

# PRINCIPAIS MÉTODOS DE AVALIAÇÃO PSICOMÉTRICA DA VALIDADE DE INSTRUMENTOS DE MEDIDA

MAIN PSYCHOMETRIC EVALUATION METHODS  
OF THE VALIDITY OF MEASURING INSTRUMENTS

Cristiane Martins Cunha<sup>a\*</sup>, Omar Pereira de Almeida Neto<sup>b\*\*</sup>, Renata Stackfleth<sup>c\*</sup>

<sup>a</sup>crismcunha@gmail.com, <sup>b</sup>omarpneto@hotmail.com, <sup>c</sup>renatastack@gmail.com

\*Universidade de São Paulo – São Paulo (SP), Brasil

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

Data de recebimento do artigo: 27/05/2015

Data de aceite do artigo: 24/09/2015

## RESUMO

**Introdução:** É cada vez mais comum a publicação de estudos que demonstram a importância da adequada avaliação psicométrica dos instrumentos de medida, como sendo um critério de boa qualidade do instrumento, no que se refere às evidências de validade e confiabilidade. A validade de um instrumento de medida reúne um conjunto de evidências que comprovam que ele está de fato mensurando o constructo que supostamente pretende medir. Este artigo tem o objetivo de apresentar os princípios básicos de avaliação psicométrica da validade dos instrumentos de medida e os principais métodos estatísticos descritos na literatura. **Desenvolvimento:** 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 dez anos, e foram excluídos artigos duplicados ou que não respondiam ao objetivo desta revisão. As técnicas de avaliação da validade quantitativas que compõem o modelo trinário (validade de constructo, validade de conteúdo e validade de critério), têm se mostrado suficientes e satisfatórias na demonstração da validade de um instrumento. **Conclusão:** De um modo geral, os autores concordam com os critérios fundamentais de avaliação da validade de um bom instrumento de medida. A validade de um instrumento de medida é um atributo psicométrico presente em um bom instrumento e precisa ser contemplado nessas categorias de estudo.

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

## ABSTRACT

**Introduction:** The publication of studies demonstrating the importance of proper psychometric evaluation of measuring instruments as a good quality instrument criterion with regard to evidence of validity and reliability is increasingly common. The validity of a measuring instrument brings together a body of evidence proving that it is in fact measuring the construct that it is supposedly measuring. This article aims to present the basic principles of psychometric evaluation of the validity of measuring instruments and the main statistical methods described in the literature. **Development:** The survey was conducted from publications in MEDLINE, LILACS, and IBECs databases and also from the descriptions in related books. The studies' analyses involved reading titles, abstracts, and articles published in the last 10 years and duplicated articles or those not responding to the purpose of this review were excluded. The techniques of evaluation of quantitative validity that make up the trinary model (construct validity, content validity, and criterion validity) has proved sufficient and satisfactory in demonstrating the validity of an instrument. **Conclusion:** In general, the authors agree with the basic criteria for evaluating the validity of a good measuring instrument. The validity of a measuring instrument is an existing psychometric attribute in a good instrument and needs to be covered in these study categories.

**Keywords:** Validity of tests; measurements, methods, and theories; questionnaires; psychometrics.

## Introdução

A psicometria é a abordagem científica que visa à mensuração e avaliação psicológica dos constructos subjetivos por meio de escalas, testes e questionários padronizados, denominados “medida psicométrica”, os quais representam uma situação experimental que avalia um determinado constructo<sup>1</sup>. É uma complexa área do conhecimento que objetiva desenvolver e aplicar técnicas de mensuração dos fenômenos psíquicos<sup>2,3</sup>. 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)<sup>3</sup>.

A TCT continua sendo bastante utilizada pela comunidade científica, visto que seu foco de avaliação direcionado à qualidade do instrumento de medida, visando garantir que estes sejam válidos e confiáveis para medir determinados comportamentos humanos<sup>2,4</sup>.

Esse modelo considera, especificamente, os parâmetros relacionados ao teste como um todo, visto que as avaliações psicométricas consideram os escores totais da medida, por meio de correlações dos itens com os escores totais do teste<sup>5,6</sup>. Assim, os testes desenvolvidos segundo este modelo são elaborados para avaliar as aptidões ou habilidades populacionais medianas, sendo necessário que a amostra seja suficientemente grande e representativa<sup>3</sup>. 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<sup>5,6</sup>.

A validade é considerada um dos principais critérios que garantem a qualidade do instrumento, pois indica que o teste realmente mede o fenômeno de interesse pretendido<sup>1,7-16</sup>. 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 produzir graves consequências sobre as conclusões estatísticas de uma pesquisa<sup>17</sup>.

Conhecer os adequados métodos psicométricos de avaliação da validade de um instrumento de medida é o primeiro passo para garantir a boa qualidade do instrumento. Desse modo, este artigo teve como objetivo apresentar os princípios básicos de avaliação psicométrica da validade dos instrumentos de medida e os principais métodos estatísticos descritos na literatura.

Para tanto, realizou-se uma pesquisa nas bases de dados Medline, LILACS e IBECs, utilizando os descritores “validade dos testes”; “medidas, métodos e teorias”; “questionários” e “psicometria”, assim como, suas correspondentes traduções para o inglês “validity of results”; “measurements, methods and theories”; “questionnaires”; “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 sobre 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 nessa ordem, publicados nos últimos dez anos, e que estavam escritos em português, inglês e espanhol. Foram excluídos artigos duplicados, os que não estavam disponíveis para baixar e os que não correspondiam com o objetivo desta revisão.

## *Princípios teórico-metodológicos da avaliação psicométrica: validade*

A validade se refere à apropriação, significação e utilidade das inferências realizadas a partir dos escores obtidos pelos instrumentos de medida<sup>1</sup>. Trata-se de uma hipótese que visa comprovar a validade do instrumento através de testes empíricos baseados em um critério externo ou padrão-ouro<sup>2,14</sup>. Desse modo, a validade envolve um conjunto de evidências que comprova que uma escala está de fato mensurando o constructo que supostamente pretende medir<sup>8</sup>.

Martins<sup>11</sup>, Silva e Ribeiro-Filho<sup>1</sup>, Pasquali<sup>2</sup> e Collares, Grec e Machado<sup>16</sup> ressaltam que a validade de uma medida não depende dos seus escores e resultados, mas do objetivo e da seleção adequada do instrumento de medida, o qual permitirá compreender o real significado obtido por meio da adequada interpretação inferencial. Vários autores referem que a validade é uma questão de julgamento e interpretação dos escores do teste, e, por isso, é sempre importante considerar o “grau de validade” do teste<sup>7,10,12,13,18</sup>.

Silva e Ribeiro-Filho<sup>1</sup> e Lamprea e Gómez-Restrepo<sup>12</sup> descrevem que a validade de um teste permite fazer inferências e testar hipóteses acerca de uma determinada variável, pois ela fornece graus de garantia de que o instrumento mede sua pretensão. Collares, Grec e Machado<sup>16</sup> e Polit e Beck<sup>15</sup> reforçam que a validade designa o grau no qual as inferências dos escores de um teste são verdadeiras e corretas. Para isso, é necessário estabelecer que o instrumento de medida utilizado se relaciona com a variável investigada, para que os resultados dessa escala permitam fazer as inferências desejadas<sup>12</sup>.

Desse modo, entende-se que um teste é válido se ele mede o fenômeno de interesse pretendido<sup>1,7-16</sup> fundamentando-se em um teste paralelo (critério) que seja comprovadamente válido e confiável<sup>7</sup>. O processo de testagem e de apresentação de evidências da validade de uma medida denomina-se validação psicométrica<sup>8,18</sup>, com o intuito de operacionalizar um determinado constructo<sup>1</sup>.

Atualmente, há várias técnicas estatísticas para se estimar a validade dos instrumentos de medida<sup>1,14,16,19</sup>. Segundo Fayers e Machin<sup>8</sup> e Kimberlin e Winterstein<sup>13</sup>, o processo de validação é longo e diversificado, e envolve a reunião de variadas evidências ao longo do tempo e em diferentes populações. No entanto, as técnicas de validação que compõem o modelo trinário (validade de constructo, validade de conteúdo e validade de critério) têm se mostrado suficientes e satisfatórias na demonstração da validade<sup>2,8,10-12,14-16</sup>. Assim, apesar de a literatura apresentar variados métodos de avaliação da validade de um instrumento de medida, estes podem ser considerados variantes das técnicas propostas pelo modelo trinário<sup>7</sup>.

É importante ressaltar que os tipos de validade não devem ser vistos como entidades diferentes e independentes umas das outras, mas sim como critérios que visam estabelecer o grau de confiança das inferências acerca de uma determinada variável investigada<sup>12</sup>.

### *Validade de conteúdo*

Segundo Pasquali<sup>3</sup>, a validade de conteúdo de um teste consiste em verificar se o teste constitui uma amostra representativa do universo finito de comportamentos. Fayers e Machin<sup>8</sup> citam que a validade de conteúdo visa avaliar e confirmar se o instrumento reflete toda a dimensão teórica do construto pretendido, identificar possíveis discrepâncias e omissões no questionário, além de avaliar a aceitabilidade, compreensão, relevância e clareza dos itens. Collares, Grec e Machado<sup>16</sup> citam que a presença desse tipo de validade demonstra que a escolha e o formato dos itens do instrumento avaliado representam consistente e adequadamente o conteúdo que se pretende medir, e é um critério de boa qualidade do instrumento.

Para Pilatti, Pedroso e Gutierrez<sup>17</sup>, a validade de conteúdo é o julgamento da proporção em que os itens de um instrumento medem uma construção teórica representada por todas as facetas importantes do conceito medido. Menezes<sup>20</sup> cita que esse termo se refere ao julgamento se o instrumento realmente abrange os diferentes aspectos do seu objeto, e não contém elementos que podem ser atribuídos a outros objetos. Para Heynes<sup>21</sup>, a validade de conteúdo de um instrumento avaliativo envolve a validação e, em alguns momentos, a refinação de um constructo, e ela se refere a um traço e não a um estado.

Várias são as perspectivas que os autores utilizam para definir a validade de conteúdo. E, de um modo geral, os autores concordam que a validade de conteúdo expressa o grau de relevância dos itens de um teste na representação de todo o universo ou dimensão de

conteúdo teórico de um determinado fenômeno investigado. Entretanto, esse tipo de validade depende do grau e representatividade do constructo adotado pelos seus avaliadores<sup>1,8,12-15,21</sup>.

A validade de conteúdo é uma avaliação qualitativa usada tanto nos processos de construção de novos instrumentos de medida, quanto nos processos de adaptação cultural de instrumentos de medida preexistentes<sup>15,19</sup>. Consiste no julgamento subjetivo por um comitê de especialistas, que avaliam a proporção na qual os itens de uma medida determinam o mesmo conteúdo e se estes são relevantes e representativos de um determinado constructo<sup>8,12,16,19</sup>.

No entanto, por ser um processo de avaliação subjetiva, e pela abstração de seus constructos, é quase impossível garantir a sua representatividade teórica nos diferentes itens. Diante disso, alguns autores ressaltam que esse método é bastante susceptível a críticas, sendo necessário utilizar procedimentos de avaliação, tanto qualitativos como quantitativos, para comprovar a validade<sup>1,19</sup>. Dentro desse tipo de validade, destaca-se uma variante da validade de conteúdo, denominada “validade aparente”, as quais caminham juntas<sup>1,8</sup>.

A validade aparente (ou validade de face) se refere à coerência aparente entre o que se quer medir a partir dos itens do instrumento de medida utilizado, os quais são avaliados subjetivamente por um corpo de juízes especialistas na área<sup>10-12,14,15,17</sup>. Trata-se de uma técnica simples e necessária, porém pouco satisfatória que demonstra à primeira vista que o teste mede o que ele sugere<sup>1,11,15,18</sup>.

Fayers e Machin<sup>8</sup> reforçam que a validade de face faz parte da validação de conteúdo, só que a sua avaliação só ocorre após a construção do instrumento. Urbina<sup>18</sup> cita que a validade aparente se refere à aparência superficial daquilo que o teste mede na perspectiva de um testando ou de qualquer outro observador leigo. Nesse contexto, Cook e Beckman<sup>10</sup> ressaltam que nessa avaliação é necessária uma abordagem sistemática e documentada que assegure que o instrumento mede o constructo pretendido, mas reforçam que a simples avaliação da “aparência” não pode ser considerada como validade e, por isso, desencorajam o uso desse termo.

Sendo uma avaliação subjetiva, a seleção do corpo de juízes do comitê de avaliadores deve ser bastante criteriosa, sobretudo no que se refere à qualificação dos membros. Para isso, devem-se levar em conta as características do instrumento, a formação, a qualificação e a disponibilidade dos profissionais indicados<sup>1,22,23</sup>. Sendo assim, recomenda-se que os juízes tenham experiência clínica, sejam pesquisadores na área temática de avaliação e que conheçam o processo metodológico utilizado na construção e/ou adaptação de instrumentos de medida. Por outro lado, recomenda-se também a inclusão de pessoas

leigas relacionadas à população do estudo, sobretudo para avaliar o grau de compreensibilidade dos itens<sup>8,24,25</sup>.

Não há consenso na literatura sobre a quantidade de membros necessários para compor o comitê de juízes, mas algumas variáveis devem ser levadas em conta, tais como as características do instrumento, o método de análise que será utilizado ou mesmo a preferência do pesquisador. No entanto, recomenda-se que este comitê apresente número ímpar de membros e que, preferencialmente, sejam entre 5 e 10 juízes. A avaliação da validade de conteúdo deve ocorrer em duas formas distintas, por meio de uma avaliação quali-quantitativa<sup>17,21</sup>.

A avaliação qualitativa da validação de conteúdo propõe avaliar a qualidade do instrumento, sua clareza, pertinência, relevância e representatividade dos itens, e se cada domínio foi adequadamente abordado pelo conjunto de itens e se todas as dimensões do constructo foram incluídas<sup>8,19</sup>. Essa avaliação também envolve avaliar as equivalências semântica, idiomática, conceitual e cultural do instrumento<sup>19</sup>. Nesse tipo de avaliação, Guillemin, Bombardier e Beaton<sup>25</sup> ressaltam que para verificar a adequada compreensibilidade de linguagem do instrumento, ele precisa ser compreendido por indivíduos com idade entre 10 e 14 anos.

Já a avaliação quantitativa envolve vários métodos distintos para quantificar o grau de concordância entre os especialistas. Dentre os métodos mais utilizados, destacam-se o método da porcentagem de concordância entre os juízes, o índice de validade de conteúdo (IVC) e o coeficiente de Kappa. No método de porcentagem de concordância entre os juízes (também chamado de “concordância interobservadores”), admite-se um percentual mínimo de concordância de 90% entre os membros do comitê<sup>22,26</sup>.

Já o cálculo do IVC mede a proporção de juízes que estão em concordância com determinados aspectos do instrumento e de seus itens, e é um método bastante utilizado na área de saúde. Utiliza-se, geralmente, uma escala tipo Likert com pontuação que varia de 1 a 4 pontos (1, sem relevância, até 4, totalmente relevante) para assinalar quanto de relevância e pertinência o item representa o seu constructo. Nesse método, avalia-se a proporção de itens com avaliação de relevância com 3 ou 4 pontos. Não há consenso na literatura quanto ao nível de concordância mínima entre os autores, no entanto, alguns autores sugerem um índice de concordância mínimo de 0,80<sup>23,27</sup>, porém os valores recomendados são maiores que 0,90<sup>15,28</sup>.

O coeficiente de Kappa é um dos índices mais utilizados em pesquisas na área da saúde, e trabalha com a razão da proporção de vezes que os juízes concordam com a proporção máxima de vezes que os juízes poderiam concordar, ambas as variáveis corrigidas pela concordância ao acaso. É aplicável quando os dados são categóricos e estão

em uma escala nominal. Seus escores variam de -1 (sem concordância) a +1 (concordância total)<sup>14</sup>.

Lamprea e Gómez-Restrepo<sup>12</sup> citam que há uma discussão frequente sobre as escalas de medida muito heterogêneas, no que se refere à sua confiabilidade e à validade de conteúdo, com o intuito de reduzir o número de itens e ainda garantir a confiabilidade do instrumento. Conforme os autores, quando ocorre essa disparidade em um instrumento, é preferível garantir a validade de conteúdo de determinado teste para manter o seu poder inferencial do que optar por um teste confiável, mas que não mede toda a dimensão teórica do constructo avaliado. Portanto, é melhor que um teste tenha boa validade de conteúdo do que altos índices de confiabilidade<sup>12</sup>.

### *Validade de constructo*

A literatura apresenta uma definição de constructo bastante heterogênea. Para Urbina<sup>18</sup>, um constructo é qualquer coisa criada pela mente humana que não seja diretamente observável. Silva e Ribeiro-Filho<sup>1</sup> referem o constructo como sendo características não observáveis de que o pesquisador se utiliza para descrever comportamentos através de um dado critério ou escala; e que, por meio da validação de constructo se obtém um modelo teórico preliminar.

Para Martins<sup>11</sup>, constructo é uma variável operacional robusta que busca representar o verdadeiro significado teórico de um conceito, a partir do grau no qual uma medida se relaciona consistentemente com outras medidas semelhantes derivadas da mesma teoria. E Vianna<sup>29</sup> define que os constructos são traços, aptidões ou características supostamente existentes e abstraídos de uma variedade de comportamentos que possuam significado psicológico (ou educacional), tais como fluência verbal, motivação, inteligência. Segundo esse autor, a operacionalização do constructo ocorre mediante a utilização de testes ou instrumentos de medida adequadamente validados. A mensuração de um constructo requer que a definição conceitual do atributo avaliado seja traduzida para uma definição operacional, que é o indicador mais concreto para representar e quantificar o constructo<sup>13</sup>.

A validade de constructo se refere ao julgamento sobre a adequação das inferências determinadas pela escala em uma variável de um determinado constructo. As evidências necessárias para esse tipo de validação são obtidas fazendo-se uma série de estudos inter-relacionados, visando à verificação empírica, por meio de testes estatísticos, das construções teóricas sobre a relação entre as variáveis a serem medidas<sup>20</sup>.

Fayers e Machin<sup>8</sup> referem que a validade de constructo é a característica psicométrica mais importante de um instrumento de medida, pois avalia o grau no qual

um instrumento mede o construto pretendido a partir da avaliação de um modelo teórico hipotético, ou seja, a relação teórica existente entre duas medidas. Pasquali<sup>2</sup> considera também que a validade de construto é a forma mais fundamental de validade dos instrumentos, pois constitui a maneira mais direta de verificar a hipótese da legitimidade da representação comportamental dos atributos não observáveis, denominados de “traços latentes”<sup>2</sup>. De um modo geral, a validade de construto reúne várias evidências sobre a adequação do modelo teórico do instrumento, a partir da comprovação da relação teórica do instrumento com um critério externo na representação do construto avaliado<sup>8,13</sup>.

Alguns autores afirmam que a validação de construto é a forma mais abrangente e complexa das categorias de evidência de validade, sendo a mais difícil de ser determinada, pois é necessária a aplicação de vários testes estatísticos empíricos para demonstrar esse tipo de validade. Ela se refere ao grau no qual uma escala mede a teoria ou a hipótese investigada<sup>1,16,20</sup>. Portanto, segundo esses autores, a validade de construto é a habilidade de uma escala para confirmar as hipóteses esperadas, sendo um julgamento sobre a adequação das inferências dos escores da escala na mensuração de um determinado construto.

A lógica do processo da validação de construto é a mesma do método científico, pois parte de previsões acerca de sua dimensão teórica e, a partir daí, testam-se hipóteses sobre as possíveis relações que podem ser estabelecidas dentro desse construto. Essas relações podem ser explicadas pelo princípio teórico estabelecido por Cronbach e Meehl<sup>30</sup> denominado de “rede nomológica”. A rede nomológica é um modelo que explica, por meio de leis, as ligações (diretas ou não) existentes entre as dimensões teóricas (construtos) e os comportamentos observáveis (Tau). Essas leis estabelecem relações entre diferentes características, entre características e construtos ou entre um construto e outro; e ditam ainda que as leis e os conceitos estão ligados a comportamentos observáveis<sup>30</sup>. Segundo esses autores, o significado de um conceito somente fica totalmente esclarecido quando se elabora uma rede de relações que demonstre que essas relações são específicas e definitivas. Portanto, a validação de construto demonstra empiricamente essas relações estabelecidas pela rede nomológica, visando garantir a concordância entre os pesquisadores na avaliação de um mesmo construto<sup>29</sup>.

Desse modo, a validade de construto procura explicar e conectar uma dimensão teórica a um determinado fenômeno comportamental, através da análise empírica, por meio de aplicação de uma série de testes estatísticos baseados, sobretudo, em correlações entre a relação teórica dos itens de uma medida e as hipóteses investigadas<sup>8,15,16,20</sup>. Silva e Ribeiro-Filho<sup>1</sup> citam que a validação

de construto é um processo contínuo e envolve várias etapas de análises tanto do conteúdo dos itens que compõem o instrumento quanto das relações dessa escala com outras que medem o mesmo construto e da sua capacidade de prever comportamentos não relacionados diretamente com os escores do instrumento de medida.

Em termos práticos, Haynes<sup>21</sup> cita em seu estudo que a validade de construto de um instrumento influencia fortemente no julgamento clínico, pois fornece informações que influenciam na tomada de decisão por meio da formação de modelos, formulação de estratégias de intervenção, predição de comportamentos futuros e avaliação dos efeitos do tratamento. Assim, quanto maior for as fontes de evidências, mais confiante e válida será a interpretação dos resultados pelo pesquisador<sup>1,5,12,13,29</sup>.

Para investigação desse tipo de validade, uma das abordagens de avaliação é a técnica por grupos conhecidos. A validade por grupos conhecidos é o método mais simples de avaliar a validade de construto e constitui um subtipo de validade de construto baseada no princípio de que em determinados grupos de sujeitos são esperadas pontuações diferentes comparadas a outros e, para isso, o instrumento deve ser sensível a essa diferença. Assim, uma escala com evidências de validade seria aquela capaz de discriminar a diferença entre os grupos em uma determinada direção prevista na formulação das hipóteses do estudo<sup>8</sup>.

Segundo Polit e Beck<sup>28</sup>, definidos os grupos conhecidos, os quais são grupos supostamente diferentes em termos do atributo-alvo, esses são avaliados pelo instrumento e após comparar os escores de cada grupo, é possível detectar possíveis diferenças nos escores do teste em virtude de determinadas características grupais. É importante ressaltar que a validade de construto por grupos conhecidos relaciona-se com a sensibilidade e responsividade do instrumento<sup>8</sup>.

A sensibilidade do instrumento relaciona-se com o tipo da amostra e com a validade de construto por grupos conhecidos, no entanto, ela apenas detecta a magnitude da diferença de algumas variáveis com valor ou relevância clínica entre os grupos. Ao passo que a validade de construto por grupos conhecidos apenas confirma uma hipótese previamente estabelecida em um grupo<sup>8</sup>. Além da investigação da validade de construto por grupos conhecidos, dispõe-se de dois outros métodos de avaliação de construto muito importantes, denominados de “validade de construto convergente e divergente” (ou discriminante).

Vianna<sup>29</sup> ressalta que a compreensão de um instrumento que mede determinado construto ganha nova dimensão quando se estabelece o grau de associação com outras medidas de construtos semelhantes. Assim, a validade de construto possibilita determinar se o teste

investigado é a melhor medida para o que se pretendeu medir, por meio da testagem empírica de hipóteses de convergência (validade de constructo convergente) e de discriminatividade (validade de constructo discriminante) em relação às diferentes medidas.

A validade de constructo convergente (ou confirmatória) se refere à presença de correlação significativa entre a medida de investigação com uma medida de dimensão teórica que mede o mesmo fator e que seja empiricamente comprovada (válida e confiável)<sup>1,8</sup>. Portanto, segundo Vianna<sup>29</sup>, a demonstração estatística da validade de constructo convergente ocorre através da correlação significativa entre medidas de constructos afins, e demonstra evidências que indicam que os testes utilizados têm significados semelhantes que correspondem o mesmo constructo. É a presença de correlação dos itens entre os mesmos traços, por meio de diferentes métodos, e por isso relaciona-se com a consistência interna do instrumento<sup>8</sup>.

A validade de constructo divergente (ou discriminante) é a determinação da validade de um teste pela demonstração que a medida investigada não se relaciona comparativamente com uma variável ou uma medida de dimensão teórica não teoricamente relacionada<sup>1,8</sup> e se refere à possibilidade de diferenciar empiricamente um constructo de outros constructos semelhantes, assim como constatar o que não está correlacionado ao constructo<sup>29</sup>, ou seja, é a comprovação de diferentes traços, por meio de testes com constructos diferentes.

A análise multitraço-multimétodo (MTMM) é uma matriz estatística que apresenta os dados estatísticos agrupados e ordenados resultantes das correlações existentes entre os itens e uma escala, de modo que os itens representam os traços e as escalas, o método. Essa matriz de correlação possibilita visualizar em uma tabela a presença de traços entre os instrumentos, mediante as correlações entre o item e a dimensão a qual pertence e entre o item e a dimensão a qual não pertence. Desse modo, a matriz multitraço-multimétodo é uma estratégia para se avaliar a validade de constructo convergente e divergente, por meio de correlações entre diferentes instrumentos. Caso haja correlação moderada entre o item e a dimensão a qual ele pertence, existe validade convergente, ao passo que se não existir correlação entre o item e outra dimensão, fala-se em validade divergente. No entanto, esse método de avaliação é dificultado quando um instrumento é muito longo. Nessas condições, há a possibilidade de aplicar a análise multitraço-escala<sup>8</sup>.

Além das correlações, a validade de constructo engloba uma série de complexas técnicas e métodos estatísticos com o intuito de avaliar a dimensionalidade e a homogeneidade dos itens na avaliação de uma determinada variável latente<sup>8</sup>. Segundo Collares, Grec e Machado<sup>16</sup>, a validade de constructo pode ser demonstrada por

meio de técnicas de análise fatorial confirmatória e de modelagem de equações estruturais, a qual denomina-se “validação fatorial”. Esse é um método de validação de constructo que visa avaliar a dimensionalidade e os componentes de um instrumento, podendo ser de forma confirmatória ou exploratória.

A análise fatorial se baseia em um conjunto de procedimentos matemáticos usados para identificar conglomerados ou agrupamentos de variáveis a partir da análise de intercorrelações, de modo que cada agrupamento (ou fator) se correlaciona altamente entre si do que com variáveis de outros agrupamentos não relacionados<sup>1</sup>. Esse método permite verificar a dimensionalidade de um instrumento de medida, fundamentando-se pelo princípio da parcimônia, ou seja, um grande número de variáveis observadas pode ser explicado por um número menor de variáveis hipotéticas, não observadas. Esse modelo estatístico também permite a avaliação quantitativa do grau de adequação dos dados à sua teoria, por meio dos índices *goodness-of-fit* e dos resíduos padronizados<sup>16</sup>.

### *Validade de critério*

A validade de critério é considerada a mais importante das validades estatisticamente determinadas, pois ela se refere à extensão na qual um instrumento demonstra uma associação com um critério ou uma medida externa independente (paralela ou externa), considerada como um padrão-ouro na área de interesse. Portanto, sua ênfase está na predição do teste avaliada por critérios externos com validades confirmadas empiricamente na avaliação do mesmo constructo que se quer medir<sup>1,8,12,13,15,16</sup>.

Segundo Fayers e Machin<sup>8</sup>, a validação de critério avalia se a medida investigada possui relação empírica com critérios externos, comprovadamente validados, que avaliam o mesmo constructo. Collares, Grec e Machado<sup>16</sup> acrescenta que as evidências de validade de critério são obtidas por meio da comparação entre o instrumento investigado e um outro instrumento (critério externo) que possui o mesmo fim, mas que já possui evidências comprovadas de sua validade.

Segundo Lamprea e Gómez-Restrepo<sup>12</sup>, a seleção do critério externo ou padrão-ouro pode ser tanto um instrumento de medida consagrado ou até mesmo uma variável clínica, a exemplo de resultados de exames anatomopatológicos (biópsias). Silva e Ribeiro-Filho<sup>1</sup> citam que o critério pode ser qualquer desempenho adequado e escalonado que seja válido, relevante e não contaminado, tais como desempenho em uma prova, sucesso de uma terapia, escore de um teste etc. De qualquer modo, o pesquisador precisa se atentar para a seleção de um

critério externo adequado, sobretudo quando se tem o intuito de validar um novo instrumento de medida. Além de apresentar boa correlação com o instrumento estudado, são necessários outros atributos favoráveis ao critério externo, tais como boa aceitabilidade e aplicabilidade para uso clínico, menor custo, menor risco para o paciente, não envolver a realização de procedimentos invasivos e ter maior simplicidade de avaliação<sup>1,12</sup>.

A verificação estatística da validade de critério ocorre mediante comprovação da associação existente entre a medida investigada e o seu critério externo. Há duas variantes desse tipo de validade, a validade concorrente e a validade preditiva, as quais diferem entre si apenas por um critério temporal de avaliação da predição<sup>1,3,8,13,16,20</sup>. Desse modo, quando a medida de investigação e o critério são aplicados simultaneamente, e objetiva obter respostas imediatas e a tomada de decisões, fala-se em validade concorrente; ou quando o critério é avaliado no futuro e visa obter predições em relação a algum fator, fala-se em validade preditiva<sup>3,8,12,15,16,20</sup>.

Urbina<sup>18</sup> descreve que a validade de critério concorrente é apropriada para escores de testes que serão empregados para determinar o status atual de uma pessoa em relação a uma determinada variável (que pode ser de desempenho, ocupacional, clínica ou diagnóstica), no momento em que o estudo está sendo conduzido. Para Silva e Ribeiro-Filho (2006), se os escores de um teste são obtidos aproximadamente ao mesmo tempo, os indicadores da relação entre os escores da escala e do critério fornecem evidência de validade concorrente. Portanto, essa validade se refere ao grau no qual a medida de investigação se relaciona com outra medida (ou critério externo) com predição do teste obtida no mesmo momento, quando as medidas são aplicadas simultaneamente<sup>1,2,8,12,13,15,18</sup>.

Em relação à validade de critério preditiva, Urbina<sup>18</sup> ressalta a utilidade desse tipo de validade para tomada de decisões baseada nas estimativas dos níveis de desempenho ou resultados comportamentais futuros. Segundo essa autora, os procedimentos de validação preditiva requerem que sejam coletados dados sobre a variável preditora e que aguardem até que os dados de critério se tornem disponíveis para que os dois conjuntos de dados possam ser correlacionados. Silva e Ribeiro-Filho<sup>1</sup> referem que os escores são usados para prever comportamentos ou desempenho em um critério predeterminado. Os escores podem ser obtidos em um momento, e as medidas do critério obtidas um tempo depois e, usualmente, após alguma estratégia interveniente ter sido implementada (a exemplo, treinamento, medicação etc.).

Dessa forma, as medidas da relação entre os escores do teste e o do critério, obtidos um tempo depois, fornecem um indicativo prognóstico e de validade preditiva do instrumento. Assim, o clínico está interessado

na utilidade dos escores do instrumento como forma de prever o comportamento futuro dos pacientes. Portanto, a validade de critério preditiva refere-se ao grau no qual a medida de investigação se relaciona com outra medida de um critério externo com predição do teste que será obtida futuramente<sup>1,2,8,12,13,15</sup>.

Segundo Urbina<sup>18</sup> em muitos casos, essas duas variantes se confundem; e nesses casos, a distinção entre elas depende muito do modo como o pesquisador faz a pergunta que se quer responder. Pasquali<sup>2</sup> ressalta que o fato da informação ser obtida simultaneamente ou futuramente não é um fator relevante para a validação de um teste, mas relevante é saber se o critério escolhido foi adequado e comprovadamente válido. Além disso, é importante que o pesquisador tenha bastante cautela na seleção do critério, pois nem todos os instrumentos são considerados critérios bons e adequados<sup>1,8</sup>.

A validade de critério é considerada a mais importante das validades determinadas estatisticamente<sup>1,8,20</sup>. Ela avalia o grau com que o instrumento de investigação discrimina indivíduos com diferentes características de acordo com um critério padrão<sup>1,21</sup>. No entanto, Pasquali<sup>2,3</sup> questiona toda essa importância desse tipo de validade. Segundo o autor, essa validade só faz sentido se existirem testes comprovadamente válidos que possam servir de critério contra o qual se quer validar um novo teste e que este novo teste tenha vantagens sobre o antigo. Assim, o questionamento do autor está no fato de, por que validar um novo teste utilizando-se de uma medida como padrão-ouro, supostamente considerada como sendo inferior? Diante disso, esse autor considera que, para a validação psicométrica de instrumentos, a avaliação da validade de constructo seja superior à validade de critério<sup>2</sup>. Por outro lado, Cook e Beckman<sup>10</sup> consideram que todos os outros tipos de validade são variações da validade de constructo e representam diferentes categorias reunidas que evidenciam a validade de constructo.

## Conclusão

Para responder alguma questão, o pesquisador pode decidir por utilizar um instrumento de medida existente ou construir um novo instrumento. Do ponto de vista do custo-benefício, sempre é 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 psicométrica<sup>13</sup>. 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<sup>8,10,13</sup>.

Desse modo, a literatura tem focado, cada vez com maior ênfase, a importância da adequada avaliação psicométrica dos instrumentos de medida, sobretudo no que se refere às evidências de sua validade e confiabilidade. E, de um 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 às suas vantagens sobre a teoria clássica dos testes<sup>4,8-10,11,13,16-19</sup>. A literatura demonstrou que as técnicas de avaliação da validade quantitativas, que compõem o modelo trinário (validade de constructo, validade de conteúdo e validade de critério), têm se mostrado suficientes e satisfatórias na constatação da validade de um instrumento.

A validade é um atributo psicométrico mais importante que precisa estar presente em um bom instrumento de medida, visto que sua ausência pode produzir graves consequências sobre as conclusões estatísticas de uma pesquisa<sup>17</sup>, porém é a mais complexa de ser avaliada. No entanto, é importante avaliar também sua confiabilidade, pois quando ela é baixa, sua medida será imprecisa, o que torna limitada a validade das inferências obtidas a partir desse instrumento<sup>10,13,15,16</sup>. Assim, um instrumento que tem boa validade, muito provavelmente também será um instrumento confiável<sup>10,11,15</sup>.

O uso de instrumentos de medida é crescente nos últimos anos, e engloba variadas áreas profissionais<sup>4</sup>. Diante disso, é necessário que os instrumentos de medida apresentem altos níveis de qualidade. Dessa forma, este artigo apresentou os principais métodos de avaliação psicométrica da validade de um instrumento de medida, ditados pela teoria clássica dos testes. A escolha do melhor método deve ser criteriosamente considerada, 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 da Esc de Enfermagem da USP. 2009a;43(especial):992-9.
3. Pasquali, L. Psicometria: Teoria dos testes na psicologia e na educação. 3ª ed. Petrópolis: Vozes; 2009b.
4. Sartes LMA, Souza-Formigoni MLO. Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. Psicologia: Reflexão e Crítica. 2013;26(2):241-50.
5. Pasquali L, Primi R. Fundamentos da teoria de resposta ao item: TRI. Avaliação Psicológica. 2003;2(2):99-110.
6. Couto G, Primi R. Teoria de resposta ao item (TRI): conceitos elementares dos modelos para itens dicotômicos. Boletim de Psicologia. 2011;LXI(134):1-15.
7. Pasquali L. Validade dos testes psicológicos: será possível reencontrar o caminho? Psicologia: Teoria e Pesquisa. 2007;23(especial):99-107.
8. Fayers PM, Machin D. Quality of life. Assessment, analysis, and interpretation. The assessment, analysis, and interpretation of patient-reported outcomes. 2nd ed. Chichester: John Wiley & Sons; 2007.
9. Primi R. Psicometria: fundamentos matemáticos da teoria clássica dos testes. Avaliação Psicológica. 2012;11(2):297-307.
10. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. The Am Journal of Med. 2006;119(2):166.e7-16.
11. Martins GA. Sobre confiabilidade e validade. RBGN. 2006;8(20):1-12.
12. Lamprea JA, Gómez-Restrepo C. Validez en la evaluación de escalas. Rev Colombiana Psiquiatr. 2007;36(2):340-8.
13. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 2008;65(23):2276-84.
14. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Delineando a pesquisa clínica. 3ª ed. Porto Alegre: Artmed; 2008.
15. Polit, DF, Beck CT. Essentials of nursing research: appraising evidence for nursing practice. 7th ed. Philadelphia: Wolters Kluwer; 2010.
16. Collares CF, Grec WLP, Machado JLM. Psicometria na garantia de qualidade da educação médica: conceitos e aplicações. Scien in Health. 2012;3(1):33-49.
17. Pilatti LA, Pedroso B, Gutierrez GL. Propriedades psicométricas de instrumentos de avaliação: um debate necessário. R.B.E.C.T. 2010;3(1):81-91.
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. Ciên Saúde Coletiva. 2011;16(7):3061-8.
20. Menezes PR. Validade e confiabilidade das escalas de avaliação em psiquiatria [Internet]. 2013 [acesso em 12 de agosto. 2015]. Disponível em: <http://hcnet.usp.br/ipq/revista/vol25/n5/conc255b.htm>.
21. Haynes SN, Richard DCS, Kubany CS. Content validity in psychological assessment: A functional approach to concepts and methods. Psychological Assessment. 1995;7(3):238-47.
22. Lynn MR. Determination and quantification of content validity. Nursing Research. 1986;35(6):381-5.
23. Grant JS, Davis LL. Selection and use of content experts



- for instrument development. *Research Nursing of Health*. 1997;20(3):269-74.
24. Tilden VP, Nelson CA, May BA. Use of qualitative methods to enhance content validity. *Nursing Research*. 1990;3(3):172-5.
  25. Guillemin F. Cross cultural adaptation and validation of health status measures. *Scand J Rheumatol*. 1995;24(2):61-3.
  26. Topf M. Three estimates of interrater reliability for nominal data. *Nursing Research*. 1986;35(4):253-45.
  27. Davis LL. Instrument review: getting the most from a panel of experts. *Appl Nurs Res*. 1992;5(4):194-7.
  28. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res Nur Health*. 2006;29(5):489-97.
  29. Vianna HM. Validade de constructo em testes educacionais. In: Vianna HM. *Testes em educação*. São Paulo: Ibrasa; 1987. p. 35-44.
  30. Cronbach LJ, Meehl PE. Construct validity and psychological tests. *Psychological Bulletin*. 1955;52:281-302.

---

**Como citar este artigo:**

Cunha CM, Neto OPA, Stackfleth R. Principais métodos de avaliação psicométrica da validade de instrumentos de medida. *Rev. Aten. Saúde*. 2016;14(47): 75-83.