Table 3 Goodness-of-fit values for the different models tested (N = 160)

	χ^2 (df)	IFI	RMSEA	CFI	AIC
Model 1					
Two oblique factors (7 + 6 items)	228.76 (64)	0.82	0.10	0.82	308.76
Model 2					
Three oblique factors (6 + 3 + 4 items)	209.54 (62)	0.84	0.10	0.84	267.54
Model 3					
Three varimax factors (5 + 3 + 5 items)	109.30 (60)	0.91	0.07	0.91	171.31

Model 1 = two-model factor structure suggested from the findings of Osman et al. and Chibnall and Tait [34,35]; model 2 = three-factor structure suggested by the findings of Van Damme et al. and by D'Eon et al. [27,33]; model 3 = three-factor structure suggested by the findings from the current study (with correlation between the error terms associated with items 4 and 5 (e4, e5) and items 4 and 12 (e4, e12); AIC = Akaike information criterion; CFI = comparative fit index; IFI = incremental fit index; RMSEA = root mean square error of approximation.

ranging from 0.73 to 0.91. The ICC (95% confidence interval) was 0.86 (0.79, 0.94) for Helplessness, 0.91 (0.80, 0.96) for Magnification, 0.73 (0.40, 0.88) for Rumination, and 0.90 (0.79, 0.95) for the total J-PCS score. The average test–retest interval for the reliability sample of 30 participants was 12.1 days (range: 7–28 days).

Associations Between the J-PCS Scores and Criterion Variables

Univariate Analyses

All four of the J-PCS scales showed significant univariate associations with the four criterion variables assessing pain intensity, pain interference, anxiety, and depression (see Table 4). The univariate associations were strongest for the PCS Helplessness scale predicting three of the criterion variables (pain intensity, pain interference, and depression), although the PCS Magnification score evidenced the strongest association with anxiety.

Regression Analyses

The results of the regression analyses predicting the criterion variables from the J-PCS total scores and subdomain scores (controlling for age, gender, and pain duration in every analyses; controlling for pain intensity in analyses

predicting pain interference, anxiety, and depression; and controlling for anxiety and depression in analyses predicting pain intensity and pain interference) are presented in Table 5. As can be seen, and consistent with the univariate analyses, the J-PCS scales made a significant contribution to the prediction of pain intensity, pain interference, anxiety, and depression. The significant effects for the J-PCS predicting pain intensity and pain interference remained, even when controlling for the demographic variables (age, gender and pain duration), pain intensity, anxiety, and depression (additional variance accounted for by the J-PCS ranged from 3% for predicting pain interference and 8% for predicting pain intensity). Similarly, the significant effects for the J-PCS scales predicting psychological functioning remained, even when controlling for the demographic variables (age, gender and pain duration) and pain intensity (additional variance accounted for by the J-PCS scales ranged from 11% for predicting depression to 25% for predicting anxiety). The J-PCS Helplessness scale made a significant independent contribution to the prediction of three of the criterion variables (pain intensity, pain interference, and depression). Especially, for pain interference, only the Helplessness scale made a significant independent contribution, whereas total PCS score did not. The J-PCS Magnification scale made a significant and unique contribution to the prediction of anxiety. In every case, consistent with the univariate analyses, the direction of the

Table 4 Correlation coefficients between the Japanese Pain Catastrophizing Scale scores and pain severity, pain interference, anxiety, and depression (N = 160)

	Japanese Pain Catastrophizing Scale			
	Rumination	Helplessness	Magnification	Total
BPI pain intensity composite score	0.20**	0.35***	0.27***	0.35***
BPI pain interference composite score	0.19**	0.47***	0.31***	0.41***
HADS anxiety	0.30***	0.42***	0.52***	0.50***
HADS depression	0.20**	0.39***	0.24**	0.36***

** $P < 0.01$; *** $P < 0.001$.
BPI = Brief Pain Inventory; HADS = Hospital Anxiety and Depression Scale.

Table 5 Hierarchical regression results predicting pain intensity, pain interference, depression, and anxiety from catastrophizing (N = 160)

Step and Variables	Total R^2	ΔR^2	F-change	Beta to Enter	t
A. Criterion: Pain intensity (BPI intensity score)					
Step 1: Demographic and pain history variables	0.02	0.02	1.10		
Age				0.13	1.64
Gender				0.03	0.36
Pain duration				−0.10	−1.35
Step 2: Depression and anxiety	0.08	0.06	5.15**		
Depression				0.08	0.80
Anxiety				0.02	0.20
Step 3: PCS subdomains scores	0.17	0.08	4.91**		
J-PCS Rumination				−0.05	−0.55
J-PCS Helplessness				0.29	2.90**
J-PCS Magnification				0.10	1.02
Step 3′: PCS total scores	0.15	0.06	11.01**	0.29	3.32**
B. Criterion: Pain interference (BPI interference score)					
Step 1: Demographic and pain history variables	0.02	0.02	1.24		
Age				0.17	2.05**
Gender				−0.01	−0.21
Pain duration				−0.03	−0.52
Step 2: Pain intensity	0.34	0.31	72.70***		
BPI intensity score				0.42	6.93***
Step 3: Depression and anxiety	0.53	0.19	30.82***		
Depression				0.32	4.71***
Anxiety				0.12	1.58
Step 4: PCS subdomains scores	0.55	0.03	2.89***		
J-PCS Rumination				−0.12	−1.80
J-PCS Helplessness				0.21	2.81**
J-PCS Magnification				−0.02	−0.23
Step 4′: PCS total scores	0.53	0.00	1.11	0.07	1.05
C. Criterion: Anxiety (HADS anxiety score)					
Step 1: Demographic and pain history variables	0.01	0.01	0.27		
Age				−0.03	−0.49
Gender				0.05	0.69
Pain duration				0.06	0.82
Step 2: Pain intensity	0.06	0.05	8.28**		
BPI intensity score				0.05	0.68
Step 3: PCS subdomains scores	0.31	0.25	18.50***		
J-PCS Rumination				0.03	0.35
J-PCS Helplessness				0.17	1.87
J-PCS Magnification				0.41	4.92***
Step 3′: PCS total scores	0.26	0.21	43.31***	0.49	6.58***
D. Criterion: Depression (HADS depression score)					
Step 1: Demographic and pain history variables	0.00	0.00	0.17		
Age				0.01	0.06
Gender				0.02	0.32
Pain duration				0.04	0.50
Step 2: Pain intensity	0.50	0.05	7.62**		
BPI intensity score				0.08	1.03
Step 3: PCS subdomains scores	0.16	0.11	6.82***		
J-PCS Rumination				−0.01	−0.16
J-PCS Helplessness				0.35	3.48***
J-PCS Magnification				0.04	0.44
Step 3′: PCS total scores	0.14	0.09	16.01***	0.32	4.00***

** $P < 0.01$; *** $P < 0.001$.
BPI = Brief Pain Inventory; J-PCS = Japanese Pain Catastrophizing Scale; HADS = Hospital Anxiety and Depression Scale.
Note: Each criterion variable was predicted with the PCS subscales and the PCS total score entered into the final step (steps 3 or 3′, or steps 4 or 4′, respectively). The beta weights listed are from the final equation after all other variables have been entered.

significant associations were in the hypothesized direction (i.e., greater catastrophizing associated with greater pain intensity, pain interference, anxiety, and depression).

Discussion

This investigation is the first to examine the factor structure of the J-PCS in a clinical sample of Japanese patients with chronic pain. It is also one of the few studies to examine the associations between the subdomains of catastrophizing assessed by the PCS and measures of pain and functioning, and the first study that we are aware of to examine the associations between the PCS subdomains and anxiety. The findings have important implications for studying the importance of catastrophizing, especially for studying the individual catastrophizing domains vs global catastrophizing as a psychological variables contributing to dysfunction in patients with chronic pain across cultures.

The results of CFA indicated that the best solution for sample of patients with chronic pain is the three-factor structure, which is the most common structure reported in the literature [30,40,41,53]. This supports a conclusion that catastrophizing—at least as measured by the PCS—is made up of three subdomains. Moreover, this solution appears to generalize across cultures, supporting its generalizability. We found a better fit for the factor structure with item 1 in the Rumination scale (as opposed to in the Helplessness scale, as suggested by the original PCS model). A closer examination of item 1 ("I worry all the time about whether the pain will end") suggests that it may reflect Rumination more than Helplessness, consistent with our findings. One study with a German translation of the PCS also indicated differences in which an item (item 12 of the PCS) loaded on factor inconsistent with the original PCS scoring. The difference in loadings found in different studies could reflect either (or both) 1) random variation between samples, regardless of culture or language, or 2) cultural differences in how the PCS items are interpreted. Further studies should examine the reliability of this finding in other samples of Japanese-speaking patients, as well as, perhaps, in samples of individuals that speak other languages for which a PCS translation is available.

The results from the regression analyses indicated that it might be more useful to examine and understand the associations between catastrophizing subdomains and measures of patient functioning, then to understand the associations between global catastrophizing and patient functioning. Specifically, like some previous studies (e.g., [11–15]), we found that the Helplessness catastrophizing domain was most closely associated with most pain-related criterion variables, independent of anxiety and depression, although the magnification subdomain was most strongly associated with anxiety in our sample. Although each subscale was correlated significantly with the indices of emotional distress, helplessness catastrophizing appears to contribute unique variance to the

prediction of most pain-related criterion variables, even when controlling for emotional distress. The extent to which our findings regarding magnification catastrophizing and anxiety would replicate in sample of patients from other cultures or who speak other languages is not clear, as no prior study has examined the individual PCS's associations with measure of anxiety.

As a group, the findings regarding the individual scales' associations and the criterion variables suggest that helplessness catastrophizing (e.g., "I feel I can't go on") may be the most important catastrophizing domain that predicts patient functioning across different languages and cultures, and that clinicians should perhaps pay closest attention to reducing this type of catastrophizing cognition, relative to other catastrophizing cognitions. The findings also raise the intriguing possibility that magnification cognitions (e.g., "I become afraid that the pain may get worse") may be particularly important to anxiety, at least in Japanese-speaking patients. Although it is not possible to draw causal conclusions from correlational data such as those collected in this study, our findings raise the intriguing possibility that when patients have a tendency to think magnifying catastrophic thoughts, treatment focused on decreasing this magnification could act to decrease their anxiety. Experimental research testing this hypothesis is warranted.

The association between catastrophic helplessness and anxiety might be weak because of the high arousal nature of anxiety symptoms. Catastrophic helplessness may elicit feelings of "hopelessness" and "giving up" rather than feelings of "helplessness" or "threat," which might explain why this type of catastrophizing is more closely related to depression than anxiety [54]. On the other hand, catastrophic magnification has been hypothesized to be most closely related to primary appraisal processes, where individuals may focus on and exaggerate the threat value of painful stimuli [1]. By focusing the patient's attention on pain sensations and exaggerating the threat value of pain symptoms, magnification cognitions may also increase anxiety-related emotional distress (e.g., feelings of "fear," "worry," "dread," and "uneasiness") associated with the pain experience.

The study has a number of limitations that should be considered when interpreting the results. This study represents the first validation of the J-PCS in a clinical mixed chronic pain sample, and the sample was selected from a specific clinic (a university-based psychosomatic medicine clinic) whose patients may differ in important ways from other clinics in Japan that serve patients with chronic pain. Patients referred to the Department of Psychosomatic Medicine in Kyushu University Hospital tend to be seen after a great deal of "doctor shopping" and to be more complex and distressed, relative to those referred to general hospital clinics. This could be one reason why the mean catastrophizing score for participants in this study (35.04) was higher than the mean PCS score (27.96) reported by Sullivan et al. [24] by patients in a multidisciplinary

treatment center specializing in the management of persistent pain disorders. It is possible, therefore, that the current findings may not generalize to all individuals with chronic pain in different pain locations and diseases who speak Japanese or who live in Japan. In addition, there were more women (70%) than men (30%) in our sample. Further study is therefore needed in other patient populations to help determine the reliability of the current findings in these other populations. Furthermore, the analyses testing the validity of the J-PCS were based on cross-sectional data, which means that causal conclusions cannot be drawn from the analyses. Thus, it cannot be determined from this study if the significant associations found were because: 1) catastrophizing influences pain interference, anxiety, and depression; 2) mood and pain interference influence catastrophizing; 3) there are ongoing bidirectional causal influences among these variables; 4) these variables are all associated with some other factor that influences them all at the same time; or 5) some combination of the above. However, the findings do support associations between catastrophizing and measured by the J-PCS and other important pain-related variables in a Japanese sample of patients with chronic pain, supporting the need for future research to examine the potential causal associations among these variables. In addition, we used only one measure of each psychological variable studied. Although these were adequate measures with proven reliability and validity, the use of additional criterion measures would have provided additional support for the study findings. Finally, the internal consistency for the PCS Magnification subscale ($\alpha = 0.69$) was marginal. Although this level of reliability is consistent with the reliability of the Magnification scale reported by Sullivan et al. [1] in their original work and by other researchers [13,29–32,40–42,53], different findings might occur when and if this catastrophizing domain is assessed more reliably. Moreover, the internal consistency for the PCS Rumination subscale ($\alpha = 0.72$) was relatively low. Although the α for the Rumination was acceptable, further work to improve the reliability of both the Magnification and Rumination subscales may also improve their predictive validity. Finally, the first and second administration of the PCS in the test–retest assessment occurred in different settings (in the clinic and at home, respectively). This difference could potentially lead to an underestimation of the stability coefficients. However, despite the difference in settings, the J-PCS scores still demonstrated adequate to excellent test–retest reliability in our sample.

Despite the study's limitations, the findings provide evidence for the greater explanatory power of the PCS subdomains, relative to the PCS total score, to predict to pain-related functioning. The results also provide additional support for the three-factor structure, reliability, and validity of the J-PCS. Because almost all of the research studying pain-related catastrophizing to date has been conducted in Western culture countries, the findings are also important because they support the psychometric properties of a measure of catastrophizing that can be used to determine the generalizability of findings from individuals from Western cultures to individuals from Japan. At this point, very little is known about the structure and correlates of the PCS items as a function of race or culture; yet, such knowledge is increasingly recognized as important [55]. Our results support the use of the J-PCS for studying catastrophizing subdomains and their correlates among individuals with chronic pain who live in Japan, an East Asian country whose ethnic group, mentality, and social circumstance differ in many ways from Western countries.

References

1 Sullivan MJ, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychol Assess 1995;7:524–32.

2 Jensen MP, Turner JA, Romano JM. Correlates of improvement in multidisciplinary treatment of chronic pain. J Consult Clin Psychol 1994;62:172–9.

3 Jensen MP, Turner JA, Romano JM. Changes in beliefs, catastrophizing, and coping are associated with improvement in multidisciplinary pain treatment. J Consult Clin Psychol 2001;69:655–62.

4 Burns JW, Glenn B, Bruehl S, Harden RN, Lofland K. Cognitive factors influence outcome following multidisciplinary chronic pain treatment: A replication and extension of a cross-lagged panel analysis. Behav Res Ther 2003;41:1163–82.

5 Burns JW, Kubilus A, Bruehl S, Harden RN, Lofland K. Do changes in cognitive factors influence outcome following multidisciplinary treatment for chronic pain? A cross-lagged panel analysis. J Consult Clin Psychol 2003;71:81–91.

6 Goodin BR, McGuire LM, Stapleton LM, et al. Pain catastrophizing mediates the relationship between self-reported strenuous exercise involvement and pain ratings: Moderating role of anxiety sensitivity. Psychosom Med 2009;71:1018–25.

7 Jamison RN, Link CL, Marceau LD. Do pain patients at high risk for substance misuse experience more pain? A longitudinal outcomes study. Pain Med 2009;10:1084–94.

8 Papaioannou M, Skapinakis P, Damigos D, et al. The role of catastrophizing in the prediction of postoperative pain. Pain Med 2009;10:1452–9.

9 Richardson EJ, Ness TJ, Doleys DM, et al. Depressive symptoms and pain evaluations among persons with

chronic pain: Catastrophizing, but not pain acceptance, shows significant effects. Pain 2009;147:147–52.

10 Shelby RA, Somers TJ, Keefe FJ, et al. Pain catastrophizing in patients with noncardiac chest pain: Relationships with pain, anxiety, and disability. Psychosom Med 2009;71:861–8.

11 Sullivan MJ, Thibault P, Simmonds MJ, et al. Pain, perceived injustice and the persistence of post-traumatic stress symptoms during the course of rehabilitation for whiplash injuries. Pain 2009;145:325–31.

12 Cusens B, Duggan GB, Thorne K, Burch V. Evaluation of the breathworks mindfulness-based pain management programme: Effects on well-being and multiple measures of mindfulness. Clin Psychol Psychother 2010;17:63–78.

13 Gilliam W, Burns JW, Quartana P, et al. Interactive effects of catastrophizing and suppression on responses to acute pain: A test of an appraisal x emotion regulation model. J Behav Med 2010;33:191–9.

14 Mounce C, Keogh E, Eccleston C. A principal components analysis of negative affect-related constructs relevant to pain: Evidence for a three component structure. J Pain 2010;11:710–7.

15 van Ittersum MW, van Wilgen CP, Groothoff JW, van der Schans CP. Is appreciation of written education about pain neurophysiology related to changes in illness perceptions and health status in patients with fibromyalgia? Patient Educ Couns 2011;85:269–74.

16 Lautenbacher S, Huber C, Baum C, et al. Attentional avoidance of negative experiences as predictor of postoperative pain ratings and consumption of analgesics: Comparison with other psychological predictors. Pain Med 2011;12:645–53.

17 Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: Relationship to patient characteristics and current adjustment. Pain 1983;17:33–44.

18 Boothby JL, Thorn BE, Stroud MW, Jensen MP. Psychosocial Factors in Pain. New York: Guilford Press; 1999.

19 Tripp DA, Nickel JC, Fitzgerald MP, et al. Sexual functioning, catastrophizing, depression, and pain, as predictors of quality of life in women with interstitial cystitis/painful bladder syndrome. Urology 2009;73:987–92.

20 Sullivan MJ, Lynch ME, Clark AJ. Dimensions of catastrophic thinking associated with pain experience and disability in patients with neuropathic pain conditions. Pain 2005;113:310–5.

21 Tripp DA, Nickel JC, Wang Y, et al. Catastrophizing and pain-contingent rest predict patient adjustment in men with chronic prostatitis/chronic pelvic pain syndrome. J Pain 2006;7:697–708.

22 Nijs J, Van de Putte K, Louckx F, Truijen S, De Meirleir K. Exercise performance and chronic pain in chronic fatigue syndrome: The role of pain catastrophizing. Pain Med 2008;9:1164–72.

23 Vienneau TL, Clark AJ, Lynch ME, Sullivan MJ. Catastrophizing, functional disability and pain reports in adults with chronic low back pain. Pain Res Manag 1999;4:93–6.

24 Sullivan MJ, Stanish W, Waite H, Sullivan M, Tripp DA. Catastrophizing, pain, and disability in patients with soft-tissue injuries. Pain 1998;77:253–60.

25 Sullivan MJ, Neish NR. Catastrophizing, anxiety and pain during dental hygiene treatment. Community Dent Oral Epidemiol 1998;26:344–9.

26 Devoulyte K, Sullivan MJ. Pain catastrophizing and symptom severity during upper respiratory tract illness. Clin J Pain 2003;19:125–33.

27 Sullivan MJ, Stanish W, Sullivan ME, Tripp D. Differential predictors of pain and disability in patients with whiplash injuries. Pain Res Manag 2002;7:68–74.

28 Desrochers G, Bergeron S, Khalife S, Dupuis MJ, Jodoin M. Provoked vestibulodynia: Psychological predictors of topical and cognitive-behavioral treatment outcome. Behav Res Ther 2009;48:106–15.

29 Tremblay I, Beaulieu Y, Bernier A, et al. Pain Catastrophizing Scale for Francophone adolescents: A preliminary validation. Pain Res Manag 2008;13:19–24.

30 Van Damme S, Crombez G, Bijttebier P, Goubert L, Van Houdenhove B. A confirmatory factor analysis of the Pain Catastrophizing Scale: Invariant factor structure across clinical and non-clinical populations. Pain 2002;96:319–24.

31 Meyer K, Sprott H, Mannion AF. Cross-cultural adaptation, reliability, and validity of the German version of the Pain Catastrophizing Scale. J Psychosom Res 2008;64:469–78.

32 Miro J, Nieto R, Huguet A. The Catalan version of the Pain Catastrophizing Scale: A useful instrument to assess catastrophic thinking in whiplash patients. J Pain 2008;9:397–406.

33 Yap JC, Lau J, Chen PP, et al. Validation of the Chinese Pain Catastrophizing Scale (HK-PCS) in patients with chronic pain. Pain Med 2008;9:186–95.

34 Matsuoka H, Sakano Y. Assessment of cognitive aspect of pain: Development, reliability, and validation of Japanese version of Pain Catastrophizing Scale. (Japanese). Jpn J Psychosom Med 2007;47:95–102.

35 Matsuoka H, Himachi M, Furukawa H, et al. Cognitive profile of patients with burning mouth syndrome in the Japanese population. Odontology 2010;98:160–4.

36 Edwards RR, Moric M, Husfeldt B, Buvanendran A, Ivankovich O. Ethnic similarities and differences in the chronic pain experience: A comparison of African American, Hispanic, and white patients. Pain Med 2005;6:88–98.

37 Forsythe LP, Thorn B, Day M, Shelby G. Race and sex differences in primary appraisals, catastrophizing, and experimental pain outcomes. J Pain 2011;12:563–72.

38 Fabian LA, McGuire L, Goodin BR, Edwards RR. Ethnicity, catastrophizing, and qualities of the pain experience. Pain Med 2011;12:314–21.

39 Hsieh AY, Tripp DA, Ji LJ, Sullivan MJ. Comparisons of catastrophizing, pain attitudes, and cold-pressor pain experience between Chinese and European Canadian young adults. J Pain 2010;11:1187–94.

40 D'Eon JL, Harris CA, Ellis JA. Testing factorial validity and gender invariance of the Pain Catastrophizing Scale. J Behav Med 2004;27:361–72.

41 Osman A, Barrios FX, Kopper BA, et al. Factor structure, reliability, and validity of the Pain Catastrophizing Scale. J Behav Med 1997;20:589–605.

42 Chibnall JT, Tait RC. Confirmatory factor analysis of the Pain Catastrophizing Scale in African American and Caucasian Workers' Compensation claimants with low back injuries. Pain 2005;113:369–75.

43 Sullivan MJ, D'Eon JL. Relation between catastrophizing and depression in chronic pain patients. J Abnorm Psychol 1990;99:260–3.

44 Uki J, Mendoza T, Cleeland CS, Nakamura Y, Takeda F. A brief cancer pain assessment tool in Japanese: The utility of the Japanese Brief Pain Inventory—BPI-J. J Pain Symptom Manage 1998;16:364–73.

45 Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. J Pain 2004;5:133–7.

46 Kalyadina SA, Ionova TI, Ivanova MO, et al. Russian Brief Pain Inventory: Validation and application in cancer pain. J Pain Symptom Manage 2008;35:95–102.

47 Radbruch L, Loick G, Kiencke P, et al. Validation of the German version of the Brief Pain Inventory. J Pain Symptom Manage 1999;18:180–7.

48 Harkapaa K. Relationships of psychological distress and health locus of control beliefs with the use of cognitive and behavioral coping strategies in low back pain patients. Clin J Pain 1991;7:275–82.

49 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361–70.

50 Herrmann C. International experiences with the Hospital Anxiety and Depression Scale—A review of validation data and clinical results. J Psychosom Res 1997;42:17–41.

51 Bollen KA, Curran PJ. Latent Curve Models: A Structural Equation Perspective, 1st edition. Hoboken, NJ: John Wiley & Sons; 2005.

52 Stevens JP. Applied Multivariate Statistics for the Social Sciences, 4th edition. Mahwah, NJ: Lawrence Erlbaum Associates; 2002.

53 Osman A, Barrios FX, Gutierrez PM, et al. The Pain Catastrophizing Scale: Further psychometric evaluation with adult samples. J Behav Med 2000;23:351–65.

54 Leonard MT, Cano A. Pain affects spouses too: Personal experience with pain and catastrophizing as correlates of spouse distress. Pain 2006;126:139–46.

55 Green CR, Anderson KO, Baker TA, et al. The unequal burden of pain: Confronting racial and ethnic disparities in pain. Pain Med 2003;4:277–94.

Original Research Articles

Pain Questionnaire Development Focusing on Cross-Cultural Equivalence to the Original Questionnaire: The Japanese Version of the Short-Form McGill Pain Questionnaire

Tatsuyuki Arimura, MA,* Masako Hosoi, MD, PhD,*† Yoshihiro Tsukiyama, DDS, PhD,§ Toshiyuki Yoshida, MA,¶ Daiki Fujiwara, MA,** Masanori Tanaka, MA,†† Ryuichi Tamura, MA,‡‡ Yasunori Nakashima, MD, PhD,‡ Nobuyuki Sudo, MD, PhD,*† and Chiharu Kubo, MD, PhD†

*Department of Psychosomatic Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka;

Departments of †Psychosomatic Medicine and

‡Anesthesiology, Kyushu University Hospital, Fukuoka;

§Department of Oral Rehabilitation, Faculty of Dental Science, Kyushu University, Fukuoka;

¶Department of Speech and Hearing Sciences, School of Rehabilitation at Fukuoka, International University of Health and Welfare, Fukuoka;

**Kitakyushu Municipal Medical Center, Kitakyushu;

††Department of School Health Science, Faculty of Education, Hirosaki University, Hirosaki;

‡‡Faculty of Humanities, Fukuoka University, Fukuoka, Japan

Reprint requests to: Masako Hosoi, MD, PhD, Department of Psychosomatic Medicine, Kyushu University Hospital, 3-1-1 Maidashi, Higashiku, Fukuoka 812-8582, Japan. Tel: +81-92-642-5320; Fax: +81-92-642-5336; E-mail: hosoi@cephal.med. kyushu-u.ac.jp.

Abstract

Objectives. The present study aimed to develop a Japanese version of the Short-Form McGill Pain Questionnaire (SF-MPQ-J) that focuses on cross-culturally equivalence to the original English version and to test its reliability and validity.

Design. Cross-sectional design.

Method. In study 1, SF-MPQ was translated and adapted into Japanese. It included construction of response scales equivalent to the original using a variation of the Thurstone method of equal-appearing intervals. A total of 147 undergraduate students and 44 pain patients participated in the development of the Japanese response scales. To measure the equivalence of pain descriptors, 62 pain patients in four diagnostic groups were asked to choose pain descriptors that described their pain. In study 2, chronic pain patients (N = 126) completed the SF-MPQ-J, the Long-Form McGill Pain Questionnaire Japanese version (LF-MPQ-J), and the 11-point numerical rating scale of pain intensity. Correlation analysis examined the construct validity of the SF-MPQ-J.

Results. The results from study 1 were used to develop SF-MPQ-J, which is linguistically equivalent to the original questionnaire. Response scales from SF-MPQ-J represented the original scale values. All pain descriptors, except one, were used by >33% in at least one of the four diagnostic groups. Study 2 exhibited adequate internal consistency and test–retest reliability, with the construct validity of SF-MPQ-J comparable to the original.

Conclusion. These findings suggested that SF-MPQ-J is reliable, valid, and cross-culturally equivalent to the original questionnaire. Researchers might consider using this scale in multicenter, multi-ethnical trials or cross-cultural studies that include Japanese-speaking patients.

Key Words. Cross-Cultural Equivalence; Measurement; Chronic Pain; SF-MPQ

Introduction

The standard Long- and Short-Form McGill Pain Questionnaires (LF-MPQ [1] and SF-MPQ [2]: © R. Melzack, 1987) have been the most widely used instruments for assessing sensory and affective dimensions of pain experience. SF-MPQ is the abbreviated version of

LF-MPQ and has been more frequently used than LF-MPQ because it is simpler. The main component of SF-MPQ consists of 15 descriptors (11 sensory and 4 affective) that are rated on a 4-point intensity scale: 0 = none, 1 = mild, 2 = moderate, 3 = severe. SF-MPQ also includes the 6-point present pain index (PPI) and a visual analog scale (VAS), which measure overall pain intensity as a single dimension [2]. Although SF-MPQ has been translated into different languages (Czech [3], Greek [4], Iranian [5], Japanese [6], Korean [7], Norwegian [8], Swedish [9], Thai [10], and Turkish [11]) for cross-cultural pain studies, only Iranian [5], Korean [7], and Turkish [11] versions of the translated SF-MPQ have considered adequate cross-cultural equivalence to the original.

For effective cross-cultural comparison, the original and translated questionnaires should be cross-culturally equivalent [12], which becomes critical for large multi-country trials or cross-cultural studies because of the need for combining and analyzing data derived from multiple language versions of same questionnaire [13]. Equivalence should include the following two dimensions: 1) linguistic equivalence, i.e., equivalence of question wording and meaning in the formulation of items, response choices, and all aspects of the instrument [12]; and 2) measurement equivalence, i.e., the psychometric properties of different language versions of the same instrument are similar [14].

A previous study [15] showed that even if questions are literally translated, it is important to consider whether the concepts are meaningful and similarly interpreted across cultures. The translated version should retain item-level characteristics, such as intervals of response choices in the response scales, as well as score-level characteristics of means, reliability, construct validity, and responsiveness [16]. The measurement equivalence also contains a scalar equivalence, i.e., numerical value on a scale that refers to the same degree, intensity, or magnitude of the construct, regardless of the population in which the respondent is a member [17]. Moreover, international translation guidelines have been proposed for achieving cross-cultural equivalence [12,16,18–22].

However, more attention should be placed on issues related to cross-cultural equivalence and translation of the SF-MPQ. Most translations of SF-MPQ [3,4,6,8–10] have not been based on the translation guidelines. Although Iranian [5], Korean [7], and Turkish [11] versions have been developed according to these guidelines, they do not emphasize testing of measurement equivalence. The lack of cross-cultural equivalence might inhibit international comparisons between samples. Due to the difference in SF-MPQ scores among culturally different samples, the differences among scores might not reflect "real" differences, or they could reflect an artifact because of lack of cross-cultural equivalence in the questionnaire as a result of poor translation.

The aims of this study were as follows: 1) to develop a Japanese version of SF-MPQ (SF-MPQ-J) that is cross-culturally equivalent to the original English version (study 1) and 2) to determine the reliability, validity, and measurement equivalence of the SF-MPQ-J (study 2). In study 2, we hypothesized that 1) the SF-MPQ-J would demonstrate adequate internal consistency and test–retest reliability comparable to the original version; and 2) the SF-MPQ-J would exhibit adequate construct validity comparable to the original version in relation to LF-MPQ-J. Study 1 introduced the following items for the development of a cross-culturally equivalent questionnaire: 1) adoption of nearly identical scale construction methods to the original SF-MPQ main components and PPI development studies [2,23]; and 2) adherence to international translation guidelines. Therefore, study 1 included five different steps that integrated the scale construction methods of the original SF-MPQ and translations that follow the International Quality of Life Assessment (IQOLA) [19] and the American Association of Orthopaedic Surgeons (AAOS) [16] guidelines. The study design is shown in Figure 1.

Methods

Study 1

This study was approved by the Kyushu University Hospital Institutional Review Board and was performed at Kyushu University Hospital. All participants provided written informed consent prior to starting the study.

Step 1

The aim of step 1 was to create Japanese response scales (PPI and intensity scale) with scalar equivalence to the response choice intervals of the original English version. To accomplish this, the Japanese response choices must meet the following requirements: 1) scale values similar to the original and 2) similar scale values across different diagnostic subgroups (i.e., students and acute and chronic pain patients). This is because the target populations of the original SF-MPQ are different diagnostic subgroups (i.e., healthy subjects such as students and acute and chronic pain patients) [2]. The Japanese SF-MPQ should also have response choices that suit all of the subgroups. However, there is evidence that different diagnostic subgroups may differently rate the intensity of pain words [24]. Therefore, the original PPI was constructed by selecting response choices with common scale values between the subgroups (i.e., 140 introductory psychology students and 20 patients) [23]. Similarly, in the construction of Japanese response scales, it is necessary to recruit different diagnostic subgroups. Therefore, we recruited students and acute and chronic patients in step 1.

Participants. The samples used in step 1 were a nonclinical sample of 147 undergraduate students enrolled in Introductory Psychology and a clinical sample of 36 pain patients. Student (99 males and 48 females) volunteers
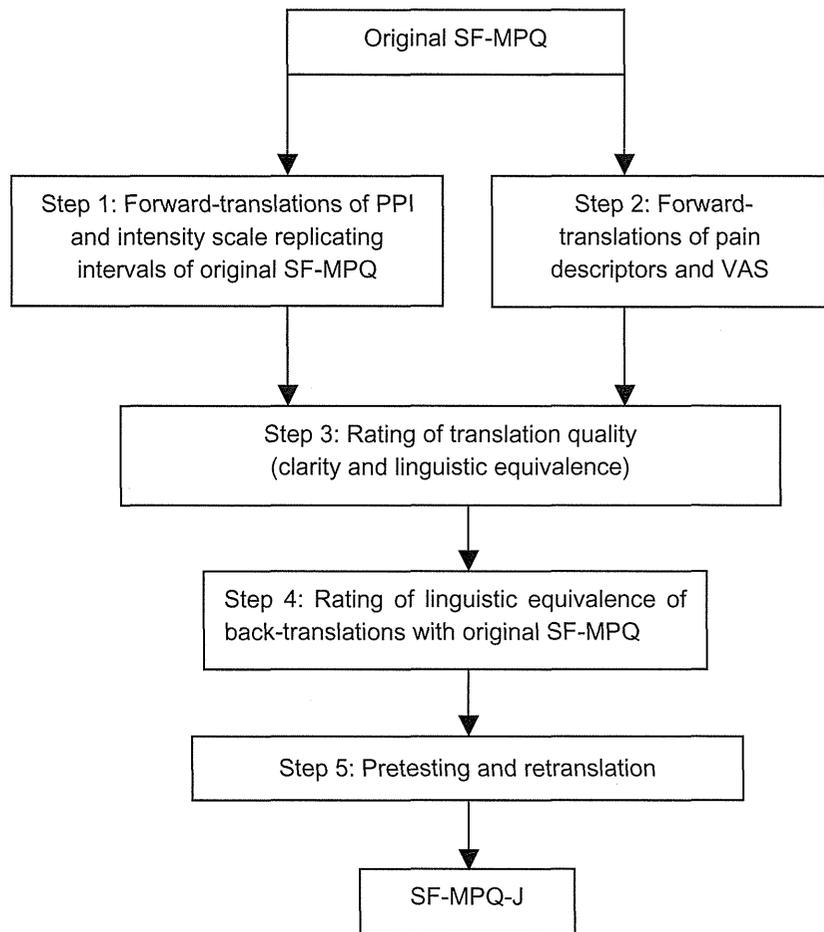
Figure 1 Flowchart showing the sequence of steps in study 1. PPI = present pain intensity; SF-MPQ = Short-Form McGill Pain Questionnaire; SF-MPQ-J = Short-Form McGill Pain Questionnaire Japanese version; VAS = visual analog scale.

participated in this study. The mean age was 18.9 years (standard deviation [SD] = 2.6). The clinical sample consisted of two subsamples: 1) 11 acute pain patients recruited from the dental clinic and 2) 25 chronic nonmalignant pain patients recruited from the Department of Psychosomatic Medicine and Pain Clinic at Kyushu University Hospital. Eligibility inclusion criteria for the clinical samples were as follows: 1) age 20 years and over; 2) symptoms of acute pain or chronic pain for at least 3 months; and 3) able to speak and read Japanese. Exclusion criteria were major cognitive or psychiatric disorders (e.g., dementia and severe depression), which could interfere with the ability to complete the self-reported questionnaire. The referring physicians determined if the participant met the inclusion/exclusion criteria without screening measures. None were excluded.

The average age of acute pain patients (three males and eight females) was 50.6 years (SD = 18.2). All acute pain patients presented with orofacial pain due to dental caries (27%), pressure on the oral mucosa by removable dentures (27%), pulpitis (18%), acute flare-up of periodontitis (18%), or aphthous stomatitis (9%). The average age of the chronic pain patients (8 males and 17 females) was 46.0 years (SD = 17.8). Average duration of pain since

time of onset was 92.8 months (SD = 88.0; range = 3–288). Primary pain locations included the head, face, and mouth (28%); upper shoulder and upper limbs (16%); lower back (12%); abdominal region (8%); lower limbs (4%), cervical region (4%); thoracic region (4%); and pain location unknown (24%), all based on standard chronic pain classifications [25].

Procedures. With permission from Dr. Ronald Melzack (the author of the original SF-MPQ) the SF-MPQ was translated into Japanese (SF-MPQ-J). Two professional translators (translators 1 and 2) were responsible for forward-translating response choices from the response scales into Japanese. A list of all reasonable translation options was independently produced and compiled. One translation was chosen for the bipolar response choices. Subsequently, participants were asked to rate intermediate response choices using the Thurstone scaling exercise [26], which is a variation of the Thurstone method of equal-appearing intervals [27].

In the case of PPI, no pain or excruciating pain ("itami-nashi" and "taerarenai") were used to define either extreme of the six-level set of PPI response choices. Translation options for the four middle response choices
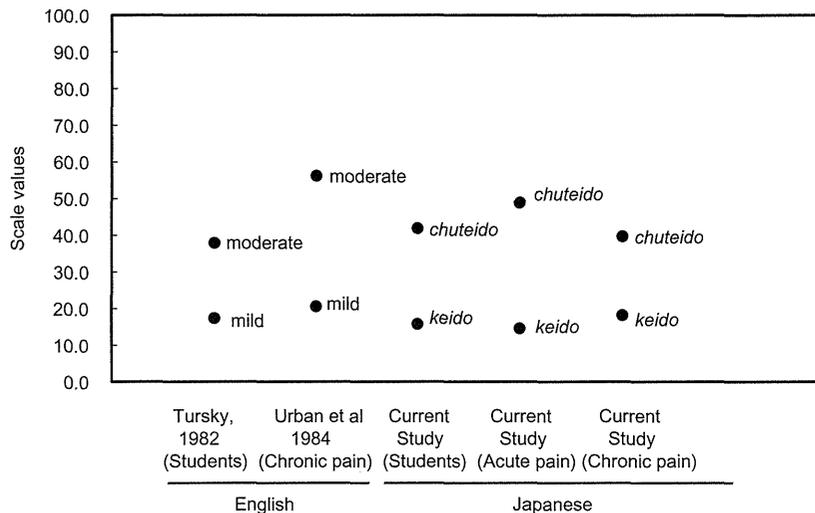
Arimura et al.

Figure 2 Graphic representation of the scale values for the anchor words ("mild" and "moderate" and "keido" and "chuteido") of the English and Japanese intensity scales on an interval scale. Note. Scaling: 0.0 = no pain; 100 = severe pain.

("mild," "discomforting," "distressing," and "horrible") were rated for interval level spacing along a 100 mm VAS. One end of the VAS represented "no pain," while the other end represented "excruciating." Participants were asked to rate the intensity of options for middle response choices on the VAS. In a similar fashion, a forward-translation of the intensity scale was also created.

Data Analysis. First, intervals among response choices from the original PPI and intensity scale were estimated. Because the original PPI was an interval scale [23], it consisted of equidistant intervals. However, because the intensity scale was an ordinal scale [2], intervals among response choices were unknown. Therefore, intervals among response choices were extrapolated from previous psychophysical scaling studies [28,29] using pain word descriptors. The left side of Figure 2 shows the estimated scale values of response choices in the intensity scale based on previous studies [28,29].

Second, mean scores for each of the potential Japanese response choices from the PPI and intensity scales were estimated from participant ratings. Finally, common response choices of the students and acute and chronic pain patients that approximately replicated intervals from the original PPI and intensity scale were selected. Data were analyzed using SPSS 16.0J for Macintosh (SPSS, Tokyo, Japan).

Step 2

Translators 1 and 2 forward-translated the pain descriptors and VAS endpoints into Japanese. Working independently, the translators produced two initial Japanese versions of pain descriptors and VAS, which were compared and compiled to produce a single, reconciled version.

Step 3

Two bilingual (Japanese and English), native Japanese speakers assessed the quality of the initial forward-

translation. Both were experienced interpreters of scientific meetings. The initial Japanese version was compared with the original English version, as well as judged (irrelevant, questionably relevant, or relevant) for translation clarity and linguistic equivalence. We reconsidered the items rated by at least one assessor as irrelevant or questionably relevant for inclusion.

Step 4

The forward-translation was back-translated into English by two professional translators (translators 3 and 4) who had no previous knowledge of SF-MPQ. Dr. Melzack, the author of SF-MPQ, compared the two back-translations with the original English version to confirm their linguistic equivalence.

Step 5

The purpose of step 5 was to perform a field test of the Japanese pain descriptors. The response rate for each descriptor was examined to detect and modify low response rate descriptors because the low response rate descriptors were estimated to be ambiguous and unidiomatic.

Participants. A total of 62 pain patients participated in the study. The patients were recruited from the Department of Psychosomatic Medicine, Dental Clinic, and Orthopedic Surgery Clinic at the Kyushu University Hospital. Eligibility inclusion criteria included the following: 1) aged 20 years and over; 2) symptoms of toothache, headache (migraine or tension-type headache), musculoskeletal pain, or pain due to rheumatoid arthritis; 3) suffering from pain for at least 3 months in the case of headache, musculoskeletal pain, or rheumatoid arthritis; and 4) able to speak and read Japanese. Exclusion criteria included 1) major cognitive or psychiatric disorders, or 2) more than two pain diagnoses. The referring physician determined if the participant met the inclusion/exclusion criteria without screening measures. None were excluded. The patient sample

544

included 10 acute pain and 52 nonmalignant, chronic pain patients. The average age of the acute pain patients (six males and four females) was 35.4 years (SD = 13.0). The average duration of pain since the time of onset was 4.3 days (SD = 3.9; range = 2–14). All acute pain patients suffered from toothache. The average age of the chronic pain patients (18 males and 34 females) was 47.9 years (SD = 16.2), and the average pain duration since the time of onset was 60.2 months (SD = 70.9; range = 3–312). A total of 12 patients suffered from headache, 29 from musculoskeletal pain, and 11 from pain due to rheumatoid arthritis.

Procedure. Forward-translation of the pain descriptors was administered by self-completion by the participants. The participants were asked to choose pain descriptors that best described their pain.

Data Analysis. The percentage of participants in each diagnostic group who chose a particular descriptor (toothache, headache, musculoskeletal pain, or pain due to rheumatoid arthritis) was calculated to detect low response rate descriptors. A low response rate descriptor was defined as a descriptor used by less than 33% of the participants from one diagnostic group, in accordance with previous studies [4,9]. Data were analyzed using SPSS 16.0J for Macintosh.

Study 2

The aim of study 2 was to examine score-level measurement equivalence, which included testing reliability, and to determine the construct validity for SF-MPQ-J.

Participants

As a part of the initial evaluation at the Department of Psychosomatic Medicine of Kyushu University Hospital, a total of 134 consecutive, nonmalignant, chronic pain patients were recruited for participation in the study. Eligibility inclusion criteria included the following: 1) aged 20 years and over; 2) symptoms of chronic pain for at least 3 months; and 3) able to speak and read Japanese. Exclusion criteria were the same as for step 1 of study 1. No one was excluded from the study due to the criteria. A total of eight participants provided incomplete data and were subsequently excluded from the analysis. There were no significant differences between the included and excluded participants in terms of socio-demographic and pain variables except for educational level. Participants who provided incomplete data were more likely to be junior high-school graduates ($X^2 = 11.97$, df = 1, $P < 0.01$). In total, 126 chronic pain patients (49 males and 77 females) were included in the final analysis, with a mean age of 47.8 years (SD = 15.3). Primary pain locations included head, face, and mouth (28.3%); lower limbs (17.3%); upper shoulder and upper limbs (21.3%); lower back (10.2%); cervical region (7.9%); abdominal region (8.7%); pelvic region (3.1%); thoracic region (0.8%); and

anal, perineal, and genital region (2.4%), all based on standard chronic pain classifications [25]. The average pain duration was 47.5 months (SD = 58.8; range = 3–312). In addition, 25% of the patients worked full-time and 15.0% received disability compensation. A total of 53% of the patients were married; 31.8% had some college education; and 52.4% were high school graduates.

Measures

Patients were asked to complete a face sheet, which included gender, age, educational level, marital status, and employment status. The patients also reported the pain site and date of pain onset. Pain duration was calculated by subtracting the patient-reported pain onset date from the date of the evaluation.

SF-MPQ-J

The SF-MPQ-J from study 1 consists of three pain scores (sensory, affective, and total scores), VAS, and PPI.

LF-MPQ-J

The standard LF-MPQ-J [30] was used to test the construct validity of SF-MPQ-J, similar to original validation studies [2,31]. The LF-MPQ is one of the most widely used assessment tools for measuring multidimensional aspects of pain. It consists of 78 pain descriptors that measure sensory, affective, evaluative, and miscellaneous dimensions of pain. The rank value of descriptors was summed to obtain a total score, as well as separate scores for sensory, affective, evaluative, and miscellaneous subscales. The LF-MPQ-J has been shown to exhibit acceptable reliability and validity for the assessment of chronic pain [30].

Numerical Rating Scale

The 11-point (0–10) NRS of pain intensity [24] was administered to test construct validity of the SF-MPQ-J. The NRS assesses current overall pain intensity (from 0 = "no pain" to 10 = "pain as bad as you can imagine") and has been shown to be valid [24]. The use of the NRS has also been recommended by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials [32].

Procedure

Patients were approached at their intake interview, on a consecutive basis, and were included in the study if they fulfilled the inclusion criteria and did not meet the exclusion criteria. All eligible patients who were approached agreed to participate. Administration of SF-MPQ-J and LF-MPQ-J was performed as previously described [1,2]. For the SF-MPQ-J, patients were told that a set of descriptors would be read to them, the patients stated whether or not the words described their present pain and, if the word did, they rated the intensity of that particular quality of pain. For the LF-MPQ-J, the descriptor sets were read to patients who were asked to choose words that best described their present pain. Administra-

tion of SF-MPQ-J and LF-MPQ-J was continuously performed in random order. Each participant was also asked to complete the NRS. For test–retest analysis, patients were asked to recomplete SF-MPQ-J. Fifteen individuals completed the SF-MPQ-J twice within a 1-week interval. Two participants were excluded due to invalid answers, leaving a final sample of 13 participants for test–retest analysis. The study was approved by the Kyushu University Hospital Institutional Review Boards, and all participants provided written informed consent prior to the start of the study.

Data Analysis and Statistics

The basic descriptive statistics included means, SDs, the floor and ceiling effect, skewness, and kurtosis of the SF-MPQ-J. Analyses revealed positively skewed distributions for sensory, affective, and total SF-MPQ-J scores. Distributions were improved by variable transformations (square-root transformation). The transformed scores were subsequently used in the following analysis, with exception of the descriptive statistics. Reliability and correlation analyses were then conducted, including internal consistency, test–retest reliability, and intercorrelations between subscales and total SF-MPQ-J scores.

Internal consistency of SF-MPQ-J was assessed using Cronbach's α. A Cronbach's α coefficient of 0.70 or above was considered satisfactory or adequate [33]. Test–retest reliability was assessed using intraclass correlation coefficients (ICCs). An ICC value of 0.70 or greater was considered satisfactory or good [33]. Intercorrelations between subscales and total scores were examined using Pearson's correlation coefficients.

Construct validity of SF-MPQ-J was determined by analyzing associations with other validated measures of the same construct [34]. The SF-MPQ-J includes two kinds of components that evaluate overall pain intensity as a single dimension (VAS and PPI) and multidimensional pain experience (main component of SF-MPQ-J). LF-MPQ-J measures multidimensional pain experience similar to the main component of SF-MPQ-J. NRS measures pain intensity similar to VAS and PPI. Therefore, LF-MPQ-J and NRS were used to determine the construct validity of SF-MPQ-J. Associations between SF-MPQ-J, LF-MPQ-J, and NRS were evaluated using Pearson's correlation coefficients. A probability level of <0.05 was established for statistical significance for all analyses. Data were analyzed using SPSS 16.0J for Macintosh (SPSS).

Results

Study 1

Step 1

According to the VAS participant rating, it was possible to situate the 26 Japanese PPI response choice options on an interval scale. The four response choices, with mutual

equal distances, were selected, and the Japanese PPI was constructed. It consisted of the following Japanese words: "itaminashi" (0), "sukoshi" (1), "kininaru" (2), "yakkaina" (3), "hidoi" (4), and "taerarenai" (5). Similarly, the Japanese intensity scale was constructed. It consisted of the following Japanese words: "nashi" (0), "keido" (1), "chuteido" (2), and "judo" (3). The right side of Figure 2 shows the scale values of response choices from the Japanese intensity scale.

Steps 2–5

The fifteen pain descriptors and endpoints of VAS were forward-translated into Japanese in Step 2. Then, the forward-translations judged as irrelevant by the assessors were modified in step 3. In step 4, the author of the SF-MPQ approved the linguistic equivalence between the back-translations and the original version. There were no items with linguistic discrepancy between them.

In step 5, "hot-burning" and "gnawing" were low response rate descriptors. Because "hot-burning" was used by 30% of toothache patients and 28% of musculoskeletal pain patients, it was not determined to be problematic. However, "gnawing" was expected to be used by 33% or more patients, particularly those with rheumatoid arthritis [35]; however, it was not used by any of the patients. Therefore, the translation of "gnawing" was revised and pilot-tested on the same group of patients. The revised translation of "gnawing" was used by 55% of the rheumatoid arthritis patients. The revision was back-translated into English and reviewed and approved by the SF-MPQ author.

Study 2

Descriptive Data, Reliability, and Subscale Intercorrelations of SF-MPQ-J

Table 1 summarizes descriptive statistics and reliability statistics for each score. The SF-MPQ-J sensory subscale, total scale, VAS, and PPI resulted in low or no floor or ceiling effects. A high floor effect was detected only in the affective subscale. All Cronbach's α coefficients were satisfactory and ranged from 0.78 to 0.87. Test–retest reliability coefficients (ICC 1.1) were also satisfactory and ranged from 0.78 to 0.91. No significant differences were found for any of the socio-demographic or pain-related variables between patients who participated in the retest and those who did not. t-Tests performed on sensory, affective, and total scores resulted in no significant score shifts during the retest interval. VAS and PPI scores significantly decreased during the retest interval (VAS; $P < 0.05$, PPI; $P < 0.05$). Intercorrelations for subscales were high (Table 2), and a significantly high correlation was determined between sensory and affective subscales ($P < 0.001$).

Construct Validity of SF-MPQ-J

Table 3 shows patterns of associations that support SF-MPQ-J construct validity. SF-MPQ-J total and

Table 1 Descriptive data and reliability coefficients of SF-MPQ-J

Subscale	Mean (SD)	Floor	Ceiling	Skewness	Kurtosis	Range	α	Test–Retest ICC (N = 13)
Sensory	10.08 (6.87)	1.6%	0.0%	0.914	0.129	0–28	0.80	0.91
Affective	3.47 (3.46)	29.4%	2.4%	0.805	–0.403	0–12	0.78	0.78
Total	13.56 (9.69)	1.6%	0.0%	0.830	–0.164	0–40	0.87	0.91
VAS	58.46 (27.40)	2.4%	5.6%	–0.358	–0.720	0–100		0.80
PPI	3.35 (1.09)	1.6%	17.5%	–0.320	0.303	0–5		0.85

Test–retest interval = 1 week.
ICC = intraclass correlation coefficient; PPI = present pain index; SF-MPQ-J = Short-Form McGill Pain Questionnaire Japanese version; VAS = visual analog scale.
Note: "floor" and "ceiling" refer to the lowest and highest possible SF-MPQ-J scores, respectively.

subscale scores significantly correlated with LF-MPQ-J total and subscale scores, as well as NRS. Total and subscale scores of SF-MPQ-J main components correlated with the corresponding LF-MPQ-J scores, which demonstrated convergent validity. For example, the SF-MPQ-J sensory subscale significantly correlated with the LF-MPQ-J sensory subscale, and the SF-MPQ-J affective subscale and total score exhibited similar associations with corresponding LF-MPQ-J scores. Evidence for the discriminant validity of the SF-MPQ-J sensory subscale was provided by a somewhat greater correlation

between SF-MPQ-J and LF-MPQ-J sensory subscales, compared with the correlation between SF-MPQ-J sensory and LF-MPQ-J affective subscales. Similarly, the SF-MPQ-J affective subscale exhibited a greater correlation with the LF-MPQ-J affective subscale than with the LF-MPQ-J sensory subscale, providing discriminant validity for this subscale. VAS and PPI exhibited a greater correlation with NRS (which reflects overall pain intensity as single dimension) than with total and subscale scores of LF-MPQ-J (which reflects multidimensional pain experience). These results supported the convergent and discriminant validity of these subscales.

Table 2 Intercorrelations of SF-MPQ-J scales

	Sensory	Affective	Total	VAS
Affective	0.72**			
Total	0.97**	0.86**		
VAS	0.50**	0.31**	0.48**	
PPI	0.59**	0.46**	0.59**	0.72**

** P < 0.01.
PPI = present pain index; SF-MPQ-J = Short-Form McGill Pain Questionnaire Japanese version; VAS = visual analog scale.

Table 3 Correlations between SF-MPQ-J, LF-MPQ, and NRS

SF-MPQ-J Score	LF-MPQ Sensory	LF-MPQ Affective	LF-MPQ Total	NRS
Sensory	0.61**	0.54**	0.67**	0.38**
Affective	0.43**	0.61**	0.57**	0.18*
Total	0.58**	0.59**	0.68**	0.40**
VAS	0.38**	0.34**	0.43**	0.76**
PPI	0.40**	0.48**	0.48**	0.57**

* P < 0.05; ** P < 0.01.
LF-MPQ = Long-Form McGill Pain Questionnaire; NRS = numeric rating scale; PPI = present pain index; SF-MPQ-J = Short-Form McGill Pain Questionnaire Japanese version; VAS = visual analog scale.

Discussion

Study 1

In study 1, SF-MPQ-J was developed by a rigorous translation method based on the IQOLA [19] and AAOS [16] guidelines and was considered to be equivalent to the original questionnaire. Equivalence included linguistics and item-level measurement equivalence, such as replication of the original response scaling values and response rate. Many guidelines exist for the translation and cultural adaptation of questionnaires [22]. For study 1, the IQOLA [19] and AAOS [16] guidelines were adopted because they contain a full spectrum of methods for achieving cross-cultural equivalence and replicating original response scaling values through a method employing equal-appearing intervals [26].

The linguistic equivalence of SF-MPQ-J was achieved by rigorous translation techniques using back-translation method by four independent professional translators in steps 1, 2, and 4. Many guidelines [12,18,19,21] have recommended the production of at least two forward-translations because a higher quality can be reached when the translation is performed by at least two independent translators [21]. In general, the consensus is that the qualification of translators is crucial and competent translators should be used [21].

In step 3, linguistic equivalence was ensured by assessment and adjustment by two bilingual assessors. Several

guidelines [16,19,21] have recommended the use of assessors for improving translation quality. In the previous studies on the translated SF-MPQ, assessors participated only in the development of the Iranian [5], Korean [7], and Norwegian [8] SF-MPQ, but were not used for other languages versions [4,6,10,11].

Measurement equivalence at the item level was achieved using two methods. First, response scales equivalent to the original scales were constructed by variation of the method of equal-appearing intervals in step 1. Response scaling values exhibited scalar equivalence to the original values (Figure 2). Results showed that only SF-MPQ-J had response scales with scalar equivalence to the original, compared with other translated versions.

Subsequently, response rates of SF-MPQ-J descriptors were verified in step 5. Almost all Japanese descriptors, with the exception of one, were used by more than 33% of the participants from one diagnostic group, which was similar to the original descriptors. These results demonstrated that the Japanese descriptors resulted in patient responses similar to the original SF-MPQ, thereby confirming measurement equivalence between the pain descriptors in SF-MPQ-J and the original SF-MPQ. The descriptor response rates were also analyzed in the Greek, Norwegian, and Swedish versions [4,8,9], but were not addressed in the remaining versions [6,7,10,11].

Study 2

Results from study 2 demonstrated that SF-MPQ-J has reliability and validity comparable to the original version; thus, the SF-MPQ-J exhibits measurement equivalence to the original English version.

Study 2 demonstrated descriptive statistics and the validity and reliability of SF-MPQ-J. Total scores exhibited positive skewness and negative kurtosis (Table 1), which is consistent with the original SF-MPQ [36,37]. These results show equivalence between the Japanese and original distributions of the SF-MPQ-J total score.

Reliability analysis revealed satisfactory reliability with SF-MPQ-J, which retained the internal consistency and test–retest reliability of the original. The internal consistency of SF-MPQ-J is supported by an adequate Cronbach's α ranging from 0.78 to 0.87 for the subscales and total score. These results were comparable to the original (ranging from 0.70 to 0.90) [36–41], as well as another Japanese version (α of total score = 0.84) [6] and other translated versions (ranging from 0.71 to 0.89) [4,7–11]. The main components of SF-MPQ-J exhibited good test–retest reliability (ICC = 0.78–0.91), although ICC was somewhat less than the original (ranging from 0.88 to 0.96) [42]. In addition, the scores were comparable to the Greek, Turkish, and Norwegian versions [8,11,43]. The reduced reliability of SF-MPQ-J may be the result of sample differences. Although our participants were mixed chronic pain patients, participants from the test–retest reliability study of the original English version were

osteoarthritis patients [42]. The Japanese VAS resulted in good reliability (ICC = 0.80), which is comparable to the English VAS (ranging from 0.69 to 0.85) [44–46] and other translated VAS results (ranging from 0.38 to 0.97) [11,43,44]. The Japanese PPI resulted in better test–retest reliability than the original (ICC = 0.54) [44] but comparable reliability to the Greek, Turkish, and Spanish PPIs [11,43,44]. The lower reliability of the original PPI might be the result of sample differences. Although our participants were chronic pain patients, participants from the Escalante et al. study [44] included patients with acute pain, such as traumatic injury.

The original SF-MPQ subscales were highly correlated with each other (0.32–0.88) [2,39]. Table 2 demonstrates that SF-MPQ-J retained the high intercorrelations of the subscales in the original. These results were consistent with other translated versions [4,5,8].

Results from the construct validity analysis replicated the correlation patterns between the SF-MPQ and LF-MPQ scores of the English-speaking samples [2,31], which provides evidence of convergent, discriminant, and construct validity (Table 3). In addition, there was a significant correlation between VAS and NRS, providing evidence for the convergent validity of VAS. Previous studies [47] have reported a similar correlation between VAS and NRS in English-speaking samples, which suggests that SF-MPQ-J performs in a manner similar to the original, as well as measurement equivalence between the Japanese and original versions. Furthermore, the present results confirmed the discriminant validity of the subscales. To the best of our knowledge, no published study of English-speaking samples has evaluated the discriminant validity of the SF-MPQ subscales [2,31], with the exception of SF-MPQ-2 [48]. Additionally, in other translated SF-MPQs, the convergent and discriminant validity of the subscales were not confirmed [3,4,6–11].

The results from the present study, however, demonstrate somewhat weaker associations between the SF-MPQ-J and their corresponding LF-MPQ-J scores, compared with previous studies [2,31]. Correlation coefficients between SF-MPQ and LF-MPQ have been shown to range from 0.52 to 0.93 in English-speaking samples [2,31]. However, the corresponding correlations ranged from 0.43 to 0.68 in the present samples. This discrepancy might be due to translation differences between SF-MPQ-J and LF-MPQ-J. Because all of the pain descriptors in the original SF-MPQ originated from LF-MPQ, SF-MPQ shares common descriptors with LF-MPQ. Therefore, SF-MPQ might highly correlate with LF-MPQ. However, SF-MPQ-J does not share common Japanese descriptors with LF-MPQ-J. Because the development of LF-MPQ-J was not based on translation guidelines, the descriptors were directly translated from the original SF-MPQ, which resulted in two different types of Japanese translations based on single pain descriptors in English. For example, "sharp" in SF-MPQ-J was translated as "surudoi," whereas the translation for "sharp" in LF-MPQ-J is "suppatokiruyouna." Many other descriptor

translations from SF-MPQ-J are different from LF-MPQ-J, which may have reduced the correlation coefficients between these questionnaires.

A strength of SF-MPQ-J is its cross-cultural equivalence to the original version, including scalar equivalence of the response scales and evidence of adequate validity and reliability. In particular, of all the translated versions only the SF-MPQ-J has response scales with scalar equivalence. Moreover, SF-MPQ-J results demonstrated the convergence and discriminative validity of SF-MPQ subscales, in contrast to other translated versions.

Limitations

Several limitations of the SF-MPQ-J exist. First, the use of 33% as the acceptable response rate of pain descriptor used in study 1 is not a gold standard. Second, samples from study 2 consisted of mixed chronic pain. The original questionnaire is valid and reliable for acute and cancer pain samples [2,31]. However, it is possible that SF-MPQ-J is not as valid, reliable, and cross-culturally equivalent as the original in those samples. Third, the treatment response of the patients who were tested with this instrument was not analyzed. Future examination of SF-MPQ-J and treatment response is needed to enable use of this instrument in treatment outcome studies. Fourth, a confirmatory factor analysis is needed for further establishment of construct validity. Fifth, differential item functioning (DIF) [12,49] was not examined for testing item-level measurement equivalence. Some SF-MPQ items might function differently in Japanese than in English. Zinke et al. [41] confirmed a cross-cultural equivalence for SF-MPQ by investigating DIF. However, future studies focused on DIF will help confirm the cross-cultural functional equivalence of SF-MPQ-J. Finally, the number of participants in the test–retest reliability analysis was small and the findings will need further confirmation with larger samples.

Conclusions

This study suggests that SF-MPQ-J developed in our study is reliable, valid, and cross-culturally equivalent to the original SF-MPQ questionnaire. Researchers might consider using this scale in multicenter, multinational trials, which have recently become increasingly important, or in cross-cultural studies that include Japanese-speaking patients.

Acknowledgments

References

1 Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. Pain 1975;1:277–99.

2 Melzack R. The short-form McGill Pain Questionnaire. Pain 1987;30:191–7.

3 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;129:1611–4.

4 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;4:275–81.

5 Najafi-Ghezeljeh T, Ekman I, Nikravesh MY, Emami A. Adaptation and validation of the Iranian version of Angina Pectoris characteristics questionnaire. Int J Nurs Pract 2008;14:470–6.

6 Yamaguchi M, Kumano H, Yamauchi Y, Kadota Y, Iseki M. The development of a Japanese version of the short-form McGill Pain Questionnaire. JJSPC 2007;14:9–14.

7 Lee H, Nicholson LL, Adams RD, et al. Development and psychometric testing of Korean language versions of 4 neck pain and disability questionnaires. Spine (Phila Pa 1976) 2006;31:1841–5.

8 Ljunggren AE, Strand LI, Johnsen TB. Development of the Norwegian short-form McGill Pain Questionnaire (NSF-MPQ). Adv Physiother 2007;9:169–80.

9 Burckhardt CS, Bjelle A. A Swedish version of the short-form McGill Pain Questionnaire. Scand J Rheumatol 1994;23:77–81.

10 Kitisomprayoonkul W, Klaphajone J, Kovindha A. Thai short-form McGill Pain Questionnaire. J Med Assoc Thai 2006;89:846–53.

11 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.

12 Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. Qual Life Res 2002;11:193–205.

13 Wild D, Eremenco S, Mear I, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: The ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. Value Health 2009;12:430–40.

14 Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: The universalist approach. Qual Life Res 1998;7:323–35.

15 Turk DC, Dworkin RH, Burke LB, et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. Pain 2006;125: 208–15.

16 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;25:3186–91.

17 Hui CH, Triandis HC. Measurement in cross-cultural psychology—A review and comparison of strategies. J Cross Cult Psychol 1985;16:131–52.

18 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.

19 Bullinger M, Alonso J, Apolone G, et al. Translating health status questionnaires and evaluating their quality: The IQOLA Project approach. International quality of life assessment. J Clin Epidemiol 1998;51: 913–23.

20 Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. Value Health 2005;8:94–104.

21 Peters M, Passchier J. Translating instruments for cross-cultural studies in headache research. Headache 2006;46:82–91.

22 Acquadro C, Conway K, Hareendran A, Aaronson N. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. Value Health 2008;11:509–21.

23 Melzack R, Torgerson WS. On the language of pain. Anesthesiology 1971;34:50–9.

24 Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, eds. Handbook of Pain Assessment. New York: The Guilford Press; 2001:15–34.

25 Merskey H, Bogduk N. Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms, 2nd edition. Seattle, WA: IASP Press; 1994.

26 Keller SD, Ware JE, Jr, Gandek B, et al. Testing the equivalence of translations of widely used response choice labels: Results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:933–44.

27 Thurstone LL. Fechner's law and the method of equal-appearing intervals. J Exp Psychol 1929;12:214–24.

28 Tursky B, Jamner LD, Friedman R. The pain perception profile—A psychophysical approach to the assessment of pain report. Behav Ther 1982;13:376–94.

29 Urban BJ, Keefe FJ, France RD. A study of psychophysical scaling in chronic pain patients. Pain 1984;20:157–68.

30 Hasegawa M, Mishima M, Matsumoto I, et al. Confirming the theoretical structure of the Japanese version of the McGill Pain Questionnaire in chronic pain. Pain Med 2001;2:52–9.

31 Dudgeon D, Raubertas RF, Rosenthal SN. The short-form McGill Pain Questionnaire in chronic cancer pain. J Pain Symptom Manage 1993;8:191–5.

32 Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.

33 Nunnally JC, Bernstein IH. Psychometric Theory, 3rd edition. New York: McGraw-Hill; 1994.

34 Jensen MP. Questionnaire validation: A brief guide for readers of the research literature. Clin J Pain 2003;19:345–52.

35 Dubuisson D, Melzack M. Classification of clinical pain description by multiple group discriminant analysis. Exp Neurol 1976;51:480–7.

36 Wright KD, Asmundson GJ, McCreary DR. Factorial validity of the short-form McGill Pain Questionnaire (SF-MPQ). Eur J Pain 2001;5:279–84.

37 Mason ST, Arceneaux LL, Abouhassan W, et al. Confirmatory factor analysis of the short Form McGill Pain Questionnaire with burn patients. Eplasty 2008;8:e54.

38 Cook AJ, Roberts DA, Henderson MD, et al. Electronic pain questionnaires: A randomized, crossover comparison with paper questionnaires for chronic pain assessment. Pain 2004;110:310–7.

39 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.

40 Truong PT, Abnousi F, Yong CM, et al. Standardized assessment of breast cancer surgical scars integrating the Vancouver Scar Scale, Short-Form McGill Pain Questionnaire, and patients' perspectives. Plast Reconstr Surg 2005;116:1291–9.

41 Zinke JL, Lam CS, Harden RN, Fogg L. Examining the cross-cultural validity of the english short-form McGill Pain Questionnaire using the matched moderated regression methodology. Clin J Pain 2010;26:153–62.

42 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.

43 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;5:109–18.

44 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.

45 Chang VT, Hwang SS, Feuerman M. Validation of the Edmonton symptom assessment scale. Cancer 2000;88:2164–71.

46 Fischer D, Stewart AL, Bloch DA, et al. Capturing the patient's view of change as a clinical outcome measure. JAMA 1999;282:1157–62.

47 Bijur PE, Latimer CT, Gallagher EJ. Validation of a verbally administered numerical rating scale of acute pain for use in the emergency department. Acad Emerg Med 2003;10:390–2.

48 Dworkin RH, Turk DC, Revicki DA, et al. Development and initial validation of an expanded and revised version of the Short-Form McGill Pain Questionnaire (SF-MPQ-2). Pain 2009;144:35–42.

49 Teresi JA, Fleishman JA. Differential item functioning and health assessment. Qual Life Res 2007; 16(suppl 1):33–42.