

PRINCIPAIS MÉTODOS DE AVALIAÇÃO PSICOMÉTRICA DA CONFIABILIDADE DE INSTRUMENTOS DE MEDIDA

MAIN PSYCHOMETRIC EVALUATION METHODS OF MEASURING INSTRUMENTS RELIABILITY

Cristiane Martins Cunha^{a*}, Omar Pereira de Almeida Neto^{b**}, Renata Stackfleth^{c*}

^acrismcunha@ymail.com, ^bomarprneto@hotmail.com, ^crenatastack@gmail.com

*Universidade de São Paulo – Ribeirão Preto (SP), Brasil

**Universidade Federal de Uberlândia – Uberlândia (MG), Brasil

Data de recebimento do artigo: 27/10/2015

Data de aceite do artigo: 04/04/2016

RESUMO

Introdução: A literatura tem focado a importância da adequada avaliação psicométrica dos instrumentos de medida como critério de boa qualidade do instrumento no que se refere às evidências de validade e à confiabilidade. Avaliar a confiabilidade é o primeiro passo para conhecer a acurácia de um instrumento de medida. **Objetivo:** Apresentar os princípios básicos de avaliação psicométrica da confiabilidade dos instrumentos de medida e os principais métodos estatísticos descritos na literatura. **Metodologia:** A pesquisa foi realizada a partir de publicações nas bases de dados Medline, LILACS e IBECs e também pelas descrições em livros da área. A análise dos estudos envolveu leitura de títulos, resumos e textos publicados nos últimos 10 anos, e foram excluídos artigos duplicados ou que não correspondiam ao objetivo desta revisão. **Resultados:** Os principais métodos de avaliação da confiabilidade são os baseados na correlação entre os itens e na consistência interna. Os principais métodos baseados em correlações são a confiabilidade teste-reteste, a confiabilidade a partir de formas equivalentes (ou paralelas) e a confiabilidade entre observadores; já os principais métodos de avaliação pela consistência interna são o método das duas metades, o coeficiente alfa de correlação (alfa de Cronbach) e o método de Kuder-Richardson. **Conclusão:** De modo geral, os autores concordam com os critérios fundamentais de avaliação da confiabilidade de um bom instrumento de medida. A confiabilidade é um atributo psicométrico presente num bom instrumento de medida e precisa ser contemplado nessas categorias de estudo.

Palavras-chave: Reprodutibilidade dos testes; medidas, métodos e teorias; questionários; psicometria.

ABSTRACT

Introduction: Literature has focused on the importance of proper psychometric evaluation of measuring instruments as a criterion for the good quality of the instrument in relation to evidence of validity and reliability. To assess the reliability is the first step to know the accuracy of a measuring instrument. **Objective:** This article presents the basic principles of psychometric evaluation of the reliability of measuring instruments and the main statistical methods described in the literature. **Method:** The survey was conducted from publications in Medline, LILACS, and IBECs databases and also from the descriptions in books regarding the field of study. The analysis of the studies involved reading titles, abstracts, and articles published in the last 10 years and excluded articles that were duplicated or not corresponding to the purpose of this review. **Results:** The main evaluation of reliability methods are based on the correlation between the items and internal consistency. The main methods based on correlations are the test-retest reliability, reliability from equivalent (or parallel) forms, and the reliability between observers; on the other hand, the main evaluation methods for internal consistency are the split-half method, the alpha correlation coefficient (Cronbach's alpha) and the Kuder-Richardson formula. **Conclusions:** In general, the authors agree with the basic criteria for evaluating the reliability of a good measurement instrument. Reliability is a psychometric attribute needed in a good measurement instrument and needs to be covered in these study categories.

Keywords: Reproducibility of results; measurements, methods, and theories; questionnaires; psychometrics.

Introdução

A psicometria é a abordagem científica que visa à mensuração e à avaliação psicológica dos constructos subjetivos por meio de escalas, testes e questionários padronizados, denominados “medida psicométrica”, os quais representam uma situação experimental que avaliam um determinado constructo¹. É uma complexa área do conhecimento que objetiva desenvolver e aplicar técnicas de mensuração dos fenômenos psíquicos^{2,3}. Atualmente, distinguem-se na psicometria dois modelos teóricos de medida: a teoria clássica dos testes (TCT) e a teoria moderna (ou teoria de resposta ao item)³.

A TCT continua sendo bastante utilizada pela comunidade científica, visto que seu foco de avaliação é direcionado à qualidade dos instrumentos de medida, buscando garantir que estes sejam válidos e confiáveis para medir determinados comportamentos humanos²⁻⁴.

A TCT considera, especificamente, os parâmetros relacionados ao teste como um todo, dado que as avaliações psicométricas consideram os escores totais da medida por meio de correlações dos itens com os escores totais do teste^{5,6}. Assim, os testes desenvolvidos segundo esse modelo são elaborados para avaliar as aptidões ou habilidades populacionais medianas, sendo necessário que a amostra seja suficientemente grande e representativa³. Além disso, nesse modelo, por adotar os escores totais de medida em uma determinada amostra, os parâmetros de dificuldade e de discriminação dos itens e o índice de erro da medida do teste são gerais para toda a amostra^{5,6}.

A confiabilidade é considerada um dos principais critérios de garantia da qualidade do instrumento, pois se refere à propriedade do teste de medir sem erros^{2,7}, focando em obter resultados consistentes e reprodutíveis na mensuração de um determinado atributo^{8,9}. É um atributo psicométrico que deve estar presente em um bom instrumento e precisa ser contemplado nessas categorias de estudo, visto que sua ausência pode trazer graves consequências sobre a validade das conclusões estatísticas de uma pesquisa¹⁰.

Conhecer os adequados métodos psicométricos de avaliação da confiabilidade de um instrumento de medida é o primeiro passo para garantir a boa qualidade do instrumento. Este artigo objetiva apresentar os princípios básicos de avaliação psicométrica da confiabilidade dos instrumentos de medida e os principais métodos estatísticos descritos na literatura.

Metodologia

Tipo de Estudo

Trata-se de uma revisão integrativa em que foram selecionados estudos da literatura nacional e

internacional. A revisão integrativa se constitui em um instrumento para a prática baseada em evidências (PBE), sendo um método voltado à atenção clínica e ao ensino, alicerçado no conhecimento e na otimização da evidência.

Esta revisão foi estruturada em cinco etapas, buscando aumentar o rigor do estudo: identificação do tema e seleção da hipótese ou questão de pesquisa para a elaboração da revisão integrativa; estabelecimento de critérios para inclusão e exclusão de estudos – amostragem ou busca na literatura; definição das informações a serem extraídas dos estudos selecionados – categorização dos estudos; avaliação dos estudos incluídos na revisão integrativa; interpretação dos resultados.

Coleta de Informações

Realizou-se um levantamento bibliográfico nas bases de dados Medline, LILACS e IBECs, utilizando os descritores em ciências da saúde (DeCS) “reprodutibilidade dos testes”; “medidas, métodos e teorias”; “questionários” e “psicometria”, assim como, suas traduções correspondentes para o inglês: “*reproducibility of results*”; “*measurements, methods, and theories*”; “*questionnaires*” e “*psychometrics*”, os quais foram interligados pelo operador booleano “*and*”.

Também foram consultadas descrições em diversos livros da área a fim de reunir o máximo de informações que descrevessem os principais métodos de avaliação da validade de instrumentos de medida. Para analisar e selecionar as publicações foi realizada a leitura de títulos, resumos e textos, respectivamente, publicados nos últimos dez anos, e que estavam escritos em português, inglês e espanhol. Foram excluídos artigos duplicados, os que não estavam disponíveis na íntegra e os que não correspondiam ao objetivo desta revisão.

Resultados

Princípios teórico-metodológicos da avaliação psicométrica: confiabilidade

A confiabilidade relaciona-se com a coerência, constância dos resultados e a confiança que o teste inspira para mensurar fenômenos, sem que haja grandes flutuações entre medidas repetidas que reflitam a presença de erros de mensuração. Isso significa que o mesmo teste medindo os mesmos sujeitos em ocasiões diferentes ou testes equivalentes medindo os mesmos sujeitos na mesma ocasião devem apresentar dados reprodutíveis e produzir resultados semelhantes^{1,2,7,11,12}.

Acrescenta-se que a confiabilidade de um teste é a diferença entre o escore observado e o seu verdadeiro valor, ou seja, um escore sem a presença de erros de medida. Portanto, admite-se que sempre em um teste haverá algum nível de erros de medida, os quais são variações métricas aleatórias que não refletem as diferenças reais dos escores de um teste^{1,7,8}.

A confiabilidade também pode ser chamada de precisão ou fidedignidade^{1,2,7,10} e objetiva estimar o nível de precisão de um teste e estabelecer uma expectativa do quão errônea a medida pode ser⁷. Ela se relaciona com os parâmetros de precisão, constância, estabilidade ao longo do tempo, equivalência, reprodutibilidade e consistência interna^{1,2,8,9,11-13}, atributos que garantem que o teste não esteja vulnerável a variadas magnitudes de fontes de erros que possam comprometer a legitimidade de seus escores e a validade de suas interpretações^{7,10}.

Portanto, verifica-se que a ausência de confiabilidade de um teste pode trazer graves consequências sobre a validade das conclusões estatísticas de uma pesquisa¹⁰. Por outro lado, apesar de a confiabilidade de um teste ser uma condição necessária, ela não é suficiente para garantir a validade de um instrumento de medida^{1,7,13,14}.

Nesse contexto, a psicometria desenvolveu uma série de modelos estatísticos que calculam os índices de confiabilidade de um teste buscando garantir a sua precisão^{1,7,10,12,15}. Desse modo, a escolha dos testes estatísticos mais adequados para mensurar o coeficiente de confiabilidade depende do método de validação e do tipo e nível de mensuração dos dados produzidos pela variável medida do instrumento^{1,10}.

Alguns fatores podem afetar a fidedignidade de um instrumento. Entre eles, cita-se que, quanto mais heterogênea for a amostra estudada, maior será a variabilidade e a variância dos escores, e com isso a confiabilidade será maior⁸. Acrescenta-se ainda que as escalas mais homogêneas (geralmente unidimensionais) são mais consistentes e isso lhes confere maior confiabilidade. E, por fim, quanto maior a variabilidade dos escores dos itens e seu poder discriminativo, maior será a confiabilidade do teste^{1,3,8}.

As estimativas de confiabilidade são expressas em termos de coeficiente de correlação, mas cada um tem significado diferente. Os parâmetros dos valores do coeficiente de confiabilidade variam conforme os autores, mas no geral, valores menores que 0,40 são considerados de baixa confiabilidade, valores entre 0,40 e 0,69 de moderada confiabilidade e maiores que 0,70 de alta confiabilidade¹.

Os principais métodos para avaliar a confiabilidade, que são baseados em correlações, são a confiabilidade teste-reteste, a confiabilidade a partir de formas equivalentes (ou paralelas) e a confiabilidade entre observadores; já os principais métodos de avaliação da confiabilidade de um teste por meio da consistência interna são o método das duas metades, o método de alfa de

Cronbach e o método de Kuder-Richardson (KR). Esses métodos estatísticos foram denominados de modelo linear clássico de avaliação da confiabilidade^{1,11,13,15,16}.

Métodos para estimar a confiabilidade baseados na correlação dos itens

A confiabilidade teste-reteste (também chamada de confiabilidade intraobservador ou método do teste repetido) é um método bem direto que se baseia no cálculo do coeficiente de estabilidade ao longo do tempo, o qual evidencia o nível de correlação existente entre as distribuições de escores num mesmo teste pelos mesmos sujeitos em duas ocasiões diferentes e independentes^{1,2,7-9,12,13,15}. Se a correlação entre os resultados das duas aplicações for significativamente positiva, o instrumento pode ser considerado confiável^{2,7,11}.

Não há consenso na literatura sobre o intervalo de tempo ideal entre as avaliações. No entanto, é importante considerar que a avaliação da estabilidade do teste sofre grande influência do período de tempo adotado entre as avaliações, sendo recomendado que o intervalo de tempo entre as avaliações não seja tão curto nem tão longo^{1,3,8,11,12,15}. Períodos longos entre as avaliações (efeito de sensibilidade ou reatividade) são suscetíveis a mudanças que podem comprometer a interpretação do coeficiente obtido, pois proporciona ao indivíduo a aquisição de novas aprendizagens, gerando subavaliação da estabilidade do teste, ao passo que períodos muito curtos podem sofrer influência de resultados contaminados pelo efeito recordatório, gerando superavaliação da estabilidade do teste^{1,11}.

Diante dessas limitações da confiabilidade teste-reteste, é esperado que o pesquisador avalie com precaução o intervalo de tempo ideal que será adotado entre as avaliações, levando em conta o tipo de variável investigada, as peculiaridades do instrumento e a possibilidade de ocorrência do efeito de aprendizagem (ou reatividade) e/ou do efeito recordatório (ou memória) ao longo do tempo. O cálculo dos escores dessas duas avaliações depende do nível de mensuração dos dados, e se espera que haja uma correlação positiva e significativa, ideal que seja acima de 0,85 entre as avaliações¹⁶.

Outro método de avaliação da confiabilidade, que se baseia na correlação entre os escores dos grupos para o cálculo do coeficiente de equivalência, é a técnica de avaliação da confiabilidade pelas formas equivalentes (ou formas paralelas). Esse é um método pouco utilizado pela dificuldade de se dispor de dois instrumentos distintos, mas que sejam similares ou equivalentes e que objetivem avaliar um mesmo atributo investigado, que serão aplicados a dois grupos de indivíduos em um curto prazo de tempo^{1,7,11-13,15}.

Na estimativa da confiabilidade pelas formas equivalentes há duas versões de um instrumento que mede

um mesmo atributo, conectados de maneira similar^{1,8}. Segundo esses autores, muitos consideram esse método de estimativa da fidedignidade como sendo o melhor indicador de confiabilidade do instrumento, porém há várias razões pragmáticas que justificam a não utilização desse método, como a dificuldade operacional de se garantir a equivalência e similaridade dos atributos psicométricos e estatísticos dos itens da escala, no que se refere aos índices de dificuldade, discriminação, média, variância etc. Além disso, é muito comum esse método ser acidentalmente contaminado pelos efeitos de memória, e ainda que esses itens sejam fraseados diferentemente, podem levar a superestimação dos coeficientes de confiabilidade¹.

Por fim, a confiabilidade de escores entre avaliadores (ou interavaliadores) é um método de avaliação subjetiva, realizada por diferentes observadores (juízes), e é guiada por medições das observações conforme critérios de avaliação pré-estabelecidos^{1,8,11,13,15}. É bastante útil em exames decisivos, sobretudo para avaliar o desempenho ou na correção de provas discursivas^{1,11,13,15} e pode ser considerado um critério de calibração do instrumento¹.

A confiabilidade de escores entre avaliadores se baseia no grau de concordância (porcentagem de concordância) ou na consistência entre os avaliadores calculados pelo nível de correlação entre as respostas dos avaliadores, por meio do cálculo da correlação linear de Pearson (ou correlação de Spearman ou de Kendal)^{11,13} e/ou coeficiente de kappa de Cohen ou coeficiente de correlação intraclasse^{1,12,15} nas medidas dos escores de confiabilidade dos avaliadores.

Métodos para estimar a confiabilidade baseados na consistência interna

A consistência interna, como uma estimativa de precisão do teste, pode ser avaliada por vários métodos estatísticos para verificar a homogeneidade e a consistência interna dos itens mediante a presença de covariância entre eles^{2,7,8}. Sua análise se baseia na correlação dos itens de uma mesma medida entre si e na correlação de cada item com o escore total dessa medida^{8,10,12-14}.

É importante ressaltar que cada item de uma escala deve contribuir para aumentar a confiabilidade da escala, e caso contrário, pode ser um critério para ser excluído do teste⁸. As estimativas de confiabilidade pela consistência interna são a média de todos os possíveis modos de subdividir o conjunto total de itens a fim de calcular o coeficiente de fidedignidade de um dado instrumento, o qual demonstra o quão bem os itens refletem o mesmo constructo e produzem resultados similares e consistentes¹.

De forma ideal, para que o teste seja considerado consistente, espera-se que os itens sejam moderadamente correlacionados entre si e que exista boa correlação de cada item com o escore total do instrumento^{8,10,14}. Esse método

exige apenas uma aplicação, e, portanto, não sofre influência da temporalidade¹.

Recomenda-se calcular a consistência interna de um teste sempre que o aplicar em uma amostra populacional, uma vez que seus índices podem variar segundo as características e peculiaridades populacionais^{1,8,14}. Ressalta-se também que o número de itens que compõem o teste também é um fator que deve ser levado em conta na avaliação da consistência interna, dado que a quantidade de itens afeta o cálculo desse coeficiente, de modo que o baixo número de itens reduz a magnitude da consistência interna. Para determinar a consistência interna de forma ideal, deve-se utilizar uma amostra de 5 a 20 participantes por item a fim de perfazer uma amostra de pelo menos 100 a 400 sujeitos. Os principais métodos de verificar a consistência interna são o método das metades partidas e o coeficiente alfa de correlação (alfa de Cronbach e o método KR-20)^{8,14}.

A técnica das metades partidas (ou duas metades ou *split-half*) consiste em avaliar a confiabilidade usando as respostas obtidas em apenas uma mesma aplicação do instrumento de medida, por meio da divisão desse mesmo instrumento em duas partes ou metades, consideradas semelhantes e comparáveis entre si para avaliar um determinado atributo^{1,7,11,13,15}.

Desse modo, convencionou-se a divisão do instrumento em duas metades comparáveis, de modo que uma das metades seja composta pelos itens pares e a outra pelos itens ímpares, mas levando em conta também a equivalência entre os itens, em termos de conteúdo e dificuldade. Em razão disso, a utilização desse método é mais indicada para ser aplicada em instrumentos multidimensionais¹.

A comparação dos escores do instrumento pelo método das metades partidas pode ser feita por meio do cálculo do coeficiente de correlação apropriado ao nível de mensuração gerado pelos dados obtidos pelas pontuações das duas metades. Se o instrumento for confiável, haverá correlação positiva e significativa entre as metades, o que é um indicador de boa consistência interna^{1,11}. No entanto, é um método bastante questionável na literatura, e por isso tem sido pouco utilizado, visto que é praticamente impossível contemplar e garantir todos os requisitos psicométricos necessários para sua utilização^{1,15}.

Além disso, o coeficiente de fidedignidade obtido pela técnica das metades partidas representa apenas a metade do instrumento, de modo que, quanto maior for o instrumento, maior será sua confiabilidade. Assim, para obter o coeficiente de confiabilidade total por esse método, é necessário aplicar uma fórmula denominada de Spearman-Braune, a qual permite a correção da subestimação do coeficiente de correlação obtido usando o método das duas metades e, com isso, obter uma estimativa mais fidedigna do instrumento total¹.

O coeficiente alfa de correlação evidencia que a medida mede apenas um atributo, e, indiretamente, indica a característica unidimensional de um teste. Nos casos dos testes que medem vários atributos (domínios), os testes multidimensionais, o alfa deve ser determinado individualmente para cada um deles³. Parte-se do princípio que os itens que compõem o instrumento e que mensuram um determinado atributo estão positivamente relacionados entre si, já que homogeneamente medem o mesmo construto^{8,10,12,14}. O coeficiente alfa pode ser calculado por meio do coeficiente KR-20 ou pelo coeficiente alfa de Cronbach.

Existe uma variante do cálculo da confiabilidade da consistência interna pelo método de alfa, o qual é denominado de coeficiente KR-20 ou fórmula 20 de Kuder-Richardson¹⁴. Seu cálculo é semelhante ao coeficiente de alfa de Cronbach, e, portanto, esses coeficientes apresentam equivalência matemática¹⁴. No entanto, esse método é indicado para ser aplicado em testes com respostas dicotômicas^{14,16}. Esse método não exige subdivisão dos itens do instrumento e requer apenas uma aplicação para seu cálculo¹.

O coeficiente de confiabilidade de alfa de Cronbach foi desenvolvido por Lee Cronbach em 1951. É considerado um dos coeficientes estatísticos de confiabilidade mais utilizados na literatura¹⁰, sendo indicado para ser aplicado em testes ou escalas politômicas¹⁴. Esse coeficiente demonstra a homogeneidade dos itens de uma escala na representação de apenas uma variável latente, o que sugere, indiretamente, o aspecto unidimensional do teste.

Desse modo, o coeficiente de alfa de Cronbach não é considerado um bom indicador de confiabilidade para testes multidimensionais, pois mensura vários atributos⁸, a menos que sejam avaliados separadamente. De maneira errônea, existe uma confusão de que a consistência interna verifica a dimensionalidade do instrumento. Entretanto, esse atributo pode ser calculado e verificado por vários outros métodos estatísticos, dentre eles a análise fatorial³.

Segundo alguns autores, o coeficiente de confiabilidade calculado pelo alfa de Cronbach é uma estimativa da proporção de variância das diferenças de desempenho (ou habilidade), e é inversamente proporcional ao erro padrão de medida. Esse coeficiente é obtido pela variância dos escores individuais e da variância da soma dos escores de cada avaliado, buscando avaliar as possíveis relações entre os itens. Seus valores variam de 0 a 1 (ou de 0 a 100%), e expressam a presença de correlações positivas entre os itens^{1,8,10,11,13}.

Portanto, o valor de alfa de Cronbach é a média de todos os coeficientes de correlação e seus valores fornecem informações sobre cada item individual. Com isso, há fortes indícios para os itens que não estão correlacionados serem excluídos do teste com intuito de aumentar sua confiabilidade^{8,11,12,14}.

Não há consenso na literatura sobre quais valores do coeficiente alfa de Cronbach são considerados ideais. Diante disso, orienta-se que o pesquisador seja bastante criterioso ao estabelecer os valores ideais de consistência interna para o instrumento investigado. De modo geral, orienta-se que sejam adotados coeficientes acima de 0,70, sendo que os valores mais recomendados estão entre 0,80 e 0,90^{8,10,13,14,16}. Em contrapartida, valores acima de 0,90 não são recomendados, dado que altos valores de alfa podem indicar a redundância dos itens, sugerindo que os itens sejam repetitivos ou que estejam em excesso na mensuração de um mesmo atributo^{1,3,13,14}.

É importante salientar que alguns fatores relacionados ao instrumento podem influenciar os valores do coeficiente de alfa de Cronbach, tais como o baixo ou alto número de itens, o alto índice de itens sem resposta (pois refletem respostas impulsivas ou relapsas) e a utilização de uma amostra pequena ou muito homogênea de testados (devido à baixa variância)^{10,14,15}.

Também vale salientar que, na tentativa de aumentar o índice de confiabilidade pelo coeficiente de alfa de Cronbach por meio da adição de itens no instrumento, recomenda-se que o pesquisador estime, previamente, a magnitude do aumento desse coeficiente pela fórmula de Spearman-Browne. Essa fórmula avalia o real benefício de se acrescentar um ou mais itens na escala visando torná-lo mais confiável⁸.

Desse modo, para obter bons índices de consistência interna em uma escala, recomenda-se que se utilizem instrumentos nem muito curtos nem muito extensos para garantir boas e adequadas respostas dos participantes e que se utilize uma amostra heterogênea para garantir uma maior variabilidade nos escores do teste^{8,14}.

Conclusões

Para responder alguma questão, o pesquisador pode decidir utilizar um instrumento de medida existente ou construir um novo instrumento. Do ponto de vista do custo-benefício, é sempre preferível e prático para o pesquisador utilizar um instrumento de medida já pronto, que seja comprovadamente válido e confiável, do que começar um longo e dispendioso processo de construção e validação psicométrica¹². Entretanto, independentemente da decisão do pesquisador, é fundamental a adequada seleção do instrumento por meio de uma criteriosa e rigorosa avaliação que garanta a sua boa qualidade, conforme as recomendações da literatura^{8,12,13}.

A literatura tem focado, cada vez com maior ênfase, na importância da adequada avaliação psicométrica dos instrumentos de medida, sobretudo no que concerne às evidências de sua validade e à confiabilidade. E, de modo geral, os autores concordam com os critérios

fundamentais para garantir um bom instrumento de medida. No entanto, alguns estudos já sinalizam sobre as novas perspectivas de avaliação psicométrica e as suas vantagens para TCT^{4,7,8,10-13,15,17-19}.

Na avaliação da confiabilidade, a literatura demonstrou que os principais métodos de avaliação da confiabilidade são os baseados na correlação entre os itens e na consistência interna. Fazem parte do primeiro grupo de métodos: a confiabilidade teste-reteste, confiabilidade a partir de formas equivalentes (ou paralelas) e confiabilidade entre observadores; já para o segundo grupo: o método das duas metades, o coeficiente alfa de correlação (alfa de Cronbach) e método de Kuder-Richardson.

A confiabilidade de um instrumento de medida é um atributo psicométrico importante que precisa estar presente em um bom instrumento de medida, posto que sua ausência pode produzir graves consequências sobre a validade das conclusões estatísticas de uma pesquisa¹⁰. Segundo Primi⁷, estimar a precisão de um instrumento de medida é um passo relativamente fácil e é comumente atingido nos estudos das propriedades psicométricas dos testes^{4,7}.

Apesar de a confiabilidade de um teste ser uma condição necessária, ela não é suficiente para garantir a validade de um instrumento de medida^{1,7,13,14}, visto que a validade de uma medida é considerada a característica mais importante do instrumento e a mais complexa de ser avaliada. No entanto, é importante ressaltar que quando sua confiabilidade é baixa, sua medida será imprecisa, o que torna limitada a validade das inferências obtidas a partir desse instrumento^{9,12,13,15}. Por outro lado, um instrumento que tem boa validade muito provavelmente também será um instrumento confiável^{11,13,15}.

O uso de instrumentos de medida é crescente nos últimos anos, e engloba diversas áreas profissionais⁴. Diante disso, é necessário que os instrumentos de medida apresentem altos níveis de qualidade. O presente artigo apresentou os principais métodos de avaliação psicométrica da confiabilidade de um instrumento de medida ditados pela TCT. A escolha do melhor método deve ser feita de forma criteriosa, levando em conta, sobretudo, as características do instrumento, a população avaliada e o tipo e nível de mensuração dos dados gerados.

Agradecimentos: À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pelo apoio para o desenvolvimento deste trabalho.

Referências

1. Silva JA, Ribeiro-Filho NP. Avaliação e mensuração da dor: pesquisa, teoria e prática. Ribeirão Preto: FUNPEC; 2006.
2. Pasquali, L. Psicometria. Rev Esc Enferm USP. 2009 dez;43:992-9.
3. Pasquali, L. Psicometria: teoria dos testes na psicologia e na educação. 3ª ed. Petrópolis: Vozes; 2009.
4. Sartes LMA, Souza-Formigoni MLO. Avanços na psicometria: da Teoria Clássica dos Testes à Teoria de Resposta ao Item. Psicol Reflex Crit. 2013;26(2):241-50.
5. Pasquali L, Primi R. Fundamentos da teoria de resposta ao item: TRI. Aval Psicol. 2003;2(2):99-110.
6. Couto G, Primi R. Teoria de resposta ao item (TRI): conceitos elementares dos modelos para itens dicotômicos. Bol Psicol. 2011;59(134):1-15.
7. Primi R. Psicometria: fundamentos matemáticos da teoria clássica dos testes. Aval Psicol. 2012;11(2):297-307.
8. Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. 2ª ed. Chichester: John Wiley & Sons; 2007.
9. Polit, DF, Beck CT. Essentials of nursing research: appraising evidence for nursing practice. 7ª ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2010.
10. Pilatti LA, Pedroso B, Gutierrez GL. Propriedades psicométricas de instrumentos de avaliação: um debate necessário. Rev Bras Ensinn Cienc Tecnol. 2010;3(1):81-91.
11. Martins GA. Sobre confiabilidade e validade. Rev Bras Gestão Negoc. 2006;8(20):1-12.
12. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 2008;65(23):2276-84.
13. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006 Feb;119(2):166.e7-16.
14. Campo-arias A, Oviedo HC. Propriedades psicométricas de uma escala: la consistencia interna. Rev Salud Publica. 2008;10(5):831-9.
15. Collares CF, Grec WLP, Machado JLM. Psicometria na garantia de qualidade da educação médica: conceitos e aplicações. Sci Health. 2012;3(1):33-49.
16. Carmines EG, Zeller RA. Reliability and validity assessment. 3ª ed. Beverly Hills: Sage Publications; 1979.
17. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Pers Assess. 2003 Feb;80(1):99-103.
18. Urbina S. Fundamentos de testagem psicológica. Porto Alegre: Artmed; 2007.
19. Alexandre NMC, Coluci MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medida. Cienc Saude Colet. 2011;16(7):3061-8.

Como citar este artigo:

Cunha CM, De Almeida Neto OP, Stackfleth R. Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. Rev. Aten. Saúde. 2016;14(49):98-103.