battery. *Journal of the International Neuropsychological Society, 15*(3), 416-425.

Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.

Wu, A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: a demonstration with TIMSS data. *Practical Assessment, Research & Evaluation, 12*(3), 1-26. Retirado de: <http://pareonline.net/getvn.asp?v=12&n=3>.

Recebido em 28/07/2012
 Reformulado em 07/02/2013
 Aprovado em 26/03/2013

Sobre o autor:

Bruno Figueiredo Damásio é psicólogo, mestre e doutorando em Psicologia (UFRGS). Tem como foco de pesquisa a Psicometria, especialmente estudos de construção, adaptação e validação de instrumentos, utilizando Modelagem de Equações Estruturais e Teoria de Resposta ao Item. Atualmente é Editor-Júnior da revista *Temas em Psicologia*, da Sociedade Brasileira de Psicologia. Em termos teóricos, interessa-se pelos seguintes temas: sentido de vida, bem-estar, qualidade de vida, *coping* e recursos psicosociais.

Contato com o autor

Rua Ramiro Barcelos, 2.600, Sala 104 – CEP 90035-003 – Porto Alegre-RS, Brasil
 E-mail: brunofd.psi@gmail.com