

Editorial: A validade nos instrumentos de avaliação [Validity in assessment instruments]

Editorial

Helena Espírito Santo, PhD (1)

(1) Instituto Superior Miguel Torga, Coimbra, Portugal

Autor para correspondência | Helena Espírito-Santo; Largo de Celas, 1, 3000-132 Coimbra, Portugal; +351 910637946; helenium@gmail.com

Face ao número crescente de artigos sobre a validação e usos de instrumentos de avaliação que chegam à RPICS, talvez seja o momento para refletir sobre a questão da validade dos mesmos. A RPICS espera que os autores nos seus artigos discutam explicitamente sobre a validade dos instrumentos na secção dos Métodos. Caso os autores utilizem um instrumento estudado previamente com o mesmo formato, com as mesmas características de sujeitos e para o mesmo objetivo, então é suficiente citar a referência prévia. Caso o instrumento seja modificado, o número de itens alterado, o objetivo diferente ou o ponto de corte se modifique, então é necessário apresentar informação suplementar sobre a adaptação. No caso de novos instrumentos criados pelos autores, é essencial explicitar o decurso da elaboração, as medidas de confiabilidade, os resultados do estudo piloto e outras informações que credibilizem o uso do instrumento de avaliação. Seguem-se alguns conceitos fundamentais no domínio da validade e confiabilidade.

Validade

A validade de um instrumento (questionários, escalas ou testes) refere-se à sua capacidade de medir realmente o que pretende medir (AERA-APA-NCM, 1999). Na validade são considerados diferentes aspetos.

A **validade de constructo** é a característica de um teste enquanto medição de um atributo ou qualidade que não tenha sido definido operacionalmente (Cronbach e Meehl, 1955).

A **validade de critério** é a capacidade de um instrumento de prever um desempenho específico de um sujeito medido através de outra ferramenta independente. Se ambas ferramentas forem administradas no mesmo momento, a validade de critério diz-se **convergente** ou concorrente. Caso os dados sejam obtidos após a recolha de informação sobre o instrumento, fala-se em **validade preditiva** (Pasquali, 2009).

A **validade do conteúdo** refere-se à capacidade do instrumento de representar um domínio finito de comportamentos (Pasquali, 2009).

Confiabilidade

A confiabilidade, fiabilidade ou fidedignidade expressa a capacidade de um instrumento de avaliação fornecer os mesmos resultados cada vez que é usado com o mesmo formato em sujeitos com características similares (e.g., idade, escolaridade, género, estado civil) e do mesmo contexto (e.g., escolar, clínico, institucional) (AERA-APA-NCM, 1999). Ou seja, diz respeito à sua consistência. A medição da confiabilidade depende do tipo de instrumento e existem muitas formas de a medir.

Uma das formas é a avaliação da sua **estrutura interna**. Uma das medidas mais usadas é o alfa de Cronbach que resulta

das correlações entre as respostas aos itens/questões do instrumento. Quanto mais alto for o seu valor e mais próximo de 1, melhor a confiabilidade (Marôco e Garcia-Marques, 2006). Outros procedimentos incluem a correlação entre as pontuações das duas metades do instrumento (*split-half* de Spearman, 1910), o procedimento de Kuder e Richardson (1937), o alfa estratificado de Cronbach, Shonenman e McKie (1965) e o estimador de máxima fiabilidade de Li, Rosenthal e Rubin (1996).

Outra forma de medir a confiabilidade é através da avaliação da capacidade de o instrumento fornecer resultados semelhantes em momentos de aplicação diferentes. Este procedimento designa-se por **teste-reteste** e é habitualmente determinado através do coeficiente de correlação de Pearson (Downing, 2004).

A confiabilidade **inter-juizes** é também uma forma de avaliação da confiabilidade e diz respeito à capacidade de o instrumento fornecer resultados semelhantes quando aplicado/cotado por avaliadores diferentes. Para o seu cálculo pode usar-se o Kapa de Cohen quando os resultados se apresentam dicotomicamente ou o Tau de Kendall (Downing, 2004).

Outra forma de avaliação diz respeito ao erro de medição que pode ser imputado a cada fator (e.g., itens, sujeitos, avaliadores) e recorre-se à ANOVA para gerar um **coeficiente de generalização** (Cook e Beckman, 2006).

Em conclusão, na RPICS defendemos a importância do uso de instrumentos válidos e fiáveis, sendo essencial a apresentação da análise das suas características psicométricas.

Referências

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education / AERA-APA-NCM. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Cook, D. A. e Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments. *American Journal of Medicine*, 119(2), 166e7–166e16.
- Cronbach, L. J. e Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.
- Downing, S. M. (2004). Reliability: on the reproducibility of assessment data. *Medical Education*, 38(9), 1006–1012.
- Kuder, G. F. e Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2, 151-160.
- Marôco, J. e Garcia-Marques, T. (2006). Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? *Laboratório de Psicologia*, 4(1), 65-90.
- Pasquali, L. (2009). *Psicometria*. Revista da Escola de Enfermagem USP, 43(Esp), 992-999.
- Cronbach, L. J., Schönemann, P. e McKie, D. (1965). Alpha coefficients for stratified parallel tests. *Educational and Psychological Measurement*, 25, 291-312.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.

RPICS

DOI: 10.7342/ismt.rpics.2017.3.1.49