

Avaliação das adaptações transculturais e propriedades de medida de questionários relacionados às disfunções do ombro em língua portuguesa: uma revisão sistemática

Assessment of cross-cultural adaptations and measurement properties of self-report outcome measures relevant to shoulder disability in Portuguese: a systematic review

Vanessa O. O. Puga¹, Alexandre D. Lopes¹, Leonardo O. P. Costa^{1,2}

Resumo

Objetivos: Avaliar os procedimentos de tradução/adaptação cultural e das propriedades de medida de questionários que avaliam dor e disfunções no ombro, os quais já foram traduzidos/adaptados para a língua portuguesa. **Métodos:** Foram realizadas buscas sistematizadas nas bases de dados eletrônicas MEDLINE, EMBASE, CINAHL, SCIELO e LILACS para identificar os estudos relevantes. Foram extraídos os dados referentes à tradução e adaptação cultural, além dos dados das propriedades de medida de cada estudo. Todos os estudos foram analisados quanto à sua respectiva qualidade metodológica de acordo com as diretrizes para adaptação cultural e para as propriedades de medida. **Resultados:** Um total de 876 estudos foi identificado nas buscas, e, desses, apenas 11 foram considerados elegíveis, sendo que eles adaptaram e/ou testaram sete instrumentos diferentes (DASH, WORC, SPADI, PSS, ASORS, ASES e UCLA). A maioria deles cumpriu adequadamente as recomendações das diretrizes de adaptação transcultural. Dois dos sete questionários não foram testados para nenhuma propriedade de medida (PSS e ASES), e apenas dois questionários (WORC e DASH) foram testados para praticamente todas as propriedades de medida, porém nem todas foram testadas adequadamente. Nenhum questionário testou por completo todas as propriedades de medida. **Conclusões:** Os processos de tradução e adaptação transcultural foram realizados de maneira adequada para a maioria dos instrumentos, porém a maioria não teve suas propriedades de medida testadas adequadamente. Recomenda-se que somente instrumentos testados para suas respectivas propriedades de medida sejam utilizados na prática clínica assim como em pesquisas.

Palavras-chave: questionários; tradução (produto); estudos de validação; ombro; fisioterapia.

Abstract

Objectives: To evaluate the quality of the adaptation procedures as well as the clinimetric testing of the shoulder disability questionnaires available in Portuguese that has occurred for each adaptation. **Methods:** Systematic literature searches on MEDLINE, EMBASE, CINAHL, SCIELO and LILACS were performed to identify relevant studies. Data on the quality of the cross-cultural adaptation procedures and clinimetric testing were extracted. All studies were evaluated according to the current guidelines for cross-cultural adaptation and measurement properties. **Results:** Seven different questionnaires adapted into Brazilian-Portuguese (DASH, WORC, SPADI, PSS, ASORS, ASES and UCLA) were indentified from eleven studies. Most of the studies performed the cross-cultural adaptation procedures following the recommendations from the guidelines. From a total of seven instruments, two were not tested for any measurement property (PSS and ASES) and two questionnaires (DASH and WORC) were evaluated for almost all of properties. None of the questionnaires were fully tested for their measurement properties. **Conclusions:** Although most of the shoulder disability questionnaires have been properly adapted into Brazilian-Portuguese, some of them were either inadequately tested or not tested at all. It is recommended that only tested instruments can be used in clinical practice, as well as in research.

Keywords: questionnaire; translation; validation studies; shoulder; physical therapy.

Recebido: 12/06/2011 – **Revisado:** 10/09/2011 – **Aceito:** 26/09/2011

¹Programa de Mestrado em Fisioterapia, Universidade Cidade de São Paulo (UNICID), São Paulo, SP, Brasil

²Musculoskeletal Division, The George Institute for Global Health, Sydney, NSW, Australia

Correspondência para: Leonardo Oliveira Pena Costa, Rua Cesário Galeno, 48/475, Tatuapé, CEP 03071-000, São Paulo, SP, Brasil, e-mail: lcosta@edu.unicid.br

Introdução

A dor na articulação do ombro é considerada a terceira condição musculoesquelética mais comum encontrada na prática médica, seguida da dor lombar e dor cervical, sendo que a prevalência de distúrbios do ombro varia entre 7% e 36% na população em geral^{1,2}. As disfunções da articulação do ombro são encontradas com frequência em trabalhadores e atletas que são expostos a movimentos repetitivos e esforços excessivos, porém alguns estudos também reportam uma alta prevalência desse problema em idosos e em indivíduos sedentários^{2,4}.

Os métodos de avaliação das lesões musculoesqueléticas têm sido modificados nos últimos anos⁵. Ao invés da avaliação ser feita somente de acordo com o exame físico, que inclui testes de força muscular, mobilidade articular e avaliação dos exames de imagem, a avaliação pode ser acompanhada por questionários ou escalas^{6,7}. Os questionários são amplamente utilizados para coletar informações de importantes desfechos clínicos, como a intensidade da dor, os níveis de qualidade de vida do paciente, a satisfação com o tratamento e a incapacidade de realizar atividades diárias^{1,8,9}. Além da utilização dos questionários na prática clínica, eles também são muito utilizados na pesquisa científica^{10,11}.

Existem vários instrumentos para avaliar pacientes com disfunções no ombro e detectar mudanças em seu quadro clínico ao longo do tempo, porém a maioria deles foi desenvolvida na língua inglesa^{9,12}. Para que esses instrumentos sejam utilizados no Brasil, recomenda-se a tradução, a adaptação cultural, assim como a realização de testes que avaliem as propriedades de medida desses instrumentos, como consistência interna, reprodutibilidade, validade e responsividade¹³⁻¹⁵. O processo de tradução e adaptação é necessário para que seja possível testar se há equivalência com a versão original, resolvendo as diferenças de costumes, linguagem e percepção da saúde entre países e culturas diferentes¹⁵. Faz-se necessário testar as propriedades de medida, mesmo que elas já tenham sido testadas no questionário original, já que pode haver diferenças culturais entre populações distintas, e também avaliar se o questionário adaptado retém as propriedades de medida da versão original¹⁴. Existem diretrizes que auxiliam a realizar, de maneira adequada, tanto a tradução e a adaptação transcultural¹³ quanto os testes das propriedades de medida¹³⁻¹⁵.

Existem questionários específicos para avaliar disfunções do ombro que já foram traduzidos e adaptados para a língua portuguesa e que tiveram algumas de suas propriedades de medida testadas, porém ainda não existe um documento que sintetize toda a informação sobre esses questionários. Sendo assim, por meio de uma revisão sistemática para identificar todos esses instrumentos, os objetivos deste estudo foram descrever e avaliar os procedimentos de tradução e adaptação cultural desses instrumentos e descrever e avaliar as propriedades de medida que foram testadas em cada um dos estudos.

Materiais e métodos

Seleção dos estudos

Para identificar os instrumentos que avaliam dor e incapacidade no ombro em língua portuguesa, foram realizadas buscas sistematizadas em cinco bases de dados eletrônicas (MEDLINE via OVID, EMBASE, CINAHL via EBSCO, SCIELO e LILACS). Os termos de busca e os operadores (*AND*, *OR* ou *NOT*) utilizados nas bases MEDLINE, EMBASE, CINAHL e SCIELO foram: (*shoulder OR impingement OR rotator cuff OR arm OR upper limb OR instability OR upper extremity*) *AND* (*questionnaire OR index OR scale OR score OR assessment OR evaluation OR selfreport OR inventory*) *AND* (*Brazil OR Brasil OR Portuguese OR Brazilian Portuguese OR Brazilian*). Os termos de busca utilizados na base LILACS foram: (*ombro OR manguito rotador OR membro superior OR instabilidade*) *AND* (*questionário OR escala OR índice OR instrumento OR escore OR avaliação*) *AND* (*Brasil OR português OR português brasileiro*). As buscas não foram limitadas por língua ou data de publicação. A última busca foi realizada em abril de 2011.

Critérios de inclusão

Foram incluídos no estudo, somente instrumentos que foram desenvolvidos para avaliação das disfunções da articulação do ombro que foram submetidos a qualquer procedimento de tradução e/ou adaptação para a língua portuguesa, podendo ser combinados com disfunções de outras articulações do membro superior. Os instrumentos de autoavaliação e instrumentos dependentes do avaliador que utilizaram medidas objetivas, como força e amplitude de movimento também foram incluídos nesta revisão. Apenas estudos que foram publicados em textos completos foram incluídos, e foram excluídos textos advindos de teses/dissertações, resumos de congressos e livros.

Extração de dados e avaliação da qualidade metodológica dos estudos elegíveis

Foram extraídos os dados referentes à tradução e adaptação cultural com o objetivo de avaliar como foram realizados os procedimentos de tradução e adaptação cultural, além dos dados referentes às propriedades de medida de cada estudo, incluindo o tamanho amostral, os dados referentes à reprodutibilidade (confiabilidade e concordância), responsividade e seus intervalos de medidas, consistência interna e validade do construto (esses dados podem ser requeridos ao autor correspondente via e-mail).

Posteriormente, cada estudo foi classificado em relação ao método de tradução e adaptação cultural de acordo com as Diretrizes de Adaptação Transcultural de Questionários¹³. O processo de tradução e adaptação inclui a tradução inicial, a síntese das traduções, a retrotradução, a revisão de um comitê de especialistas e o teste da versão pré-final. A qualidade de cada etapa foi classificada como positiva (+) quando o procedimento realizado estava de acordo com o critério de qualidade adotado; duvidosa (?), quando o método era realizado de forma questionável; negativa (-), quando o procedimento era realizado corretamente, porém com número insuficiente de tradutores e/ou retrotradutores; ou nula (0), quando não havia informações suficientes para qualificar cada etapa (Tabela 1).

As propriedades de medida foram classificadas por meio dos Critérios de Qualidade para Propriedades de Medida de Questionários da Área de Saúde¹⁴, e a avaliação foi restrita somente aos itens relevantes para os instrumentos avaliados. Outros itens originais dos Critérios de Qualidade, como a validade do conteúdo ou validade de face e a interpretabilidade, somente são relevantes durante o desenvolvimento do questionário original. Da mesma forma, o item validade do critério só pode ser considerado quando há um instrumento de comparação que seja padrão-ouro, o que não é o caso dos instrumentos de avaliação do ombro, portanto esses três itens não foram considerados aplicáveis neste estudo.

Os itens avaliados neste estudo foram a validade do construto, a consistência interna, a reprodutibilidade (concordância e confiabilidade), a responsividade e os efeitos

de teto e piso. Assim, como nos critérios adotados para a adaptação transcultural, a qualidade de cada etapa dos critérios de qualidade das propriedades de medida também foi classificada como positiva (+) quando os procedimentos de cada etapa eram realizados de acordo com o critério de qualidade adotado; duvidosa (?), quando os métodos ou o delineamento do estudo eram questionáveis; negativa (-), quando os dados referentes a cada propriedade clinimétrica tinham valores maiores ou menores que os definidos pelos critérios adotados, mesmo que o delineamento ou método do estudo estivessem adequados; ou nula (0), quando não havia informação suficiente para qualificar cada propriedade de medida (Tabela 2). A extração de dados e as avaliações foram realizadas por um avaliador e, em seguida, conferidas por um revisor independente que checou todos os dados. Não houve nenhuma divergência de opiniões entre o avaliador e o revisor independente, que se reuniram e discutiram os dados avaliados.

Resultados

Foram encontrados 876 estudos nas buscas, sendo que apenas 11 foram considerados elegíveis para a análise dos dados (Figura 1). Desses, sete instrumentos diferentes que foram traduzidos/adaptados para a língua portuguesa foram identificados, sendo eles: DASH (*Disabilities of the Arm, Shoulder and Hand*)¹⁶, SPADI (*Shoulder Pain and Disability Index*)¹⁷, WORC (*Western Ontario Rotator Cuff Index*)¹⁸, ASES (*American*

Tabela 1. Diretrizes para os procedimentos de adaptação transcultural de questionários¹³ (adaptado por Costa et al.³¹).

Etapa	Descrição	Critérios de qualidade
Tradução	Dois (ou mais) tradutores devem traduzir o questionário de forma independente. Preferencialmente a língua nativa dos tradutores deve ser a língua-alvo da tradução.	+ Tradução realizada por dois ou mais tradutores independentes; ? Processo de tradução questionável; - Tradução realizada por somente um tradutor; 0 Não há informações sobre o processo de tradução.
Síntese das traduções	Os tradutores devem sintetizar todas as traduções e produzir uma versão consensual.	+ Síntese realizada; ? Delineamento questionável; 0 Não há informações sobre a síntese ou a tradução foi realizada somente por um tradutor.
Retrotradução	Tradutores que não tenham conhecimento do questionário original devem traduzir a versão consensual das traduções novamente para a língua original do questionário.	+ Retrotradução realizada por dois ou mais tradutores independentes. ? Procedimento de retrotradução questionável. - Retrotradução realizada por somente um tradutor. 0 Não há informações sobre o processo de retrotradução.
Análise do comitê	Um comitê de especialistas deve analisar todas as versões do questionário e desenvolver o que será considerada a versão pré-final do questionário.	+ Foi claramente relatada a existência de um comitê de especialistas. ? Delineamento questionável. 0 Não há informações sobre o comitê de especialistas.
Pré-teste da versão pré-final	A versão pré-final deve ser testada em membros da população-alvo.	+ Foi realizado o pré-teste. ? Desenho duvidoso. 0 Não há informações sobre o pré-teste.

+classificação positiva; -classificação negativa; 0=não há informações disponíveis; ?=não está claro.

Tabela 2. Critérios de Qualidade para Propriedades de Medida de Questionários da Área de Saúde¹⁴ (adaptado por Costa et al.³¹).

Propriedade clinimétrica	Definição	Critérios de qualidade
Consistência interna	Consistência interna é uma medida de homogeneidade de uma (sub) escala de um questionário. Ela indica o grau em que os itens de uma (sub) escala estão relacionados entre si, mensurando o mesmo construto. A análise fatorial deve ser aplicada para determinar a dimensionalidade das escalas, determinando assim quantas dimensões são compostas pelos itens do instrumento.	+ Análise fatorial realizada em uma amostra com tamanho adequado (Sete sujeitos por item ou no mínimo 100 sujeitos no total) E Alfa de Cronbach calculado por dimensão E Alfa de Cronbach entre 0,70 e 0,95; ? Não há análise fatorial OU delineamento ou método questionável; - Alfa de Cronbach <0,70 ou >0,95, apesar de o delineamento e método estarem adequados; 0 Não há informação sobre a consistência interna.
Validade do construto	A validade do construto examina a extensão de cada pontuação de um questionário específico relacionado com outras medidas similares, de forma que seja consistente com as hipóteses relacionadas com os conceitos que estão sendo medidos.	+ Hipóteses específicas foram formuladas E, pelo menos, 75% dos resultados estão de acordo com estas hipóteses; ? Delineamento ou método questionável (ex. hipóteses não foram formuladas); - Menos de 75% das hipóteses foram confirmadas, apesar de o delineamento e método estarem adequados; 0 Não há informações sobre a validade do construto.
Reprodutibilidade	O grau em que medidas repetidas em indivíduos estáveis (teste-reteste) fornecem respostas semelhantes. A reprodutibilidade é um termo guarda-chuva para confiabilidade e concordância (descrição abaixo)	
Confiabilidade	Avalia até que ponto os indivíduos podem ser distinguidos entre si, apesar dos erros de medida (erro relativo da medida).	+ CCI ou Kappa $\geq 0,70$; ? Delineamento ou método duvidoso (ex. intervalo de tempo entre as medidas não-mencionado ou não-justificado); - CCI ou Kappa <0,70, apesar de o delineamento e método estarem adequados; 0 Não há informações sobre a confiabilidade.
Concordância	A concordância mede o quão próximo duas ou mais medidas repetidas estão uma das outras (erro absoluto da medida).	+ MMI < MMD ou MMI fora do LOC ou argumentos convincentes de que a concordância é aceitável; ? Delineamento ou método duvidoso ou MMI não definido E sem argumentos convincentes de que a concordância é aceitável; - MMI \geq MMD ou MMI igual ou dentro do LOC, apesar de o desenho e método estarem adequados; 0 Não há informações sobre a concordância.
Responsividade	A capacidade do questionário em detectar mudanças clínicas ao longo do tempo no construto que está sendo medido.	+ MMD individual ou MMD grupo < MMI OU MMI fora dos LOC OU razão da responsividade > 0,96 OU área abaixo da curva $\geq 0,70$; ? Método ou delineamento questionável OU tamanho amostral <50 participantes OU falhas metodológicas graves; - MMD individual ou MMD grupo \geq MMI OU MMI dentro dos limites da concordância OU razão da responsividade $\leq 0,96$ OU área abaixo da curva <0,70, apesar do delineamento e método adequados; 0 Não há informações sobre a responsividade.
Efeitos de teto e piso	O número de entrevistados que alcançaram o máximo ou mínimo escore possível.	+ $\leq 15\%$ dos entrevistados alcançaram o máximo ou mínimo escore possível; ? Método ou delineamento questionável OU tamanho da amostra <50 OU falhas metodológicas graves; - >15% dos entrevistados alcançaram o máximo ou mínimo escore possível, apesar de o delineamento e métodos estarem adequados; 0 Não há informação sobre a efeitos de teto e piso.

+ = classificação positiva; ? = delineamento ou método duvidoso; - = classificação negativa; 0 = não há informação disponível. Delineamento ou método duvidoso = falta de uma descrição clara do delineamento ou dos métodos do estudo, amostra menor do que 50 participantes ou qualquer importante problema metodológico no desenho ou na execução do estudo. MMI = Mudança minimamente importante; MMD = Menor mudança detectável; LOC = Limites de concordância; CCI = Coeficiente de Correlação Intraclasse; DP = Desvio-padrão.



Figura 1. Fluxograma da revisão sistemática.

Shoulder and Elbow Surgeons Questionnaire)¹⁹, PSS (*Penn Shoulder Score*)²⁰, ASORS (*Athletic Shoulder Outcome Rating Scale*)²⁰ e *Modified-UCLA (Modified-University of California at Los Angeles Shoulder Rating Scale)*²¹. Dos sete instrumentos traduzidos/adaptados para a língua portuguesa, cinco foram testados em relação às suas propriedades de medida (DASH^{16,22}, WORC^{18,22-24}, SPADI¹⁷, ASORS²⁵ e *Modified-UCLA*²⁴). Somente o PSS²⁰ e o ASES¹⁹ não apresentavam nenhuma das propriedades de medida testadas.

A Tabela 3 apresenta as avaliações das traduções/adaptações de acordo com as Diretrizes de Adaptação Transcultural de Questionários¹³. Dos sete instrumentos encontrados nos 11 estudos avaliados, todos realizaram corretamente as etapas de tradução, síntese, análise do comitê de especialistas e o pré-teste. A etapa de retrotradução não foi realizada adequadamente para os instrumentos WORC e ASORS, pois foi realizada somente por um único tradutor, não seguindo os critérios de qualidade adotados, segundo os quais essa etapa deveria ser realizada por dois ou mais tradutores independentes.

A Tabela 4 apresenta as avaliações de todas as propriedades de medida avaliadas de acordo com os Critérios de Qualidade para Propriedades de Medida de Questionários da Área de Saúde¹⁴. A confiabilidade foi a propriedade de medida mais testada, já que, dos cinco instrumentos que testaram as propriedade de medida, quatro deles a testaram, (somente o *Modified-UCLA* não a testou), além de terem realizado o teste de maneira correta e de obterem valores do Coeficiente de Correlação Intraclasse (CCI) $\geq 0,70$ ^{16,17,23,25}. Já a

Tabela 3. Análise das adaptações transculturais dos questionários de dor e incapacidade no ombro em língua portuguesa de acordo com as Diretrizes do Processo de Adaptação Transcultural dos Questionários¹³.

Estudos	Tradução	Síntese	Retrotradução	Revisão do comitê de especialistas	Pré-teste
DASH ¹⁶	+	+	+	+	+
UCLA ²¹	+	+	+	+	+
WORC ¹⁸	+	+	-*	+	+
WORC ²³	N/A	N/A	N/A	N/A	N/A
DASH ²⁶	N/A	N/A	N/A	N/A	N/A
WORC e DASH ²⁴	N/A	N/A	N/A	N/A	N/A
WORC, DASH e UCLA ²⁴	N/A	N/A	N/A	N/A	N/A
ASES ¹⁹	+	+	+	+	+
PSS ²⁰	+	+	+	+	+
SPADI ¹⁷	+	+	+	+	+
ASORS ²⁵	+	+	-*	+	+

DASH=Disabilities of the Arm, Shoulder and Hand; UCLA=University of California Los Angeles Shoulder Rating Scale; WORC=Western Ontario Rotator Cuff Index; ASES=American Shoulder and Elbow Surgeons Questionnaire; PSS=Penn Shoulder Score; ASORS=Athletic Shoulder Outcome Rating Scale; SPADI=Shoulder, Pain and Disability Index. N/A=Não se aplica - Não foi realizada a tradução e adaptação cultural, somente os testes clinicométricos. Os questionários utilizados nesses estudos já haviam sido traduzidos anteriormente em outros estudos. * Retrotradução realizada apenas por um tradutor.

Tabela 4. Testes das propriedades de medida das versões em língua portuguesa dos questionários de dor e incapacidade do ombro de acordo com os Critérios de Qualidade para Propriedades de Medida de Questionários da Área de Saúde.

Estudos	Reprodutibilidade (Concordância)	Reprodutibilidade (Confiabilidade)	Consistência interna	Responsividade	Validade do construto	Efeitos de Teto e piso	Notas
DASH ¹⁶	0	+	0	0	?	0	Hipóteses não foram formuladas
UCLA ²¹	0	0	0	0	0	0	
WORC ¹⁸	0	0	0	0	0	0	
WORC ²³	+	+	?	0	?	0	*Delineamento questionável da consistência interna. Hipóteses não foram formuladas.
DASH ²⁶	0	0	+	0	0	0	
WORC e DASH ²²	0	0	0	0	?	0	Hipóteses não foram formuladas.
WORC, DASH e UCLA ²²	0	0	0	?	0	0	Tamanho amostral menor que 50 participantes.
ASES ¹⁹	0	0	0	0	0	0	
PSS ²⁰	0	0	0	0	0	0	
ASORS ²⁵	0	+	0	0	?	0	Hipóteses não foram formuladas.
SPADI ¹⁷	0	+	?	0	0	0	Não realizou análise fatorial

DASH=Disabilities of the Arm, Shoulder and Hand; UCLA=University of California Los Angeles Shoulder Rating Scale; WORC=Western Ontario Rotator Cuff Index; ASES=American Shoulder and Elbow Surgeons Questionnaire; PSS=Penn Shoulder Score; ASORS=Athletic Shoulder Outcome Rating Scale; SPADI=Shoulder, Pain and Disability Index. * Consistência interna avaliada intra e interexaminador.

concordância foi testada somente no WORC, no mesmo estudo em que foi testada a confiabilidade desse instrumento, também apresentando valores adequados²³. A consistência interna foi testada no DASH²⁶, no SPADI¹⁷ e no WORC²³, sendo que os dois primeiros apresentaram níveis aceitáveis (Alfa de Cronbach entre 0,70 e 0,95), e o último apresentou um delineamento questionável, pois a consistência interna foi testada intra e interexaminador, não seguindo os critérios sugeridos pela diretriz¹⁴. Somente o DASH realizou a análise fatorial. A validade do construto não foi testada adequadamente em nenhum dos instrumentos encontrados nesta revisão (DASH^{16,22}, WORC^{23,24} e ASORS²⁵) devido ao fato de as hipóteses entre as correlações dos instrumentos/escalas não terem sido formuladas *a priori*. A responsividade foi testada em um estudo que utilizou os instrumentos DASH, WORC e UCLA²⁴, mas não pôde ser considerada adequada, já que o tamanho amostral foi de apenas 30 pacientes. Os efeitos de teto e piso não foram testados em nenhum dos instrumentos identificados em nossa revisão.

O DASH e o WORC tiveram quase todas as suas propriedades de medida testadas, o WORC somente não testou os efeitos de teto e piso²²⁻²⁴, e o DASH, além dos efeitos de teto e piso, também não testou a concordância^{20,23,24,26}. Tanto o SPADI quanto o ASORS tiveram duas de suas propriedades de medida testadas. O SPADI testou adequadamente tanto a confiabilidade quanto a consistência interna¹⁷, e o ASORS testou adequadamente a confiabilidade, porém não testou adequadamente a validade do construto²⁵. O UCLA somente testou a responsividade, mas foi avaliada de forma inadequada²⁴.

Discussão

Os objetivos do estudo foram descrever e avaliar os procedimentos de tradução e adaptação cultural e das propriedades de medida de instrumentos que avaliam dor e incapacidade no ombro, os quais foram traduzidos e adaptados para a língua portuguesa por meio de uma revisão sistemática que os identificou. Foram encontrados sete instrumentos diferentes que avaliam disfunções na articulação do ombro que foram traduzidos/adaptados para a língua portuguesa, e cinco deles tiveram algumas de suas propriedades de medida testadas. Os processos de tradução/adaptação foram avaliados de acordo com as Diretrizes de Adaptação Transcultural de Questionários¹³, e as propriedades de medida foram avaliadas de acordo com os Critérios de Qualidade para Propriedades de Medida de Questionários da Área de Saúde¹⁴.

Os instrumentos DASH¹⁶, SPADI¹⁷, UCLA²¹, PSS²⁰ e ASES¹⁹ seguiram as diretrizes de tradução/adaptação cultural¹³ e tiveram todas as suas etapas testadas corretamente; o WORC¹⁸ seguiu o protocolo específico de equivalência linguística sugerido pelos autores da versão original do instrumento, seguindo os critérios definidos pelo *MAPI Research Institute*²⁷ e, portanto, teve a etapa de retrotradução realizada por um único tradutor. Os autores deste estudo relataram que as diretrizes propostas por Guillemin, Bombardier e Beaton²⁸ e Beaton et al.¹³ podem apresentar processos de difícil realização, dependendo da população testada, complexidade das etapas, longa duração e alto custo, por isso seguindo o protocolo sintetizado, ressaltando que todas as versões em desenvolvimento do WORC em outras línguas

utilizavam o mesmo protocolo. O ASORS²⁵ seguiu as diretrizes propostas por Guillemain, Bombardier e Beaton²⁸, porém não realizou a etapa de retrotradução de acordo com essa diretriz, traduzindo-a somente por um tradutor. Todas as outras etapas foram realizadas corretamente em ambos os questionários.

Nenhum questionário testou por completo todas as propriedades de medida, dois dos sete não foram testados para nenhuma propriedade de medida (PSS e ASES), e dois (WORC e DASH) foram testados para praticamente todas as propriedades de medida, no entanto nem todas foram testadas adequadamente. A confiabilidade foi a propriedade de medida mais testada, sendo que todos os instrumentos que a testaram^{16,17,23,25} aplicaram os testes de maneira correta, com tamanho amostral adequado, e utilizaram os intervalos de medidas adequados¹⁴, porém nenhum dos estudos mencionou qual o tipo do CCI utilizado para mensurar a confiabilidade. Somente os estudos que testaram a confiabilidade do SPADI¹⁷ e do ASORS²⁵ descreveram seus intervalos de confiança (IC), ambos a 95%. É extremamente importante informar o tipo de CCI utilizado nos diferentes testes, uma vez que CCIs diferentes podem apresentar resultados completamente distintos, o que pode subestimar (ou superestimar) a confiabilidade, dependendo do tipo de CCI utilizado²⁹. A concordância é uma importante propriedade de medida, calculando o grau em que medidas repetidas em indivíduos estáveis fornecem respostas semelhantes¹⁴. Infelizmente, nesta revisão, encontrou-se somente um estudo²³ que testou a concordância. A concordância é mais fácil de ser clinicamente interpretada do que a confiabilidade devido ao fato de ela ser expressa nas unidades de medida do instrumento, já a confiabilidade apresenta seus índices (CCI ou Kappa) expressos em uma escala que varia de 0 a 1, os quais, em muitos casos, são de interpretação mais abstrata. Para se testar a reprodutibilidade por completo, o ideal seria testar tanto a confiabilidade (erro relativo da medida) quanto a concordância (erro absoluto da medida), porém a maioria dos estudos testou somente a confiabilidade.

A validade do construto foi testada em três instrumentos (DASH^{16,22}, WORC²³ e ASORS²⁵). Todos os estudos utilizaram os testes de Correlação de Pearson, correlacionando um questionário específico com outras medidas similares, porém é importante que sejam formuladas hipóteses antes de serem realizados os testes de validade de construto, e essas hipóteses devem especificar tanto a magnitude quanto a direção da correlação esperada¹⁴. Não foram formuladas hipóteses em nenhum dos estudos avaliados que testaram a validade do construto^{16,22,23,25}, o que é necessário, pois, sem as hipóteses específicas, os riscos de viés podem ser altos, já que seria mais fácil desenvolver uma explicação alternativa para baixas correlações ao invés de concluir que o questionário pode não ter altos índices de validade do construto¹⁴.

A consistência interna foi avaliada para os instrumentos DASH²⁶, WORC²³ e SPADI¹⁷. Somente o estudo que avaliou a consistência interna do DASH²⁶ utilizou o teste Alfa de Cronbach em combinação com a análise fatorial, identificando três fatores diferentes. Essa análise é importante, pois tal procedimento pode identificar quantas escalas estão presentes em um questionário e, se existir mais de uma escala, o Alfa de Cronbach deverá ser calculado para cada subescala separadamente¹⁴. Os estudos que avaliaram o WORC²³ e o SPADI¹⁷ utilizaram o Alfa de Cronbach, mas não utilizaram a análise fatorial.

A responsividade foi avaliada somente em um estudo que testava os instrumentos DASH, WORC e UCLA²⁴, com um intervalo de medidas adequado (três meses), porém com um tamanho amostral reduzido (30 pacientes). A responsividade representa a habilidade de um questionário em detectar mudanças clínicas no tempo, podendo ser medida pela responsividade interna (tamanho do efeito) ou externa (teste de Correlação de Pearson e/ou pela construção de curvas ROC)³⁰. Nenhum estudo desta revisão sistemática testou a responsividade por completo. Além disso, nenhum estudo testou os efeitos de teto e piso e, conseqüentemente, não se sabe se os instrumentos avaliados podem falhar em detectar melhora ou piora em alguns pacientes^{14,30}.

Outros estudos relacionados à revisão sistemática de instrumentos de medida e análise de suas propriedades de medida confirmam os nossos achados de que há uma clara necessidade de pesquisas futuras que avaliem, por completo, todas as suas propriedades de medida, para que se possa escolher o melhor questionário para cada situação^{31,32}. Em uma avaliação das propriedades de medida de questionários que avaliam disfunções do ombro na língua inglesa, feita por meio de uma revisão sistemática⁹, foram encontrados diferentes métodos de avaliação das propriedades de medida, além de falhas nas avaliações da validade do construto, consistência interna e responsividade. Na maioria dos instrumentos avaliados, não foram formuladas hipóteses relacionadas à magnitude e à direção das correlações esperadas com os outros instrumentos, não foram utilizadas as análises fatoriais, mas, quando utilizadas, algumas das análises não confirmaram as dimensões que o questionário se propunha a medir. A responsividade foi testada, na maioria das vezes, em uma amostra de tamanho inadequado ($n < 43$), e a maioria dos estudos não descreveu adequadamente o método do estudo e/ou a análise dos dados.

Os mesmos problemas foram encontrados em uma revisão sistemática sobre adaptações transculturais e testes das propriedades de medida de um instrumento que avalia a qualidade e a intensidade da dor (*McGill Pain Questionnaire*)³¹, sendo que, mesmo com 44 diferentes versões do questionário, representando 26 diferentes línguas/culturas, na maioria das vezes, os testes não foram realizados ou foram realizados de forma

inadequada, resultados esses compatíveis também com os encontrados na revisão sistemática de instrumentos de medida para dor lombar³², em que a maioria dos questionários avaliou a confiabilidade e a validade do construto dos instrumentos, porém a consistência interna, a responsividade e os efeitos de teto e piso não foram testados para a maioria deles.

É importante ressaltar que existem outras diretrizes que podem ser utilizadas como guias metodológicos para avaliação dos procedimentos de tradução/adaptação e dos testes das propriedades de medida, as quais discordam da necessidade de todos os passos seguidos nas diretrizes utilizadas neste estudo, porém optou-se por seguir tais critérios de qualidade por serem os mais recentes, atualizados e amplamente aceitos na literatura. Todos os estudos de tradução e adaptação analisados nesta revisão foram realizados após a publicação das Diretrizes de Adaptação Transcultural de Questionários¹³, e apenas um estudo, que testou as propriedades de medida (confiabilidade e validade do construto) do DASH¹⁶, foi realizado antes da publicação dos Critérios de Qualidade para Propriedades de Medida de Questionários da Área de Saúde¹⁴.

Utilizaram-se todos os recursos durante a busca sistematizada para que fossem identificados todos os estudos sobre instrumentos de avaliação das disfunções do ombro em português-brasileiro. Apesar de as buscas terem sido realizadas nas bases de dados mais utilizadas (inclusive bases de dados locais, como LILACS e SCIELO), alguns estudos podem não ter sido capturados, uma vez que algumas revistas brasileiras e portuguesas podem não estar indexadas em nenhuma das bases utilizadas, o que se pode interpretar como uma possível limitação desta revisão. Além disso, pode haver dados em relação às propriedades de medida dos instrumentos identificados que não tenham sido considerados nos métodos e, portanto, não tenham sido publicados, porém podem estar presentes nas versões originais de dissertações e teses. Outra possível limitação é o fato de não se ter realizado a avaliação dos dados referentes à tradução/adaptação e aos testes das propriedades

de medida por dois avaliadores, impossibilitando a mensuração da concordância entre eles, porém optou-se por realizar a avaliação por um avaliador e checar os dados por meio de um revisor, considerando improvável que a acurácia das informações fosse diferente.

Baseados nos dados desta revisão, pode-se dizer que o DASH^{16,22,24,26} e o WORC^{18,22-24} foram os instrumentos mais adequados para ser utilizados em relação aos outros questionários testados, pois foram os que tiveram mais propriedades de medida apropriadamente testadas em relação aos outros. Cabe ressaltar que o DASH¹⁶, o SPADI¹⁷, o UCLA²¹, o PSS²⁰ e o ASES¹⁹ são instrumentos desenvolvidos para avaliar qualquer disfunção do ombro; o WORC¹⁸ avalia somente indivíduos com disfunções no manguito rotador, e o ASORS²⁵ avalia o ombro do atleta. Portanto, ao fim desta revisão, pode-se dizer que o WORC¹⁸ seria o instrumento mais adequado para avaliar pacientes com disfunções no manguito rotador e o DASH¹⁶ seria o instrumento mais adequado para avaliar qualquer outra disfunção da articulação do ombro.

Conclusão

A relevância deste estudo reside na observação da importância de seguir diretrizes adequadas na realização tanto da tradução/adaptação quanto dos testes das propriedades de medida, além de mostrar que, sem a realização de testes adequados que possam determinar um critério de qualidade para os instrumentos de medida, não é possível escolher o melhor instrumento para se utilizar na prática clínica e/ou pesquisa científica. Sem a adequada realização dos testes das propriedades de medidas, os fisioterapeutas, médicos e profissionais da área da saúde devem ter cuidado ao interpretar os escores dos questionários adaptados que não testaram essas propriedades de medida. São necessários estudos futuros que realizem os testes de todas as propriedades de medida.

Referências

- Romeo AA, Bach BR Jr, O'Halloran KL. Scoring systems for shoulder conditions. *Am J Sports Med.* 1996;24(4):472-6.
- Green S, Buchbinder R, Hetrick S. Physiotherapy interventions for shoulder pain. *Cochrane Database Syst Rev.* 2003;(2):CD004258.
- Andersen JH, Haahr JP, Frost P. Risk factors for more severe regional musculoskeletal symptoms: a two-year prospective study of a general working population. *Arthritis Rheum.* 2007;56(4):1355-64.
- Eltayeb S, Staal JB, Kennes J, Lamberts PH, de Bie RA. Prevalence of complaints of arm, neck and shoulder among computer office workers and psychometric evaluation of a risk factor questionnaire. *BMC Musculoskelet Disord.* 2007;8:68.
- Kirkley A, Griffin S. Development of disease-specific quality of life measurement tools. *Arthroscopy.* 2003;19(10):1121-8.
- Higginson IJ, Carr AJ. Measuring quality of life: Using quality of life measures in the clinical setting. *BMJ.* 2001;322(7297):1297-300.
- Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ.* 2002;324(7351):1417.
- Fitzpatrick R, Fletcher A, Gore B, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care. I. Applications and issues in assessment. *BMJ.* 1992;305(6861):1074-7.
- Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis.* 2004;63(4):335-41.
- Velloso FSB, Barra AA, Dias RC. Functional performance of upper limb and quality of life after sentinel lymph node biopsy of breast cancer. *Rev Bras Fisioter.* 2011;15(2):146-53.

11. Assumpção A, Pagano T, Matsutani LA, Ferreira EA, Pereira CA, Marques AP. Quality of life and discriminating power of two questionnaires in fibromyalgia patients: Fibromyalgia Impact Questionnaire and Medical Outcomes Study 36-Item Short-Form Health Survey. *Rev Bras Fisioter.* 2010;14(4):284-9.
12. Croft P. Measuring up to shoulder pain. *Ann Rheum Dis.* 1998;57(2):65-6.
13. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25(24):3186-91.
14. 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.* 2007;60(1):34-42.
15. Maher CG, Latimer J, Costa LOP. The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments. *Rev Bras Fisioter.* 2007;11(4):245-52.
16. Orfale AG, Araújo PMP, Ferraz MB, Natour J. Translation into Brazilian Portuguese, cultural adaptation and evaluation of the reliability of the Disabilities of the Arm, Shoulder and Hand Questionnaire. *Braz J Med Biol Res.* 2005;38(2):293-302.
17. Martins J, Napolis BV, Hoffman CB, Oliveira AS. Versão brasileira do Shoulder Pain and Disability Index: tradução, adaptação cultural e confiabilidade. *Rev Bras Fisioter.* 2010;14(6):527-36.
18. Lopes AD, Stadniky SP, Masiero D, Carrera EF, Ciconelli RM, Griffin S. Tradução e adaptação cultural do WORC: um questionário de qualidade de vida para alterações do manguito rotador. *Rev Bras Fisioter.* 2006;10(3):309-15.
19. Knaut LA, Moser AD, Melo Sde A, Richards RR. Translation and cultural adaptation to the portuguese language of the American Shoulder and Elbow Surgeons Standardized Shoulder assessment form (ASES) for evaluation of shoulder function. *Rev Bras Reumatol.* 2010;50(2):176-89.
20. Napolis BV, Hoffman CB, Martins J, Oliveira AS. Tradução e adaptação cultural do Penn Shoulder Score para a Língua Portuguesa: PSS-Brasil. *Rev Bras Reumatol.* 2010;50(4):389-97.
21. Oku EC, Andrade AP, Stadniky SP, Carrera EF, Tellini GG. Tradução e adaptação cultural do Modified-University of California at Los Angeles Shoulder Rating Scale para a língua portuguesa. *Rev Bras Reumatol.* 2006;46(4):246-52.
22. Lopes AD, Furtado RV, Silva CA, Yi LC, Malfatti CA, Araújo SA. Comparison of self-report and interview administration methods based on the Brazilian versions of the Western Ontario Rotator Cuff Index and Disabilities of the Arm, Shoulder and Hand Questionnaire in patients with rotator cuff disorders. *Clinics (São Paulo).* 2009;64(2):121-5.
23. Lopes AD, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, Dos Reis FB. Validity and reliability of the Western Ontario Rotator Cuff Index (WORC) for use in Brazil. *Clin J Sport Med.* 2008;18(3):266-72.
24. Lopes DA, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, Baldy dos Reis F. Comparison of the responsiveness of the Brazilian version of the Western Ontario Rotator Cuff Index (WORC) with DASH, UCLA and SF-36 in patients with rotator cuff disorders. *Clin Exp Rheumatol.* 2009;27(5):758-64.
25. Leme L, Saccol M, Barbosa G, Ejnisman B, Faloppa F, Cohen M. Validação, reprodutibilidade, tradução e adaptação cultural da escala Athletic Shoulder Outcome Rating Scale para a língua portuguesa. *RBM Rev Bras Med.* 2010;67(Suppl 3):29-38.
26. Cheng HM, Sampaio RF, Mancini MC, Fonseca ST, Cotta RM. Disabilities of the arm, shoulder and hand (DASH): factor analysis of the version adapted to Portuguese/Brazil. *Disabil Rehabil.* 2008;30(25):1901-9.
27. Acquadro C, Conway K, Girouard C, Mear I. Linguistic Validation Manual for Patient Reported Outcomes (PRO) Instruments, 2004. Available from: http://www.mapi-research.fr/i_02_manu.htm
28. 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):1417-32.
29. Krebs DE. Declare your ICC type. *Phys Ther.* 1986;66(9):1431.
30. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol.* 2000;53(5):459-68.
31. Costa LCM, Maher CG, McAuley JH, Costa LO. Systematic review of cross-cultural adaptations of McGill Pain Questionnaire reveals a paucity of clinimetric testing. *J Clin Epidemiol.* 2009;62(9):934-43.
32. Costa LO, Maher CG, Latimer J. Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. *Spine (Phila Pa 1976).* 2007;32(9):1028-37.