

**A confiabilidade no desenvolvimento e avaliação de instrumentos de medida na área da saúde***Reliability in the development and evaluation of measurement instruments in the health field**Confiabilidad en el desarrollo y evaluación de instrumentos de medición en el área de salud*Neusa Maria Costa Alexandre<sup>1</sup>, Cristiane Helena Gallasch<sup>2</sup>, Maria Helena Melo Lima<sup>3</sup>, Roberta Cunha Matheus Rodrigues<sup>4</sup>

<sup>1</sup> Enfermeira, Doutora em Enfermagem. Professora Associada do Departamento de Enfermagem da Faculdade de Ciências Médicas (FCM) da Universidade de Campinas (UNICAMP), Campinas, SP, Brasil. E-mail: [neusalex@fcm.unicamp.br](mailto:neusalex@fcm.unicamp.br).

<sup>2</sup> Enfermeira, Doutora em Enfermagem. Campinas, SP, Brasil. E-mail: [cristiane\\_gallasch@yahoo.com.br](mailto:cristiane_gallasch@yahoo.com.br).

<sup>3</sup> Enfermeira, Doutora em Biologia Funcional e Molecular. Professora Doutora do Departamento de Enfermagem da FCM/UNICAMP. Campinas, SP, Brasil. E-mail: [melolima@fcm.unicamp.br](mailto:melolima@fcm.unicamp.br).

<sup>4</sup> Enfermeira, Doutora em Enfermagem. Professora Livre-Docente do Departamento de Enfermagem da FCM/UNICAMP. Campinas, SP, Brasil. E-mail: [robertar@fcm.unicamp.br](mailto:robertar@fcm.unicamp.br).

**RESUMO**

A avaliação de intervenções e tratamentos atualmente procura englobar variáveis subjetivas, destacando a importância da perspectiva do paciente sobre a doença e os efeitos do tratamento. Estas medidas devem ser confiáveis, satisfazendo propriedades básicas, como validade, confiabilidade, sensibilidade e responsividade que representam um amplo e extenso repertório metodológico para avaliação. O presente estudo teve como objetivo realizar uma revisão de literatura sobre o conceito e métodos de avaliação da confiabilidade de instrumentos de medidas da área de saúde, com levantamento bibliográfico nas principais bases de dados desta área. Após seleção e análise dos textos, as informações foram organizadas, apresentando bases conceituais e métodos utilizados para avaliar a confiabilidade destes instrumentos, incluindo estabilidade, equivalência e consistência interna. O estabelecimento da confiabilidade é essencial para demonstrar a qualidade metodológica de instrumentos de medida. Além disso, as demais propriedades psicométricas de um instrumento também devem ser verificadas, garantindo a fidedignidade do processo de avaliação.

**Descritores:** Estudos de Validação; Psicometria; Reprodutibilidade dos Testes; Questionários.

**ABSTRACT**

The evaluation of interventions and treatments currently seeks to comprise subjective variables, highlighting the importance of the patient's perspective regarding the disease and treatment effects. These measurements must be reliable and meet basic properties such as validity, reliability, sensitivity and responsiveness, which represent a broad and extensive methodological evaluation repertoire. The objective was to perform a literature review regarding the concept and methods of evaluating the reliability of measurement instruments, through a bibliographic search on the main databases in the health field. After selecting and analyzing the texts, the information was organized, and a presentation was made of the conceptual bases and methods used to evaluate the reliability of those instruments, including stability, equivalence and internal consistency. It is essential to establish reliability to demonstrate the methodological quality of measurement instruments. Furthermore, the other psychometric properties of an instrument must be verified, thus guaranteeing the trustworthiness of the evaluation process.

**Descriptors:** Validation Studies; Psychometrics; Reproducibility of Results; Questionnaires.

**RESUMEN**

La evaluación de intervenciones y tratamientos actualmente procura incluir variables subjetivas, destacando la importancia de la perspectiva del paciente sobre la enfermedad a efectos del tratamiento. Estas medidas deben ser confiables, satisfaciendo propiedades básicas como validez, confiabilidad, sensibilidad y responsividad, que representan un extenso repertorio metodológico para evaluación. El presente estudio objetivó realizar una revisión de literatura sobre el concepto y métodos de evaluación de confiabilidad de instrumentos de medidas del área de salud, con relevamiento bibliográfico en las principales bases del área. Luego de selección y análisis de los textos, las informaciones fueron organizadas, presentando bases conceptuales y métodos utilizados para evaluación de confiabilidad de estos instrumentos, incluyendo estabilidad, equivalencia y consistencia interna. El establecimiento de confiabilidad es esencial para demostrar la calidad metodológica de instrumentos de medición. Además, las otras propiedades psicométricas de un instrumento deben también ser verificadas, garantizando que el proceso de evaluación sea fidedigno.

**Descriptor:** Estudios de Validación; Psicometría; Reproducibilidad de Resultados; Cuestionarios.

## INTRODUÇÃO

A avaliação de intervenções e tratamentos tem procurado englobar variáveis subjetivas, envolvendo diversas medidas, como incapacidade, severidade e intensidade dos sintomas, bem como aspectos como a satisfação com o tratamento e efeitos adversos de substâncias medicamentosas<sup>(1-2)</sup>. Cada vez mais se destaca a importância de considerar a perspectiva do paciente sobre aspectos da doença e dos efeitos do tratamento<sup>(3)</sup>.

Há um constante aumento de estudos para o desenvolvimento de instrumentos de avaliação que refletem as perspectivas dos pacientes, e que podem verificar a eficácia de tratamentos, influenciando as decisões no cuidado à saúde<sup>(1,4-5)</sup>. Essas avaliações podem incluir sintomas, funcionalidade, bem-estar, qualidade de vida, percepções sobre o tratamento, a satisfação com o atendimento recebido e com a comunicação profissional<sup>(3)</sup>. Além disso, questionários podem ser utilizados também para auditorias, avaliação de serviços de saúde, ensaios clínicos e avaliação de custos<sup>(6)</sup>.

A literatura tem descrito medidas e escalas que podem verificar inúmeras dimensões. Muitas destas representam abstrações, ou seja, construções teóricas que objetivam organizar e atribuir significados a percepções que não podem ser mensuradas por atributos ou indicadores, sendo chamadas de construtos teóricos<sup>(2,7)</sup>.

O desenvolvimento de escalas ou instrumentos traz um meio para a avaliação comportamental do construto, proporcionando a medida do mesmo. Essas medidas podem ser entendidas como mecanismos que associam conceitos abstratos com indicadores observáveis e mensuráveis<sup>(8)</sup>. Por isso, a importância de se determinar o quanto um específico indicador realmente representa um conceito teórico a ser medido<sup>(9)</sup>, uma vez que os instrumentos de medida devem ser confiáveis e apropriados para determinada população<sup>(5)</sup>.

Pesquisadores têm também alertado sobre a relevância de uma correta avaliação das qualidades psicométricas dos instrumentos de coleta de dados<sup>(10-12)</sup>. Essas qualidades são particularmente importantes ao se optar pelo desenvolvimento ou realização da adaptação cultural de instrumentos que serão usados tanto em pesquisas como na prática clínica<sup>(13)</sup>. Todas as medidas, seja uma avaliação da qualidade de vida ou a mensuração da pressão arterial, devem satisfazer algumas

propriedades básicas tais como: validade, confiabilidade, sensibilidade, praticabilidade e responsividade<sup>(14-15)</sup>. A determinação dessas propriedades é necessária para verificar a qualidade metodológica de instrumentos utilizados na avaliação do estado de saúde de pacientes<sup>(16)</sup>.

Pesquisas e estudos da área de saúde devem estabelecer e documentar a validade e a confiabilidade dos instrumentos de medida utilizados na coleta de dados<sup>(17)</sup>, destacando a importância da utilização dessas duas formas de avaliação das qualidades psicométricas de um instrumento de medida.

A confiabilidade verifica a homogeneidade, redundância ou heterogeneidade de um instrumento, por meio da capacidade de reproduzir resultados, mesmo que em diferentes condições, como, por exemplo, na utilização de diferentes itens para um grupo semelhante de indivíduos (consistência interna), ao longo do tempo (teste-reteste), ou entre indivíduos em diferentes ocasiões (intra-observadores)<sup>(18)</sup>.

Considerando o vasto conteúdo sobre avaliação de propriedades psicométricas, questiona-se: quais as mais recentes publicações relacionadas à avaliação da confiabilidade como propriedade psicométrica de um instrumento de medida? Quais aspectos têm sido considerados mais relevantes para realização desta avaliação?

Dentro deste contexto, o presente estudo teve como objetivo realizar uma revisão de literatura, apresentando bases conceituais e métodos de avaliação da confiabilidade de instrumentos de coleta de dados utilizados na área de saúde.

## MÉTODOS

Trata-se de uma revisão de literatura<sup>(19)</sup> que seguiu os passos metodológicos estabelecidos pela literatura para elaboração deste tipo de estudo, incluindo elaboração de pergunta norteadora, busca na literatura, coleta de dados, análise crítica para seleção, estudos e levantamentos dos dados de interesse, resultados e discussão e síntese dos resultados<sup>(19)</sup>.

Para o levantamento bibliográfico, optou-se por bases de dados das Ciências da Saúde, incluindo a Biblioteca Virtual em Saúde – Enfermagem (BDENF) e o Portal de Revistas de Enfermagem; o Scientific Electronic Library Online (SciELO); a Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS); a National

Library of Medicine-USA (MEDLINE), o International Nursing Index (INI) e o Cumulative Index to Nursing and Allied Health Literature (CINAHL) e SciVerse Scopus.

Ainda dentro do levantamento bibliográfico, utilizaram-se critérios de tipo de publicação, considerando apenas textos na íntegra, de periódicos e livros de metodologia de pesquisa e estatística; como também idioma, utilizando artigos nacionais e internacionais, publicados em português ou inglês.

Os descritores em português pesquisados por meio dos Descritores em Ciências da saúde (DeCS/BIREME) foram: *Estudos de Validação; Tradução; Validade dos Testes; Reprodutibilidade dos Testes e Questionários*. Em inglês a pesquisa dos descritores foi realizada na National Library of Medicine's Medical Subject Headings (MeSH/Medical Subject Heading Terms), utilizando-se:

*Validation studies; Translations, Psychometrics; Questionnaire design e Reproducibility of results.*

Após seleção dos textos, foi organizada a revisão de modo a apresentar as bases conceituais e os métodos de medidas usados para avaliar a confiabilidade, com ênfase em sua aplicação na área da saúde.

## RESULTADOS E DISCUSSÃO

Seguindo os critérios descritos anteriormente para levantamento bibliográfico, e após a análise crítica das publicações encontradas, foram selecionadas e utilizadas 25 publicações, sendo quatro de livros publicados em inglês, duas de livros em português, 16 artigos de periódicos publicados na língua inglesa e três de publicação na língua portuguesa brasileira, demonstrados nos Quadros 1 e 2.

**Quadro 1:** Síntese dos resultados encontrados para análise de confiabilidade – Livros.

Referência	Ano	País	Contribuição conceitual – Confiabilidade			
			Bases conceituais	Estabilidade	Equivalência	Consistência interna
Fayers PM, Machin D. Assessment, analysis and interpretation <sup>(14)</sup> .	2007	Inglaterra		X		X
Contandriopoulos AP, Champagne F, Potvin L, Denis JL, Boyle P. Saber preparar uma pesquisa <sup>(20)</sup> .	1999	Brasil	X		X	
Dempsey PA, Dempsey AD. Using nursing research <sup>(21)</sup> .	1996	Estados Unidos	X			X
Polit DF, Hungler BP. Fundamentos de pesquisa em enfermagem <sup>(22)</sup> .	1995	Brasil	X	X	X	X
Burns N., Grove SK. The practice of nursing research: conduct, critique & utilization <sup>(23)</sup> .	1997	Estados Unidos	X	X	X	X
LoBiondo-Wood G, Harber J. Nursing Research: methods, critical appraisal, e and utilization <sup>(24)</sup> .	1998	Estados Unidos		X	X	X

**Quadro 2:** Síntese dos resultados encontrados para análise de confiabilidade – artigos completos.

Referência	Ano	Periódico	Contribuição conceitual – Confiabilidade			
			Bases conceituais	Estabilidade	Equivalência	Consistência interna
Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient for reliability and validity of patient-reported outcome measures? <sup>(1)</sup>	2007	Value Health.	X			
Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research <sup>(2)</sup> .	2008	Am. J. Health. Syst. Pharm	X		X	
Turner RR, Quittner L, Parasuraman BM, Kallich JD, Cleeland CS, 2007, Estados Unidos. Patient-Reported Outcomes: Instrument development and selection issues <sup>(4)</sup> .	2007	Value Health.	X			
Braga CG, da Cruz DA. Contribuições da psicometria para a avaliação de respostas psicossociais na enfermagem <sup>(7)</sup> .	2006	Rev. Esc. Enferm. USP	X			
Terwee CB, Bot SDM, Boer MR, van der Windt DAWM, Knol DL., Dekker J et al. Quality criteria were proposed for measurement properties of health status questionnaires <sup>(16)</sup> .	2007	J. Clin. Epidemiol.		X		X
Selby-Harrington ML, Mehta SM, Jutsum V, Riortella-Muller R, Quade D. Reporting of instrument validity and reliability in selected clinical nursing journals <sup>(17)</sup> .	1994	J. Prof. Nurs.	X			X
DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, Savov SM, Kostas-Polston E. A psychometric toolbox for testing validity and reliability <sup>(25)</sup> .	2007	J. Nurs. Scholarsh.	X	X	X	X
Salmond SS. Evaluating the reliability and validity of measurement instruments <sup>(26)</sup> .	2008	Orthop. Nurs.	X		X	X
Fitzner K. Reliability and validity: a quick review <sup>(27)</sup> .	2007	Diabetes Educ.	X			
Roberts P, Priest H, Traynor M. Reliability and validity in research <sup>(28)</sup> .	2006	Nurs. Stand.	X			
De Vet HCW, Terwee CB, Bouter LM. Current challenges in clinimetrics <sup>(29)</sup> .	2003	J. Clin. Epidemiol.	X			
Alexandre NMC, Coluci, MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medida <sup>(15)</sup> .	2011	Cienc. Saude Colet.		X		X
Dimitrov D, Rumrill . Reliability in rehabilitation measurement <sup>(30)</sup> .	2001	Work.		X		X
Switzer GE, Wisniewski SR, Belle SH, Dew MA, Schultz R. Selecting, developing, and evaluating research instruments <sup>(31)</sup> .	1999	Soc. Psychiatry. Psychiatr. Epidemiol.		X	X	X
Houser J. Precision, reliability, and validity: essential elements of measurement in nursing research <sup>(32)</sup> .	2008	J. Spec. Pediatr. Nurs.			X	X
Keszei A, Novak M, Streiner DL. Introduction to health measurement scales <sup>(33)</sup> .	2010	J. Psychosom. Res.				X
Rattray J, Jones MC. Essential elements of questionnaire design and development <sup>(34)</sup> .	2007	J. Clin. Nurs.				X
Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires <sup>(35)</sup> .	2007	Optom. Vis. Sci.				X
Mota DDCF, Pimenta CAM. Avaliação e mensuração de variáveis psicossociais: desafio para pesquisa e clínica de enfermagem <sup>(36)</sup> .	2007	Rev. Gaúcha Enferm.				X

## SÍNTESE DOS RESULTADOS

### Confiabilidade – bases conceituais

A confiabilidade é a capacidade em reproduzir um resultado de forma consistente no tempo e no espaço, ou com observadores diferentes<sup>(20)</sup>. Pode ser considerada a capacidade de um instrumento mensurar um atributo de uma forma consistente<sup>(25)</sup>. Se um instrumento é confiável, ele não deve variar sob certas condições ambientais. Deve também apresentar resultados estáveis e consistentes ao longo do tempo ou quando utilizado por dois diferentes pesquisadores<sup>(26)</sup>. A confiabilidade tem o objetivo de promover pesquisas confiáveis e fidedignas, ou seja, diminuindo a chance de distorção dos dados ou da ocorrência de erros<sup>(2,4,7,27-28)</sup>. A confiabilidade é uma qualidade necessária, mas que deve ser avaliada em conjunto com outros componentes, como a validade<sup>(25)</sup>.

Na área de Ciências Sociais, o conceito de confiabilidade deve ser aplicado em todas as medidas como inventários de personalidades, testes educacionais, pesquisas, questionários de atitudes, entre outros. Deve também fazer parte da avaliação de instrumentos usados na clínica como entrevistas, exame físico, exames de laboratório, radiológicos e histológicos<sup>(29)</sup>. No entanto, a literatura tem ressaltado que a confiabilidade não é uma propriedade estática do instrumento, devendo ser reavaliada em diferentes populações<sup>(17)</sup>.

Diferentes contextos em pesquisa podem afetar a confiabilidade, como o tipo e número de pacientes estudados, o método de administração do instrumento, período considerado, análise estatística, entre outros<sup>(1)</sup>. Por isso, recomenda-se que ao se publicar estudos sobre o desenvolvimento e avaliações de questionários e escalas deve-se fornecer algumas informações detalhadas que irão auxiliar a confirmar os atributos psicométricos desses instrumentos. Dentre essas informações destaca-se: o tipo de sujeitos, o tamanho amostral, o método de administração do instrumento, o procedimento de pontuação, a análise estatística e a evidência de estabilidade dos achados<sup>(1)</sup>.

A confiabilidade é geralmente expressa em forma numérica denominada coeficiente. Um instrumento de medida que apresenta uma confiabilidade perfeita deve ter um coeficiente de +1<sup>(21)</sup>. Os coeficientes de confiabilidade que são designados como um  $r$ , variam de 0.00 a um máximo de 1.00. Quanto mais elevada a pontuação, mais estável é o instrumento de medida<sup>(22)</sup>.

Valores satisfatórios de confiabilidade indicam que um instrumento tem medida consistente. É um modo de demonstrar ou comunicar o rigor do processo científico e a fidedignidade das informações, descrevendo o quanto um teste se mostra particular e irá reproduzir, em diferentes circunstâncias, resultados similares, demonstrando que nada se alterou<sup>(25,27-28)</sup>.

Os seguintes procedimentos são utilizados para sua avaliação: estabilidade (teste-reteste), equivalência (inter-observadores) e consistência interna<sup>(23)</sup>. A escolha de um desses procedimentos vai depender do tipo de instrumento de medida<sup>(37)</sup>.

### Estabilidade

Geralmente denominada confiabilidade teste-reteste. Consiste na administração do mesmo instrumento aos mesmos sujeitos sob condições semelhantes em duas ou mais ocasiões<sup>(24)</sup>. O uso desse procedimento requer que o fato a ser medido deva permanecer o mesmo em duas avaliações e que qualquer alteração no valor do escore é devido a um erro aleatório<sup>(23)</sup>. É indicado para avaliar características relativamente estáveis<sup>(22,30)</sup>.

O intervalo entre as administrações depende do fenômeno que está sendo avaliado. Há controvérsias sobre o tempo necessário entre a aplicação do teste original e o reteste. Estudos apresentam intervalos que variam entre horas, até meses. É importante considerar um intervalo longo o suficiente para que o paciente não recorde as respostas do primeiro momento, porém não tão extenso a ponto de modificar o objeto de estudo<sup>(1,14,25)</sup>, como por exemplo, sintomas de uma determinada patologia. Apesar de apresentarem uma sugestão de uma a duas semanas, autores ressaltam que não se deve fixar um tempo, mas o importante é descrever e justificar a escolha do período utilizado<sup>(16)</sup>.

Para análise estatística dos dados, recomenda-se o uso do Coeficiente de Correlação Intraclasse (ICC) para medidas contínuas e o Coeficiente de Kappa para as dicotômicas ou ordinais<sup>(16,31)</sup>. Sugere-se um valor de confiabilidade de 0,70 com o ICC ou o Coeficiente Kappa, com uma amostra de pelo menos 50 sujeitos<sup>(16)</sup>. Recomenda-se também valores de 0,70 a 0,90 quando trata-se de estudos que envolvem medidas de personalidade, interesse e atitudinais<sup>(15)</sup>.

## Equivalência

É a concordância entre observadores que utilizam um mesmo instrumento de medida ou a concordância entre formas alternadas de um instrumento<sup>(26)</sup>. Dessa forma, a equivalência pode ser verificada por dois métodos. A consistência de formas paralelas consiste em um mesmo instrumento com diferentes apresentações produzir os mesmos resultados<sup>(17,26)</sup>. Muitas medidas relacionadas com as áreas de educação e psicologia têm diferentes versões de instrumentos que avaliam o mesmo constructo<sup>(26)</sup>. Nesse procedimento, os mesmos sujeitos devem completar os dois instrumentos no mesmo período de tempo. Esse método é geralmente usado em estudos educacionais, não sendo muito empregado em pesquisa clínica<sup>(23,25)</sup>.

O outro método, denominado avaliação da consistência interavaliadores ou interobservadores, é a capacidade de um teste aplicado por múltiplos avaliadores alcançar os mesmos resultados<sup>(2, 21-23, 26-29)</sup>. O objetivo desse procedimento é a avaliação da equivalência do instrumento quanto a sua capacidade de mensurar os mesmos traços nos mesmos sujeitos<sup>(22)</sup>. É apropriado para dados que envolvem apreciações ou julgamentos dos avaliadores como em entrevistas e observações<sup>(31)</sup>.

Uma forma é quando dois ou mais indivíduos observam e registram determinado evento simultaneamente, de forma independente, usando um protocolo de estudo<sup>(23)</sup>. Os observadores devem ser treinados ou orientados e precisa ser usado com instrumentos observacionais estruturados<sup>(22,24)</sup>. Assim, esse tipo de confiabilidade depende de definições precisas das variáveis envolvidas e requer observadores bem treinados<sup>(2)</sup>. Uma outra maneira é quando um mesmo avaliador observa e registra um evento em duas ocasiões. Para uma avaliação adequada é necessário observar pelo menos 10 sujeitos ou eventos<sup>(23)</sup>.

Os valores da confiabilidade devem sempre serem apresentados quando os dados são coletados por meio de observação ou quando utilizam dois ou mais avaliadores<sup>(23)</sup>.

Pode ser determinada pela correlação dos escores de dois ou mais observadores ou pelo coeficiente de concordância, ou *agreement*, das apreciações de avaliadores<sup>(2)</sup>. Para analisar emprega-se o coeficiente de confiabilidade para verificar o poder da concordância entre os escores dos dois observadores. Os valores

podem variar de 0.00 a 1.00, sendo que os valores mais elevados indicam um maior grau de equivalência. Há publicações que indicam que o valor da confiabilidade interobservadores não deve ser abaixo de 0.80<sup>(27)</sup>, enquanto outros autores recomendam que esse valor deve ser, no mínimo, de 0.85<sup>(32)</sup>.

## Consistência interna

A consistência interna verifica a consistência entre os vários itens que compõem um instrumento, tendo como base a correlação média entre esses itens<sup>(15)</sup>. Este procedimento verifica, então, a extensão que todos os itens mensuram o mesmo constructo<sup>(16,23)</sup>. Assim, um instrumento apresenta consistência interna ou homogeneidade quando todas as suas subpartes mensuram o mesmo conceito<sup>(22)</sup>. As questões da escala devem se correlacionar ou serem complementares umas com as outras<sup>(24)</sup>. Se a escala é formada por diferentes domínios mensurando diferentes características, a consistência interna deve ser avaliada separadamente para cada parte<sup>(25,30)</sup>.

Pode ser avaliada por meio de quatro métodos: metade dividida, correlações totais, coeficiente de Kuder-Richardson e alfa de Cronbach<sup>(24)</sup>.

A metade dividida é uma técnica de divisão em metades, sendo considerado o método mais antigo. Compreende dividir uma escala em duas metades e fazer uma comparação<sup>(26)</sup>. No caso das duas partes estarem mensurando o mesmo atributo, o coeficiente de confiabilidade deve ser elevado<sup>(22)</sup>. Pode-se dividir as questões em números pares e ímpares, dividir simplesmente entre a primeira e a segunda metade, ou selecionar e comparar as questões de uma forma aleatória<sup>(24)</sup>. Esse método é efetivo para questionários longos<sup>(21)</sup>.

O método das correlações totais mensura a relação entre cada uma das questões e a escala total. Encontra-se uma correlação para cada questão. Questões que não apresentam uma confiabilidade adequada pode ser eliminada do instrumento<sup>(24)</sup>.

O coeficiente de Kuder-Richardson (KR-20) é usado para instrumentos que têm formato com resposta dicotômica, como por exemplo, respostas de sim ou não<sup>(24)</sup>.

Finalmente, o método estatístico Alfa de Cronbach compara cada questão de uma escala simultaneamente uma com a outra<sup>(21)</sup> e mede a correlação média entre

todos os itens<sup>(33)</sup>, devendo ter um valor maior ou igual a 0,70<sup>(25-26,32)</sup>.

Encontra-se também que valores acima de 0,70 são aceitáveis, acima de 0,80 são bons e superando 0,90 são excelentes<sup>(14)</sup>. Outros autores defendem que deve apresentar um valor superior a 0,70 para questionários em desenvolvimento ou 0,80 para os que já estão sendo largamente utilizados<sup>(34)</sup>. Como o alfa de Cronbach é determinado pela média dos coeficientes de correlação entre os itens, valores excepcionalmente elevados podem sugerir redundância<sup>(35)</sup>. O valor vai também depender do número de questões que compõem o instrumento, aumentando com uma elevação dos itens<sup>(31,33)</sup>. É muito utilizado no desenvolvimento de escalas. Neste caso, pode ser usado para decidir as questões que devem ser eliminadas<sup>(35)</sup>. Se o alfa de Cronbach muda pouco quando um item é omitido, este item pode ser removido do questionário<sup>(14)</sup>.

Por medir as respostas dos sujeitos da pesquisa em curso, deve ser verificada cada vez que o instrumento for utilizado<sup>(25)</sup>. A confiabilidade também pode ser nomeada como precisão, fidedignidade e constância<sup>(7,36)</sup>.

Na análise das publicações relacionadas à avaliação de todas as propriedades psicométricas de instrumentos de medida na área da saúde, há destaque na utilização da confiabilidade e da validade.

A determinação da confiabilidade é um passo essencial ao se desenvolver ou selecionar um específico instrumento de medida para ser utilizado na prática clínica ou em pesquisas. O estabelecimento dessa propriedade é necessário para demonstrar a qualidade metodológica de instrumentos utilizados na avaliação do estado de saúde de pacientes<sup>(16)</sup>.

Porém, para correta avaliação e resultados fidedignos, os passos metodológicos devem ser rigorosos, assim como a análise estatística dos resultados obtidos. Foi encontrada na literatura concordância entre

os autores quanto à realização dos procedimentos de avaliação incluindo: estabilidade (teste-reteste), homogeneidade e equivalência (inter-observadores). Não há variação significativa entre os autores em relação aos escores utilizados na avaliação estatística desses resultados.

## CONSIDERAÇÕES FINAIS

Estudos sobre o desenvolvimento, adaptação cultural e avaliações de questionários e escalas devem fornecer informações sobre os diferentes contextos que podem afetar a confiabilidade, tais como o tipo de sujeitos, o tamanho amostral, o método de administração do instrumento, o procedimento de pontuação, a análise estatística e a evidência de estabilidade dos achados. Essas informações irão auxiliar a confirmar as qualidades psicométricas desses instrumentos.

Este estudo buscou demonstrar, por meio de revisão da literatura, os passos metodológicos e parâmetros estatísticos para avaliação da confiabilidade de um instrumento de medidas na área da saúde, mecanismo essencial a fim de garantir resultados fidedignos na realização de pesquisas e na avaliação de sujeitos com a utilização destas ferramentas.

A avaliação das propriedades psicométricas de instrumentos de medida em saúde é reconhecida como metodologia essencial para obtenção de dados confiáveis. A verificação da confiabilidade não deve ser definida como único elemento. Deve-se também discutir outras propriedades, como validade e responsividade, que não foram objeto deste estudo.

Apesar da avaliação da confiabilidade ser extensa, envolvendo diferentes procedimentos de avaliação, as demais propriedades psicométricas de um instrumento, como validade e responsividade, também devem ser verificadas para garantir a fidedignidade do processo de avaliação.

## REFERÊNCIAS

1. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health*. 2007;10 Suppl 2:S94-S105.
2. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* [Internet]. 2008 [acesso em: 30 set 2013];65(23):2276-84. Disponível em: <http://dx.doi.org/10.2146/ajhp070364>.
3. Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B. Patient-reported outcomes: conceptual issues. *Value Health*. 2007;10 Suppl 2:S66-75.
4. Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS. Patient-reported outcomes: instrument development and selection issues. *Value Health*. 2007;10 Suppl 2:S86-93.
5. Snyder CF, Watson ME, Jackson JD, Cella D, Halyard MY. Patient-reported outcome instrument selection: designing a measurement strategy. *Value Health*. 2007;10 Suppl 2:S76-85.
6. Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care. I: Applications and

- issues in assessment. *BMJ* [Internet]. 1992 [acesso em: 30 set 2013];305(6861): 1074-7. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1883623/>.
7. Braga CG, Cruz DALM. Contribuições da psicometria para a avaliação de respostas psicossociais na enfermagem. *Rev Esc Enferm USP* [Internet]. 2006 [acesso em: 30 set 2013];40(1):98-104. Disponível em: <http://dx.doi.org/10.1590/S0080-62342006000100014>.
8. Carmines EG, Zeller RA. Reliability and validity assessment. 11<sup>th</sup> ed. Beverly Hills: Sage; 1988.
9. Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research* [Internet]. 2003 [acesso em: 30 set 2013];27(2):94-105. Disponível em: <http://dx.doi.org/10.1093/swr/27.2.94>.
10. Nunnally JC, Bernstein I. Psychometric theory. 3<sup>rd</sup> ed. New York: McGraw-Hill; 1994.
11. Dekker J, Dallmeijer AJ, Lankhorst GJ. Clinimetrics in rehabilitation medicine: current issues in developing and applying measurement instruments. *J Rehabil Med* [Internet]. 2005 [acesso em: 30 set 2013];37(4):193-201. Disponível em: <http://dx.doi.org/10.1080/16501970510035179>.
12. Galea MP. Introducing clinimetrics. *Australian Journal of Physiotherapy* [Internet]. 2005 [acesso em: 30 set 2013];51(3):139-40. Disponível em: [http://dx.doi.org/10.1016/S0004-9514\(05\)70019-2](http://dx.doi.org/10.1016/S0004-9514(05)70019-2).
13. Alexandre NMC, Guirardello EB. Adaptação cultural de instrumentos utilizados em salud ocupacional. *Rev Panam Salud Publica* [Internet]. 2002 [acesso em: 30 set 2013];11(2):109-11. Disponível em: <http://dx.doi.org/10.1590/S1020-49892002000200007>.
14. Fayers PM, Machin D. Quality of life. Assessment, analysis and interpretation. Chichester: John Wiley & Sons; 2007.
15. Alexandre NMC, Coluci MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Cien Saude Colet* [Internet]. 2011 [acesso em: 30 set 2013];16(7):3061-8. Disponível em: <http://dx.doi.org/10.1590/S1413-81232011000800006>.
16. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* [Internet]. 2007 [acesso em: 30 set 2013];60(1):34-42. Disponível em: <http://dx.doi.org/10.1016/j.jclinepi.2006.03.012>.
17. Selby-Harrington ML, Mehta SM, Jutsum V, Riportella-Muller R, Quade D. Reporting of instrument validity and reliability in selected clinical nursing journals, 1989. *J Prof Nurs*. 1994;10(1):47-56.
18. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL et al. COSMIN checklist manual [Internet]. Amsterdam: VU University Medical Center; 2012 [acesso em: 30 set 2013]. Disponível em <http://www.cosmin.nl/images/upload/File/COSMIN%20checklis%20manual%20v9.pdf>.
19. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* [Internet]. 1997 [acesso em: 30 set 2013];126(5):376-80. Disponível em: <http://dx.doi.org/10.7326/0003-4819-126-5-199703010-00006>.
20. Contandriopoulos AP, Champagne F, Potvin L, Denis JL, Boyle P. Saber preparar uma pesquisa. 3<sup>a</sup> ed. São Paulo: Hucitec/Abrasco; 1999.
21. Dempsey PA, Dempsey AD. Using nursing research. 5<sup>th</sup> ed. Philadelphia: Lippincott; 1996.
22. Polit DF, Hungler BP. Fundamentos de pesquisa em enfermagem. 3<sup>a</sup> ed. Porto Alegre: Artes Médicas; 1995.
23. Burns N., Grove SK. The practice of nursing research: conduct, critique & utilization. 3<sup>rd</sup> ed. Philadelphia: Saunders Company; 1997.
24. LoBiondo-Wood G, Harber J. Nursing Research: methods, critical appraisal and utilization. 4<sup>th</sup> ed. St Louis: Mosby-Year Book; 1998.
25. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ et al. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh* [Internet]. 2007 [acesso em: 30 set 2013];39(2):155-64. Disponível em: <http://dx.doi.org/10.1111/j.1547-5069.2007.00161.x>.
26. Salmund SS. Evaluating the reliability and validity of measurement instruments. *Orthop Nurs* [Internet]. 2008 [acesso em: 30 set 2013];27(1):28-30. Disponível em: <http://dx.doi.org/10.1097/01.NOR.0000310608.00743.54>.
27. Fitzner K. Reliability and validity: a quick review. *Diabetes Educ* [Internet]. 2007 [acesso em: 30 set 2013];33(5):775-80. Disponível em: <http://dx.doi.org/10.1177/0145721707308172>.
28. Roberts P, Priest H. Reliability and validity in research. *Nurs Stand*. 2006;20(44):41-5.
29. de Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. *J Clin Epidemiol* [Internet]. 2003 [acesso em: 30 set 2013];56(12):1137-41. Disponível em: <http://dx.doi.org/10.1016/j.jclinepi.2003.08.012>.
30. Dimitrov D, Rumrill P, Fitzgerald S, Hennessey M. Reliability in rehabilitation measurement. *Work*. 2001;16(2):159-64.
31. Switzer GE, Wisniewski SR, Belle SH, Dew MA, Schultz R. Selecting, developing, and evaluating research instruments. *Soc Psychiatry Psychiatr Epidemiol* [Internet]. 1999 [acesso em: 30 set 2013];34(8):399-409. Disponível em: <http://dx.doi.org/10.1007/s001270050161>.
32. Houser J. Precision, reliability, and validity: essential elements of measurement in nursing research. *J Spec Pediatr Nurs* [Internet]. 2008 [acesso em: 30 set 2013];13(4):297-9. Disponível em: <http://dx.doi.org/10.1111/j.1744-6155.2008.00171.x>.
33. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. *J Psychosom Res* [Internet]. 2010 [acesso em: 30 set 2013];68(4):319-23. Disponível em: <http://dx.doi.org/10.1016/j.jpsychores.2010.01.006>.
34. Rattray J, Jones MC. Essential elements of questionnaire design and development. *J Clin Nurs* [Internet]. 2007 [acesso em: 30 set 2013];16(2):234-43. Disponível em: <http://dx.doi.org/10.1111/j.1365-2702.2006.01573.x>.
35. Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optom Vis Sci* [Internet]. 2007 [acesso em: 30 set 2013];84(8):663-74. Disponível em: <http://dx.doi.org/10.1097/OPX.0b013e318141fe75>.
36. Mota DDCF, Pimenta CAM. Avaliação e mensuração de variáveis psicossociais: desafio para pesquisa e clínica de enfermagem. *Rev Gaucha Enferm* [Internet]. 2007 [acesso em: 30 set 2013];28(3):309-14. Disponível em: <http://seer.ufrgs.br/RevistaGauchadeEnfermagem/article/view/4674>.
37. Morales AR, Zárate LEM. Epidemiología Clínica. Investigación clínica aplicada. 20<sup>a</sup> ed. Bogotá: Editorial Médica Internacional; 2005.

Artigo recebido em 23/10/2012.

Aprovado para publicação em 11/06/2013.

Artigo publicado em 30/09/2013.