

Construção de instrumentos de medida na área da saúde

Construction of measurement instruments in the area of health

Marina Zambon Orpinelli Coluci¹
Neusa Maria Costa Alexandre¹
Daniela Milani¹

Abstract *Measurement instruments are an integral part of clinical practice, health evaluation and research. These instruments are only useful and able to present scientifically robust results when they are developed properly and have appropriate psychometric properties. Despite the significant increase of rating scales, the literature suggests that many of them have not been adequately developed and validated. The scope of this study was to conduct a narrative review on the process of developing new measurement instruments and to present some tools which can be used in some stages of the development process. The steps described were: I-The establishment of a conceptual framework, and the definition of the objectives of the instrument and the population involved; II-Development of the items and of the response scales; III-Selection and organization of the items and structuring of the instrument; IV-Content validity, V-Pre-test. This study also included a brief discussion on the evaluation of the psychometric properties due to their importance for the instruments to be accepted and acknowledged in both scientific and clinical environments.*

Key words *Measurements, Methods and theories, Questionnaires, Validation studies, Reproducibility of the tests*

Resumo *Instrumentos de medida são partes integrantes da prática clínica, da avaliação em saúde e de pesquisas. Esses instrumentos só são úteis e capazes de apresentar resultados cientificamente robustos quando são desenvolvidos de maneira apropriada e quando apresentam boas qualidades psicométricas. A literatura aponta que, apesar do aumento significativo no número de escalas de avaliação, muitas não têm sido desenvolvidas e validadas adequadamente. O presente estudo teve como objetivos realizar uma revisão narrativa sobre o processo de elaboração de novos instrumentos e apresentar algumas ferramentas que podem ser utilizadas em algumas etapas do seu processo de desenvolvimento. As etapas descritas foram: I-Estabelecimento da estrutura conceitual e definição dos objetivos do instrumento e da população envolvida; II-Construção dos itens e das escalas de respostas; III-Seleção e organização dos itens e estruturação do instrumento; IV-Validade de conteúdo; e V-Pré-teste. Uma discussão breve sobre a avaliação das propriedades psicométricas foi incluída devido à importância que esta tem para que os instrumentos sejam aceitos e reconhecidos tanto no meio científico como no meio clínico.*

Palavras-chave *Medidas, Métodos e teorias, Questionários, Estudos de Validação, Reprodutibilidade dos testes*

¹ Faculdade de Enfermagem, Universidade Estadual de Campinas. Cidade Universitária Zeferino Vaz, Barão Geraldo. 13083-887 Campinas SP Brasil. mazorpinelli@yahoo.com.br

Introdução e justificativa

Questionários são instrumentos integrantes da prática clínica, da avaliação em saúde e de pesquisas¹. Estes instrumentos exercem grande influência nas decisões sobre o cuidado, tratamento e/ou intervenções e na formulação de programas de saúde e de políticas institucionais^{1,2}.

Com o atual e crescente interesse pelo cuidado à saúde, pesquisadores e organizações internacionais têm desenvolvido instrumentos importantes e de grande aplicabilidade¹⁻³. Os aspectos avaliados e/ou mensurados por estes instrumentos são diversos: dor, qualidade de vida, capacidade funcional, estado de saúde, vitalidade e limitações, adesão ao tratamento, fatores emocionais e psicossociais, dentre outros^{2,4-8}.

Pesquisadores apontam que os instrumentos para avaliação só são úteis e capazes de apresentar resultados cientificamente robustos quando demonstram boas propriedades psicométricas³. Apesar do aumento significativo do número de escalas de avaliação e/ou questionários, muitos não são desenvolvidos e validados de forma apropriada^{1,9-10}.

O desenvolvimento integral de um novo instrumento de mensuração em saúde é complexo, consome vários recursos e requer a mobilização de capacidades e de conhecimentos de diversas áreas¹¹. Por este motivo, antes de se desenvolver novos instrumentos, recomenda-se que o pesquisador esteja ciente sobre os questionários já existentes. Estes, muitas vezes, podem atender às mesmas finalidades pretendidas ou similares¹¹.

Outra recomendação é a de que haja preferência por realizar a adaptação cultural de questionários previamente desenvolvidos e validados em outros idiomas em detrimento de construir novos instrumentos^{12,13}. Esta é uma alternativa facilitadora para a troca de informações e divulgação do conhecimento entre a comunidade científica¹¹⁻¹³.

Porém, quando o desenvolvimento de um novo instrumento é realmente necessário, os pesquisadores e profissionais da área da saúde precisam estar cientes de que devem seguir uma metodologia adequada a fim de que esse novo instrumento seja apropriado e confiável^{2,14-16}.

Neste sentido, discussões sobre o processo de desenvolvimento de novos questionários e escalas na área de saúde são bastante relevantes. O presente estudo tem como objetivos: discutir de forma sistematizada as principais etapas a serem seguidas na elaboração de instrumentos de medida na área de saúde e apresentar algumas

ferramentas que podem ser utilizadas como auxílio nas etapas do desenvolvimento desses novos instrumentos.

Método

Trata-se de uma pesquisa bibliográfica que pode ser definida como narrativa¹⁷. O levantamento bibliográfico foi realizado nas seguintes bases de dados das Ciências da Saúde: Biblioteca Virtual em Saúde – Enfermagem (BDENF), Portal de Revistas de Enfermagem, SciELO, Lilacs (Literatura Latino - Americana e do Caribe em Ciências da Saúde), Medline (National Library of Medicine-USA), International Nursing Index (INI) e Cumulative Index to Nursing and Allied Health Literature (CINAHL).

No estudo foram incluídas publicações nacionais e internacionais disponíveis nos idiomas português e inglês a partir de 1990 e que abordssem temas relacionados ao desenvolvimento de instrumentos de medida.

Os seguintes descritores utilizados em português foram pesquisados na Bireme (DeCS – Terminologia em Saúde): “estudos de validação”, “validade dos testes”, “reprodutibilidade dos testes” e “questionários”. Para os descritores em inglês utilizou-se a National Library of Medicine’s Medical Subject Headings (MeSH = Medical Subject Heading Terms). Estes foram os seguintes: “validation studies”, “psychometrics”, e “questionnaire design”.

Resultados e discussão

Devido ao aumento da importância dos instrumentos de medida em saúde, o entendimento sobre o processo e as técnicas utilizadas para o desenvolvimento destas é crucial².

Com o objetivo de melhorar a qualidade dos instrumentos, autores sugerem e descrevem etapas e métodos padronizados e sistemáticos que devem ser utilizados durante esse processo^{3,18,19}.

De uma forma geral, a literatura destaca as seguintes etapas a serem seguidas no processo de construção de instrumentos: I-Estabelecimento da estrutura conceitual; II-definição dos objetivos do instrumento e da população envolvida; III-Construção dos itens e das escalas de resposta; IV-Seleção e organização dos itens; V-Estruturação do instrumento; VI-Validade de Conteúdo; e VII-Pré-teste.

Cada uma dessas etapas será discutida de

maneira mais detalhada nas seções seguintes. A Figura 1 ilustra esse processo e apresenta de maneira resumida os recursos, critérios e/ou recomendações que geralmente são utilizados em cada etapa.

Com relação ao processo de avaliação das propriedades psicométricas de um instrumento, não há um consenso na literatura sobre este pertencer ao processo de desenvolvimento do instrumento ou não. Com a justificativa de que a utilidade de um instrumento de medida no meio científico e clínico só é reconhecida após este ter suas propriedades psicométricas avaliadas, este estudo também inclui uma breve discussão sobre a avaliação dessas propriedades.

As etapas do processo de construção de instrumentos de medidas na área da saúde

Estabelecimento da estrutura conceitual e definição dos objetivos do instrumento e da população envolvida

Para a construção de instrumentos de medida, é fundamental que os objetivos sejam estabelecidos e que estes tenham conexão com os conceitos a serem abordados^{20,21}. A caracterização/definição da população-alvo também é importante porque serve para justificar a relevância da criação de um instrumento específico¹⁵.

A elaboração da estrutura conceitual, também conhecida como definição operacional do

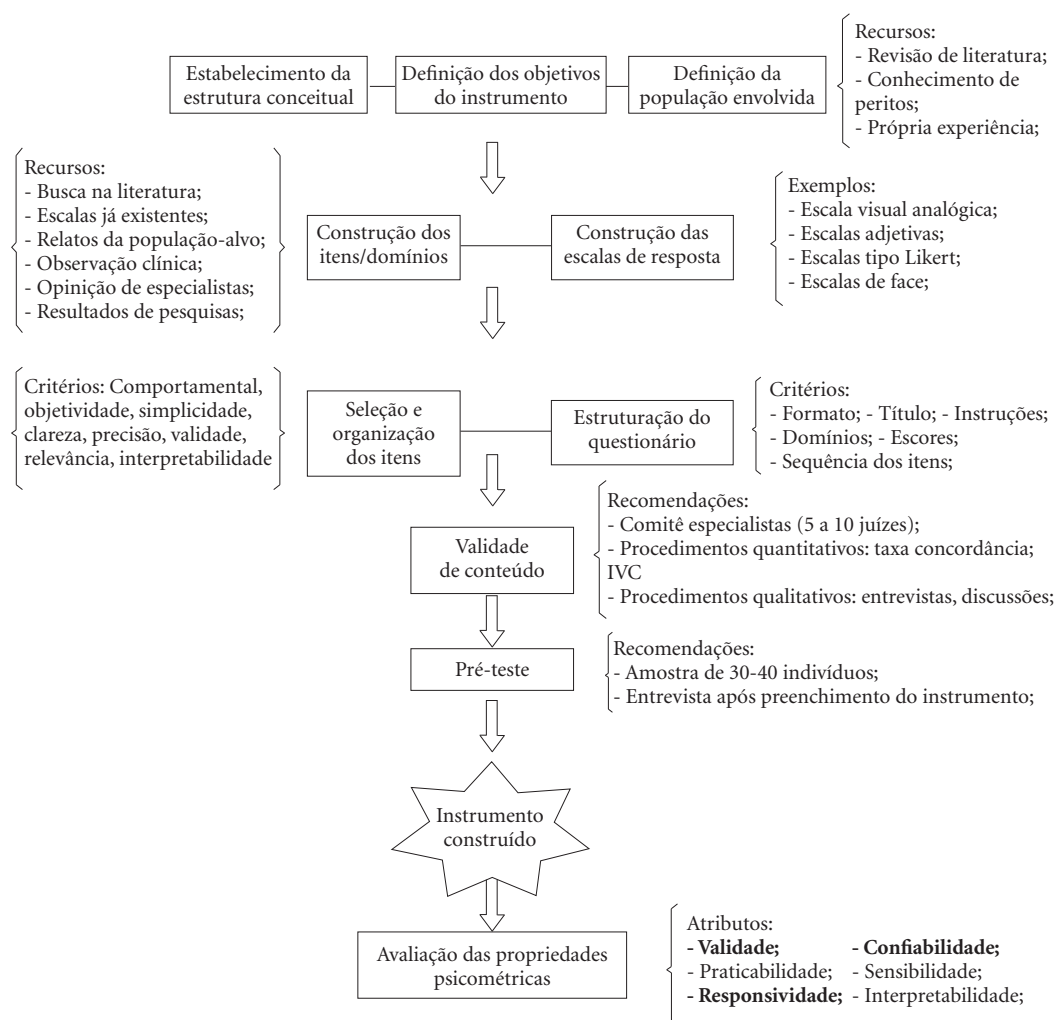


Figura 1. As etapas do processo de desenvolvimento de instrumentos de medida e os recursos, critérios, recomendações e/ou atributos geralmente mais utilizados em cada etapa.

constructo e de sua dimensionalidade, é a etapa responsável por definir o contexto do instrumento e sustentar o desenvolvimento dos domínios e itens²². Esta é considerada uma etapa bastante importante porque “quanto melhor e mais completa for a especificação do constructo, melhor será a garantia de que o instrumento será útil e válido”²². Por este motivo, pesquisadores apontam para a importância das definições operacionais do constructo de forma detalhada²².

Construção dos itens e das escalas de resposta

Os itens de uma escala não devem ser construídos ao acaso. Eles devem ser elaborados ou selecionados em função das definições operacionais do constructo que já foram analisadas na etapa anterior²².

Vários são os recursos a partir dos quais esses itens podem ser construídos: busca na literatura, questionários já existentes, relatos da população-alvo, observação clínica, opinião de especialistas, resultados de pesquisa, teorias, dentre outros^{3,11,15,23}. Cada um desses recursos possui pontos fortes e fracos dependendo do tipo, do objetivo e da aplicação da escala a ser desenvolvida³.

A busca na literatura junto às bases de dados nacionais e internacionais é comumente utilizada como principal recurso nas pesquisas de desenvolvimento de instrumentos de medidas. Esta estratégia visa dar ciência ao pesquisador quanto aos instrumentos que já existem¹¹.

A utilização de instrumentos já existentes também é apontada por pesquisadores^{3,11} como um recurso útil, pois, na maioria das vezes, seus itens já foram testados quanto às qualidades psicométricas³. A utilização destes itens em detrimento de construir novos pode economizar tempo e trabalho do pesquisador^{3,11}.

Outro recurso que pode ser significativo e fonte excelente de itens é a experiência da população-alvo^{3,11}. Dentre as técnicas utilizadas para obtenção de dados da população de forma sistemática, citam-se a realização de grupo focal e entrevistas em profundidade^{3,11,23}.

A observação clínica pode ser útil quando utilizada previamente à opinião de especialistas e também é uma forma de reunir informações de forma sistemática. Geralmente este recurso está associado à teoria, resultados laboratoriais e achados clínicos^{3,11}.

A opinião de especialistas também é um recurso bastante utilizado e influencia consideravelmente na geração de itens mais potenciais para a escala¹¹. Dentre as vantagens, está o fato dos es-

pecialistas serem escolhidos cuidadosamente e, portanto, representarem o que há de mais recente no conhecimento da área^{3,11}. Recomenda-se a participação de três a dez especialistas^{3,11}.

Com relação aos resultados de pesquisas, os mais úteis são as revisões sistemáticas e estudos^{3,11}.

Além da construção dos itens, o desenvolvimento e escolha de um método para obtenção das respostas também é imprescindível¹¹. As escalas de respostas aos itens podem assumir muitas formas e a escolha do método deve ser determinada pela natureza da pergunta realizada¹¹.

Dentre as técnicas utilizadas para a formulação de escalas de resposta, as mais comuns são as de estimativa direta, como a escala visual analógica, as escalas adjetivas, as escalas tipo Likert, as escalas de faces, entre outras¹¹.

Cada um dos tipos de escala tem suas vantagens e desvantagens¹¹.

Seleção e organização dos itens, e estruturação do instrumento

Uma vez realizada a pesquisa bibliográfica, a consulta aos especialistas da área e aos representantes da população-alvo, o próximo passo é definir as suas dimensões/domínios de forma que se construa a variedade dos itens segundo os constructos²⁴.

Geralmente, nem todos os itens criados estão em concordância com a proposta do pesquisador e nem todos possuem bom desempenho. Por isso, o pesquisador deve estar atento aos diversos tipos de critérios existentes para selecionar os itens adequados do novo instrumento^{3,11}.

Dentre os critérios mais comumente utilizados estão: o critério comportamental, o critério da objetividade, da simplicidade, da clareza, da precisão, da validade, da relevância e da interpretabilidade²¹⁻²³. O objetivo da adoção desses critérios é eliminar qualquer item que esteja ambíguo, incompreensível, com termos vagos, com duplas perguntas, com jargões e/ou que remeta a juízo de valores, dentre outros^{3,11,21-23}.

Com relação à quantidade de itens de um instrumento, este é um quesito não consensual²². Estudos sugerem que a representação de um constructo é conseguida com uma quantidade de 20 itens aproximadamente²². No entanto, para outra vertente, aponta-se que o início da construção de um instrumento deve possuir pelo menos o triplo de itens do instrumento final.

Outros estudos^{22,23} se baseiam na teoria dos traços latentes para determinar o número de itens que farão parte do instrumento. De acordo com essa teoria, não se deve começar o instru-

mento com mais de 10% do número de itens desejado no instrumento final²².

A estruturação do questionário é o estágio que visa consolidar as etapas anteriores, ou seja, visa organizar os itens em seus respectivos domínios e estabelecer o formato geral do instrumento. Deve-se considerar o título, as instruções, as escalas de respostas, os escores, entre outros.

A estrutura e a sequência do questionário contribuem significativamente para reduzir o esforço físico e/ou mental dos respondentes e asseguram que todos os termos sejam tratados de forma que o interesse do respondente seja mantido até o final do questionário²¹.

Um princípio utilizado na estruturação do questionário é o de que os itens estejam em uma ordem lógica. O direcionamento desses itens sempre que possível deve ser do item mais geral até o mais específico; no sentido do menos pessoal e menos delicado para o mais pessoal e mais delicado²¹.

O procedimento de validade de conteúdo

Após ser estruturado e organizado, o novo instrumento provavelmente ainda contemplará mais itens do que ele necessariamente apresentará em seu formato final¹¹. Esse instrumento ainda precisa ser testado quanto à hipótese de que os itens escolhidos representam e/ou contemplam adequadamente os domínios do constructo desejado^{3,11,20,22,24-26}.

O procedimento de escolha é a avaliação de conteúdo. Esta é essencial no processo de desenvolvimento de novos instrumentos de medidas porque representa o início de mecanismos para associar conceitos abstratos com indicadores observáveis e mensuráveis^{11,24-27}.

A avaliação de conteúdo deve ser realizada por um comitê composto por cinco a dez juízes especialistas na área do instrumento de medida^{28,29}. A avaliação por juízes pode envolver procedimentos qualitativos e quantitativos²⁹⁻³². O processo é iniciado com o convite aos membros do comitê de juízes.

Deve-se redigir aos especialistas uma carta constando o motivo da escolha daquele sujeito como juiz e a relevância dos conceitos envolvidos e do instrumento como um todo³³.

Todo o procedimento para o julgamento da validade de conteúdo deve ser descrito de forma estruturada e entregue aos juízes. Cada um deles deve realizar uma avaliação inicial independente antes de se reunir com os demais membros³⁴. Para isso, são necessárias instruções específicas a

partir das quais esses juízes poderão determinar a validade de conteúdo³³.

É importante que o comitê de especialistas participe em dois estágios distintos. O primeiro estágio é aquele em que os juízes realizam uma avaliação para a fase de especificação dos domínios. O segundo estágio é aquele para o qual realizam uma avaliação na fase de desenvolvimento dos itens³⁴. Os especialistas devem receber instruções específicas em cada estágio sobre como avaliar cada item, como avaliar o instrumento como um todo e como preencher o questionário que orienta a avaliação.

No primeiro estágio de avaliação, os membros do comitê devem julgar os domínios e os itens criados para interpretação das respostas do questionário. Inicialmente é realizada a avaliação dos domínios, determinando a abrangência destes. Isto é, os juízes devem avaliar se cada domínio ou conceito foi adequadamente coberto pelo conjunto de itens e se todas as dimensões foram incluídas³¹. Os especialistas deverão verificar se o conteúdo está apropriado aos respondentes, se a estrutura do domínio e seu conteúdo estão corretos, e se o conteúdo contido no domínio é representativo³⁴. Nesta fase, sugestões quanto à inclusão ou a eliminação de itens podem ser feitas³⁵. As instruções para análise da validade de conteúdo podem ser redigidas como no Quadro 1.

Após a avaliação dos juízes, pode-se utilizar a taxa de concordância do comitê, que é obtida pelo cálculo da porcentagem em cada domínio. Este é realizado por meio da seguinte fórmula³¹:

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

Essa taxa é interpretada considerando que um resultado maior ou igual a 90% de concordância, significa que os domínios estão adequados²⁶. Quando o resultado for menor que 90%, o domínio precisa ser discutido e alterado.

No segundo estágio, os juízes devem avaliar cada item individualmente. Além disso, essa avaliação deve ser realizada com relação ao formato, ao título, às instruções, aos domínios, aos escores dos domínios (ou do instrumento) e à análise (interpretação) dos escores, considerando a clareza e/ou pertinência de cada aspecto a ser avaliado.

Com relação à clareza, a orientação é feita no sentido de que seja avaliada a redação dos itens. Ou seja, se estes itens estão redigidos de forma que o conceito esteja compreensível e se expressem adequadamente o que se espera medir³⁶.

Quadro 1. Modelo de instruções para primeira análise da validade de conteúdo.

INSTRUÇÕES PARA ANÁLISE DO INSTRUMENTO	AVALIAÇÃO DA VALIDADE DE CONTEÚDO				
<p>Para realizar a avaliação do conteúdo do instrumento _____, descrevemos abaixo os conceitos envolvidos no estudo.</p>	<p>1º Passo – Especificação dos domínios: Avalie se cada domínio do instrumento foi adequadamente coberto pelo conjunto de itens.</p>				
<p>(Descrever conceitos importantes relacionados ao instrumento com base na literatura a fim de nortear a avaliação do juiz).</p>	<p>DOMÍNIO 1: (Coloque o domínio 1 completo para avaliação) Cada item do Domínio 1 realmente expressa seu conteúdo.</p>				
<p>A avaliação do instrumento envolve 2 fases: 1) avaliação dos domínios e 2) avaliação dos itens.</p>	<table border="1"> <tr> <td>CONCORDO</td> <td>NÃO CONCORDO</td> </tr> <tr> <td> </td> <td> </td> </tr> </table>	CONCORDO	NÃO CONCORDO		
CONCORDO	NÃO CONCORDO				
<p>1. Avaliação dos domínios:</p>	<p>Comentários: _____</p>				
<p>Pedimos que avalie, primeiramente, os domínios. Verifique se a estrutura do domínio e seu conteúdo estão corretos, se o conteúdo contido no domínio é representativo e se está apropriado aos respondentes. Portanto, considere o conceito de abrangência conforme descrito abaixo na sua avaliação:</p>	<p>▪ Os itens do Domínio 1 devem permanecer nesse domínio.</p>				
<p>❖ <i>Abrangência:</i> verificar se cada domínio ou conceito foi adequadamente coberto pelo conjunto de itens.</p>	<table border="1"> <tr> <td>CONCORDO</td> <td>NÃO CONCORDO</td> </tr> <tr> <td> </td> <td> </td> </tr> </table>	CONCORDO	NÃO CONCORDO		
CONCORDO	NÃO CONCORDO				
<p>Durante essa fase, você poderá sugerir a inclusão ou exclusão de itens nos domínios e opinar se os itens realmente pertencem ao domínio correspondente.</p>	<p>Comentários: _____</p>				
<p>2. Avaliação dos itens:</p>	<p>(Realizar essa avaliação para cada domínio do seu questionário)</p>				
<p>Na segunda etapa, pedimos para que avalie cada item separadamente, considerando os conceitos de clareza e pertinência/representatividade conforme descrito:</p>	<p>2º Passo – Avaliação dos itens: Avalie cada item quanto à clareza (redação dos itens, se eles foram redigidos de forma que o conceito esteja compreensível e se expressa adequadamente o que se espera medir) e à representatividade (notar se os itens realmente refletem os conceitos envolvidos, se são relevantes e, se são adequados para atingir os objetivos propostos).</p>				
<p>❖ <i>Clareza:</i> avaliar a redação dos itens, ou seja, verificar se eles foram redigidos de forma que o conceito esteja compreensível e se expressa adequadamente o que se espera medir;</p>	<p>Questões – Domínio 1:</p>				
<p>❖ <i>Pertinência ou representatividade:</i> notar se os itens realmente refletem os conceitos envolvidos, se são relevantes e, se são adequados para atingir os objetivos propostos.</p>	<p>1. (Redigir a questão)</p>				
<p>Utilize a escala sobre concordância para avaliar estes critérios, assinalando um X no campo correspondente. Abaixo de cada escala, deixamos espaços para que possa redigir sugestões para melhorar o item, sugerir inclusão e/ou eliminação de itens, ou fazer comentários. Além disso, poderá visualizar o novo instrumento em anexo.</p>	<p>▪ O item 1 do instrumento é claro, está compreensível.</p>				
<p> </p>	<table border="1"> <tr> <td>CONCORDO</td> <td>NÃO CONCORDO</td> </tr> <tr> <td> </td> <td> </td> </tr> </table>	CONCORDO	NÃO CONCORDO		
CONCORDO	NÃO CONCORDO				
<p> </p>	<p>Comentários: _____</p>				
<p> </p>	<p>▪ O item 1 é representativo ao conceito explorado, é relevante.</p>				
<p> </p>	<table border="1"> <tr> <td>CONCORDO</td> <td>NÃO CONCORDO</td> </tr> <tr> <td> </td> <td> </td> </tr> </table>	CONCORDO	NÃO CONCORDO		
CONCORDO	NÃO CONCORDO				
<p> </p>	<p>Comentários: _____</p>				
<p> </p>	<p>(Realizar essa avaliação para cada item de todos os domínios do seu questionário)</p>				

Quanto à pertinência ou representatividade, os juízes devem notar se os itens realmente refletem os conceitos envolvidos, se estes são relevantes e se são e/ou estão adequados para atingir os objetivos propostos^{36,37}. Além disso, os especialistas podem reavaliar a abrangência dos domínios e do instrumento como um todo. Sugestões e/ou comentários para melhorar o item poderão ser feitos em espaços deixados especificamente para essa finalidade^{31,35}.

Neste segundo estágio recomenda-se que a concordância dos membros do comitê seja verificada de forma quantitativa por meio do Índice de Validade de Conteúdo (IVC)^{26,29,38}. Esse índice mede a proporção ou porcentagem de juízes que estão em concordância sobre determinados aspectos do instrumento e de seus itens. Uma ferramenta a ser utilizada para avaliação dos juízes nessa etapa pode ser baseada no Quadro 2.

O IVC é calculado com a utilização de uma escala tipo Likert de 4-pontos ordinais. Para avaliar a relevância/representatividade do item, os juízes poderão escolher as seguintes respostas: 1 = não relevante ou não representativo, 2 = item necessita de grande revisão para ser representativo, 3 = item necessita de pequena revisão para ser representativo, ou 4 = item relevante ou representativo^{26,35,38}. A abrangência, a clareza e a pertinência serão avaliadas com a mesma escala e podem apresentar opções mais curtas, como: 1 = não claro, 2 = pouco claro, 3 = bastante claro, 4 = muito claro^{26,32,39}.

Assim, o cálculo é feito a partir da somatória das respostas “3” e “4” de cada juiz em cada item do questionário e divide-se esta soma pelo número total de respostas, como segue^{29,32}.

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

Os itens que recebem pontuação “1” ou “2” devem ser revisados ou eliminados. A taxa de concordância aceitável entre os juízes para avaliação dos itens individualmente deve ser superior a 0,78⁴⁰. Para a verificação da validade do novo instrumento de uma forma geral, deve haver uma concordância mínima de 0,80³⁶ e, preferencialmente, superior a 0,90⁴⁰.

Nos dois estágios, a avaliação de conteúdo do instrumento é quantitativa. Esta é inicialmente realizada pelos juízes de forma individual, independente e em um período pré-estabelecido. Em seguida deve ser feita uma discussão em grupo que corresponde ao procedimento qualitativo^{31,32,34}.

A abordagem de se usar pelo menos dois métodos, geralmente quantitativo e qualitativo, é denominada triangulação metodológica⁴¹.

A fase qualitativa é representada por um processo interativo entre pesquisadores e os membros do comitê por meio de entrevistas e discussões para clarificar pontos controversos³⁶.

Nessas discussões/entrevistas, todas as sugestões e comentários dos membros do comitê são expostos e anotados pelo pesquisador. Os documentos devolvidos pelos juízes e suas propostas de modificações emitidas durante a reunião devem ser avaliadas e revisadas.

Posteriormente, uma nova versão do instrumento que contemple as sugestões apresentadas pelo comitê pode ser entregue novamente a cada membro. Os juízes deverão revisar seus critérios de avaliação para que seja finalizado o processo de validade de conteúdo e a versão para pré-teste seja obtida.

O procedimento de pré-teste

O pré-teste (também referido como análise semântica dos itens) tem como objetivo verificar se todos os itens são compreensíveis para todos os membros da população a qual o instrumento se destina²².

Esta etapa deve ser realizada em uma amostra de 30-40 indivíduos da população-alvo. Cada sujeito deve completar o questionário e em seguida ser entrevistado individualmente com relação ao entendimento dos itens e das palavras e quanto ao preenchimento das respostas²².

Algumas modificações poderão ser necessárias na versão final do instrumento. Caso sejam alterações significativas, estas devem ser submetidas à avaliação pelos membros do comitê de juízes novamente.

Terminada essa etapa, o instrumento de medida estará pronto para ter suas propriedades psicométricas avaliadas²².

Avaliação das propriedades psicométricas

A robustez dos resultados de um estudo depende muito do instrumento utilizado. Se a qualidade deste estudo é apropriada, os resultados são válidos e o questionário pode ser uma ferramenta útil para novas pesquisas ou para a prática clínica⁴². No entanto, quando a qualidade do estudo é inadequada, os resultados não são confiáveis e a qualidade e utilidade do questionário utilizado no estudo permanecem obscuras⁴².

Quadro 2. Modelo de instruções para segunda análise da validade de conteúdo.

<p>INSTRUÇÕES PARA ANÁLISE DO INSTRUMENTO</p> <p>Para realizar a avaliação do conteúdo do instrumento _____, descrevemos abaixo os conceitos envolvidos no estudo. (Descrever conceitos importantes relacionados ao instrumento com base na literatura a fim de nortear a avaliação do juiz).</p> <p>Pedimos que avalie o título, o formato (lay-out), as instruções, cada item separadamente, e o escore do instrumento (cálculo e classificação), considerando os conceitos de clareza e pertinência/representatividade conforme descrito:</p> <p>❖ <i>Clareza</i>: avaliar a redação, ou seja, verificar se o conceito pode ser bem compreendido e se expressa adequadamente o que se espera medir;</p> <p>❖ <i>Pertinência ou representatividade</i>: notar se há relação com os conceitos envolvidos, se é relevante e se atinge os objetivos propostos.</p> <p>Em seguida, avalie cada domínio e o instrumento como um todo, determinando sua abrangência:</p> <p>❖ <i>Abrangência</i>: verificar se cada domínio foi adequadamente coberto pelo conjunto de itens e se todas as dimensões foram incluídas.</p> <p>Utilize a escala de 1 a 4 para avaliar estes critérios, assinalando um X no campo correspondente. Abaixo de cada escala, deixamos espaço para que possa redigir sugestões ou fazer comentários.</p> <p>O novo instrumento encontra-se em anexo.</p> <p>A próxima etapa será uma reunião no dia _____, às _____, no local _____, quando participarão todos os integrantes do comitê de especialistas, a pesquisadora e a orientadora, com o objetivo de para clarificar pontos controversos e produzir uma versão final do questionário.</p> <p style="text-align: center;">AVALIAÇÃO DA VALIDADE DE CONTEÚDO</p> <p>I. Avalie o título quanto à <u>clareza</u> (verificar se expressa adequadamente o que se espera medir).</p> <p>TÍTULO: (inserir o título do questionário)</p> <ul style="list-style-type: none"> ▪ O título do instrumento é claro e expressa a medida? <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <tr><td style="width: 50%;">1 = não claro</td><td style="width: 50%;"></td></tr> <tr><td>2 = pouco claro</td><td></td></tr> <tr><td>3 = bastante claro</td><td></td></tr> <tr><td>4 = muito claro</td><td></td></tr> </table>	1 = não claro		2 = pouco claro		3 = bastante claro		4 = muito claro		<p>Comentários: _____</p> <hr/> <p>II. Avalie o formato (<i>lay-out</i>) quanto à <u>clareza</u> (verificar se o formato é compreensível) e à adequação.</p> <p>FORMATO DO INSTRUMENTO</p> <ul style="list-style-type: none"> ▪ O formato do instrumento? <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <tr><td style="width: 50%;">1 = não claro</td><td style="width: 50%;"></td></tr> <tr><td>2 = pouco claro</td><td></td></tr> <tr><td>3 = bastante claro</td><td></td></tr> <tr><td>4 = muito claro</td><td></td></tr> </table> <p>Comentários: _____</p> <hr/> <p>III. Avalie as instruções quanto à <u>clareza</u> (verificar se a redação está correta e se expressa adequadamente o que se espera medir).</p> <p>INSTRUÇÕES: (descrever as instruções do questionário)</p> <ul style="list-style-type: none"> ▪ As instruções do instrumento são claras? <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <tr><td style="width: 50%;">1 = não claras</td><td style="width: 50%;"></td></tr> <tr><td>2 = pouco claras</td><td></td></tr> <tr><td>3 = bastante claras</td><td></td></tr> <tr><td>4 = muito claras</td><td></td></tr> </table> <p>Comentários: _____</p> <hr/> <p>IV. Avalie cada item quanto à <u>clareza</u> (verificar se a redação está correta, se a redação permite compreender o conceito e se expressa adequadamente o que se espera medir) e à <u>representatividade</u> (notar se há relação com os conceitos envolvidos, se é relevante e se atinge os objetivos propostos).</p> <p>QUESTÕES – Domínio 1:</p> <p>1. (Redigir a questão)</p> <ul style="list-style-type: none"> ▪ O item 1 do instrumento é claro, está compreensível? <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <tr><td style="width: 50%;">1 = não claro</td><td style="width: 50%;"></td></tr> <tr><td>2 = pouco claro</td><td></td></tr> <tr><td>3 = bastante claro</td><td></td></tr> <tr><td>4 = muito claro</td><td></td></tr> </table> <p>Comentários: _____</p> <hr/>	1 = não claro		2 = pouco claro		3 = bastante claro		4 = muito claro		1 = não claras		2 = pouco claras		3 = bastante claras		4 = muito claras		1 = não claro		2 = pouco claro		3 = bastante claro		4 = muito claro	
1 = não claro																																	
2 = pouco claro																																	
3 = bastante claro																																	
4 = muito claro																																	
1 = não claro																																	
2 = pouco claro																																	
3 = bastante claro																																	
4 = muito claro																																	
1 = não claras																																	
2 = pouco claras																																	
3 = bastante claras																																	
4 = muito claras																																	
1 = não claro																																	
2 = pouco claro																																	
3 = bastante claro																																	
4 = muito claro																																	

continua

Quadro 2. continuação

<p>▪ O item 1 é representativo ao conceito explorado, é relevante?</p> <table border="1"> <tbody> <tr><td>1 = não claro</td><td></td></tr> <tr><td>2 = pouco claro</td><td></td></tr> <tr><td>3 = bastante claro</td><td></td></tr> <tr><td>4 = muito claro</td><td></td></tr> </tbody> </table> <p>Comentários: _____</p> <p>(Realizar essa avaliação para cada item de todos os domínios do seu questionário)</p> <p>V. Avalie o cálculo do escore de cada domínio do instrumento e o cálculo do escore total quanto à <u>clareza</u> (verificar se é compreensível).</p> <p>ESCORE DOMÍNIOS (Descrever como se calcula o escore dos domínios)</p> <p>▪ O cálculo do escore dos domínios é claro, está compreensível?</p> <table border="1"> <tbody> <tr><td>1 = não claro</td><td></td></tr> <tr><td>2 = pouco claro</td><td></td></tr> <tr><td>3 = bastante claro</td><td></td></tr> <tr><td>4 = muito claro</td><td></td></tr> </tbody> </table> <p>Comentários: _____</p> <p>ESCORE TOTAL (se houver) (Descrever como se calcula o escore total do instrumento)</p> <p>▪ O cálculo do escore total é claro, está compreensível?</p> <table border="1"> <tbody> <tr><td>1 = não claro</td><td></td></tr> <tr><td>2 = pouco claro</td><td></td></tr> <tr><td>3 = bastante claro</td><td></td></tr> <tr><td>4 = muito claro</td><td></td></tr> </tbody> </table> <p>Comentários: _____</p> <p>VI. Avalie a classificação desenvolvida para análise do escore quanto à <u>clareza</u> (verificar se está compreensível e se expressa adequadamente o que se espera medir) e à <u>representatividade</u> (notar se há relação com os conceitos envolvidos, se é relevante e se atinge os objetivos propostos).</p> <p>ANÁLISE DO ESCORE (classificação) (Descrever como os resultados dos escores devem ser analisados e classificados)</p> <p>▪ A classificação baseada no escore é clara?</p>	1 = não claro		2 = pouco claro		3 = bastante claro		4 = muito claro		1 = não claro		2 = pouco claro		3 = bastante claro		4 = muito claro		1 = não claro		2 = pouco claro		3 = bastante claro		4 = muito claro		<table border="1"> <tbody> <tr><td>1 = não clara</td><td></td></tr> <tr><td>2 = pouco clara</td><td></td></tr> <tr><td>3 = bastante clara</td><td></td></tr> <tr><td>4 = muito clara</td><td></td></tr> </tbody> </table> <p>Comentários: _____</p> <p>▪ A classificação baseada no escore é representativa, é relevante?</p> <table border="1"> <tbody> <tr><td>1 = não representativa</td><td></td></tr> <tr><td>2 = necessita grande revisão para ser representativa</td><td></td></tr> <tr><td>3 = necessita pouca revisão para ser representativa</td><td></td></tr> <tr><td>4 = representativa</td><td></td></tr> </tbody> </table> <p>Comentários: _____</p> <p>VII. Avalie cada domínio do instrumento considerando a <u>abrangência</u> (se cada domínio ou conceito foi adequadamente coberto pelo conjunto de itens). (Coloque o domínio 1 completo para avaliação)</p> <p>▪ O Domínio 1 é abrangente?</p> <table border="1"> <tbody> <tr><td>1 = não abrangente</td><td></td></tr> <tr><td>2 = necessita grande revisão para ser abrangente</td><td></td></tr> <tr><td>3 = necessita pouca revisão para ser abrangente</td><td></td></tr> <tr><td>4 = abrangente</td><td></td></tr> </tbody> </table> <p>Comentários: _____</p> <p>Alguns itens devem ser removidos ou inseridos? _____</p> <p>(Realizar essa avaliação para cada domínio do seu questionário)</p> <p>VIII. Avalie o instrumento como um todo, ou seja, todos os domínios, considerando a <u>abrangência</u> (verificar se todas as dimensões foram incluídas). O instrumento é abrangente?</p> <table border="1"> <tbody> <tr><td>1 = não abrangente</td><td></td></tr> <tr><td>2 = necessita grande revisão para ser abrangente</td><td></td></tr> <tr><td>3 = necessita pouca revisão para ser abrangente</td><td></td></tr> <tr><td>4 = abrangente</td><td></td></tr> </tbody> </table> <p>Comentários: _____</p>	1 = não clara		2 = pouco clara		3 = bastante clara		4 = muito clara		1 = não representativa		2 = necessita grande revisão para ser representativa		3 = necessita pouca revisão para ser representativa		4 = representativa		1 = não abrangente		2 = necessita grande revisão para ser abrangente		3 = necessita pouca revisão para ser abrangente		4 = abrangente		1 = não abrangente		2 = necessita grande revisão para ser abrangente		3 = necessita pouca revisão para ser abrangente		4 = abrangente	
1 = não claro																																																									
2 = pouco claro																																																									
3 = bastante claro																																																									
4 = muito claro																																																									
1 = não claro																																																									
2 = pouco claro																																																									
3 = bastante claro																																																									
4 = muito claro																																																									
1 = não claro																																																									
2 = pouco claro																																																									
3 = bastante claro																																																									
4 = muito claro																																																									
1 = não clara																																																									
2 = pouco clara																																																									
3 = bastante clara																																																									
4 = muito clara																																																									
1 = não representativa																																																									
2 = necessita grande revisão para ser representativa																																																									
3 = necessita pouca revisão para ser representativa																																																									
4 = representativa																																																									
1 = não abrangente																																																									
2 = necessita grande revisão para ser abrangente																																																									
3 = necessita pouca revisão para ser abrangente																																																									
4 = abrangente																																																									
1 = não abrangente																																																									
2 = necessita grande revisão para ser abrangente																																																									
3 = necessita pouca revisão para ser abrangente																																																									
4 = abrangente																																																									

Portanto, ao desenvolver e/ou selecionar instrumentos que serão utilizados em pesquisas e/ou na prática clínica, é muito importante que estes sejam ou tenham sido avaliados quanto às suas propriedades psicométricas⁴³.

A literatura descreve vários atributos que podem ser testados no processo de avaliação das propriedades psicométricas de um instrumento^{14,44,45}: validade (*validity*), confiabilidade (*reliability*), praticabilidade (*practicability*), sensibilidade (*sensitivity*), responsividade (*responsiveness*) e interpretabilidade (*interpretability*).

A seleção dos atributos que devem ser avaliados, bem como o método a ser utilizado vai depender do tipo e objetivos de cada instrumento. Dentre os atributos citados, aqueles mais comumente utilizados são a validade, a confiabilidade e a responsividade⁴².

A validade é a capacidade de um instrumento medir com precisão o fenômeno a ser estudado^{14,19,27,46}. A avaliação deste atributo pode ser feita de várias maneiras: validade de conteúdo, validade de constructo e validade de critério^{1,3,47}. Cada uma dessas maneiras avalia aspectos diferentes do instrumento e deve ser pensada como parte de um processo³.

A confiabilidade é a capacidade do instrumento em reproduzir um resultado de forma consistente no tempo e no espaço ou com observadores diferentes^{14,27}. Ou seja, confiabilidade refere-se à quão estável, consistente ou preciso é o instrumento^{19,28}. Os procedimentos utilizados para a avaliação da confiabilidade também são diversos. Dentre eles, os mais utilizados são: consistência interna (homogeneidade) e estabilidade (confiabilidade teste-reteste, confiabilidade interobservadores ou intraobservadores)³.

A responsividade é a capacidade de um instrumento detectar mudanças sobre o constructo a ser medido ao longo do tempo^{14,42}. Este atributo se refere à validade no contexto longitudinal. A diferença entre a validade e responsividade é que a primeira se refere à validade de um escore

simples e a responsividade se refere à validade de mudanças nesse escore⁴².

Com relação às propriedades psicométricas, é importante ressaltar que estas não são estáticas do instrumento. Ou seja, elas podem variar de acordo com a mudança da população de estudo³.

Além disso, outros fatores que também podem influenciar na avaliação das propriedades psicométricas são: modo de administração (entrevista, telefone, ou autoaplicado), tipo de população-alvo, tamanho amostral, dentre outros^{11,48}.

Conclusões

Este estudo realizou uma discussão de forma sistematizada sobre os procedimentos recomendados para a construção de novos instrumentos de saúde a serem utilizados tanto na assistência como em pesquisas.

Compreender, analisar e seguir o processo descrito neste estudo é essencial para pesquisadores e profissionais da área de saúde, que estejam preocupados em construir e utilizar instrumentos de medidas cada vez mais confiáveis e apropriados.

Colaboradores

MZO Coluci trabalhou na concepção, no desenvolvimento dos modelos para avaliação da validade de conteúdo, na redação e revisão final do artigo. D Milani trabalhou na concepção, redação e revisão final e crítica do artigo. NMC Alexandre trabalhou na concepção, no delineamento do estudo e na revisão crítica do artigo.

Referências

1. Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence* 2011; 5:279-290.
2. Marx RG, Bombardier C, Hogg-Johnson S, Wright JG. Clinimetric and psychometric strategies for development of a health measurement scale. *J Clin Epidemiol* 1999; 5(2):105-111.
3. Keszei A, Novak M, Streiner DL. Introduction to health measurement scales. *J Psychosom Res* 2010; 68(4):319-323.
4. Rattray J, Jones MC. Essential elements of questionnaire design and development. *J Clin Nurs* 2007; 16(2):234-243.
5. Gallasch CH, Alexandre NMC, Amick III B. Cross-cultural adaptation, Reliability, and validity of the Work Role Functioning Questionnaire to Brazilian Portuguese. *J Occup Rehabil* 2007; 17(4):701-711.
6. Vigatto R, Alexandre NMC, Correa Filho HR. Development of a Brazilian Portuguese version of the Oswestry Disability Index. Cross-cultural adaptation, reliability, and validity. *Spine* 2007; 32(4):481-486.
7. Nakajima KM, Rodrigues RCM, Gallani MCBJ, Alexandre NMC, Oldridge N. Psychometric properties of MacNew Heart Disease Health-related Quality of Life Questionnaire: Brazilian version. *J Adv Nurs* 2009; 65(5):1084-1094.
8. Toledo R, Alexandre NMC, Rodrigues RCM. Psychometric evaluation of a Brazilian Portuguese version of the Spitzer Quality of Life Index in patients with low back pain. *Rev Lat Am Enfermagem* 2008; 16(6):943-950.
9. Cano SJ, Hobart JC, Fitzpatrick R, Bhatia K, Thompson AJ, Warner TT. Patient-based outcomes of cervical dystonia: a review of rating scales. *Mov Disord* 2004; 19(9):1054-1059.
10. Chen CM, Cano SJ, Klassen AF, King T, McCarthy C, Cordeiro PG, Morrow M, Pusic AL. Measuring quality of life in oncologic breast surgery: a systematic review of patient-reported outcome measures. *Breast J* 2010; 16(6):58-97.
11. Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use.* 4th ed. New York: Oxford University Press; 2008.
12. Alexandre NMC, Guirardello EB. Adaptación cultural de instrumentos utilizados em salud ocupacional. *Rev Panam Salud Publica* 2002; 11(2):109-111.
13. Hutchinson A, Bentzen N, Konig-Zanhn C. *Cross cultural health outcome assessment: a user's guide.* The Netherlands: ERGHO; 1996.
14. Terwee CB, Bot SDM, Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60(1):34-42.
15. Turner R, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcomes: instrument development and selection issues. *Value Health* 2007; 10(Supl. 2):S86-S93.
16. 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.
17. Rother ET. Revisão sistemática X revisão narrativa. *Acta paul enfer* 2007; 20(2):V-VI.
18. 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.
19. Pittman J, Bakas T. Measurement and instrument design. *J Wound Ostomy Continence Nurs* 2010; 37(6):603-607.
20. Snyder CF, Watson ME, Jackson JD, Cella D, Halyard MY, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcome instrument selection: designing a measurement strategy. *Value Health* 2007; 10(Supl. 2):S76-S85.
21. Günther H. Como Elaborar um Questionário. *Serie Planejamento de pesquisa nas ciências sociais* 2003; 1:1-15.
22. Pasquali L. Princípios de elaboração de escalas psicológicas. *Rev Psiq Clin* 1998; 25(5):206-213.
23. Cardoso CS, Bandeira M, Ribeiro ALP, Oliveira GL, Caiáffa WT. Escalas de satisfação com o atendimento às doenças cardiovasculares: Cardiosatis- usuário equipe. *Cien Saude Colet* 2011; 16(Supl. 1):1401-1407.
24. Sireci SG. The construct of content validity. *Soc Indic Res* 1998; 45(1-3):83-117.
25. Fagarasanu M, Kumar S. Measurement instruments and data collection: a consideration of constructs and biases in ergonomics research. *Int J Ind Ergon* 2002; 30(6):355-369.
26. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. *West J Nurs Res* 2003; 25(5):508-518.
27. Contandriopoulos AP. *Saber preparar uma pesquisa.* 3^a ed. São Paulo, Rio de Janeiro: Hucitec, Abrasco; 1999.
28. Polit DF, Beck CT, Hungler BP. *Fundamentos em pesquisa em enfermagem: métodos, avaliação e utilização.* 5^a ed. Porto Alegre: Artmed Editora; 2004.
29. Alexandre NMC, Coluci, MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medida. *Cien Saude Colet* 2011; 16(7):3061-3067.
30. Burns N, Grove SK. *The practice of nursing research: conduct, critique & utilization.* 3^a ed. Philadelphia: Saunders Company; 1997.
31. Tilden VP, Nelson CA, May BA. Use of qualitative methods to enhance content validity. *Nurs Res* 1990; 39(3):172-175.
32. Hyrkas K, Appelqvist-Schmidlechner K, Oksa L. Validating an instrument for clinical supervision using an expert panel. *Int J Nurs Stud* 2003; 40:619-625.
33. Salmond SS. Evaluating the reliability and validity of measurement instruments. *Orthop Nurs* 2008; 27(1):28-30.
34. Berk RA. Importance of expert judgment in content-related validity evidence. *West J Nurs Res* 1990; 12(5):659-671.
35. 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.
36. Grant JS, Davis LL. Selection and use of content experts for instrument development. *Res Nurs Health* 1997; 20(3):269-274.

37. McGilton K. Development and psychometric evaluation of supportive leadership scales. *Can J Nurs Res* 2003; 35(4):72-86.
38. Summers S. Establishing the Reliability and Validity of a New Instrument: Pilot Testing. *J Post Anesth Nurs* 1993; 8(2):124-127.
39. 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):155-164.
40. 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(5):489-497.
41. Morse JM. Approaches to qualitative-quantitative methodological triangulation. *Nurs Res* 1991; 40(1):120-123.
42. Terwee CB, Schellingerhout JM, Verhagen AP, Koes BW, de Vet HC. Methodological quality of studies on the measurement properties of neck pain and disability questionnaires: a systematic review. *J Manipulative Physiol Ther* 2011; 34(4):261-272.
43. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther* 2008; 88(2):156-175.
44. Kimberling CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008; 65(23):2276-2284.
45. 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):155-164.
46. Roberts P, Priest H, Traynor M. Reliability and validity in research. *Nurs Stand* 2006; 20(44):41-45.
47. Pasquali L. Psicometria. *Rev Esc Enf USP* 2009; 43(esp):992-999.
48. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient for reliability and validity of patient-reported outcome measures? *Value Health* 2007; 10(Supl. 2):S94-S105.

Artigo apresentado em 15/03/2013

Aprovado em 14/04/2013

Versão final apresentada em 16/04/2013