

Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas

Content validity in the development and adaptation processes of measurement instruments

Neusa Maria Costa Alexandre¹
Marina Zambon Orpinelli Coluci¹

Abstract *This study sought to conduct a review of content validity, which is an important phase of processes of construction and adaptation of measurement instruments. Research of comprehensive literature was conducted by means of a review of national and international databases. Initially, a description of the conceptual basis and the measurement methods used in content validity was made, with emphasis on its application in the healthcare area. It was seen that controversy exists in the literature on the terminology and concept of content validity. The recommended procedures used to check content validity during the construction and adaptation processes of instruments were described, especially the judges' assessment, which can involve qualitative and quantitative procedures. The number, selection and qualification of these judges were also described. The different methods used to quantify the level of agreement among the experts were verified, mainly the Content Validity Index (CVI). This study described aspects of the content validity process, one of the procedures to be considered by healthcare researchers and professionals who are interested in using reliable and appropriate measurements and instruments scales for given population groups.*

Key words *Validity of tests, Validation studies, Questionnaires, Methods, Reproducibility of results*

Resumo *Este estudo teve por objetivo realizar uma revisão sobre validade de conteúdo, fase importante nos processos de construção e adaptação de instrumentos de medida. Foi realizada pesquisa bibliográfica definida como integrativa, por meio de pesquisa a bases de dados nacionais e internacionais. Iniciou-se a descrição das bases conceituais e os métodos de medidas usados em validade de conteúdo, com ênfase na área da saúde. Verificou-se que existem controvérsias na literatura sobre a terminologia e o conceito da validade de conteúdo. Foram descritos os procedimentos recomendados para realizar a validade de conteúdo durante os processos de construção e de adaptação de instrumentos, particularmente a avaliação por juízes, o que pode envolver procedimentos qualitativos e quantitativos. Descreveu-se número, seleção e qualificação desses juízes. Os diferentes métodos para quantificar o grau de concordância entre os especialistas foram analisados, principalmente o Índice de Validade de Conteúdo (IVC). Este estudo descreveu aspectos do processo de realização da validade de conteúdo, um dos procedimentos a serem considerados por pesquisadores e profissionais da área de saúde preocupados em utilizarem medidas e instrumentos confiáveis e apropriados para determinada população.*

Palavras-chave *Validade dos testes, Estudos de validação, Questionários, Métodos, Reprodutibilidade dos testes*

¹ Departamento de Enfermagem, Faculdade de Ciências Médicas, Universidade Estadual de Campinas. Cidade Universitária Zeferino Vaz, Barão Geraldo. 13083-970 Campinas SP. nesalex@fcm.unicamp.br

Introdução

Ao planejar o método de coleta de dados, deve-se pensar em procedimentos que garantam indicadores confiáveis. A decisão vai depender do desenho da pesquisa e da seleção de instrumentos de medidas adequados e precisos. Um número crescente de questionários e escalas está atualmente disponível na área de saúde.

Dessa forma, cada vez mais a literatura tem alertado os pesquisadores sobre uma correta avaliação das qualidades dos instrumentos de coleta de dados¹⁻³. Os mais importantes atributos desses instrumentos são: validade (validity), confiabilidade (reliability), praticabilidade (usability), sensibilidade (sensitivity) e responsividade (responsiveness)⁴⁻⁶. Na realidade, as características de validade e de confiabilidade são particularmente importantes ao se escolher, desenvolver ou realizar a adaptação cultural de instrumentos que serão usados tanto em pesquisas como na prática clínica^{5,7-9}.

Na área de ciências sociais, o conceito de validade deve ser aplicado em todas as medidas como inventários de personalidades, testes educacionais, surveys, questionários de atitudes, entre outros¹⁰. Deve também fazer parte da avaliação de instrumentos usados na clínica como entrevistas, exame físico, exames de laboratório, radiológicos e histológicos⁵.

A validade verifica se o instrumento mede exatamente o que se propõe a medir¹¹. Isto é, avalia a capacidade de um instrumento medir com precisão o fenômeno a ser estudado¹². Pode-se considerar um instrumento válido quando ele consegue avaliar realmente seu objetivo, como, por exemplo, ansiedade e não outro constructo como o estresse¹³. Pode ser avaliada por meio dos seguintes métodos: validade de conteúdo, validade relacionada a um critério e validade do constructo.

A confiabilidade é a capacidade em reproduzir um resultado de forma consistente no tempo e no espaço, ou com observadores diferentes¹². Os seguintes procedimentos são utilizados para sua avaliação: estabilidade (teste-reteste), homogeneidade e equivalência (inter-observadores)¹⁴.

Sensibilidade é a habilidade de a medida identificar diferenças entre pacientes ou grupo de pacientes com afecções leves e severas⁶. A responsividade avalia a capacidade do instrumento em detectar melhora ou piora nas condições de pacientes, isto é, consegue verificar mudanças⁶. A praticabilidade refere-se aos aspectos práticos da utili-

zação de instrumentos. Podem-se levar em conta a facilidade e o tempo de administração, a interpretação dos resultados e os custos envolvidos⁴.

Estudos e discussões sobre essa temática são escassos particularmente em países da América Latina e Caribe, apesar de já se notar a utilização de avaliações psicométricas no desenvolvimento de instrumentos na área de saúde em países em desenvolvimento^{15,16}.

Considera-se relevante a divulgação e a discussão dos diferentes métodos de se avaliar as propriedades psicométricas de instrumentos de coleta de dados utilizados na área de saúde, tanto na assistência como na pesquisa. Dessa forma, o objetivo do presente estudo foi realizar uma revisão sobre a validade de conteúdo.

Método

Trata-se de uma pesquisa bibliográfica que pode ser definida como integrativa¹⁷. Para o levantamento bibliográfico, optou-se pelas bases de dados das Ciências da Saúde como a Biblioteca Virtual em Saúde – Enfermagem (BDENF) e o Portal de Revistas de Enfermagem; o SCIELO; a LILACS (Literatura Latino - Americana e do Caribe em Ciências da Saúde); a MEDLINE (National Library of Medicine-USA); o International Nursing Index (INI) e o Cumulative Index to Nursing and Allied Health Literature (CINAHL).

Para seleção das pesquisas, foram utilizados alguns critérios. O tipo de publicação considerado foram os periódicos e os livros de metodologia de pesquisa e estatística. A procedência e o idioma deveriam ser de artigos nacionais e internacionais, publicados em português e inglês. Os descritores em português pesquisados na BIREME (DeCS - Terminologia em Saúde) foram: estudos de validação; tradução; validade dos testes e reprodutibilidade dos testes. Em inglês, a pesquisa foi feita na National Library of Medicine's Medical Subject Headings (MeSH = Medical Subject Heading Terms), utilizando-se: validation studies; translations and psychometrics.

Discussão

Essa revisão foi organizada de modo a apresentar as bases conceituais e os métodos de medidas usados em validade de conteúdo, com ênfase em sua aplicação na área da saúde.

Abordagem conceitual

A avaliação de conteúdo é um passo essencial no desenvolvimento de novas medidas porque representa o início de mecanismos para associar conceitos abstratos com indicadores observáveis e mensuráveis¹⁸.

Encontram-se controvérsias na literatura sobre a terminologia e o conceito da validade de conteúdo^{10,19}. Para alguns autores, consiste em julgar em que proporção os itens selecionados para medir uma construção teórica representam bem todas as facetas importantes do conceito a ser medido¹². Refere-se em verificar a extensão em que os itens de uma medida determinam o mesmo conteúdo²⁰. Outros autores tentam explicar dessa forma: as perguntas do instrumento são representativas dentro do universo de todas as questões que poderiam ser elaboradas sobre esse tópico²¹.

A validade de conteúdo é também definida de uma forma mais abrangente. Isto é, avalia o grau em que cada elemento de um instrumento de medida é relevante e representativo de um específico constructo com um propósito particular de avaliação¹⁹. Segundo os mesmos autores, os elementos de um instrumento são todos os aspectos do processo de medida que podem afetar a coleta de dados. No caso de um questionário, consideram-se como elementos as instruções, o formato das respostas e os itens de forma individual.

A validade de conteúdo é fundamental no processo de desenvolvimento e adaptação de instrumento de medidas¹⁰. No entanto, apresenta limitações por ser um processo subjetivo. Dessa forma, sua utilização não elimina a necessidade de aplicação de outras medidas psicométricas adicionais²⁰.

Abordagens para realizar a validade de conteúdo

Construção e adaptação cultural de instrumentos de medida

Nessa fase, abordam-se inicialmente os procedimentos recomendados para realizar a validade de conteúdo durante os processos de construção e de adaptação de instrumentos de medida.

Para alguns autores, a validade de conteúdo compreende somente a avaliação por um comitê de especialistas^{4,7}. No entanto, pesquisadores têm descrito que a validade de conteúdo é um processo de julgamento sendo composto por duas partes distintas. A primeira envolveria o desenvolvimento do instrumento e, posteriormente, a

avaliação desse por meio da análise por especialistas^{22,23}. Assim, pode-se considerar que a validade de conteúdo de instrumentos seria também garantida pelo procedimento de elaboração dos mesmos²⁴. Para avaliar a validade de conteúdo recomenda-se também utilizar procedimentos quantitativos e qualitativos¹⁹.

Para construir a variedade de itens, o pesquisador deve inicialmente definir o constructo de interesse e suas dimensões por meio de pesquisa bibliográfica e consulta a estudiosos da área e a representantes da população de interesse^{14,25}. Autores têm defendido que esse estágio de desenvolvimento de instrumentos deve englobar três fases: identificação dos domínios, a formação dos itens e a construção do instrumento^{23, 26-28}. A fase de elaboração de questionários e escalas tem sido objeto de publicações específicas indicando diferentes metodologias²⁹⁻³¹.

Autores sugerem e descrevem etapas e métodos padronizados e sistemáticos que deveriam ser usados durante esse processo com o objetivo de melhorar a qualidade dos instrumentos de medidas¹⁹. Essas etapas diferem dependendo do que o instrumento quer avaliar. Segundo Lynn²³, para medidas cognitivas deve-se seguir os seguintes passos: identificação da totalidade dos domínios; construção dos itens e por fim, organização desses itens. Ainda pensando no processo de delineamento dos itens, não se pode deixar de mencionar a questão da adaptação cultural de instrumentos de medidas³¹. A adaptação de um instrumento para outra linguagem é um processo complexo. A realização de uma simples tradução não pode ser efetuada devido às diferenças culturais e de linguagem. Deve-se levar em consideração o idioma, o contexto cultural e o estilo de vida. Assim, para realizar uma adaptação de um instrumento de uma linguagem para outra, deve-se levar em conta aspectos técnicos, linguísticos e semânticos³². A literatura internacional tem procurado propor normas para assegurar adaptações de instrumentos de uma forma metodologicamente correta. A tendência atual é padronizar normas internacionais para assegurar a qualidade de instrumentos adaptados. Segundo esses protocolos, os seguintes passos são essenciais nesse processo: tradução inicial; síntese; tradução de volta para a língua original; revisão por um comitê e realização de um pré-teste³³⁻³⁷.

Avaliação por um comitê de especialistas

Alguns autores defendem que o segundo estágio do procedimento da validade de conteúdo consiste na avaliação do instrumento por especi-

alistas. Procura-se abordar esse procedimento no processo de construção de questionários e escalas e durante a realização de uma adaptação cultural.

Durante o desenvolvimento de instrumento, um dos pontos discutidos nessa avaliação é o número e a qualificação desses juízes. A literatura apresenta controvérsias sobre esse ponto. Lynn²³ recomenda um mínimo de cinco e um máximo de dez pessoas participando desse processo. Outros autores sugerem de seis a vinte sujeitos, sendo composto por um mínimo de três indivíduos em cada grupo de profissionais selecionados para participar¹⁹. Nessa decisão, deve-se levar em conta as características do instrumento, a formação, a qualificação e a disponibilidade dos profissionais necessários^{23,27}.

Em relação à seleção, deve-se levar em consideração a experiência e a qualificação dos membros desse comitê. Recomenda-se, descrever os critérios utilizados nessa seleção²³. Entre esses critérios, a literatura destaca: ter experiência clínica; publicar e pesquisar sobre o tema; ser perito na estrutura conceitual envolvida e ter conhecimento metodológico sobre a construção de questionários e escalas^{27,38}. Indica-se também a inclusão de pessoas leigas potencialmente relacionadas com a população do estudo^{20,39}. A inclusão de pessoas leigas asseguraria a correção de frases e termos que não estão muito claros.

Quando for um processo de adaptação cultural, sugere-se a formação de um comitê multidisciplinar³⁵. Neste caso, recomenda-se que deve ser composto por pessoas bilíngües e por especialistas na área de conhecimento do instrumento, que precisam ser informados sobre as medidas e conceitos envolvidos³³. Pode também ser composto por profissionais da área de saúde, professor de línguas, especialista da metodologia e pelos tradutores envolvidos no processo³⁷. Alguns autores recomendam a inclusão de pessoas leigas^{32,34}. Quando se tratar de adaptação entre culturas por população imigrante, sugere-se que indivíduos representativos dessa população estejam envolvidos.

A avaliação por juízes pode envolver procedimentos qualitativos e quantitativos^{14,39,40}. O processo deve ser iniciado com o convite aos membros do comitê de juízes. Esses especialistas devem receber uma carta explicativa e um questionário desenvolvido especificamente para essa avaliação²⁷. Alguns autores sugerem a inclusão de um questionário para uma breve caracterização desses especialistas²⁰.

A carta deve explicar porque o sujeito foi escolhido como juiz e a relevância dos conceitos

envolvidos e do instrumento como um todo^{23,27}. Recomenda-se também incluir o objetivo do estudo, a descrição do instrumento e a sua pontuação e interpretação, e uma explicação sobre a forma de resposta^{20,38}. Pode conter também definições conceituais que deram origem ao instrumento, as dimensões envolvidas e o modelo de medida usado. Isto é, fornecer as bases conceituais e teóricas do instrumento³⁸. Se for importante, pode também incluir informações sobre o contexto e a população envolvida. Essa carta deve considerar o nível educacional dos membros do comitê quando incluir sujeitos leigos²⁰.

Berk²⁸ relata que essa avaliação deve ser feita em dois momentos distintos, com orientações específicas para cada fase. Primeiro sugere realizar uma avaliação na fase de especificação dos domínios e depois na fase de desenvolvimento dos itens. De uma forma geral, recomenda-se que os especialistas recebam instruções específicas sobre como avaliar cada item, o instrumento como um todo e como preencher o questionário que orienta a avaliação¹⁴.

Os juízes devem inicialmente avaliar o instrumento como um todo, determinando sua abrangência. Isto é, se cada domínio ou conceito foi adequadamente coberto pelo conjunto de itens e se todas as dimensões foram incluídas³⁹. Nesta fase, podem sugerir a inclusão ou a eliminação de itens²⁰.

Precisam também analisar os itens individualmente verificando sua clareza e pertinência. Em relação à clareza, deve-se avaliar a redação dos itens, se eles foram redigidos de forma que o conceito esteja compreensível e se expressa adequadamente o que se espera medir²⁷. Pertinência ou representatividade significa notar se os itens realmente refletem os conceitos envolvidos, se são relevantes e, se são adequados para atingir os objetivos propostos^{27,41}. Deve-se deixar um espaço para que os especialistas possam redigir sugestões para melhorar o item ou fazer comentários^{20,39}.

Essa avaliação pode ser inicialmente feita de forma individual e independente pelos juízes, seguida por uma discussão em grupo que corresponde ao procedimento qualitativo^{28,39,40}. Alguns autores também defendem um processo interativo entre pesquisadores e os membros do comitê, sugerindo o emprego de entrevistas e discussões para clarificar pontos controversos²⁷. A abordagem de se usar pelo menos dois métodos, geralmente quantitativo e qualitativo, é denominada triangulação metodológica⁴².

No caso de adaptação cultural, esse comitê deve assegurar que a versão final seja totalmente

compreensível e avaliar a sua equivalência cultural. Há a necessidade de existir equivalências semântica, idiomática, conceitual e experimental³⁴.

Equivalência semântica é relativa ao significado das palavras (vocabulário, gramática); equivalência idiomática refere-se às expressões idiomáticas e coloquiais (por exemplo, sentir-se em casa); equivalência experimental aborda situações coerentes com o contexto cultural (por exemplo: usar o carro / usar o transporte público; dificuldade para usar um garfo); e, finalmente, a equivalência conceitual se refere ao conceito explorado (por exemplo: explorar as maneiras que as diferentes populações entendem saúde)³³. Dentro deste contexto, o comitê pode modificar as orientações iniciais e o formato do instrumento, alterar ou rejeitar itens inadequados e até desenvolver novos itens. Para verificar se o instrumento está realmente em uma linguagem compreensível para uma grande parte de pessoas, recomenda-se que ele precisa ser entendido por indivíduos de 10 a 14 anos³³.

Medidas quantitativas para avaliar validade de conteúdo

Publicações têm apresentado métodos diferentes para quantificar o grau de concordância entre os especialistas durante o processo de avaliação da validade de conteúdo de um instrumento. Dentre esses, destaca-se:

Porcentagem de concordância

Método empregado para calcular a porcentagem de concordância entre os juízes. É a medida mais simples de concordância interobservadores⁴³. A fórmula utilizada está descrita a seguir^{39, 44}.

$$\% \text{ concordância} = \frac{\text{número de participantes que concordaram}}{\text{número total de participantes}} \times 100$$

As vantagens desse procedimento é proporcionar informações úteis que são facilmente calculadas. No entanto, apresenta limitações que restringem sua utilização⁴⁴. Autores têm usado esse método na fase inicial para auxiliar na determinação dos itens^{27, 39}. Ao usar esse método, deve-se considerar como uma taxa aceitável de concordância de 90% entre os membros do comitê^{22, 44}.

Índice de validade de conteúdo (IVC)

Compreende um método muito utilizado na área de saúde^{40,41,45}. Mede a proporção ou porcentagem de juízes que estão em concordância sobre determinados aspectos do instrumento e de seus itens. Permite inicialmente analisar cada item individualmente e depois o instrumento como um todo. Este método emprega uma escala tipo Likert com pontuação de um a quatro. Para avaliar a relevância/representatividade, as respostas podem incluir: 1 = não relevante ou não representativo, 2 = item necessita de grande revisão para ser representativo, 3 = item necessita de pequena revisão para ser representativo, 4 = item relevante ou representativo^{20,23}. Outros autores sugerem opções mais curtas. Por exemplo: 1 = não claro, 2 = pouco claro, 3 = bastante claro, 4 = muito claro^{18,25,40}. Outra opção seria: 1= irrelevante a 4= extremamente relevante²⁵.

O escore do índice é calculado por meio da soma de concordância dos itens que foram marcados por “3” ou “4” pelos especialistas²⁷. Os itens que receberam pontuação “1” ou “2” devem ser revisados ou eliminados. Dessa forma, o IVC tem sido também definido como “a proporção de itens que recebe uma pontuação de 3 ou 4 pelos juízes”¹⁸. A fórmula para avaliar cada item individualmente fica assim:

$$\text{IVC} = \frac{\text{número de respostas "3" ou "4"}}{\text{número total de respostas}}$$

Para avaliar o instrumento como um todo, não existe um consenso na literatura. Polit e Beck²² recomendam que os pesquisadores devem descrever como realizaram o cálculo. Esses autores apresentam três formas que podem ser usadas. Uma é definida como a “média das proporções dos itens considerados relevantes pelos juízes”. A outra é a “média dos valores dos itens calculados separadamente, isto é soma-se todos os IVC calculados separadamente e divide-se pelo número de itens considerados na avaliação”. Finalmente, a última forma seria dividir o “número total de itens considerados como relevantes pelos juízes pelo número total de itens”. Deve-se também estipular a taxa de concordância aceitável entre os juízes. Autores defendem que no processo de avaliação dos itens individualmente, deve-se considerar o número de juízes. Com a participação de cinco ou menos sujeitos, todos devem concordar para ser representativo. No caso de seis ou

mais, recomenda-se uma taxa não inferior a 0,78^{22,23}. Para verificar a validade de novos instrumentos de uma forma geral, alguns autores sugerem uma concordância mínima de 0,80^{27,38}. No entanto, neste caso os valores recomendados devem ser de 0,90 ou mais²².

Coefficiente de kappa

O coeficiente de kappa (k) também tem sido recomendado para avaliar medidas de concordância entre avaliadores na área de saúde^{44,45}. O coeficiente de kappa de concordância é a razão da proporção de vezes que os juizes concordam (corrigido por concordância devido ao acaso) com a proporção máxima de vezes que os juizes poderiam concordar (corrigida por concordância devido ao acaso)^{43,46}. É um índice útil quando vários avaliadores categorizam cada grupo de objetos ou sujeitos em categorias nominais. Des-

sa forma, é aplicável quando os dados são categóricos e estão em uma escala nominal⁴⁶. Os valores de kappa variam de -1 (ausência total de concordância) a 1 (concordância total)⁴³.

Conclusão

Este estudo descreve aspectos do processo de realização da validade de conteúdo, uma fase importante do desenvolvimento e adaptação de questionários e escalas. Ressalta-se que esse é somente o início de um processo que deve englobar também outras medidas para avaliar outros tipos de validade e a confiabilidade.

Compreender esses procedimentos é essencial para pesquisadores e profissionais da área de saúde preocupados em utilizar cada vez mais medidas e instrumentos confiáveis e apropriados para determinada população.

Colaboradores

NMC Alexandre e MZO Coluci participaram igualmente de todas as etapas de elaboração do artigo.

Referências

- Marx RG, Bombardier C, Hogg-Johnson S, Wright JG. Clinimetric and psychometric strategies for development of a health measurement scale. *J Clin Epidemiol* 1999; 52(2):105-111.
- Selby-Harrington ML, Mehta SM, Jutsum V, Riportella-Muller R, Quade D. Reporting of instrument validity and reliability in selected clinical nursing journals. *J Prof Nurs* 1994; 10(1):47-56.
- Salmond SS. Evaluating the reliability and validity of measurement instruments. *Orthop Nurs* 2008; 27(1):28-30.
- Dempsey PA, Dempsey AD. *Using nursing research*. 5th ed. Philadelphia: Lippincott; 1996.
- De Vet HCW, Terwee CB, Bouter LM. Current challenges in clinimetrics. *J Clin Epidemiol* 2003; 56:1137-1141.
- Fayers PM, Machin D. *Quality of life. Assessment, analysis and interpretation*. Chichester (UK): John Wiley & Sons; 2000.
- Fitzner K. Reliability and validity. *Diabetes Educ* 2007; 33(5):775-780.
- Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient for reliability and validity of patient-reported outcome measures? *Value Health* 2007; 10(Supl. 2):S94-S105.
- Esperidião M, Trad, LAB. Avaliação de satisfação de usuários. *Cien Saude Colet* 2005; 10(Supl.):303-312.
- Sireci SG. The construct of content validity. *Soc Indic Res* 1998; 45:83-117.
- Roberts P, Priest H, Traynor M. Reliability and validity in research. *Nurs Stand* 2006; 20(44):41-45.
- Contandriopoulos AP, Champagne F, Potvin L, Denis JL, Boyle P. *Saber preparar uma pesquisa*. 3ª ed. São Paulo: Hucitec, Abrasco; 1999.
- LoBiondo-Wood G, Harber J. *Nursing Research: methods, critical appraisal, e and utilization*. 4th ed. St Louis: Mosby-Year Book; 1998.
- Burns N, Grove SK. *The practice of nursing research: conduct, critique & utilization*. 3ª ed. Philadelphia: Saunders Company; 1997.
- Mendonça KMPP, Guerra RO. Desenvolvimento e validação de um instrumento de medida da satisfação do paciente com a fisioterapia. *Rev Bras Fisioter* 2007; 11(5):369-376.
- Habib ES, Magalhães LC. Criação de questionário para detecção de comportamentos atípicos em bebês. *Rev Bras Fisioter* 2007; 11(3):177-183.
- Beya S, Nicoll LH. Writing an integrative review. *AORN J* 1998; 67(4):877-880.
- Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. *West J Nurs Res* 2003; 25(5):508-518.
- Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. *Psychol Assess* 1995; 7(3):238-247.
- Rubio DM, Berg-Weger M, Tebb SS, Lee S, Rauch S. Objectifying content validity: conducting a content validity study in social work research. *Soc Work Res* 2003; 27(2):94-105.
- Polit DF, Hungler BP. *Fundamentos de pesquisa em enfermagem*. 3ª ed. Porto Alegre: Artes Médicas; 1995.
- Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health* 2006; 29:489-497.
- Lynn MR. Determination and quantification of content validity. *Nurs Res* 1986; 35(6):382-385.
- Pasquali L. *Psicometria*. 2ª ed. Petrópolis (RJ): Editora Vozes; 2004.
- DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, Savoy SM, Kostas-Polston E. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh* 2007; 39(2):155-64.
- Carmines EG, Zeller RA. *Reliability and validity assessment*. 11th ed. Beverly Hills (CA): Sage; 1988.
- Grant JS, Davis LL. Selection and use of content experts for instrument development. *Res Nurs Health* 1997; 20(3):269-274.
- Berk RA. Importance of expert judgment in content-related validity evidence. *West J Nurs Res* 1990; 12(5):659-671.
- Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985; 38(1):27-36.
- Pasquali L. Princípios de elaboração de escalas psicológicas. *Rev Psiquiatr Clin* 1998; 25(5):206-13.
- Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use*. 2nd ed. New York: Oxford University Press; 2002.
- Hunt SM, Bucquet D, Niero M, Wiklund I, McKenna S. Cross-cultural adaptation of health measures. *Health Policy* 1991; 19(44):33-44.
- Guillemin F, Bombardier C, Beaton, D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46(12):1417-1432.
- Guillemin F. Cross-cultural adaptation and validation of health status measures. *Scand J Rheumatol* 1995; 24(2):61-63.
- Hutchinson A, Bentzen N, Konig-Zanhn C. *Cross cultural health outcome assessment: a user's guide*. The Netherlands: ERGHO; 1996.
- Alexandre NMC, Guirardello EB. Adaptação cultural de instrumentos utilizados em saúde ocupacional. *Rev Panam Salud Publica* 2002; 11(2):109-111.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; 25(24):3186-3191.
- Davis LL. Instrument review: getting the most from a panel of experts. *Appl Nurs Res* 1992; 5(4):194-197.
- Tilden VP, Nelson CA, May BA. Use of qualitative methods to enhance content validity. *Nurs Res* 1990; 39(3):172-175.
- Hyrkäs K, Appelqvist-Schmidlechner K, Oksa L. Validating an instrument for clinical supervision using an expert panel. *Int J Nurs Stud* 2003; 40(6):619-625.
- McGilton K. Development and psychometric evaluation of supportive leadership scales. *Can J Nurs Res* 2003; 35(4):72-86.
- Morse JM. Approaches to qualitative-quantitative methodological triangulation. *Nurs Res* 1991; 40(1):120-123.

43. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. *Delineando a pesquisa clínica*. 2ª ed. Porto Alegre: Artmed; 2003.
44. Topf M. Three estimates of interrater reliability for nominal data. *Nurs Res* 1986; 35(4):253-245.
45. Wynd CA, Schaefer MA. The Osteoporosis Risk Assessment Tool: establishing content validity through a panel of experts. *Appl Nurs Res* 2002; 16(2):184-188.
46. Siegel S, Castellan HJ. *Estatística não paramétrica para ciências do comportamento*. 2ª ed. Porto Alegre: Artmed; 2006.

Artigo apresentado em 06/07/2009

Aprovado em 08/11/2009

Versão final apresentada em 29/11/2009