ISSN 1413-3555
Rev Bras Fisioter, São Carlos
©Revista Brasileira de Fisioterapia

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

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

Vanessa O. O. Puga[1], Alexandre D. Lopes[1], Leonardo O. P. Costa[1,2]

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.

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.

Received: 06/12/2011 – **Revised:** 09/10/2011 – **Accepted:** 09/26/2011

[1] Masters Program in Physical Therapy, Universidade Cidade de São Paulo (UNICID), São Paulo, SP, Brazil
[2] Musculoskeletal Division, The George Institute for Global Health, Sydney, NSW, Australia
Correspondence to: Leonardo Oliveira Pena Costa, Rua Cesário Galeno, 48/475, Tatuapé, CEP 03071-000, São Paulo, SP, Brasil, e-mail: lcosta@edu.unicid.br

Introduction

Shoulder pain is considered the third most common musculoskeletal condition. The prevalence of shoulder disorders ranges from 7 to 36% in the general population[1,2]. Shoulder joint dysfunctions are often observed in workers and athletes who are exposed to repetitive movements and excessive strain, but some studies also report a high prevalence of this problem in elderly and sedentary people[2-4].

Assessment methods for musculoskeletal injuries have been modified in recent years[5]. Instead of a purely physical examination, including tests of muscular strength, joint mobility and imaging exams, self-report questionnaires or scales were incorporated[6,7]. Questionnaires are widely used to collect information related to important clinical outcomes, such as pain intensity, quality of life, satisfaction with treatment and disability[1,8,9]. In addition to the use of questionnaires in clinical practice, they are also widely used in research[10,11].

There are several instruments for assessing patients with shoulder dysfunctions and detecting changes in their clinical condition over time, but most of them were developed in English[9,12]. In order to be used in Brazil, they must be translated, cross-culturally adapted and tested for their measurement properties such as the internal consistency, reproducibility, validity and responsiveness[13-15].

The translation and cross-cultural adaptation process is necessary to verify the equivalence with the original version and to resolve cultural, linguistic and health perception differences between different countries and cultures[15]. Even if an instrument's measurement properties have already been tested in its original version, it is necessary to retest them after translation, since cultural differences may affect the results. Testing is also necessary to determine if the adapted questionnaire retains the measurement properties of the original version[14]. Guidelines for evaluating the appropriateness translation and cross-cultural adaptation[13] and clinimetric testing[13-15] have been developed in order to help researchers to conduct optimal studies on these topics.

Some questionnaires specifically designed to assess shoulder dysfunctions have already been translated and cross-culturally adapted into Portuguese and have had some of their measurement properties tested; however, a synthesis of the information contained in these questionnaires is still not available in the literature. Therefore the objectives of this study were to describe and to assess the translation and to cross-cultural adaptation procedures of these instruments and to describe and to assess the measurement properties tested in each of the studies.

Methods

Study selection

In order to identify the instruments available in Portuguese for assessing shoulder pain and disability, systematic searches were performed in five electronic databases (MEDLINE through OVID, EMBASE, CINAHL through EBSCO, SCIELO and LILACS). The search terms and Boolean operators (*AND, OR* or *NOT*) used in the databases MEDLINE, EMBASE, CINAHL and SCIELO were: *(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 self report OR inventory) AND (Brazil OR Brasil OR Portuguese OR Brazilian Portuguese OR Brazilian)*. The search terms (in Portuguese) used in the LILACS database were: *(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)*. The searches were not limited by language or publication date. The final search was performed in April 2011.

Inclusion criteria

Only instruments developed for the assessment of shoulder joint dysfunctions that had been translated and/or cross-culturally adapted into Portuguese were included in the study, regardless of whether they were combined with the assessment of other upper limb joints. Self-report instruments and evaluator-dependent instruments using only objective measurements, such as strength or range of motion, were also included in this review. Only full-text papers were included; theses/dissertations, abstracts from conferences and books were excluded from this systematic review.

Data extraction and assessment of methodological quality of eligible studies

Data regarding the translation and cross-cultural adaptation were extracted in order to assess the design of these procedures. We also extracted data relating to the measurement properties of each study, including the sample size, reproducibility (reliability and agreement), responsiveness (and time interval between measures), internal consistency and construct validity (raw data can be requested from the corresponding author via email).

Subsequently, the translation and cross-cultural adaptation methods of each study were classified according to the

Guidelines for the Process of Cross-cultural Adaptation of Self-report Measures[13]. The translation and adaptation process includes a initial translation, a synthesis of translations, back-translation, reviews by an Expert Committee and the pre-test version of the instrument. The quality of each step was classified as either positive (+) when the procedure was performed in accordance with the quality criteria, doubtful (?) when the description of method was unclear, negative (-) when the procedure was performed correctly but with an insufficient quantity of translators and/or back-translators or, finally, zero (0) when there was not enough information to evaluate each step (Table 1).

The measurement properties were classified according to the *Quality Criteria for Measurement Properties of Health Status Questionnaires*[14] and the evaluation was restricted only to those items relevant to the instruments evaluated. Other original items of the Quality Criteria such as content (or face) validity and interpretability are relevant for the development of the original questionnaire. Similarly, the item criterion validity should be considered only when there a gold standard available for comparison, which is not the case for shoulder assessment instruments, so these three items were not included in our review.

The items assessed in this review were construct validity, internal consistency, reproducibility (agreement and reliability), responsiveness and ceiling and floor effects. Similarly to the cross-cultural adaptation criteria, the quality of each of the measurement properties was also classified as positive (+) when the procedures for each stage were performed in accordance with the criteria, doubtful (?) when the methods or the design of the study were unclear, negative (-) when the data for each clinimetric property included greater or smaller values than those defined by the criteria in spite of appropriate approach or methods, or zero (0) when there was no information to qualify each measurement property (Table 2). Data extraction and the assessments were carried out by one rater and then checked by an independent reviewer, who reviewed all data. There was no disagreements between the rater and the independent reviewer who met and discussed the data.

Results

A total of 876 studies were retrieved from the searches but only 11 were considered eligible for data analysis (Figure 1). From these, seven different instruments that had been translated and cross-culturally adapted into Portuguese were identified: 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 Shoulder and Elbow Surgeons Questionnaire*)[19], PSS (*Penn Shoulder Score*)[20], ASORS (*Athletic Shoulder Outcome Rating Scale*)[20] and *Modified*-UCLA (*Modified-University of California at Los Angeles Shoulder Rating Scale*)[21]. Of the seven instruments translated and cross-culturally adapted into Portuguese, five had been tested for at least one measurement property: (DASH[16,22], WORC[18,22-24], SPADI[17],

Table 1. Guidelines for the process of cross-cultural adaptation of self-report measures[13] (adapted from Costa et al.[31]).

Steps	Description	Rating Scheme
Translation	Two (or more) translators should independently translate the original questionnaire. The translators should preferably be native speakers to the target language.	+ Translation performed by at least two independent translators ? Doubtful translation procedure - Translation performed by only one translator 0 No information about translation
Synthesis	The translators should synthesize the multiple translations to produce a consensus of the translations.	+ Performed synthesis ? Doubtful design 0 No information about synthesis OR translation performed by only one translator
Back translation	Translators, blinded to the original questionnaire, should translate the consensus translation back into the original language.	+ Back translation performed by at least two independent translators ? Doubtful back translation procedure - Back translation performed by only one translator 0 No information about back translation
Expert committee	The expert committee should consolidate all the versions of the questionnaire and develop what would be considered the prefinal version of the questionnaire for testing.	+ Clearly reported the existence of an expert committee ? Doubtful design 0 No information about expert committee
Pretesting	The prefinal questionnaire undergoes pilot testing with members of the target population.	+ Performed pretesting ? Doubtful design 0 No information about pretesting

+=positive rating; -=negative rating; 0=no information available; ?=unclear.

Table 2. Quality criteria for measurement properties of health status questionnaires[14] (adapted from Costa et al.[31]).

Property	Definition	Quality Criteria
Internal consistency	Internal consistency is a measure of the homogeneity of a (sub) scale. It indicates the extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct. Factor analysis should be applied to determine the dimensionality of the item-this is, to determine whether or not they formed only one overall dimension or more than one.	+ Factor analyses performed on adequate sample size (7 x # items and ≥100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95; ? No factor analysis OR doubtful design or method; - Cronbach's alpha(s) <0.70 or >0.95, despite adequate design and method; 0 No information found on internal consistency.
Construct validity	Content validity examines the extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured.	+ Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses; ? Doubtful design or method (e.g., no hypotheses); - Less than 75% of hypotheses were confirmed, despite adequate design and methods; 0 No information found on construct validity.
Reproducibility	The degree to which repeated measurements in stable persons (test retest) provide similar answers.	
Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	+ ICC or Kappa ≥0.70; ? Doubtful design or method (e.g., time interval not mentioned); - ICC or Kappa <0.70, despite adequate design and method; 0 No information found on reliability
Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	+ MIC <SDC or MIC outside the LOA or convincing arguments that agreement is acceptable; ? Doubtful design or method or (MIC not defined AND no convincing arguments that agreement is acceptable); - MIC ≥SDC or MIC equals or inside LOA, despite adequate design and method; 0 No information found on agreement.
Responsiveness	The ability of a questionnaire to detect clinically important change over time in the concept being measured. A predefine hypotheses about the relation of change in the instrument to corresponding changes in reference measures should be postulated.	+ Smallest detectable change individual or Smallest detectable change group < Minimal important change OR Minimal important change outside the limits of agreement OR Responsiveness ratio >1.96 OR Area under the curve≥0.70; ? Doubtful design or method OR sample size < 50 OR methodological flaws; - Smallest detectable change individual or Smallest detectable change group ≥ Minimal important change OR Minimal important change equals or inside limits of agreement OR Responsiveness ratio ≤1.96 OR Area under the curve <0.70, despite adequate design and methods; 0 No information found on responsiveness.
Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	+ ≤15% of the respondents achieved the highest or lowest possible scores; ? Doubtful design or method OR sample size <50 OR methodological flaws; - >15% of the respondents achieved the highest or lowest possible scores, despite adequate design and methods; 0 No information found on interpretation.

+=positive rating; ?=doubtful design or method; -=negative rating; 0=no information available. Doubtful design or method=lacking of a clear description of the design or methods of the study, sample smaller than 50 subjects, or any important methodological weakness in the design or execution of the study. MIC=minimal important changes; SDC=smallest detectable change; LOA=limits of agreement; ICC=interclass correlation coefficient; SD=standard deviation.

ASORS[25] and *Modified*-UCLA[24]). Finally, the PSS[20] and the ASES[19] had not been tested for any measurement property.

Table 3 presents the ratings of the translations and cross-cultural adaptations according to the *Guidelines for the Process of Cross-cultural Adaptation of Self-report Measures*[13]. For the seven instruments found in the 11 included studies, the steps regarding translation, synthesis, analysis by a committee of experts and the pre-tests had all been correctly performed. The back-translation step had not been adequately performed for the WORC and ASORS as it was performed by a single

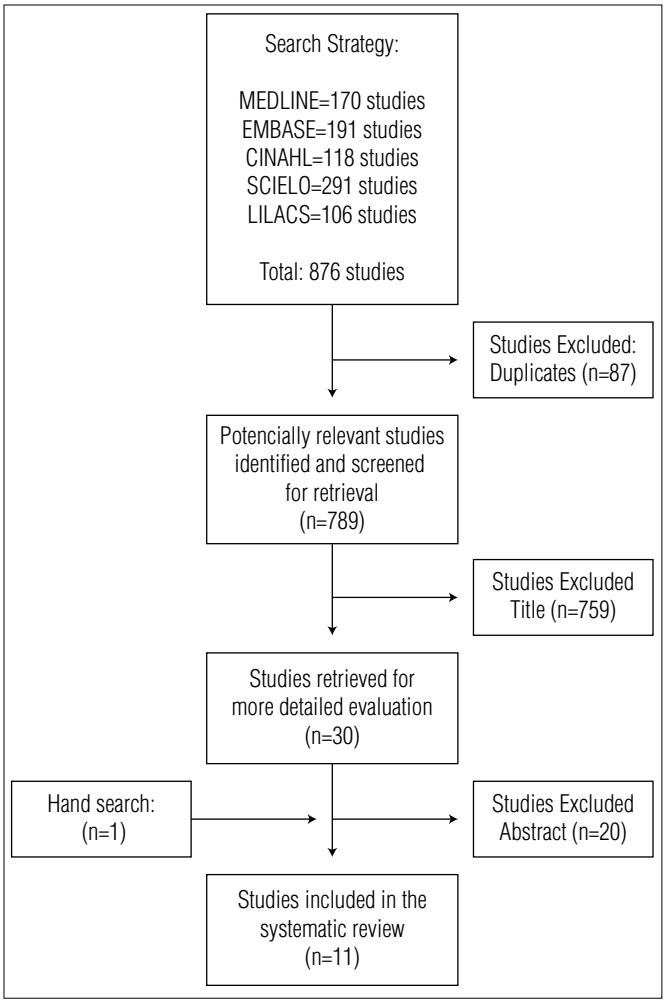

Figure 1. Flow diagram of the literature search.

translator instead of two or more independent translators as recommended by the guidelines.

Table 4 presents the ratings of all evaluated measurement properties according to the *Quality Criteria for Measurement Properties of Health Status Questionnaires*[14]. Reliability was the most frequently tested measurement property, included in four of the five studies that tested measurement properties. These four studies performed the test appropriately and obtained Intraclass Correlation Coefficient (ICC) estimates ≥0.70[16,17,23,25]. On the other hand, agreement was tested only for the WORC in a study that also tested its reliability, and it presented appropriate estimates[23]. Internal consistency was tested for DASH[26], SPADI[17] and WORC[23]; the first two presented acceptable levels (Cronbach's alpha ranging between 0.70 and 0.95), but the WORC's design was questionable, since the internal consistency was tested by intra- and interrater reliability rather than the criteria suggested by the guidelines[14]. Factorial analysis was performed only for the DASH. Construct validity was not properly tested in any of the instruments found in this review (DASH[16,22], WORC[23,24] and ASORS[25]) due to the fact that the hypothesis regarding the correlations of the instruments/scales had not been formulated a priori. Responsiveness was tested in one study that involved DASH, WORC and UCLA[24], but it could not be considered appropriate due to a small sample size (i.e. lower than 50 participants). Ceiling and floor effects were not tested in any of the instruments identified in our systematic review.

Table 3. Cross-cultural adaptations of the shoulder questionnaires adapted into Brazilian-Portuguese that used the translation-based approach related to the Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures.

Studies	Translation	Synthesis	Back translation	Expert committee review	Pretesting
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=Not applicable – The cross cultural adaptations was not performed, only the clinimetric tests. The questionnaires used in these studies have been previously translated in other studies.
* Back translation performed by only one translator.

Table 4. Measurement properties of the shoulder questionnaires adapted into Brazilian-Portuguese related to Quality Criteria for Measurement Properties of Health Status Questionnaires.

Studies	Reproducibility (Agreement)	Reproducibility (Reliability)	Internal Consistency	Responsiveness	Construct Validity	Celing and floor effects	Notes
DASH[16]	0	+	0	0	?	0	Hypotheses were not formulated
UCLA[21]	0	0	0	0	0	0	
WORC[18]	0	0	0	0	0	0	
WORC[23]	+	+	?	0	?	0	* Doubtful design or method. Hypotheses were not formulated
DASH[26]	0	0	+	0	0	0	
WORC e DASH[22]	0	0	0	0	?	0	Hypotheses were not formulated
WORC, DASH e UCLA[22]	0	0	0	?	0	0	Sample size <50
ASES[19]	0	0	0	0	0	0	
PSS[20]	0	0	0	0	0	0	
ASORS[25]	0	+	0	0	?	0	Hypotheses were not formulated
SPADI[17]	0	+	?	0	0	0	No factor analysis were performed

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. *No factor analysis was performed.

Almost all of the measurement properties were tested for the DASH and WORC; only ceiling and floor effects were not tested for the WORC[22-24], whereas both ceiling and floor effects and agreement were not tested for the DASH[20,23,24,26].

Two measurement properties were tested for both SPADI and ASORS: reliability and internal consistency were tested properly for SPADI, and reliability was tested adequately in ASORS[17], but construct validity was inadequately conducted[25]. Only responsiveness was evaluated for UCLA, but it was rated as doubtful[24].

Discussion

The objectives of this study were to describe and assess the translation and cross-cultural adaptation procedures of these instruments and to describe and assess the measurement properties tested in each of the studies. Seven different instruments (from 11 studies) were retrieved and. at least one measurement property was tested in five of them. The translation and adaptation processes were assessed in accordance with the *Guidelines for the Process of Cross-cultural Adaptation of Self-report Measures*[13]; the measurement properties were assessed according to the *Quality Criteria for Measurement Properties of Health Status Questionnaires*[14].

The guidelines for cross-cultural adaptations[13] was followed for DASH[16], SPADI[17], UCLA[21], PSS[20] and ASES[19] with all steps having been adequately performed. The specific protocol for linguistic equivalence suggested by the authors of the original WORC[18] was followed, according to the criteria of the MAPI Research Institute[27] and therefore back-translation was performed by a single translator. The authors of the Portuguese WORC study reported that the guidelines proposed by Guillemim, Bombardier and Beaton[28] and Beaton et al.[13] involved difficult processes and due to the population tested, the complexity of following each of the steps, the long duration to perform all steps and high costs they followed a synthesized protocol, pointing out that all versions of WORC under development in other languages followed the same procedures. The guidelines proposed by Guillemin, Bombardier and Beaton[28] were followed for the ASORS[25], but the guideline-recommended back-translation step was not fully performed because only one translator was involved. All other steps were carried out adequately in both questionnaires.

None of the questionnaires have tested all measurement properties. In fact, no measurement properties were evaluated in two of the seven instruments (PSS and ASES). Testing was conducted for virtually all measurement properties for the WORC and DASH, however these tests did not always occurred in accordance with the guidelines.

Reliability was the most frequently tested property. The tests were applied correctly in all studies[16,17,23,25], including adequate sample size and adequate measurement intervals[14], however none of the studies mentioned which type of ICC had been used while testing reliability. Only the studies that tested the reliability of SPADI[17] and ASORS[25] described their 95%

confidence intervals. It is extremely important to report the type of ICC used in different tests, since different ICCs might lead to completely distinct results, under- or overestimating the estimates of reliability[29].

Agreement is an important measurement property and reflects the degree to which repeated measures applied to stable patients provide similar answers[14]. Unfortunately our review identified only one study[23] that tested agreement. Agreement is more easily interpreted clinically than reliability due to the fact that it is expressed on the instrument's units of measurement. Reliability, on the other hand, presents its indexes (ICC or Kappa) expressed on a scale ranging from 0 to 1, which, in many cases, require a more difficult interpretation. In order to fully test reproducibility it would be ideal to test both the reliability (relative error of measurement) and the agreement (absolute error of the measure), unfortunately most studies tested reliability only.

Construct validity was tested in three studies (DASH[16,22], WORC[23] and ASORS[25]). All of them used Pearson Correlation tests, which involve correlating a questionnaire with other similar instruments. However, hypotheses must be formulated a priori and they must specify both the magnitude and direction of the expected correlation[14]. Such formulated hypotheses were found in none of the included studies that tested construct validity[16,22,23,25]. A specific a priori hypothesis is necessary because it would be easier to develop an alternative explanation for low correlations than admit that the questionnaire's levels of construct validity is compromised[14].

Internal consistency was tested for DASH[26], WORC[23] and SPADI[17]. Only the study that assessed the internal consistency of DASH[26] used Cronbach's alpha in combination with factorial analysis, identifying three different factors. This analysis is important because such a procedure can identify how many factors are present in a questionnaire and if there is more than one factor Cronbach's alpha should be calculated for each factor separately[14]. The studies that evaluated WORC[23] and SPADI[17] used Cronbach's alpha but not factorial analysis.

Responsiveness was evaluated only in one study that tested DASH, WORC and UCLA[24] measurement properties. This study used an appropriate between-measure interval (i.e. three months) but recruited a small sample size (30 patients). Responsiveness represents the ability of a questionnaire to detect clinical changes over time and can be measured by internal (measured by effect sizes) or external responsiveness (measured by correlation tests and/or the construction of receiver operator characteristics curves)[30]. No study in this systematic review fully tested responsiveness. Moreover, no study tested ceiling and floor effects and, consequently, it is not known whether the evaluated instruments would fail to detect patient improvement or deterioration[14,30].

Other systematic reviews of measurement instruments confirm our findings that there is a clear need to complete the evaluation of the instrument's measurement properties so that the best choice of questionnaire for specific situations can be made[31,32]. In a systematic review of English language shoulder disability questionnaires[9], different methods of assessing measurement properties were found, in addition to flaws in the construct validity, internal consistency and responsiveness assessments.

In most evaluated instruments, hypotheses about the expected magnitude and direction of correlations with other instruments had not been formulated. Factorial analyses were not generally used but, even when present the dimensions that the questionnaire intended to measure were lacking in some cases. Responsiveness was frequently tested, although with an inadequate sample size (n<43) and finally most studies did not adequately describe the study methods and/or data analysis.

The same types of problems were also found in a systematic review of cross-cultural adaptations and measurement properties of the McGill Pain Questionnaire, which is an instrument for assessing the quality and intensity of pain[31]. Among the 44 different versions of the questionnaire, representing 26 different languages/cultures, it was frequently observed that the tests had been not performed or were carried out inappropriately. This is also similar with the findings of a systematic review of low back pain measurement instruments[32], in which the majority of questionnaires evaluated the reliability and construct validity of the instruments but not the internal consistency, responsiveness or ceiling and floor effects.

It is important to note that there are other guidelines available for evaluating the translation and cross-cultural adaptation process and measurement properties of instruments. Their steps diverge from those found in the guidelines selected for this study. However, we decided to use these specific criteria due to fact that these guidelines are the most updated and widely accepted in the literature. All of the translation and adaptation studies examined in this review were conducted after publication of the *Guidelines for the Process of Cross-cultural Adaptation of Self-report Measures*[13], and only one study, which tested DASH measurement properties (reliability and construct validity)[16] was developed prior to the publication of the *Quality Criteria for Measurement Properties of Health Status Questionnaires*[14].

We have used all efforts in our searches in order to identify and retrieve all relevant questionnaires that were adapted into Brazilian-Portuguese. Even though the most

commonly used databases (including local databases, such as LILACS and SCIELO) were selected, some studies may have not been detected since some Brazilian and Portuguese journals may not be indexed in these databases. Accordingly, this could be considered a limitation for the present review.

In addition there may be original versions of dissertations and theses with unpublished data regarding measurement properties of the identified instruments. Another possible limitation of our review was that we have not used two raters to extract and classify the data regarding translation and adaptation processes and measurement properties. However, it is unlikely that there were major discrepancies in the accuracy of the information.

Based on the results from this review, we can state that the DASH[16,22,24,26] and the WORC[18,22-24] were the most appropriately developed and tested questionnaires. It should be pointed out that the DASH[16], SPADI[17], UCLA[21], PSS[20] and ASES[19] are tools developed to assess any shoulder dysfunction, whereas the WORC[18] evaluates only people with rotator cuff dysfunctions and the ASORS[25] was developed for athletes. Thus, it can be said that the WORC[18] is the most appropriate instrument for evaluating patients with rotator cuff dysfunctions and the DASH[16] is the most appropriate instrument for evaluating other shoulder joint dysfunctions.

Conclusion

The relevance of this study lies in the importance of following appropriate guidelines for performing both the translation and cross-cultural adaptation of instruments and for testing their measurement properties. Without testing these properties the decision to choose the most relevant instrument for use in clinical practice and/or research becomes difficult. When measurement properties have not been properly tested, physical therapists, physicians and any other healthcare professionals should be careful when interpreting its results. Further studies are needed to fully test of the properties of these instruments.

References

1. Romeo AA, Bach BR Jr, O'Halloran KL. Scoring systems for shoulder conditions. Am J Sports Med. 1996;24(4):472-6.
2. Green S, Buchbinder R, Hetrick S. Physiotherapy interventions for shoulder pain. Cochrane Database Syst Rev. 2003;(2):CD004258.
3. 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.
4. 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.
5. Kirkley A, Griffin S. Development of disease-specific quality of life measurement tools. Arthroscopy. 2003;19(10):1121-8.
6. Higginson IJ, Carr AJ. Measuring quality of life: Using quality of life measures in the clinical setting. BMJ. 2001;322(7297):1297-300.
7. 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.
8. Fitzpatrick R, Fletcher A, Gore B, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care. I. Applications and issues in assesment. BMJ. 1992;305(6861):1074-7.
9. 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.
10. 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, Napoles 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. Traducao e adaptação cultural do WORC: um questionario 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. Napoles BV, Hoffman CB, Martins J, Oliveira AS. Tradução e adaptação cultural do Penn Shoulder Score para a Lingua Portuguesa: PSS-Brasil. Rev Bras Reumatol. 2010;50(4):389-97.
21. Oku EC, Andrade AP, Stadiniky 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 lingua 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, Girourdet 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.