ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 903−912

Journal of Clinical Epidemiology

The Brazilian-Portuguese versions of the McGill Pain Questionnaire were reproducible, valid, and responsive in patients with musculoskeletal pain

Lucíola da C. Menezes Costa[a,*], Christopher G. Maher[a], James H. McAuley[h], Mark J. Hancock[b], Warley de Melo Oliveira[c], Daniel Camara Azevedo[d], Ludmilla Motta Andrade Freitas Pozzi[e], André Roberto Scarpelli Pereira[f], Leonardo Oliveira Pena Costa[a,g]

[a]*The George Institute for Global Health, The University of Sydney, Sydney, Australia*
[b]*Back Pain Research Group, Faculty of Health Sciences, The University of Sydney, Sydney, Australia*
[c]*Faculdade Pitágoras, Belo Horizonte CEP 30140-061, Brazil*
[d]*Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte CEP 30535-901, Brazil*
[e]*Vita Fisio Physiotherapy Home Care*
[f]*Private Physiotherapist*
[g]*Universidade Cidade de São Paulo, CEP 03071-000, São Paulo, Brazil*
[h]*Prince of Wales Medical Research Institute, Randwick 2031, Australia*

Accepted 21 December 2010

Abstract

Objective: To cross-culturally adapt the Short Form of the McGill Pain Questionnaire (SF-MPQ) into Brazilian-Portuguese and test the clinimetric properties of the newly developed SF-MPQ and the previously cross-culturally adapted Brazilian-Portuguese Long Form of the McGill Pain Questionnaire (LF-MPQ).

Study Design and Setting: The SF-MPQ was translated and adapted into Brazilian-Portuguese following recommendations from current guidelines. Both SF-MPQ and LF-MPQ were administered in a prospective longitudinal design to 203 patients with a range of musculoskeletal conditions to evaluate their clinimetric properties.

Results: Both questionnaires demonstrated high levels of internal consistency (Cronbach α range = 0.70−0.79), reliability (intraclass correlation coefficient$_{2,1}$ range = 0.69−0.85), and agreement (standard error of the measurement range = 0.80−6.92). We observed positive and moderate-to-high correlations among the SF-MPQ, the LF-MPQ, and the Numerical Rating Scale (Pearson r ranging from 0.49 to 0.68). No ceiling or floor effects were detected. Both versions demonstrated acceptable levels of responsiveness (effect size range = 0.30−0.60; correlations range = 0.23−0.51; and area under the curve range = 0.56−0.76).

Conclusions: The Brazilian-Portuguese versions of the MPQ were found to be reproducible, valid, and responsive for the assessment of pain in patients with musculoskeletal conditions. © 2011 Elsevier Inc. All rights reserved.

Keywords: McGill Pain Questionnaire; Clinimetric properties; Cross-cultural adaptation; Reproducibility; Validity; Responsiveness

1. Introduction

The McGill Pain Questionnaire (MPQ) was developed by Melzack [1] to assess the sensory, affective, and evaluative dimensions of pain. The MPQ is the most widely used instrument for measuring the quality and intensity of pain, and it has been translated and adapted into 17 different languages [2], including a Long Form Brazilian-Portuguese (Long Form of the MPQ [LF-MPQ]) version [3]. However, the LF-MPQ usually takes 15−20 minutes to complete, making it impractical for use in many clinical settings. Additionally, in clinical trials research where outcomes other than pain are also collected, the MPQ provides excessive detail, which may be unnecessary and time consuming. Thus in 1987, Melzack [4] developed the Short Form of the MPQ (SF-MPQ).

The SF-MPQ was developed for use in specific research settings when time to obtain information from patients is limited and when more information is desired than that provided by simple pain intensity measures, such as the visual analog scale (VAS) or present pain intensity (PPI). The SF-MPQ consists of 15 representative words from the sensory

* Corresponding author. The George Institute for Global Health, The University of Sydney, PO Box M201, Missenden Road, Sydney, NSW 2050, Australia. Tel.: +61-2-9657-0396; fax: +61-2-9657-0302.

E-mail address: lmenezes@georgeinstitute.org.au (L. da C. Menezes Costa).

0895-4356/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi: 10.1016/j.jclinepi.2010.12.009

What is new?

- Although the McGill Pain Questionnaire (MPQ) has been adapted into a large number of languages, the clinimetric testing of the adapted questionnaires was generally poorly performed.
- The Brazilian-Portuguese versions of the MPQ were found to be reproducible and valid for the assessment of patients with musculoskeletal pain.
- These measures will enable international comparisons to be performed and encourage researchers to include Portuguese speakers in their clinical studies.

($n = 11$) and affective ($n = 4$) categories of the standard long form. The 6-point intensity scale and a VAS are included to provide indices of overall pain intensity.

Similar to the long-form version of the MPQ, the SF-MPQ has been translated into 13 languages [2] and used to assess the pain experience of several types of patients. The SF-MPQ has not been translated and adapted into Brazilian-Portuguese, and the clinimetric properties of the LF-MPQ [3] have not been tested. Cross-cultural adaptations of existing English language questionnaires enable clinicians and researchers to assess a patient's pain in their own cultural context as well as allow comparisons of research from studies conducted in non-English speaking countries [5]. However, after the procedure of cross-cultural adaptation, it is crucial to test the clinimetric properties of the adapted questionnaire in the target population to ensure that the new version is reproducible, valid, and responsive.

Portuguese is the sixth most spoken language in the world; however, there are just a few outcome measures relevant to musculoskeletal pain adapted into Portuguese. Brazil is currently in the 15th place in the world ranking of scientific publications [6] with a large potential for performing studies in musculoskeletal pain, and it could be argued that this research gap must be filled.

A recent systematic review of cross-cultural adaptation of the MPQ [2] revealed that although the MPQ has been adapted into a large number of languages, the clinimetric testing of the adapted questionnaires was generally poorly performed. Therefore, the aims of this study were to translate and to cross-culturally adapt the SF-MPQ into Brazilian-Portuguese and to test the clinimetric properties of the newly developed SF-MPQ and the previously developed LF-MPQ [3].

2. Methods

2.1. Overview of study design

The study was carried out in two stages: the first stage was to develop a Brazilian-Portuguese translation and cross-cultural adaptation of the SF-MPQ and the second stage was to test the clinimetric properties of the Brazilian-Portuguese versions of the newly adapted SF-MPQ and the previously adapted LF-MPQ [3].

2.2. Translation and cross-cultural adaptation

The translation and cross-cultural adaptation of the SF-MPQ were performed by following the recommendations from the *Guidelines for the process of cross-cultural adaptation of self-report measures* [7] as follows:

1. Translation: Two independent translators translated the SF-MPQ from English to Brazilian-Portuguese.
2. Synthesis of the translation: The two translators then synthesized the results of the independent translations and prepared the Brazilian-Portuguese version of the SF-MPQ.
3. Back translation: Using the final synthesized version, two new independent translators, who were blind to the original version, translated the questionnaire back to English.
4. An expert committee comprising the translators, health professionals, and experts in the field of musculoskeletal pain reviewed all translations, discussed possible discrepancies, and developed the final version of the SF-MPQ to be tested in Brazil.

The translators were asked to achieve equivalence of the source (English) and target (Portuguese) words and sentences (we cross-culturally adapted not only the items of the SF-MPQ but also response options and instructions of the instrument) taking into consideration four areas: (1) *Semantic equivalence* (i.e., does the words mean the same thing?); (2) *Idiomatic equivalence*. Colloquialisms are difficult to translate. The committee was asked to formulate an equivalent expression in the target version; (3) *Experiential equivalence*. Items are seeking to capture and experience of daily life; however, often in a different country or culture, a given task may simply not be experienced (even if it is translatable). The questionnaire item would have to be replaced by a similar item that is in fact experienced in the target culture; and (4) *Conceptual equivalence*. Often words hold different conceptual meaning between cultures. The committee must examine the source and back-translated questionnaires for all such equivalences.

2.3. Testing the clinimetric properties

This stage was performed on 203 patients with a range of musculoskeletal conditions recruited from different physiotherapy clinics (both public and private) in Belo Horizonte, Brazil. To be eligible, the participants had to be able to speak, read, and write in Brazilian-Portuguese; able to provide written consent to participate in the study; aged between 18 and 80 years; and currently seeking treatment for their musculoskeletal pain.

The study sample size was determined according to the *Quality Criteria for Health Status Questionnaires* [8], which suggests that at least 50 patients are necessary for an appropriate analysis of construct validity, reproducibility, responsiveness, and ceiling/floor effects and a minimum of 100 patients are required to perform internal consistency analysis. To provide more precise estimates, we recruited 203 participants who completed the LF-MPQ, the SF-MPQ, the Global Perceived Effect (GPE) scale, and the Pain Numerical Rating Scale (NRS) at the beginning of treatment (baseline). From those, the first 100 participants completed the same questionnaires 24−48 hours later (to test reproducibility) and after 2 weeks of treatment or discharge, whichever happened first (to test responsiveness). The 24- to 48-hour interval for testing reproducibility was chosen because significant change in pain status was unlikely in such a short period of time. The 2-week interval for testing responsiveness was chosen because some improvement is expected in this period of time. Given that we have collected both the LF-MPQ and the SF-MPQ at the same time, the order of presentation of the questionnaires may influence the pattern of answers and the correlation between them [4]. To avoid this issue, the participants were assigned to complete the questionnaires in a counterbalanced order, that is, 50% of patients answered the LF-MPQ first followed by SF-MPQ and the remaining ones answered the SF-MPQ first followed by the LF-MPQ.

Ethical approval for the study was obtained from the University of Sydney Human Research Ethics Committee and the Pontifícia Universidade Católica de Minas Gerais/ Brazil Ethics Committee, and each participant gave informed consent before testing.

2.4. Description of the measures

2.4.1. Brazilian-Portuguese LF-MPQ

The LF-MPQ [3] contains 78 pain descriptors within 20 groups of words divided into four categories (sensory, affective, evaluative, and miscellaneous). Each group contains from two to five descriptors that have an assigned rank value of 1−5 reflecting the level of intensity in each subclass. The Pain Rating Index (PRI) consists of the sum of the rank values of the words chosen by the patient to obtain a score separately for the sensory, affective, evaluative, and miscellaneous words, in addition to providing a total score. The score system is calculated as follows: (1) The *Pain Rating Index Sensory* (PRI-S) consists of the sum of groups 1−10 with the total score ranging from 0 to 41; (2) The *Pain Rating Index Affective* (PRI-A) consists of the sum of groups 11−15 with the total score ranging from 0 to 14; (3) The *Pain Rating Index Evaluative* consists of the rank value of group 16 with the total score ranging from 0 to 5; (4) The *Pain Rating Index Miscellaneous* (PRI-M) consists of the sum of groups 17−20 with the total score ranging from 0 to 18; and (5) The *Pain Rating Index Total* (PRI-T) consists of the sum of all 20 groups (1−20) with the total score ranging from 0 to 78. Another way of scoring these 20 groups is by summing the Number of Words Chosen (NWC) with the score ranging from 0 to 20. The LF-MPQ also includes the Present Pain Intensity (PPI), which is a 6-point verbal rating scale that ranges from 0 (no pain) to 5 (excruciating pain).

2.4.2. Brazilian-Portuguese SF-MPQ

The SF-MPQ contains 15 descriptors of the pain sensation (11 sensory and 4 affective), with each descriptor ranked on a 4-point rating scale (0 = none, 1 = mild, 2 = moderate, 3 = severe). Three measures of pain experience based on the sensory and affective descriptors can be obtained: (1) The PRI-T consists of the sum of all 15 descriptors with the total score ranging from 0 to 45; (2) The PRI-S consists of the sum of descriptors 1−11 with the total score ranging from 0 to 33; and (3) The PRI-A consists of the sum of descriptors 12−15 with the total score ranging from 0 to 12. Another way to score these 15-item descriptors is by counting the NWC with the score ranging from 0 to 15. The SF-MPQ also includes the VAS and the PPI of the standard LF-MPQ.

2.4.3. The GPE scale

The GPE is a global measure of change [9]. The GPE is an 11-point scale that ranges from −5 (vastly worse) through 0 (no change) to +5 (complete recovery). The GPE was used as an external criterion to measure overall change [10].

2.4.4. The Pain NRS

The NRS is an 11-point scale that ranges from 0 (no pain) to 10 (worst possible pain). The NRS was used as an external criterion for measuring pain intensity. Both the GPE and the NRS have been cross-culturally adapted into Brazilian-Portuguese and clinimetrically tested in a previous study (NRS reliability intraclass correlation coefficient$_{2,1}$ [ICC$_{2,1}$] = 0.94, 95% CI [confidence interval]: 0.90, 0.96; responsiveness effect size [ES] = 1.16, 95% CI: 1.03, 1.28. GPE reliability ICC$_{2,1}$ = 0.90, 95% CI: 0.84, 0.93; responsiveness ES = 0.99, 95% CI: 0.89, 1.09) [10].

2.5. Analysis

To assess the clinimetric properties of the adapted questionnaires (LF-MPQ and SF-MPQ), the following tests were conducted:

2.5.1. Internal consistency

Internal Consistency was evaluated using Cronbach α, which was calculated for each dimension separately. We also calculated the Cronbach α *if an item was deleted*. Cronbach α *if an item was deleted* is an estimate of a new Cronbach α index if one of the items of the questionnaire or its subscales is deleted; this analysis is very useful to identify possible redundancy or heterogeneity of the items of a questionnaire. Cronbach α values are considered adequate if they range between 0.70 and 0.95 [8].

2.5.2. Reproducibility

Reproducibility was tested using a test–retest design, and it was evaluated by using two measures of reproducibility: reliability (relative measurement error) and agreement (absolute measurement error). Reliability was evaluated using the ICC (type 2, 1—agreement) with 95% CIs. The ICC has been interpreted as follows: less than 0.40 poor reproducibility; 0.40–0.75 moderate reproducibility; 0.75–0.90 substantial reproducibility, and greater than 0.90 excellent reproducibility [8]. To evaluate agreement, two measures were used: the standard error of the measurement ($SEM_{agreement}$) and the smallest detectable change (SDC). The SEM was calculated by taking the square root of the error variance (i.e., SEM = $\sqrt{}$error variance). The error variance was obtained using the maximum likelihood method for variance components. The SEM reflects the error of the instrument itself. The SDC was calculated using the formula $SDC = 1.96 \times \sqrt{2} \times SEM$. The SDC reflects the smallest within-person change in a score that, with $P < 0.05$, can be interpreted as a "real" change, above measurement error in one individual. The percentage of the SEM related to the total score of a questionnaire is suggested by Ostelo et al. [11] as an important indicator of agreement, and it should be interpreted as follows: $\leq$5% very good; $>$5% and $\leq$10% good; $>$10% and $\leq$20% doubtful; and $>$20% negative.

2.5.3. Construct validity

Construct validity was evaluated by measuring the correlation between the Brazilian-Portuguese version of the NRS, the Brazilian-Portuguese SF-MPQ, and the LF-MPQ at baseline using Pearson *r*. A score of 0.70 has being recommended for instruments that measure the same construct. When similar constructs are compared, scores lower than 0.70 are acceptable [8].

2.5.4. Floor and ceiling effects

Potential floor and ceiling effects were measured by calculating the percentage of patients indicating the minimum or maximum possible scores in both questionnaires. Floor and ceiling effects are considered to be present if more than 15% of respondents achieved the highest or lowest possible total score (floor and ceiling effects are not related to individual items) [8].

2.5.5. Internal responsiveness

Internal responsiveness was assessed by calculating the ES with 84% CIs. We calculated 84% CIs for direct comparison of the ES. We choose 84% CIs because nonoverlapping 84% CIs are equivalent to a *Z* test of means at the 0.05 level [12,13]. Responsiveness indicates the sensitivity of a questionnaire to measure true change with higher scores preferred [8]. To facilitate decisions regarding the clinical importance of the observed change in the measure, some benchmarks have been proposed. A value of 0.20 or less represents a change of approximately one-fifth of the baseline standard deviation and is considered small. A value of 0.50 is considered moderate, whereas a value of 0.80 or greater is viewed as large [14].

2.5.6. External responsiveness

External responsiveness of the Brazilian-Portuguese LF-MPQ and SF-MPQ was assessed by first correlating GPE rates to change scores of Brazilian-Portuguese LF-MPQ and SF-MPQ and by second constructing receiver operating characteristic (ROC) curves using dichotomized GPE ratings to categorize subjects that *improved* and *did not improve*. The GPE cutoff to categorize improvement was patients who scored 3 points or more on the GPE scale. The analysis is based on the area under the curve (AUC), and values of 0.70 or more are considered responsive [8].

We specified these hypotheses a priori:

1. The Brazilian-Portuguese SF-MPQ and LF-MPQ will have acceptable internal consistency (Cronbach α between 0.70 and 0.95).
2. The Brazilian-Portuguese SF-MPQ and LF-MPQ will correlate positively with the Pain NRS; the magnitude of these correlations is expected to be moderate to high ($r > 0.30$) [15].
3. The Brazilian-Portuguese SF-MPQ will correlate positively with the LF-MPQ; the magnitude of these correlations is expected to be high ($r \geq 0.70$).
4. The Brazilian-Portuguese SF-MPQ and LF-MPQ will demonstrate high test–retest reliability (ICC ≥ 0.70) and agreement (the percentage of the SEM related to the total score of the questionnaires $\leq$10%).
5. The Brazilian-Portuguese SF-MPQ and LF-MPQ will be able to detect changes over the time (i.e., will be responsive) if the changes occur. The magnitude of the ESs and the correlations used to test responsiveness are expected to be moderate (i.e., ES ≥ 0.50 and correlations $\geq$0.30). We also considered that the AUC should achieve at least 0.70 to consider the questionnaire as responsive [14].

3. Results

A total of 203 eligible patients with musculoskeletal pain completed all questionnaires at time 1 (baseline), and from those the first 100 patients answered the same questionnaires at all three time points. Few data (0.05%) on questionnaire items were missing across the three time points. The distribution of the missing data suggests random missing values around the scale items. Therefore, we replaced the missing values with the within-subject median value. Only one patient did not complete the whole SF-MPQ at baseline. Patients presented a wide range of musculoskeletal conditions, with 23% related to back pain, 8% related to neck pain, 38% related to lower limb conditions, 29% related to upper limb conditions, and 3% were patients with fibromyalgia. Table 1 shows the characteristics of the study participants.

The distribution of the total scores of the SF-MPQ and LF-MPQ is described in Fig. 1.

3.1. Cross-cultural adaptation of the SF-MPQ

The final Brazilian-Portuguese version of the SF-MPQ is described in the Appendix. All 15 items were preserved; therefore, the scoring system of the Brazilian-Portuguese remains the same as the original English version. From a total of 15 items of the SF-MPQ, 11 were adapted from all translators with 100% agreement (i.e., no consensus was needed) reflecting exact equivalence between English and Portuguese. On the other hand, disagreements were observed for four items (i.e., gnawing, tender, splitting, and sickening) and a more complex adaptation was necessary; which was resolved by the final meeting among the translators and specialists in the field of musculoskeletal pain.

3.2. Internal consistency

Both LF-MPQ and SF-MPQ demonstrated adequate internal consistency by achieving Cronbach α higher than 0.70 for each subscale with the exception of the LF-MPQ miscellaneous subscale (Cronbach $\alpha = 0.55$) (Table 2). Analyzing alpha "if item deleted" demonstrated that no individual items from the LF-MPQ or SF-MPQ contribute more to the construct than the others (LF-MPQ alpha "if the item deleted" range, 0.83–0.84 and SF-MPQ alpha "if the item deleted" range, 0.83–0.85).

3.3. Reproducibility

3.3.1. Reliability

The reliability coefficients for the LF-MPQ ranged from PPI $ICC_{2,1} = 0.46$, 95% CI: 0.29, 0.60 to PRI-T $ICC_{2,1} = 0.80$, 95% CI: 0.72, 0.86, whereas the SF-MPQ had its reliability coefficients ranging from 0.55 to 0.69 (Table 2).

3.3.2. Agreement

Because each subscale of the questionnaires has different lengths, caution must be taken when interpreting agreement. For example, the SDC for the PRI-T of the LF-MPQ and SF-MPQ are 19.17 and 17.26, respectively. However, the PRI-T score of the LF-MPQ has 79 points, whereas

Table 1
Characteristics of study participants

Patient characteristics	Baseline ($n = 203$)	24–48 hr after baseline ($n = 203$)	2 wk after baseline ($n = 100$)
Age (yr)	42.5 (15.5)		41.7 (15.1)
Male gender (%)	95 (47)		44 (44)
Musculoskeletal pain duration (wk)	20 (44)		20 (42)
Education level, *n* (%)			
Elementary school	63 (31)		35 (35)
School certificate	31 (15)		12 (12)
High school certificate	59 (29)		38 (38)
Currently at university	14 (7)		1 (1)
Bachelor degree	19 (9)		9 (9)
Postgraduate certificate	17 (8)		6 (6)
Working situation, *n* (%)			
Yes	130 (64)		70 (70)
No	73 (36)		30 (30)
Exercising regularly, *n* (%)	72 (36)		63 (63)
Smoker, *n* (%)	15 (7)		12 (12)
Currently taking medication for musculoskeletal pain, *n* (%)	53 (26)		35 (35)
Self-rated health, *n* (%)			
Poor	6 (3)		2 (2)
Fair	36 (18)		21 (21)
Good	87 (43)		42 (42)
Very good	54 (27)		25 (25)
Excellent	20 (10)		10 (10)
LF-MPQ total score (0–78) at baseline	26.5 (12.9)	25.4 (13.5)	19.7 (13.5)
SF-MPQ total score (0–45) at baseline	13.7 (8.8)	12.4 (9.0)	10.1 (9.1)
Pain numerical rating intensity scale (0–10) at baseline	5.4 (2.4)	5.0 (2.7)	3.9 (2.8)
Global perceived effect (−5 to +5) at baseline	1.2 (2.8)	0.8 (2.7)	2.4 (2.4)

Abbreviations: LF-MPQ, Long Form of the McGill Pain Questionnaire; SF-MPQ, Short Form of the McGill Pain Questionnaire.

Continuous data are presented as mean (standard deviation), musculoskeletal pain duration is presented as median (interquartile range), and categorical data are presented as *n* (%).

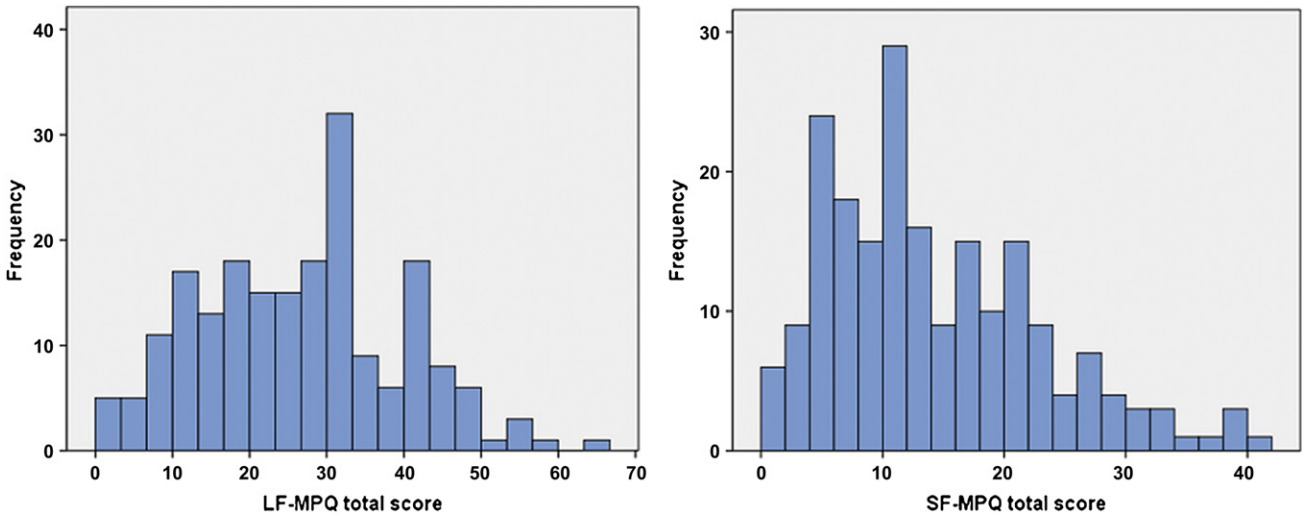

Fig. 1. Distribution of the LF-MPQ and SF-MPQ total scores at baseline. LF-MPQ, Long Form of the McGill Pain Questionnaire; SF-MPQ, Short Form of the McGill Pain Questionnaire.

the SF-MPQ has 46 points; therefore, for measuring "true change" in these scales, patients have to improve at least 24.3% in the LF-MPQ and 37.5% in the SF-MPQ (Table 2).

3.4. Construct validity

A correlation matrix amongst the LF-MPQ, SF-MPQ, and pain NRS is presented in Table 3. We observed positive, moderate-to-high, and statistically significant correlations among these measures, which demonstrates good construct validity.

3.5. Floor and ceiling effects

No floor and ceiling effects were detected for either the total scores or the subscores (i.e., VAS, NWC, PPI, and NRS) of the measures at baseline. The distribution of responses at baseline is shown in the Fig. 1.

3.6. Responsiveness

Table 4 presents the internal and external responsiveness of the measures. In terms of internal responsiveness, we observed no differences between the SF-MPQ and LF-MPQ as the 84% CIs overlapped in all comparisons among the ESs. The ESs values ranged from, for example, PPI SF-MPQ = 0.35, 84% CI: 0.19, 0.51 to, for example, PRI-T LF-MPQ = 0.55, 84% CI: 0.42, 0.69. In terms of external responsiveness measured by the correlation of the change scores of the MPQ with GPE scale at discharge, we found small to moderate correlations (Table 4). Finally we

Table 2
Internal consistency and reproducibility of the Brazilian-Portuguese versions of the LF-MPQ and SF-MPQ

	Internal consistency (n = 203)	Reproducibility (n = 103)		
		Reliability	Agreement	
Instruments	Cronbach α (range of "α if item deleted")	$ICC_{2,1}$ (95% CI)	SEM	SDC
LF-MPQ				
PRI Total (0–78)	n/a	0.80 (0.72, 0.86)	6.92	19.17
PRI Sensory (0–41)	0.70 (0.66–0.70)	0.77 (0.67, 0.84)	4.71	13.05
PRI Affective (0–14)	0.79 (0.72–0.79)	0.77 (0.67, 0.84)	2.03	5.62
PRI Evaluative (0–5)	n/a	0.52 (0.36, 0.64)	1.18	3.27
PRI Miscellaneous (0–18)	0.55 (0.43–0.54)	0.70 (0.58, 0.79)	2.38	6.59
NWC (0–20)	n/a	0.80 (0.72, 0.86)	2.68	7.42
PPI (0–5)	n/a	0.46 (0.29, 0.60)	0.87	2.41
SF-MPQ				
PRI Total (0–45)	n/a	0.69 (0.57, 0.78)	6.23	17.26
PRI Sensory (0–33)	0.78 (0.75–0.79)	0.68 (0.55, 0.77)	4.87	13.49
PRI Affective (0–12)	0.76 (0.62–0.78)	0.64 (0.52, 0.74)	2.18	6.04
NWC (0–15)	n/a	0.66 (0.53, 0.76)	2.34	6.48
VAS (0–10)	n/a	0.75 (0.65, 0.82)	2.42	6.70
PPI (0–5)	n/a	0.55 (0.40, 0.67)	0.80	2.22

Abbreviations: LF-MPQ, Long Form of the McGill Pain Questionnaire; SF-MPQ, Short Form of the McGill Pain Questionnaire; ICC, intraclass correlation coefficient; CI, confidence interval; SEM, standard error of the measurement; SDC, smallest detectable change; PRI, Pain Rating Index; NWC, Number of Words Chosen; PPI, Present Pain Intensity; VAS, visual analog scale; n/a, not applicable.

Table 3
Pearson correlation among the LF-MPQ, SF-MPQ, and NRS

	LF-MPQ		SF-MPQ			
Instruments	PRI Total	PPI	PRI Total	PPI	VAS	NRS
LF-MPQ						
PRI Total	1	0.35	0.68	0.39	0.44	0.49
PPI	0.35	1	0.41	0.83	0.42	0.34
SF-MPQ						
PRI Total	0.68	0.41	1	0.50	0.62	0.55
PPI	0.39	0.83	0.50	1	0.51	0.42
VAS	0.44	0.42	0.62	0.51	1	0.65
NRS	0.49	0.34	0.55	0.42	0.65	1

Abbreviations: LF-MPQ, Long Form of the McGill Pain Questionnaire; SF-MPQ, Short Form of the McGill Pain Questionnaire; NRS, Numerical Rating Scale; PRI, Pain Rating Index; PPI, Present Pain Intensity; VAS, visual analog scale.

All correlations were significant ($P < 0.0001$).

observed AUCs ranging from 0.56 (95% CI: 0.45, 0.68) for the PRI-A of the SF-MPQ to 0.76 (95% CI: 0.66, 0.85) for the PRI-T score of the LF-MPQ. Most of the subscales were not able to distinguish patients who improved from those who did not improve.

4. Discussion

Our study aimed to cross-culturally adapt and test the clinimetric properties of the Brazilian-Portuguese version of the SF-MPQ and to test the clinimetric properties of the Brazilian-Portuguese version of the LF-MPQ [3]. The cross-cultural adaptation of a questionnaire requires a process of translation, back translation, and checking to ensure that the content and meaning of the original items of the questionnaire have been adequately retained for the new target language. We tested both Brazilian-Portuguese versions using the same structure that Melzack [1,4] had proposed for the original English versions. The results of this study indicate that these versions of the LF-MPQ and SF-MPQ are reproducible and valid for measuring the quality of pain in Portuguese-speaking patients with musculoskeletal conditions, making them suitable for use in routine health care and research studies.

We used a previous Brazilian-Portuguese version of the LF-MPQ [3], but we decided against using the items from the Brazilian-Portuguese LF-MPQ to compile the new SF-MPQ because not only the items but also the response options and the instructions must be adapted. Although the participants from the final committee compared the versions of the SF-MPQ with the LF-MPQ and four discrepancies were observed (items stabbing, cramping, heavy, and tender), these participants considered the words from the SF-MPQ more appropriate in terms of semantic and conceptual equivalence as well as considered these items from the short form easier for patients to understand.

Originally, the PRI of the LF-MPQ has been postulated by Melzack and Wall to assess three components of pain (i.e., sensory, affective, and evaluative); similarly the PRI of the SF-MPQ was claimed to assess two components of pain (i.e., sensory and affective). To fully test the internal consistency of a questionnaire, it is necessary to perform a confirmatory factor analysis to confirm the number of dimensions present in the questionnaire followed by a Cronbach α analysis for each dimension. However, the dimensions of the original English version of the LF-MPQ and the SF-MPQ were not

Table 4
Internal and external responsiveness of the measures

	Internal responsiveness	External responsiveness	
Instruments	Effect size (84% CI)	Correlations of the change scores with GPE at discharge	AUC—GPE cutoff for improvement = 3 or better (95% CI)
LF-MPQ			
PRI Total	0.55 (0.42, 0.69)	0.51	0.76 (0.66, 0.85)
PRI Sensory	0.46 (0.32, 0.60)	0.48	0.73 (0.63, 0.82)
PRI Affective	0.40 (0.25, 0.47)	0.34	0.68 (0.57, 0.79)
PRI Evaluative	0.34 (0.19, 0.49)	0.28	0.65 (0.54, 0.76)
PRI Miscellaneous	0.60 (0.44, 0.75)	0.43	0.72 (0.62, 0.82)
NWC	0.51 (0.38, 0.65)	0.49	0.72 (0.62, 0.82)
PPI	0.35 (0.19, 0.51)	0.34	0.62 (0.51, 0.74)
SF-MPQ			
PRI Total	0.47 (0.35, 0.59)	0.39	0.68 (0.58, 0.79)
PRI Sensory	0.49 (0.36, 0.61)	0.39	0.71 (0.61, 0.82)
PRI Affective	0.30 (0.18, 0.42)	0.23	0.56 (0.45, 0.68)
NWS	0.46 (0.32, 0.60)	0.39	0.69 (0.58, 0.79)
VAS	0.54 (0.38, 0.69)	0.36	0.64 (0.53, 0.75)
PPI	0.33 (0.18, 0.48)	0.32	0.63 (0.52, 0.74)
NRS	0.76 (0.57, 0.94)	0.43	0.69 (0.59, 0.80)

Abbreviations: CI, confidence interval; GPE, Global Perceived Effect; AUC, area under the curve; LF-MPQ, Long Form of the McGill Pain Questionnaire; PRI, Pain Rating Index; NWC, Number of Words Chosen; PPI, Present Pain Intensity; SF-MPQ, Short Form of the McGill Pain Questionnaire; VAS, visual analog scale; NRS, Numerical Rating Scale.

All correlations were significant ($P < 0.05$).

empirically derived, and factor analytic studies of the PRI of both versions have challenged the original three-factor and two-factor structure, respectively [16–20]. Because of the inconsistency of results from previous studies with regard to the factorial structure of the MPQ, the present study aimed to test only the original versions of the MPQ respecting the original theoretical structure. We are currently testing the factorial structure of the Brazilian-Portuguese versions of the MPQ, which will be reported in a separate manuscript.

Both the LF-MPQ and the SF-MPQ achieved adequate internal consistency with Cronbach α ranging from 0.70 to 0.79 for the LF-MPQ and from 0.76 to 0.78 for the SF-MPQ. The subscale PRI-M is a formative subscale of the LF-MPQ, and as a consequence the items of this subscale are not supposed to be highly correlated; for this reason, we have not calculated the internal consistency of this subscale.

In terms of reproducibility, both versions achieved moderate-to-substantial reliability and agreement. Interestingly, we found that the LF-MPQ was consistently more reproducible than the SF-MPQ. A possible explanation of this is that it may be easier for patients to nominate a particular word when answering the LF-MPQ rather than rating a four-point Likert scale when answering the SF-MPQ. The 24- to 48-hour interval chosen in our study could also inflate the results in favor to the LF-MPQ (because it may be easier to remember words rather than numbers in such a short period of time). However, a longer interval between the tests would also bias the results because true changes in pain in patients with musculoskeletal conditions receiving treatment would become more likely. Our reliability estimates for the SF-MPQ are lower than those previously reported for the English version [21] (Brazilian-Portuguese SF-MPQ $ICC_{2,1} = 0.69$ [95% CI: 0.57, 0.78] vs. English SF-MPQ $ICC_{1,1} = 0.96$ [95% CI: 0.94, 0.98]), but our estimates for both SF-MPQ and LF-MPQ are very similar to those reported on other international versions of the MPQ [2]. It is important to consider that most measurement studies that have tested the MPQ used different time intervals and different ICC types and, therefore, a direct comparison is not straightforward.

We tested agreement by calculating both SEM and SDC. After calculating the percentage of the SEM related to the total score, we observed 13.8% and 8.8% for the SF-MPQ and LF-MPQ representing doubtful and good agreement, respectively, for the measures. The estimates of agreement found in our study are similar to those found for the Norwegian version of the SF-MPQ. The SEM values for the PRI-T, PRI-S, and PRI-A score for the Norwegian version in musculoskeletal patients were 4.28, 3.19, and 1.73, respectively [22], compared with 6.23, 4.87, and 2.18 in our study. The SEM values for the PRI-T were 6.92 for the LF-MPQ and 6.23 for the SF-MPQ, and the SDC values were 19.17 and 17.26. The clinical interpretation of this is that the measurement error of the total scores of the LF-MPQ and the SF-MPQ are around 6.92 and 6.23 units of the MPQ, respectively, and there would need to be at least 19.17 and 17.26 units of improvement in the PRI-T, respectively, to be sure that a true change had occurred.

We measured both internal and external responsiveness of the pain measures by using a test–retest design at a 2-week interval. We classified the ability of the measures in detecting changes over time (internal responsiveness) as moderate [14] (PRI-T LF-MPQ ES = 0.55 (84% CI: 0.42, 0.69), PRI-T SF-MPQ ES = 0.47 (84% CI: 0.35, 0.59), and NRS ES = 0.76 (84% CI: 0.57, 0.94). We tested external responsiveness in two ways. First, we observed moderate correlations between the pain measures and the GPE scale at discharge. The strongest correlation was with the PRI-T of the LF-MPQ ($r = 0.51$, $P < 0.0001$). Second, we tested the ability of the pain measures to distinguish patients who improved from those who did not improve by using ROC curves; again the PRI-T of the LF-MPQ was the most responsive with an AUC = 0.76 (95% CI: 0.66, 0.85). Our hypotheses for responsiveness were confirmed for internal and external responsiveness for the LF-MPQ (PRI-T), as well as for external responsiveness measured by correlations for the SF-MPQ (PRI-T), but not for internal responsiveness and AUC. Few studies have previously evaluated the responsiveness of the MPQ [2]; our findings are consistent with those reported for the English version of the LF-MPQ (Spearman rho = 0.55, AUC = 0.69) [23] as well as the Korean (ES = 0.69 [84% CI: 0.60, 0.78], $r = 0.21$ [$P < 0.001$] and AUC = 0.63 [95% CI: 0.55, 0.72]) [24] and Norwegian (AUC = 0.61 [95% CI: 0.44, 0.77]) [22] versions of the SF-MPQ. Interestingly, the Norwegian [22], English [23], and Korean [24] studies found that the pain intensity scales such as VAS were more responsive than the MPQ, whereas we found that the pain NRS was better than the MPQ in terms of internal responsiveness but not for external responsiveness.

We observed positive and moderate-to-high correlations among the PRI-T LF-MPQ, PRI-T SF-MPQ, and Pain NRS, which demonstrate good construct validity of these measures. The pattern of results confirms our prespecified hypotheses with regard to size and direction of the correlations, with the exception of the PRI-T LF-MPQ. Finally, we did not observe any ceiling or floor effects for any of the measures.

Our results can inform the clinical decision as to whether to use the SF-MPQ or the LF-MPQ. First, if we consider a face-to-face comparison of the clinimetric properties of the MPQ, the LF-MPQ is better than the SF-MPQ with regard to reproducibility and responsiveness; however, the differences between the reliability and agreement coefficients, correlations, ESs, and AUCs in most of the cases were small. On the other hand, the SF-MPQ is simpler and quicker for patients to complete and, therefore, the SF-MPQ could be recommended for busy practices or in situations in which timing is an issue.

The results of this study provide an important set of relevant pain questionnaires for patients who speak Portuguese; these measures were cross-culturally adapted and clinimetrically tested following the recommendations of the current

guidelines on the topic [7,8]. As a consequence, clinicians and researchers from Portuguese-speaking countries have important self-report outcome measures for their musculoskeletal patients and research participants. Additionally, multicultural English-speaking countries such as the United States, the United Kingdom, and Australia can recruit Portuguese speakers for their clinical studies, and it will be possible to develop multicenter trials in Portuguese-speaking countries (in this case, it is suggested to perform Differential Item Functioning analysis using "language" as a group variable before pooling data of the MPQ collected with different language versions) [25]. With these questionnaires available, it is now possible to increase the amount of musculoskeletal research conducted in Brazil.

The Brazilian-Portuguese versions of the MPQ developed for this study were found to be reproducible and valid for the assessment of patients with musculoskeletal pain. The patients were recruited from both private and public physiotherapy practices in different regions of one of the largest cities of Brazil; therefore, the sample recruited for this study is likely to provide an adequate representative of patients with musculoskeletal pain who are seeking treatment for their conditions. With the new era of evidence-based practice, not only are questionnaires with proper clinimetric properties needed, but also high-quality cross-cultural adaptation of outcome measures is required. Our results suggest that clinicians and researchers from Portuguese-speaking countries could rely upon on the set of pain measures tested in this study.

Appendix

Versão curta do questionário de dor McGill

*Por favor, leia cada palavra abaixo e decida se ela descreve a dor que você sente. Se a palavra **não** descreve a sua dor, assinale NENHUMA, e vá para o próximo item. Se a palavra descreve a sua dor, quantifique essa sensação, escolhendo as opções leve, moderada ou severa.*

	Nenhuma	Leve	Moderada	Severa
1. Latejante	0	1	2	3
2. Em fiscadas	0	1	2	3
3. Em fincada	0	1	2	3
4. Aguda	0	1	2	3
5. Cólica	0	1	2	3
6. Pressionante	0	1	2	3
7. Em queimação	0	1	2	3
8. Dolorida	0	1	2	3
9. Pesada	0	1	2	3
10. Dolorida à palpação	0	1	2	3
11. Cortante	0	1	2	3
12. Cansativa - Exaustiva	0	1	2	3
13. Nauseante	0	1	2	3
14. Amedrontadora	0	1	2	3
15. Cruel - Punitiva	0	1	2	3

Por favor, marque na escala como, no geral, sua dor se apresentou nos **últimos dias**.

Nenhuma Dor _____ Pior Dor Possível

Qual a intensidade da sua dor ***agora***?

0 Sem dor _____
1 Leve _____
2 Desconfortante _____
3 Angustiante _____
4 Horrível _____
5 Excruciante _____

References

[1] Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975;1:277−99.

[2] Menezes Costa LdC, Maher CG, McAuley JH, Costa LOP. Systematic review of cross-cultural adaptations of McGill Pain Questionnaire reveals a paucity of clinimetric testing. J Clin Epidemiol 2009;62:934−43.

[3] Varoli FK, Pedrazzi V. Adapted version of the McGill Pain Questionnaire to Brazilian Portuguese. Braz Dent J 2006;17(4):328−35.

[4] Melzack R. The short-form McGill Pain Questionnaire. Pain 1987;30:191−7.

[5] 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.

[6] Thompson ISI. Top 20 country rankings in all essential science indicators, 2007. 2007. Available at http://in-cites.com/countries/2007allfields.html. Accessed February 12, 2009.

[7] 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.

[8] 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;60:34−42.

[9] Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17:163−70.

[10] Costa LOP, Maher CG, Latimer J, Ferreira PH, Ferreira ML, Pozzi GC, et al. Clinimetric testing of three self-report outcome measures for low back pain patients in Brazil: which one is the best? Spine 2008;33:2459−63.

[11] Ostelo RW, de Vet HC, Knol DL, van den Brandt PA. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. J Clin Epidemiol 2004;57:268−76.

[12] Tryon WW. Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: an integrated alternative method of conducting null hypothesis statistical tests. Psychol Methods 2001;6(4):371−86.

[13] Pengel LHM, Refshauge KMP, Maher CGP. Responsiveness of pain, disability, and physical impairment outcomes in patients with low back pain. Spine 2004;29:879−83.

[14] Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53:459−68.

[15] Innes E, Straker L. Validity of work-related assessments. Work 1999;13(2):125−52.

[16] Burckhardt CS, Bjelle A. A Swedish version of the short-form McGill Pain Questionnaire. Scand J Rheumatol 1994;23(2):77−81.

[17] Turk DC, Rudy TE, Salovey P. The McGill Pain Questionnaire reconsidered: confirming the factor structure and examining appropriate uses. Pain 1985;21:385−97.

[18] Prieto EJ, Hopson L, Bradley LA, Byrne M, Geisinger KF, Midax D, et al. The language of low back pain: factor structure of the McGill pain questionnaire. Pain 1980;8:11−9.

[19] Wright KD, Asmundson GJ, McCreary DR. Factorial validity of the short-form McGill pain questionnaire (SF-MPQ). Eur J Pain 2001;5(3):279−84.

[20] Beattie PF, Dowda M, Feuerstein M. Differentiating sensory and affective-sensory pain descriptions in patients undergoing magnetic resonance imaging for persistent low back pain. Pain 2004;110:189−96.

[21] Grafton KV, Foster NE, Wright CC. Test-retest reliability of the Short-Form McGill Pain Questionnaire: assessment of intraclass correlation coefficients and limits of agreement in patients with osteoarthritis. Clin J Pain 2005;21:73−82.

[22] 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(7):917−25.

[23] Scrimshaw SV, Maher C. Responsiveness of visual analogue and McGill pain scale measures. J Manipulative Physiol Ther 2001;24(8):501−4.

[24] 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.

[25] Zumbo BD. Three generations of DIF analyses: considering where it has been, where it is now, and where it is going. Lang Assess Q 2007;4(2):223−33.