

REVIEW ARTICLE

Systematic review of cross-cultural adaptations of McGill Pain Questionnaire reveals a paucity of clinimetric testing

Luciola da Cunha Menezes Costa*, Christopher G. Maher, James H. McAuley,
Leonardo Oliveira Pena Costa

Musculoskeletal division, The George Institute for International Health, PO Box M201, Missenden Road, Sydney, NSW 2050, Australia

Accepted 24 March 2009

Abstract

Objectives: The objectives of this study were to identify the available cross-cultural adaptations of the McGill Pain Questionnaire (MPQ), to describe the clinimetric testing that has occurred for each adaptation and to evaluate both the quality of the adaptation procedures and the clinimetric testing for each version.

Study Design and Setting: This study is a systematic review. Searches of the MEDLINE, EMBASE, and CINAHL databases were used to identify relevant studies. Data on the quality of the adaptation procedures and clinimetric testing were extracted using current guidelines.

Results: Forty-four different versions of the MPQ were identified representing 26 different languages/cultures. Regardless of the method of cross-cultural adaptation, clinimetric testing of the adapted questionnaires was generally poorly performed and for 18 versions no clinimetric testing has been undertaken.

Conclusions: Although the MPQ has been adapted into a large number of languages, because of inadequate testing most of the adaptations have unknown clinimetric properties. This situation means that users should be cautious when interpreting scores from adapted questionnaires. © 2009 Elsevier Inc. All rights reserved.

Keywords: McGill Pain Questionnaire; Cross-cultural adaptation; Translation; Psychometric properties; Clinimetrics; Pain

1. Introduction

The McGill Pain Questionnaire (MPQ) [1] is the most widely used multidimensional instrument for measuring the quality and intensity of pain [2–5]. In its original form, the MPQ consists of 78 pain descriptors within 20 groups of words divided into four categories (sensory, affective, evaluative, and miscellaneous). Each subclass contains from two to six descriptors that have an assigned value of 1–6 reflecting the level of intensity in that subclass. In addition, the MPQ includes a five-point intensity scale. A short form of the McGill Pain Questionnaire (SF-MPQ) [6] has also been developed. It consists of 15 representative words from the sensory ($n = 11$) and affective ($n = 4$) categories of the standard long form. The five-point intensity scale and a visual analog scale are included to provide indices of overall pain intensity.

The development of the original version of the MPQ, took place in two phases: in the first phase physicians and university graduates were asked to classify 102 words

obtained from the clinical literature into small groups describing different qualities of pain [7]. In the second phase of development, groups of physicians, patients, and students were asked to assign an intensity value for each word within subclasses, using a five-point numerical scale ranging from the least to the worst type of pain. A mean rating and standard deviation of each word was calculated, and the words within subclasses were ranked based on their mean ratings [1].

Because of the relevance of the original MPQ in clinical practice and research, versions of the MPQ have been developed worldwide for different language/cultural settings. The process used to develop these new versions of the MPQ has not been uniform. Some authors have used a cross-cultural adaptation approach [8] (e.g., the Turkish [9] and Korean [10] SF-MPQ), others have constructed a new questionnaire following similar methodological procedures to those used in the development of the original MPQ (e.g., the Finnish Pain Questionnaire [11] and the Italian Pain Questionnaire [12]) and some versions [13–15] have used a “mix” of the two procedures described above. Regardless of the approach used, the main objective of the cross-cultural adaptation is to develop equivalent versions of the MPQ that enable clinicians and researchers to assess

* Correspondent author. Tel.: +61-2-9657-0396; fax: +61-2-9657-0302.

E-mail address: lmenezes@george.org.au (L.daC. Menezes Costa).

What is new?

- Although MPQ has been cross-culturally adapted into 26 different languages, the clinimetric testing of these adapted versions was generally poorly performed.
- Our study shown that without proper clinimetric testing it is impossible to adequately understand the usage of the MPQ in different cultures/settings.
- The use of the most non-English versions of the MPQ should be undertaken with caution and international comparisons from studies in non-English countries remain a problem.
- Future research examining the clinimetric properties of non-English versions of the MPQ is urgently needed.

a patient's pain in their own cultural context as well as to allow comparisons of research from studies conducted in non-English speaking countries. After the revision of any type of outcome measure it is essential to test the clinimetric properties, such as internal consistency, reproducibility, validity, and responsiveness, of the new version. The clinimetric testing provides important information on the value of the new instrument to the potential user.

At present, the number and quality of cross-cultural adaptations of the MPQ are unclear. The aims of this study were to identify the available cross-cultural adaptations of the MPQ, to describe the clinimetric testing that has occurred for each adaptation and to evaluate both the quality of the adaptation procedures and the clinimetric properties for each version.

2. Methods

2.1. Study selection

To identify versions of the MPQ developed for non-English languages, two independent search strategies were carried out on MEDLINE, CINAHL, and EMBASE databases for the period from 1966 to 19/02/2009. The terms for the first search were *McGill Pain Questionnaire (plus its variations, e.g., Melzack Pain Questionnaire) AND validation OR translation OR cross-cultural adaptation OR version* and the terms of the second search were *McGill Pain Questionnaire (plus its variations) AND 50 different languages (e.g., German) or equivalent in native tongue (e.g., Deutsche)*. The results of the two searches were combined in an *Endnote X* software file. Additionally, hand searches of journals, references lists, and textbooks related to pain were performed comprehensively.

2.2. Inclusion criteria

Studies were considered eligible for inclusion if they related to cross-cultural adaptation in a specific language and were published as a full manuscript in a peer-reviewed journal. There were no language restrictions and all non-English papers were translated by accredited professionals or native speakers.

2.3. Data extraction and quality assessment

The data were extracted to describe all cross-cultural adaptation procedures (i.e., how the translation procedures were performed) and all clinimetric properties relevant to cross-cultural adaptation (i.e., reproducibility, internal consistency, responsiveness, construct validity, and ceiling/floor effects) from each adaptation. Additionally, the cross-cultural adaptation procedures were rated using the *Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures* [8]. The process of translating instruments into a new language and culture includes initial translation, synthesis, back translation, expert committee review, and pilot testing of draft translation (Table 1). The final step in the adaptation process is the assessment of the clinimetric properties of the new questionnaire.

The clinimetric properties were rated by the *Quality Criteria for Psychometric Properties of Health Status Questionnaire* [16] with the evaluation restricted to the subset of items relevant to cross-cultural adaptation. The original items *content validity* and *interpretability* are only relevant on the original development of a questionnaire and the item *criterion validity* is only possible to be considered when there is a gold standard available which is not the case for pain; and therefore these three items were considered not applicable for the purpose of this study. The Terwee criteria form a checklist that considers both the methodological quality of the clinimetric testing and the results from the clinimetric tests and so is somewhat different from scales used to measure the methodological quality of clinical trials [17] and therefore the criteria do not provide a single summary score because the various clinimetric properties are distinct. Instead a table is used to comprehensively describe the quality of testing and the clinimetric results (Table 2). This approach has already been used in previous systematic reviews for questionnaires for low back pain [18], vision impairments [19], and shoulder disability [20].

The data extraction and ratings were performed by the first author (L.C.M.C.) and then double-checked by an independent reviewer (L.O.P.C.). Inconsistent ratings were resolved by consensus.

3. Results

From the search strategies, 1,687 potentially relevant studies were found. From these, only 53 studies were considered eligible for data analysis (see Fig. 1).

Table 1
Guidelines for the process of cross-cultural adaptation of self-report measures [8]

Steps	Description	Rating scheme
Translation	Two (or more) translators should independently translate the source 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, blind 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.

From the 53 eligible studies, 44 different versions of the MPQ were identified (there is more than one study reporting the clinimetric testing of the Spanish [14], Japanese [21], Dutch [22], Czech [23], Greek [24], and Norwegian [25] versions). The 44 versions represent 26 different language/cultures as there are multiple versions for the same language (4 Dutch, 4 German (being one in German-Berne [15]), 4 Norwegian, 3 Greek, 2 Brazilian-Portuguese, 2 Danish, 2 Italian, 2 Chinese, 2 Korean, 2 Turkish, 2 Japanese, and 2 Spanish versions (being one developed for Mexican-Americans [26] and another developed in Spain [14] and tested in Argentina, Costa Rica, Mexico, Panama, and Spain [27]).

Table 3 describes the international pain questionnaires that followed the same methods undertaken by Melzack and colleagues to develop the original English language version of the MPQ [1,7], 14 classified as long form and 2 as short form. None of the new long-form versions preserved the number of pain descriptors of the original English version, the number of pain descriptors of these new questionnaires of the long form varied from 42 to 176.

From the two new short-form versions, just one preserved the original number of descriptors [25].

Table 4 describes the procedures used to develop the 29 different versions of the MPQ (15 long form, 14 short form) that used a translation-based approach. From these 29 versions only 11 (4 long form, 7 short form) rated positively in more than 3 out of 5 steps recommended by the guidelines [8]. The process of cross-cultural adaptation includes initial translation, synthesis, back translation, expert committee review, and pilot test. In our review, we found that 53%, 47%, 27%, 27%, and 20% of these steps were followed in the long-form version, respectively, whereas 71%, 64%, 43%, 36%, and 36% of these steps were followed

in the short form. Only the Brazilian-Portuguese [28], the Korean [10], the Turkish [9], and the Farsi [29] versions followed all recommendations. The number of pain descriptors ranged from 29 to 78 words for the long form and ranged from 13 to 20 words for the short form. The original structure of the English version of the MPQ (78 descriptors for the long form and 15 descriptors for the short form) was preserved in 79.3% of the non-English versions developed using the translation-based approach.

We found from the data extraction of the clinimetric testing that there are large differences in sample sizes used in clinimetric testing ranging from 12 to 409 patients. In general, almost all versions were poorly tested and 18/44 versions (all long forms) were not assessed for any of the measurement properties. The clinimetric property most frequently tested was construct validity (10 long-form versions and 11 short-form versions). Construct validity was assessed by comparing the MPQ scores with many different constructs ranging from self-report measures such as pain intensity, depression, anger, and anxiety scales to functional impairment tests (e.g., Karnofsky performance test) and laboratory markers of disease status (e.g., erythrocyte sedimentation rate). Most of the studies did not prespecify the hypothesis for using these different measures and the results were reported in different ways (e.g., expressing results only by *P*-values or only by correlation coefficients) making judgments about construct validity difficult.

Reproducibility was assessed in 11 studies; with intraclass correlation values for reliability of the Pain Rating Index Total score ranging from 0.72 to 0.97. Some studies used different statistical tests for measuring reproducibility such as Spearman's correlation [30], Pearson's *r* [31,32], or Kappa coefficient [33]. Measures of agreement were performed in

Table 2
Quality criteria for measurement properties of health status questionnaires [16]

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 ($7 \times \#$ 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 \geq 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} \geq 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	+ $\leq 15\%$ 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.

only one study [34]. Unfortunately, confidence intervals for reproducibility statistics were infrequently reported.

It is suggested that internal consistency should be tested by Cronbach's alpha and factor analysis [16]. To check the questionnaire for subscales (multidimensionality), factor analysis should be carried out and if a questionnaire

consists of multiple subscales, the Cronbach's alpha should be calculated for each subscale. From 53 studies, only 8 (6 long form and 2 short form) performed both Cronbach's alpha tests and factor analysis. However, most of these studies found Cronbach's alpha values for each subscale lower than 0.70. For the remaining studies, 1 performed the

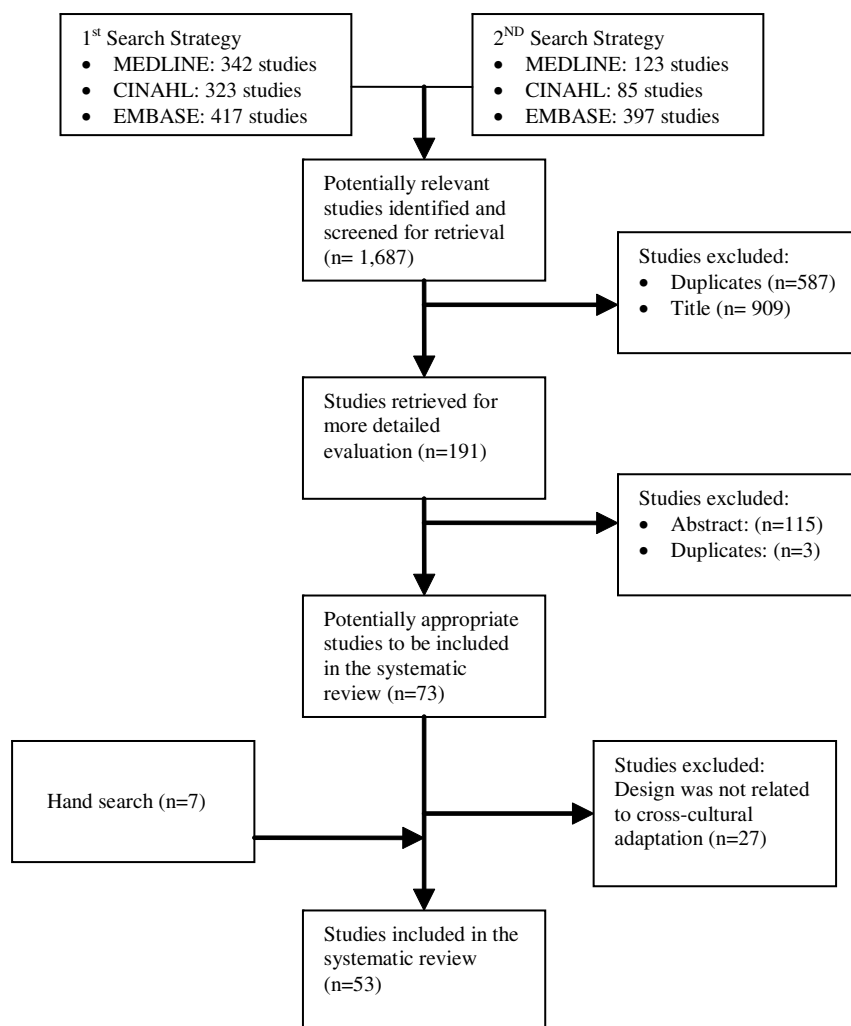


Fig. 1. Quality of reporting of meta-analyses (QUORUM) statement flow diagram of the literature search.

Kuder–Richardson test, 7 performed Cronbach’s alpha only, 2 performed factor analysis only, and 35 studies (25 long form and 10 short form) did not provide any information about internal consistency testing.

Responsiveness was evaluated in 11 studies (4 long form and 7 short form); however, only 2 studies presented hypotheses about the relationship of change in the instrument to corresponding changes in reference measures [10,34]. Nine studies presented only *t* values and *P* values. Ceiling and floor effects were not tested in any study.

The Table 5 shows the quality assessment of the 53 studies of the 44 versions of the MPQ, summarizing each criterion as having a positive, doubtful, negative, or no information. None of the versions demonstrated positive ratings for all properties. Only six versions received positive rating for construct validity (three long form and three short form), three for reproducibility (all short forms), one for internal consistency (German long form [35]), one for responsiveness (Korean short form [10]), and none for ceiling/floor effects. The versions with the most complete clinimetric testing were the long-form German [35] and Japanese [21,36,37] versions

and the short form of the Turkish [9], Swedish [31], and Norwegian [25,34] versions.

4. Discussion

We have presented a systematic review of the international versions of the MPQ. A comprehensive search strategy found a total of 44 versions from 26 different languages. The versions were evaluated on the basis of two existing criteria for quality of cross-cultural adaptation [8] and clinimetric properties [16]. We found different methods of adaptation with few authors following accepted guidelines for cross-cultural adaptation. Clinimetric testing was infrequently performed and when conducted was typically inadequate. Additionally, it was observed that the majority of the poorest adaptations were developed before the publication of the first version of the guidelines for cross-cultural adaptation [38] providing some indirect evidence of the importance of guidelines for helping the development of better self-report instruments.

Table 3
Questionnaires developed using similar methods to the original MPQ [1,7]

Name of the questionnaire	Language	Number of pain descriptors
Long-form MPQ		
Arabic pain words	Arabic [40]	Over 100 ^a
MPQ—Dutch version	Dutch [47]	65
176 Pain descriptive words	Dutch [48]	176
The Finnish Pain Questionnaire	Finnish [11]	56
The Questionnaire Douleur Saint-Antoine	French [13]	61
German language (Berne) Pain Questionnaire	German [15]	51
Italian Pain Questionnaire	Italian [12]	42
Norwegian Pain Questionnaire	Norwegian [45]	106
Not named	Norwegian [41]	43
Polish Pain Language	Polish [46]	58
Spanish version of the MPQ	Spanish [14,27,49,50]	65
Slovak version of the McGill-Melzack's questionnaire of pain	Slovak [39]	60
Urdu pain words	Urdu [44]	106
Welsh medium-chronic pain descriptors	Welsh [43]	137
Short-form MPQ		
Brief, descriptive Danish Pain Questionnaire	Danish [42]	18
Short-form MPQ (NSF-MPQ)	Norwegian [25,34]	15

^a There was no precise number of pain descriptors in the original manuscript.

The procedures used to cross-culturally adapt the MPQ differed greatly between studies, and the reasons for this are unclear. Because pain descriptors may be unique to the country, language, or culture from which they are drawn; some authors [11–15,25,27,34,39–50] argue that the list of pain descriptors should be created by repeating the steps that Melzack and Torgerson [1,7] performed in developing the original MPQ English version. Another argument for this approach is the view that some languages have more or less pain descriptors than the English language. Indeed, our review provides some evidence for this view as the adapted questionnaires vary quite widely in the number of words used to describe pain, ranging from 42 [12] to 176 [48] items. Although the creators of these adaptations consider this method as the most adequate, it does pose problems with scoring and interpretation of research trials that have used these questionnaires as an outcome measure.

On the other hand, some authors [9,10,21–24,26,28,30–33,35–37,51–67] have decided to adapt the original English version using a translation-based approach. In this approach, it is argued that a simple translation of an English language pain descriptor may not convey the same meaning in that language/culture as it does in English. It is therefore necessary that the group of translators/researchers who

perform these procedures should also culturally adapt the descriptors because of the variations and nuances of each language. In some cases, this may require that they exclude, add, or change the original descriptors. A good example is the Brazilian-Portuguese version where the authors found that two descriptors from the original English version have the same meaning in Brazilian-Portuguese (the English words *pulsing* and *beating* are both equivalent to the Portuguese word *pulsátil*). Accordingly, the authors removed one of the synonyms that resulted in the Brazilian-Portuguese version having 77 items, instead of 78 [28].

The procedures for translation and cross-cultural adaptation have already been well described in guidelines for cross-cultural adaptations [8,38,68]. We believe that adhering to these guidelines is likely to maximize linguistic, structure (scoring system), and cultural equivalence in health-related instruments. The development of equivalent versions of the same questionnaire will stop the creation of a wide number of new instruments in the literature and will also facilitate the exchange of information from studies across different countries and cultures.

Regardless of the method of adaptation, the key finding of this study was the lack of clinimetric testing in almost all international versions. After adapting or creating a questionnaire, examining clinimetric properties in a clinically relevant sample is the only way to ensure that the instrument is useful and able to measure the construct that it is supposed to measure. The clinimetric testing of the adapted or new questionnaire varied substantially between studies. Major differences in sample sizes, test–retest periods, and statistical analyses were found. Construct validity was the most frequently tested property. It is important to formulate hypotheses before validity testing and these hypotheses should specify both magnitude and direction of the expected correlation [16,69]; this is necessary because without specific hypothesis the risk of bias is high because it is tempting to develop an alternative explanation for low correlations instead of concluding that the questionnaire may not be valid [16]. Only six studies provided appropriate information about the hypotheses (Table 5).

Many authors presented only Cronbach's alpha values when making the argument that their version has sufficient internal consistency. Internal consistency should be assessed by using Cronbach's alpha in combination with factor analysis. Factor analysis is needed because this procedure explores how many scales (dimensions) are present on the questionnaire (e.g., the English version of the MPQ has four scales: sensory, affective, evaluative, and miscellaneous). If factor analysis (either exploratory or confirmatory) yields multiple scales, Cronbach's alpha should be calculated for each subscale separately [16,19]. As the MPQ is proposed to be a multidimensional measure of pain, this criterion is relevant for all MPQ versions; however, only six studies have adhered to this criterion.

Responsiveness represents the ability of an instrument to detect true change in a patient's status; moreover, the

Table 4

Cross-cultural adaptations of the MPQ that used the translation-based approach relate to the *Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures* [8]

Language	Translation	Synthesis	Back translation	Expert committee review	Pretesting	Number of pain descriptors
Long-form MPQ						
Brazilian-Portuguese [28]	+	+	+	+	+	77
Brazilian-Portuguese [62]	+	+	0	+	+	78
Chinese [55]	?	0	0	0	0	78
Danish [53]	+	+	0	0	0	78
Dutch [22,66]	0	0	0	0	0	63
Dutch [67]	?	0	0	0	0	78
German [57]	+	+	0	0	0	78
German [65]	+	+	0	0	0	78
German [35]	+	0	0	0	0	29
Greek [32]	+	+	+	?	0	78
Italian [61]	–	0	0	0	0	78
Japanese [21,36,37]	?	?	+	0	0	78
Japanese [63]	–	0	0	0	0	35
Norwegian [56]	–	0	–	+	+	78
Spanish [26]	+	+	+	+	0	78
Short-form MPQ						
Amharic [52]	?	?	?	?	0	15
Chinese [54]	?	0	0	0	?	15
Czech [23,58]	+	0	0	0	0	15
Greek [24]	+	+	+	0	+	15
Greek [30]	+	+	+	0	+	15
Hebrew [64]	+	+	0	0	0	15
Hindi [52]	?	?	0	0	0	13
Korean [10]	+	+	+	+	+	15
Korean [60]	+	+	0	+	0	15
Swedish [31]	+	+	+	+	0	15
Thai [33]	+	+	0	0	0	15
Turkish [9]	+	+	+	+	+	15
Turkish [59]	?	?	?	?	0	20

+ = Positive rating; – = negative rating; 0 = no information available; ? = unclear; N/A = not applicable.

questionnaire should be able to distinguish clinically important change from measurement error [16]. Although 11 studies (4 long form and 7 short form) tested responsiveness, most of them (9/11) tested internal responsiveness (e.g., effect sizes and *t*-tests) rather than external responsiveness, which considers an external or independent measure of change. Although the internal responsiveness approach tells us little about the ability to the instrument to detect clinically relevant change, it must be acknowledged that there is an evident lack of clarity about the definition of responsiveness in the literature [69]. No version was tested for ceiling and floor effects. The consequence of ceiling and floor effects is that the instrument can fail to detect deterioration or improvement in certain patients [16].

In our study, 10 non-English papers were found [13,22,23,28,35,39,46,58,62,66] and the data extraction was undertaken after appropriate translation by native speakers of the language or professional translators. It is possible that studies published in local or non-indexed journals could be missed by our search strategy. Additionally some authors changed the name of the questionnaire after performing the adaptation, such as the “Italian Pain Questionnaire” [12] or the “Arabic Pain Words [40].” This potentially

decreases the likelihood that these versions will be retrieved by our systematic search strategies. We strongly recommend that the name of the original questionnaire should be maintained to facilitate its usage by clinicians and researchers.

The lack of clinimetric testing is a critical issue for an important and widely used questionnaire such as the MPQ and clearly more effort in this area is needed. Recently, many published randomized controlled trials from non-English speaking countries [70–73] have reported the MPQ as a primary outcome measure, assuming that the international versions retained similar clinimetric properties to the original English language version. This assumption cannot be made without appropriate testing. Without proper clinimetric testing is impossible to adequately understand the usage of the MPQ in different cultures/settings and therefore international comparisons from studies in non-English countries remain a problem.

Acknowledgments

Luciola Menezes holds a PhD scholarship from the Faculty of Medicine, The University of Sydney, Australia. Christopher Maher holds a research fellowship from the

Table 5

Clinimetric testing of the MPQ versions relate to *Quality Criteria for Psychometric Properties of Health Status Questionnaire* [16]

Language	Reproducibility		Internal consistency	Responsiveness	Construct validity	Ceiling and floor effects
	Reliability	Agreement				
Long-form MPQ						
Arabic [40]	0	0	0	0	0	0
Brazilian-Portuguese [28]	0	0	0	0	0	0
Brazilian-Portuguese [62]	0	0	0	0	0	0
Chinese [55]	0	0	0	0	0	0
Danish [53]	?	0	0	0	0	0
Dutch [47]	0	0	0	?	0	0
Dutch [48]	0	0	0	0	0	0
Dutch [67]	0	0	0	0	0	0
^a Dutch [22]	?	0	?	0	?	0
^a Dutch [66]	?	0	?	0	0	0
Finnish [11]	0	0	0	0	0	0
French [13]	0	0	0	0	0	0
German [57]	0	0	0	0	0	0
German [65]	0	0	0	0	0	0
German (Berne) [15]	0	0	0	0	?	0
German [35]	0	0	+	0	+	0
Greek [32]	?	0	?	?	?	0
Italian [12]	0	0	0	0	0	0
Italian [61]	0	0	0	0	0	0
Japanese [63]	0	0	0	0	0	0
^b Japanese [36]	0	0	–	0	+	0
^b Japanese [37]	0	0	–	0	0	0
^b Japanese [21]	?	0	–	0	+	0
Norwegian [56]	0	0	0	?	?	0
Norwegian [45]	0	0	?	0	0	0
Norwegian [41]	0	0	0	0	0	0
Polish [46]	0	0	0	0	0	0
Slovak [39]	0	0	0	0	0	0
Spanish [26]	?	0	0	0	0	0
^c Spanish [14]	0	0	0	?	?	0
^c Spanish [49]	0	0	–	0	0	0
^c Spanish [50]	0	0	?	0	?	0
^c Spanish [27]	0	0	0	0	?	0
Urdu [44]	0	0	0	0	0	0
Welsh [43]	0	0	0	0	0	0
Short-form MPQ						
Amharic [51]	0	0	0	0	+	0
Chinese [54]	0	0	0	0	?	0
^d Czech [58]	0	0	?	0	–	0
^d Czech [23]	0	0	0	0	–	0
Danish [42]	0	0	0	0	+	0
Farsi [29]	0	0	0	0	?	0
^e Greek [24]	0	0	–	0	0	0
^e Greek [30]	?	0	0	?	0	0
Hebrew [64]	0	0	0	?	–	0
Hindi [52]	0	0	0	?	0	0
Korean [10]	0	0	–	+	?	0
Korean [60]	0	0	0	–	0	0
^f Norwegian [25]	0	0	?	0	?	0
^f Norwegian [34]	+	+	0	?	0	0
Swedish [31]	+	0	?	?	?	0
Thai [33]	?	0	–	0	0	0
Turkish [9]	+	0	–	0	+	0
Turkish [59]	0	0	?	0	0	0

a,b,c,d,e,f Same version.

+ = 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.

National Health and Medical Research Council of Australia; Leonardo Costa is a PhD student supported by CAPES, Brazilian Government, and Pontifícia Universidade Católica de Minas Gerais, Brazil.

The authors would like to thank the contribution from Dr. Ronald Melzack, A/Prof Raymond Ostelo, João Carlos Aydos, Pierre Trudelle, and Thomas Pawlus.

References

- [1] Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975;1(3):277–99.
- [2] Agrawal RP, Choudhary R, Sharma P, Sharma S, Beniwal R, Kaswan K, et al. Glyceryl trinitrate spray in the management of painful diabetic neuropathy: a randomized double blind placebo controlled cross-over study. *Diabetes Res Clin Pract* 2007;77(2):161–7.
- [3] Castel A, Perez M, Sala J, Padrol A, Rull M. Effect of hypnotic suggestion on fibromyalgic pain: comparison between hypnosis and relaxation. *Eur J Pain* 2007;11:463–8.
- [4] Harden RN, Remble TA, Houle TT, Long JF, Markov MS, Gallizzi MA. Prospective, randomized, single-blind, sham treatment-controlled study of the safety and efficacy of an electromagnetic field device for the treatment of chronic low back pain: a pilot study. *Pain Pract* 2007;7(3):248–55.
- [5] Passard A, Attal N, Benadhira R, Brasseur L, Saba G, Sichere P, et al. Effects of unilateral repetitive transcranial magnetic stimulation of the motor cortex on chronic widespread pain in fibromyalgia. *Brain* 2007;130(Pt 10):2661–70.
- [6] Melzack R. The short-form McGill Pain Questionnaire. *Pain* 1987;30(2):191–7.
- [7] Melzack R, Torgerson WS. On the language of pain. *Anesthesiology* 1971;34(1):50–9.
- [8] Beaton DE, Bombardier C, Guillemin F, Ferraz M. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
- [9] Yakut Y, Yakut E, Bayar K, Uygur F. Reliability and validity of the Turkish version short-form McGill Pain Questionnaire in patients with rheumatoid arthritis. *Clin Rheumatol* 2007;26:1083–7.
- [10] Lee H, Nicholson LL, Adams RD, Maher CG, Halaki M, Bae SS. Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. *Spine* 2006;31:1841–5.
- [11] Ketovuori H, Pontinen PJ. A pain vocabulary in Finnish—the Finnish Pain Questionnaire. *Pain* 1981;11:247–53.
- [12] De Benedittis G, Massei R, Nobili R, Pieri A. The Italian Pain Questionnaire. *Pain* 1988;33(1):53–62.
- [13] Boureau F, Luu M, Doubrere JF, Gay C. Construction of a self evaluation pain questionnaire with a description list. Comparison with the Melzack's McGill Pain Questionnaire. *Therapie* 1984;39(2):119–29.
- [14] Lázaro C, Bosch F, Torrubia R, Baños J. The development of a Spanish Questionnaire for assessing pain: preliminary data concerning reliability and validity. *Eur J Psychol Assess* 1994;10(2):145–51.
- [15] Radvila A, Adler RH, Galeazzi RL, Vorkauf H. The development of a German language (Berne) Pain Questionnaire and its application in a situation causing acute pain. *Pain* 1987;28(2):185–95.
- [16] Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, 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.
- [17] Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro Scale for rating quality of randomized controlled trials. *Phys Ther* 2003;83(8):713–21.
- [18] Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain—searching for international cross-cultural adaptations. *Spine* 2007;32:1028–37.
- [19] De Boer MR, Moll AC, De Vet HCW, Terwee CB, Volker-Dieben HJM, van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt* 2004;24:257–73.
- [20] Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker JW, de Vet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis* 2004;63:336–41.
- [21] Hasegawa M, Hattori S, Ishizaki K, Suzuki S, Goto F. The McGill Pain Questionnaire, Japanese version, reconsidered: confirming the reliability and validity. *Pain Res Manag* 1996;1(4):233–7.
- [22] van Lankveld W, van Pad Bosch P, van de Putte L, van der Staak C, Naring G. Pain in rheumatoid arthritis measured with the visual analogue scale and the Dutch version of the McGill Pain Questionnaire. *Ned Tijdschr Geneesk* 1992;136(24):1166–70.
- [23] Solcova I, Jakoubek B, Sykora J, Hnik P. Characterization of vertebrogenic pain using the short form of the McGill Pain Questionnaire. *Cas Lek Cesk* 1990;129(51):1611–4.
- [24] Georgoudis G, Watson PJ, Oldham JA. The development and validation of a Greek version of the short-form McGill Pain Questionnaire. *Eur J Pain* 2000;4(3):275–81.
- [25] Ljunggren AE, Strand LI, Johnsen TB. Development of the Norwegian short-form McGill Pain Questionnaire (NSF-MPQ). *Adv Physiother* 2007;9:169–80.
- [26] Escalante A, Lichtenstein MJ, Rios N, Hazuda HP. Measuring chronic rheumatic pain in Mexican Americans: cross-cultural adaptation of the McGill Pain Questionnaire. *J Clin Epidemiol* 1996;49:1389–99.
- [27] Lazaro C, Caseras X, Whizar-Lugo VM, Wenk R, Baldiaceda F, Bernal R, et al. Psychometric properties of a Spanish version of the McGill Pain Questionnaire in several Spanish-speaking countries. *Clin J Pain* 2001;17(4):365–74.
- [28] Varoli FK, Pedrazzi V. Adapted version of the McGill Pain Questionnaire to Brazilian Portuguese. *Braz Dent J* 2006;17(4):328–35.
- [29] Najafi-Ghezaljah T, Ekman I, Nikravesh MY, Emami A. Adaptation and validation of the Iranian version of Angina Pectoris characteristics questionnaire. *Int J Nurs Pract* 2008;14(6):470–6.
- [30] Georgoudis G, Oldham JA, Watson PJ. Reliability and sensitivity measures of the Greek version of the short form of the McGill Pain Questionnaire. *Eur J Pain* 2001;5(2):109–18.
- [31] Burckhardt CS, Bjelle AA. Swedish version of the short-form McGill Pain Questionnaire. *Scand J Rheumatol* 1994;23(2):77–81.
- [32] Mystakidou K, Parpa E, Tsilika E, Kalaidopoulou O, Georgaki S, Galanos A, et al. Greek McGill Pain Questionnaire: validation and utility in cancer patients. *J Pain Symptom Manage* 2002;24(4):379–87.
- [33] Kitisomprayoonkul W, Klaphajone J, Kovindha A. Thai Short-form McGill Pain Questionnaire. *J Med Assoc Thai* 2006;89:846–53.
- [34] Strand LI, Ljunggren AE, Bogen B, Ask T, Johnsen TB. The short-form McGill Pain Questionnaire as an outcome measure: test-retest reliability and responsiveness to change. *Eur J Pain* 2008;12:917–25.
- [35] Geissner E. Measuring pain via questionnaires: several results concerning the validity of a modified German version of the McGill Pain Questionnaire. *Z Klin Psychol* 1988;17(4):334–40.
- [36] Hasegawa M, Hattori S, Mishima M, Matsumoto I, Kimura T, Baba Y, et al. The McGill Pain Questionnaire, Japanese version, reconsidered: confirming the theoretical structure. *Pain Res Manag* 2001a;6(4):173–80.
- [37] Hasegawa M, Mishima M, Matsumoto I, Sasaki T, Kimura T, Baba Y, et al. Confirming the theoretical structure of the Japanese version of the McGill Pain Questionnaire in chronic pain. *Pain Med* 2001b;2(1):52–9.
- [38] 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;46:1417–32.
- [39] Bartko D, Kondas M, Janco S. Quantification of pain in neurology. The Slovak version of the McGill-Melzack pain questionnaire. *Cesk Neurol Neurochir* 1984;47(2):113–21.

- [40] Harrison A. Arabic pain words. *Pain* 1988;32:239–50.
- [41] Ljunggren AE. Descriptions of pain and other sensory modalities in patients with lumbago-sciatica and herniated intervertebral discs. Interview administration of an adapted McGill Pain Questionnaire. *Pain* 1983;16(3):265–76.
- [42] Perkins FM, Werner MU, Persson F, Holte K, Jensen TS, Kehlet H. Development and validation of a brief, descriptive Danish Pain Questionnaire (BDDPQ). *Acta Anaesthesiol Scand* 2004;48:486–90.
- [43] Roberts G, Kent B, Prys D, Lewis R. Describing chronic pain: towards bilingual practice. *Int J Nurs Stud* 2003;40:889–902.
- [44] Salim M, Chaudhary MA. Urdu pain words. *J Coll Phys Surg Pak* 1999;9(10):463–6.
- [45] Strand LI, Wisnes AR. The development of a Norwegian Pain Questionnaire. *Pain* 1991;46(1):61–6.
- [46] Szczudlik A. Proposal for the creation of polish “pain language”. *Neurol Neurochir Pol* 1983;17:547–52.
- [47] Vanderiet K, Adriaensen H, Carton H, Vertommen H. The McGill Pain Questionnaire constructed for the Dutch language (MPQ-DV). Preliminary data concerning reliability and validity. *Pain* 1987;30:395–408.
- [48] Verkes RJ, Van der Kloot WA, Van der Meij J. The perceived structure of 176 pain descriptive words. *Pain* 1989;38:219–29.
- [49] Masedo AI, Esteve MR. On the affective nature of chronic pain. *Psicothema* 2002;14:511–5.
- [50] Masedo AI, Esteve R. Some empirical evidence regarding the validity of the Spanish version of the McGill Pain Questionnaire (MPQ-SV). *Pain* 2000;85:451–6.
- [51] Aboud FE, Hiwot MG, Arega A, Molla M, Samson S, Seyoum N, et al. The McGill Pain Questionnaire in Amharic: Zwai Health Center patients’ reports on the experience of pain. *Ethiop Med J* 2003;41(1):45–61.
- [52] Ahuja S, Saluja V, Bhattacharya A. A Modified Short form McGill Pain Questionnaire for evaluation of post-operative pain and behavioural response to pain relief. *J Anaesthesiol Clin Pharmacol* 1999;15(2):149–53.
- [53] Drewes AM, Helweg-Larsen S, Petersen P, Brennum J, Andreasen A, Poulsen LH, et al. McGill Pain Questionnaire translated into Danish: experimental and clinical findings. *Clin J Pain* 1993;9(2):80–7.
- [54] Hsieh LL, Kuo CH, Yen MF, Chen TH. A randomized controlled clinical trial for low back pain treated by acupressure and physical therapy. *Prev Med* 2004;39:168–76.
- [55] Hui YL, Chen AC. Analysis of headache in a Chinese patient population. *Ma Zui Xue Za Zhi* 1989;27(1):13–8.
- [56] Kim HS, Schwartz-Barcott D, Holter IM, Lorensen M. Developing a translation of the McGill Pain Questionnaire for cross-cultural comparison: an example from Norway. *J Adv Nurs* 1995;21:421–6.
- [57] Kiss I, Muller H, Abel M. The McGill Pain Questionnaire-German version. A study on cancer pain. *Pain* 1987;29(2):195–207.
- [58] Knotek P, Solcova I, Zalsky M. Czech version of the short form McGill Pain Questionnaire: restandardization. *Bolest* 2002;5(3):169–72.
- [59] Kuguoglu S, Aslan FE, Olgun N. Turkish version of the McGill Melzack Pain Questionnaire Form (MPQF). *Agri-Istanbul* 2003;15(1):47–51.
- [60] Lee MC, Essoka G. Continuing education. Patient’s perception of pain: comparison between Korean-American and Euro-American obstetric patients. *J Cult Divers* 1998;5(1):29–40.
- [61] Maiani G, Sanavio E. Semantics of pain in Italy: the Italian version of the McGill Pain Questionnaire. *Pain* 1985;22(4):399–405.
- [62] Pimenta CA, Teixeira MJ. Proposal to adapt the McGill Pain Questionnaire into Portuguese. *Rev Esc Enferm USP* 1996;30:473–83.
- [63] Satow A, Nakatani K, Taniguchi S. Analysis of perceptual characteristics of pain describing in words caused by occupational cervicobrachial disorder and similar disease. *J Psychol Res* 1988;30(3):132–43.
- [64] Sloman R, Rosen G, Rom M, Shir Y. Nurses’ assessment of pain in surgical patients. *J Adv Nurs* 2005;52(2):125–32.
- [65] Stein C, Mendl G. The German counterpart to McGill Pain Questionnaire. *Pain* 1988;32:251–5.
- [66] van der Kloot WA, Oostendorp RA, van der Meij J, van den Heuvel J. The Dutch version of the McGill Pain Questionnaire: a reliable pain questionnaire. *Ned Tijdschr Geneesk* 1995;139(13):669–73.
- [67] Vermote R, Ketelaer P, Carton H. Pain in multiple sclerosis patients. A prospective study using the McGill Pain Questionnaire. *Clin Neurol Neurosurg* 1986;88(2):87–93.
- [68] van Widenfelt BM, Treffers PD, de Beurs E, Siebelink BM, Koudijs E. Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clin Child Fam Psychol Rev* 2005;8(2):135–47.
- [69] Terwee CB, Dekker JW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003;12:349–62.
- [70] Brockow T, Wagner A, Franke A, Offenbacher M, Resch KL. A randomized controlled trial on the effectiveness of mild water-filtered near infrared whole-body hyperthermia as an adjunct to a standard multimodal rehabilitation in the treatment of fibromyalgia. *Clin J Pain* 2007;23(1):67–75.
- [71] Ciaramella A, Grosso S, Poli P, Gioia A, Inghirami S, Massimetti G, et al. When pain is not fully explained by organic lesion: a psychiatric perspective on chronic pain patients. *Eur J Pain* 2004;8(1):13–22.
- [72] Oksuz E, Mutlu ET, Malhan S. Nonmalignant chronic pain evaluation in the Turkish population as measured by the McGill Pain Questionnaire. *Pain Pract* 2007;7(3):265–73.
- [73] Rocha-Filho PAS, Gherpelli JLD, de Siqueira JTT, Rabello GD. Post-craniotomy headache: characteristics, behaviour and effect on quality of life in patients operated for treatment of supratentorial intracranial aneurysms. *Cephalalgia* 2008;28(1):41–8.