

## AVALIAÇÃO E MENSURAÇÃO DE VARIÁVEIS PSICOSSOCIAIS: desafio para pesquisa e clínica de enfermagem

Dálete Delalibera Correa de Faria MOTA<sup>a</sup>  
Cibele Andruccioli de Mattos PIMENTA<sup>b</sup>

### RESUMO

A avaliação e mensuração de variáveis psicossociais são um desafio. Construtos psicossociais são complexos e há desconfiança sobre a validade em se utilizar instrumentos padronizados para sua avaliação; há grande variedade de instrumentos e confusão entre os termos utilizados para descrever testes de validade e fidedignidade. Trata-se de artigo de atualização com referencial teórico da psicometria. Os objetivos foram: analisar as características gerais e psicométricas de instrumentos de avaliação de fenômenos subjetivos, estabelecer paralelo entre termos similares e apresentar guia para a escolha dos melhores instrumentos. Aprimorar instrumentos de medida levará ao aperfeiçoamento da pesquisa e clínica de enfermagem.

**Descritores:** Avaliação. Psicometria. Medidas. Validade dos testes. Reprodutibilidade dos testes.

### RESUMEN

*La evaluación y la medida de variables psicosociales son un desafío. Los constructos psicosociales son complejos y existe desconfianza sobre la validez de usar instrumentos estandarizados para su evaluación. Hay una gran variedad de instrumentos y confusión de los términos que describen las pruebas de validez y fiabilidad. Este artículo adopta la referencia teórica de la psicometría. Su propósito fue analizar las características generales y psicométricas de los instrumentos de evaluación de fenómenos subjetivos, establecer un paralelo entre términos similares y presentar una guía para la selección de los instrumentos. Realzar instrumentos de medida llevará al perfeccionamiento de la investigación y atención de enfermería.*

**Descriptores:** Evaluación. Psicometría. Medidas. Validez de las pruebas. Reproducibilidad de resultados.

**Título:** Evaluación y medición de variables psicosociales: un desafío para la investigación y la clínica en enfermería.

### ABSTRACT

*Assessment and measurement of psychosocial variables are challenging. These constructs are complex, and there is distrust as to the validity of using standardized instruments for their measurement. There is a large variety of instruments and much confusion regarding the terms used to describe the validity and reliability tests. This paper adopts the theoretical reference of psychometrics. Its purpose is to update knowledge on general and psychometric characteristics of instruments for the assessment of subjective phenomena, to establish a parallel between similar terms, and to present a guide for choosing instruments. The enhancement of measurement instruments will improve nursing research and attention.*

**Descriptors:** Evaluation. Psychometrics. Measures. Validity of tests. Reproducibility of results.

**Title:** Assessment and measurement of psychosocial variables: a challenge for nursing research and attention.

<sup>a</sup> Enfermeira. Mestre em Enfermagem na Saúde do Adulto e Doutoranda do Programa de Pós-Graduação em Enfermagem na Saúde do Adulto da Escola de Enfermagem da Universidade de São Paulo (USP).

<sup>b</sup> Enfermeira. Doutora em Enfermagem. Professora Titular do Departamento de Enfermagem Médico-Cirúrgica da Escola de Enfermagem da USP.

## 1 INTRODUÇÃO

O cuidado com a saúde das pessoas envolve atenção para diversos aspectos de natureza psicossocial. Identificar a presença de tais aspectos, conhecer sua magnitude e compreender suas relações é fundamental para o adequado atendimento por qualquer profissional de saúde.

Fenômenos subjetivos podem e devem ser mensurados. Sua mensuração é essencial na pesquisa científica e na prática clínica e constituem-se num desafio. Questionários, inventários, testes e escalas, instrumentos de lápis e papel utilizados para mensurar atributos subjetivos às vezes são considerados inadequados, por anularem a subjetividade das pessoas e distorcerem os resultados. Para esses argumentos há respostas.

A mensuração de fenômenos subjetivos é feita de forma indireta – é possível conhecer a altura da torre pelo tamanho da sua sombra<sup>(1)</sup>. Em algumas situações já é bastante conhecida e aceita, como por exemplo, na avaliação do quociente de inteligência e do humor, mas é freqüente o conceito de que mensuração de fenômenos objetivos é livre de erros. No entanto, a toda medida corresponde uma margem de erro, mesmo em se tratando de fenômenos objetivos. O que varia é a magnitude do erro. Utilizar instrumentos de medida padronizados permite conhecer o tamanho do erro de avaliação, facilita a explicitação do tamanho do erro, além de permitir o aperfeiçoamento do instrumento.

Questiona-se a possibilidade dos instrumentos de medida para avaliação de fenômenos subjetivos anularem a individualidade das pessoas, o que não pode ser totalmente negado, nem aceito. Reconhecem-se na humanidade aspectos verdadeiramente peculiares a cada indivíduo, por outro lado, observam-se pontos semelhantes, ou seja, características universais, comuns a todos. O método de construção de instrumentos padronizados envolve a amostragem de comportamentos (características de um dado fenômeno) e representatividade da amostra, o que tende a assegurar que as características universais dos indivíduos relacionados a um determinado fenômeno sejam contempladas. A construção de um instrumento de medida é fruto do somatório da experiência de várias pessoas. É possível, porém, explorar as características individuais de outras formas, por exemplo, utilizando-se análises qualitativas.

Outro aspecto, fonte de crítica aos instrumentos de medida de fenômenos subjetivos, é a indisposição ocasional do indivíduo; uma pessoa poderia ser avaliada em um dia ruim e isso resultaria em dados falseados. Cabe considerar que o indivíduo bom tende a ser melhor que fraco mesmo num dia ruim<sup>(1)</sup> e que a representatividade da amostra (origem, tamanho e aleatoriedade) tende a minimizar o erro.

Há outras vantagens atribuídas aos instrumentos de medida: os investigados são expostos a estímulos controlados (mesmas perguntas, mesmas condições de abordagem, etc.); os dados são rapidamente conhecidos, podendo ser comparados e o seu uso acelera a aquisição de experiência. Alguém pode aprender mais rapidamente a avaliar com precisão determinado atributo subjetivo, por exemplo, a depressão, por meio de um instrumento, ao invés de conseguir dados apenas a partir de sua experiência.

Visando tornar os instrumentos de medida válidos e fidedignos, muito já se produziu na área da psicometria, mas esse assunto ainda precisa ser melhor divulgado entre os pesquisadores de modo geral e, especificamente, na área de enfermagem. Na literatura observa-se confusão no uso dos termos, quer seja pelo uso de classificações diferentes, pela tradução inadequada ou por desconhecimento, o que dificulta o aprendizado e predispõe a erros.

Os objetivos deste texto são apresentar os principais conceitos relacionados à validade, fidedignidade e outras características de um instrumento de medida, com base no referencial da psicometria, e apresentar um guia contendo critérios para a escolha dos melhores instrumentos.

## 2 CARACTERÍSTICAS GERAIS E PSICOMÉTRICAS DOS INSTRUMENTOS

A decisão por determinado instrumento de medida deve ser calcada na análise de suas características, tais como conceituação do fenômeno, descrição do método de construção, índices de validade e fidedignidade e propósito para o qual o instrumento foi construído, entre outros aspectos. Essas características assim como as informações acerca da validade e fidedignidade serão apresentadas a seguir<sup>(2-4)</sup>.

Os instrumentos podem variar conforme o número de fenômenos avaliados e dimensões in-

cluídas. Os instrumentos que avaliam uma gama de fenômenos ou conjunto de outros sintomas são denominados genéricos, e os que avaliam apenas um fenômeno ou sintoma, são conhecidos como instrumentos específicos. É desejável o uso de instrumentos específicos para avaliação de sintomas ainda pouco conhecidos que necessitam maior exploração. Os instrumentos que avaliam uma única dimensão são denominados unidimensionais, e os que avaliam mais de uma dimensão, multidimensionais. Os unidimensionais não permitem uma descrição completa do fenômeno. É freqüente que a definição de um sintoma/fenômeno seja multidimensional e o ideal seria que a avaliação contemplasse todas as suas dimensões.

A conceituação do objeto, ou seja, a definição do fenômeno investigado é fator importante e deve ser explicitado no instrumento para que o investigador e investigado estejam cientes do que se quer conhecer. A definição conceitual do fenômeno justifica o conteúdo do instrumento e evidencia a teoria que o fundamenta<sup>(2)</sup>. Observa-se que grande número de instrumentos não apresenta a definição do fenômeno, o que torna quase impossível saber o porquê da inclusão de determinados itens. Para certos fenômenos não há completo consenso quanto à definição, o que pode ser causa de confusão ao se comparar resultados de avaliações que deveriam estar mensurando o mesmo fenômeno.

Grande parte dos instrumentos para a avaliação de fenômenos subjetivos é de auto-relato, isto é, avaliam exclusivamente o relato do sujeito. O indivíduo é quem vai descrever a sua experiência. O auto-relato é de grande importância quando se trabalha com sintoma subjetivo. As escalas de auto-relato são consideradas econômicas quanto ao tempo despendido, porém apresentam desvantagens pela restrição do uso em indivíduos com menor grau de instrução, pela exigência de que as pessoas sejam cooperativas e não tenham limitações que impeçam sua aplicação de forma adequada<sup>(5)</sup>. Acrescido a isso, observa-se que as pessoas têm dificuldades para avaliar a gravidade de seus próprios sintomas. Às vezes, por razões como quererem ser aceitos ou chamarem a atenção, supervalorizam ou subestimam a queixa (desejabilidade social).

Outros quesitos na elaboração dos instrumentos são sua extensão, tempo necessário para

respondê-los e linguagem. Esses três aspectos estão intimamente relacionados e devem ser analisados cuidadosamente. Linguagem complexa, instrumentos extensos e que demandam muito tempo para sua aplicação tendem a ser fatigantes e desestimulantes e podem ser causa de erros e de menor fidedignidade. Recomenda-se o uso de instrumentos simples e curtos ou medianamente extensos (8 a 30 itens).

Quanto a finalidade, um instrumento classifica-se em discriminativa, preditiva e avaliativa<sup>(6)</sup>. Os discriminativos são usados para distinguir os indivíduos que apresentam o fenômeno em estudo daqueles que não o apresentam. Por exemplo, no caso de avaliação de fadiga, diferencia indivíduos fatigados dos não fatigados. Os instrumentos preditivos são usados para prever quais indivíduos desenvolverão uma condição específica ou em que situações um resultado específico será alcançado. Por exemplo, prever qual pessoa apresenta risco para desenvolver úlcera por pressão. Os avaliativos são utilizados para medir a magnitude da mudança do fenômeno em um indivíduo, num determinado período de tempo. No caso da dor, é o que avalia quão intensa é a dor.

A explicitação da finalidade dos instrumentos demonstra objetividade na construção do instrumento. E, segundo a finalidade do instrumento, diferentes testes de validade são apropriadas.

A validade é uma propriedade psicométrica dos instrumentos, considerada indispensável. Há várias formas de se analisar a validade de um instrumento e as mais comumente usadas são por julgamento ou de modo empírico. Em todas, o objetivo é averiguar se o instrumento está medindo o fenômeno que se propôs medir. Não existe uma estratégia única para sua confirmação e a escolha das estratégias dependerá da finalidade do instrumento.

A validade por julgamento permite a análise da validade de conteúdo e validade aparente. As validades de conteúdo e aparente são avaliações subjetivas realizadas por *experts* no assunto. O primeiro permite verificar se o instrumento contém todos os componentes e domínios relevantes relacionados ao fenômeno. O segundo, também denominado de validade de face, permite avaliar se o instrumento mede as qualidades desejadas<sup>(2,3)</sup>. A validade por julgamento possibilita, basicamente, concluir se determinado instrumento está apro-

priado para alcançar seu objetivo. Vale ressaltar que esse tipo de estratégia não é suficiente para garantir a validade de um instrumento, pois não há dados obtidos de modo empírico.

As estratégias de modo empírico permitem a análise da validade de critério, da validade discriminante e da validade de construto. A validade de critério busca avaliá-la por meio da correlação existente entre o instrumento que se quer testar e outro instrumento que possua atributos iguais ou similares, geralmente um instrumento “padrão-ouro”. Validade de critério é dividida em validade concorrente e preditiva<sup>(2,3)</sup>. A concorrente é avaliada pela aplicação de dois instrumentos equivalentes ao mesmo tempo, nos mesmos sujeitos e seguida da comparação dos resultados. A preditiva refere-se à acurácia de um instrumento em prever um evento futuro.

É freqüente o uso da técnica de validação convergente e de validação divergente. A convergente é averiguada pela existência de correlação positiva entre o construto mensurado pelo instrumento em estudo e outras variáveis com as quais tal construto deveria estar relacionado<sup>(4)</sup>. A divergente é verificada pela comparação de resultados obtidos por dois instrumentos que avaliam fenômenos opostos, distintos ou fenômenos independentes. Busca-se a correlação negativa ou a não correlação entre os fenômenos, por exemplo, fadiga e energia, fadiga e condição social, respectivamente<sup>(7)</sup>.

A validade discriminante é analisada pela comparação entre dois ou mais grupos populacionais nos quais se espera diferença quanto à quantidade do fenômeno de interesse detectável pelo instrumento em estudo. Os conceitos de sensibilidade e especificidade estão associados a essa estratégia. A sensibilidade refere-se à capacidade do instrumento em identificar corretamente o fenômeno em indivíduos que o apresentam. Especificidade refere-se à capacidade do instrumento em determinar corretamente os indivíduos que não apresentam o fenômeno.

A validade de construto, também conhecida como validade de conceito, determina a adequação do instrumento à definição conceitual do fenômeno investigado<sup>(2)</sup>. Essa validade permite tanto a avaliação da teoria que fundamenta a elaboração do instrumento, como possibilita a avaliação do próprio instrumento<sup>(2,3)</sup>. Ao contrário dos outros

testes de validade, a validade de construto ocorre num processo contínuo, pois à medida que se avança no conhecimento sobre o fenômeno, novas hipóteses são construídas e novos testes precisam ser empregados<sup>(2)</sup>. Para se avaliar essa validade, realiza-se análise fatorial ou análise por hipótese.

A análise fatorial permite determinar quais construtos (domínios) compõem o fenômeno de interesse. Analisa-se a magnitude da relação entre as diversas variáveis empíricas (itens) que se julga compor um domínio (construto). Supõe-se que, se os itens estão interrelacionados, é porque têm uma “causa” comum que produz essa interrelação<sup>(4)</sup>. A carga fatorial para um item ser confirmado em um determinado domínio varia entre 0,30<sup>(8)</sup>, 0,40<sup>(9)</sup> e 0,65<sup>(10)</sup>.

A análise por hipótese refere-se às estratégias de validade discriminante e de critério<sup>(4)</sup>. Pode-se dizer que várias estratégias para se analisar a validade por meio de correlações (validade de critério, discriminante, convergente e divergente) são apenas componentes da validade de construto pois, na verdade, toda validade busca a validade de construto.

Há tendência de se reduzir os testes de validade dos instrumentos à validade de construto<sup>(4)</sup>. É importante lembrar que qualquer estratégia isolada é insuficiente para provar que um instrumento é válido<sup>(2)</sup>. Ainda, a validação de instrumentos é um processo, isto é, a avaliação das propriedades psicométricas de um instrumento ocorre a cada situação em que o instrumento é aplicado e, gradativamente, ajustado.

É importante lembrar a questão do número de sujeitos na amostra de validação do instrumento. Quando a amostra é grande e variável, a tendência é de se aumentar o índice de fidedignidade<sup>(4)</sup>. Recomenda-se, como regra geral, o uso da maior amostra possível e sugere-se o uso de 10 sujeitos por item do instrumento<sup>(11)</sup>. Esse critério (utilizar 10 sujeitos por item) é atendido por boa parte dos instrumentos.

Fidedignidade, também conhecida por confiabilidade e precisão, é outra propriedade de extrema importância para um instrumento ter boa qualidade psicométrica. Essa propriedade refere-se ao quão consistente, exato e estável é um instrumento<sup>(2)</sup>. A fidedignidade é a propriedade que garante que o instrumento mede o objeto ao qual

se propõe, de forma reproduzível<sup>(3)</sup>. As principais formas de se obter a fidedignidade de um instrumento são por meio do cálculo da consistência interna, da estabilidade, da utilização de formas alternativas ou através da confiabilidade inter-observadores.

A consistência interna é obtida através de uma única administração do instrumento avaliando-se a correlação entre o resultado de um item e o resultado obtido nos demais. Nos casos em que os testes são destinados a avaliar mais de um atributo (subescalas ou domínios), a consistência interna deve ser determinada para cada um deles. O meio mais comum para calcular essa correlação é através do Alpha de Cronbach. Atribui-se valor de 0 a 1, sendo que 0 indica “sem correlação” e 1 indica “correlação perfeita”. Coeficientes em torno de 0,90 são os mais adequados, coeficientes em torno de 0,80 são considerados razoáveis e abaixo de 0,70 são considerados insuficientes como demonstração de fidedignidade de um instrumento<sup>(4)</sup>. Na literatura, porém, são aceitos coeficientes acima de 0,50 como razoáveis<sup>(12)</sup>. Outra maneira de se calcular a consistência interna é através da fórmula de número 20 de Kuder-Richardson (KR-20). É utilizada quando os itens do instrumento são dicotômicos<sup>(4)</sup>. Os valores variam entre 0 e 1, e 1 é o valor máximo possível para a consistência interna.

O teste-reteste é a estratégia utilizada para analisar a estabilidade ou reprodutibilidade de uma medida. É obtido através da aplicação do instrumento em duas ocasiões diferentes, em um mesmo indivíduo, nas quais se espera que o fenômeno não tenha se alterado e que o resultado do instrumento foi mantido. O intervalo ideal entre as aplicações do instrumento é difícil de ser determinado. Se o intervalo for curto, a memória poderá interferir e, se for longo, fatores intrínsecos ou extrínsecos ao sujeito podem afetar a resposta<sup>(4)</sup>.

Ao contrário do teste-reteste que busca averiguar se o fenômeno permaneceu inalterado, a responsividade, outra propriedade psicométrica, avalia se o instrumento é sensível à mudança do fenômeno<sup>(6)</sup>. Testes que avaliam a responsividade são aplicados quando se quer mostrar que o instrumento é capaz de detectar alteração no fenômeno. Em assim sendo, o teste da responsividade é indispensável para instrumentos com fi-

nalidade avaliativa, visto que instrumentos com essa finalidade buscam analisar a sensibilidade do instrumento à mudança do fenômeno. Os resultados da responsividade podem sofrer a influência do número de sujeitos, assim como no teste-reteste.

Formas alternativas são demonstradas por resultados idênticos ou muito similares obtidos por meio da utilização de formas paralelas de um instrumento, ou seja, pela equivalência de duas medidas. Quanto mais próximo o coeficiente de correlação de 1, mais preciso é o teste e quanto mais distante do coeficiente 1, pouca ou nenhuma equivalência/concordância existe, prejudicando a qualidade psicométrica de fidedignidade<sup>(2)</sup>.

A confiabilidade inter-observadores, também chamada de confiabilidade inter-juízes, é averiguada por meio da aplicação do instrumento por dois observadores diferentes, de modo independente, buscando o grau de associação entre eles<sup>(2)</sup>. Quanto maior for o acordo entre os achados dos observadores, maior a fidedignidade.

A partir do descrito, elaborou-se guia para a escolha de um instrumento para avaliação de fenômenos subjetivos:

- a) possuir definição do fenômeno;
- b) ser medianamente extenso, com uso de linguagem simples e baseado em relato; do entrevistado (instrumento de auto-relato);
- c) avaliar múltiplas dimensões do fenômeno;
- d) ser fidedigno; apresentar, minimamente, consistência interna adequada;
- e) explicitar se a finalidade é discriminativa, avaliativa e/ou preditiva;
- f) ser válido; apresentar dados sobre a validade de conteúdo, validade de construto e, dependendo da finalidade do instrumento, dados sobre validade discriminante ou responsividade;
- g) estar adaptado para a cultura do país e para a população.

### 3 CONSIDERAÇÕES FINAIS

O conhecimento sobre um fenômeno avança à medida que os instrumentos para sua avaliação são aprimorados e, à medida que os instrumentos tornam-se mais válidos e fidedignos, aprimora-se o conhecimento sobre o fenômeno. A

construção de um instrumento de medida é fruto do desejo de pesquisadores de avaliar um conceito sob uma nova perspectiva, abordar novas dimensões ainda não investigadas ou aperfeiçoar um instrumento já existente. Se antigamente fenômenos subjetivos tais como fadiga, dor, auto-eficácia, depressão e qualidade de vida, entre outros, eram considerados imensuráveis, atualmente dispomos de diversos instrumentos para a sua avaliação, os quais devem ser incorporados à clínica e à pesquisa em enfermagem, visando aperfeiçoá-las.

### REFERÊNCIAS

- 1 Medeiros EB. Medidas psico&lógicas: introdução à psicometria. Rio de Janeiro: Ediouro; 1999.
- 2 McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. New York: Oxford University; 1996.
- 3 Streiner DL, Norman GR. Heath measurement scales: a practical guide to their development and use. New York: Oxford Medical; 1995.
- 4 Pasquali L. Psicometria: teoria e aplicações. Brasília (DF): Editora Universidade de Brasília; 1997.
- 5 Marcolan JF. Análise comparativa das escalas psicométricas de depressão: um subsídio para a avaliação clínica do enfermeiro psiquiátrico [tese de Doutorado]. São Paulo: Escola de Enfermagem, Universidade de São Paulo; 2002. 163 f.
- 6 Kirshner B, Guyatt G. A methodologic framework for assessing health indices. *Journal of Chronic Diseases* 1985;38:27-36.
- 7 Clark-Carter D. Doing quantitative psychological research: from design to report. Sussex: Psychological; 1999.
- 8 Kline P. An easy guide to factor analysis. London: Routledge; 1994.
- 9 Schwartz JE, Jandorf L, Krupp LB. The measurement of fatigue: a new instrument. *Journal of Psychosomatic Research* 1993;37(7):753-62.
- 10 Okuyama T, Akechi T, Kuhaya A, Okamura H, Shima Y, Maruguchi M, *et al.* Development and validation of the Cancer Fatigue Scale: a brief, three-dimensional, self-rating scale for assessment of fatigue in cancer patients. *Journal of Pain Symptom Manage* 2000;19(1):5-14.
- 11 Kerlinger FN. Foundations of behavioral research. New York: Holt, Rinehard and Winston; 1986.
- 12 Bowling A. Measuring disease: a review of disease-specific quality of life measurement scales. Philadelphia: Open University; 1995.

---

**Endereço da autora/Author's address:**  
Dálate Delalibera Corrêa de Faria Mota  
Rua 03 de Maio, 130, Aptº. 82, Vila Clementino  
04.044-020, São Paulo, SP  
E-mail: [dalate.mota@globocom](mailto:dalate.mota@globocom)

Recebido em: 12/06/2006  
Aprovado em: 17/11/2006