SPINE Volume 25, Number 24, pp 3186–3191
©2000, Lippincott Williams & Wilkins, Inc.

Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures

Dorcas E. Beaton, BScOT, MSc, PhD,*†‡§ Claire Bombardier, MD, FRCP,*§÷¶#
Francis Guillemin, MD, MSc,** and Marcos Bosi Ferraz, MD, MSc, PhD††

With the increase in the number of multinational and multicultural research projects, the need to adapt health status measures for use in other than the source language has also grown rapidly.[1,4,27] Most questionnaires were developed in English-speaking countries,[11] but even within these countries, researchers must consider immigrant populations in studies of health, especially when their exclusion could lead to a systematic bias in studies of health care utilization or quality of life.[9,11]

The cross-cultural adaptation of a health status self-administered questionnaire for use in a new country, culture, and/or language necessitates use of a unique method, to reach equivalence between the original source and target versions of the questionnaire. It is now recognized that if measures are to be used across cultures, the items must not only be translated well linguistically, but also must be adapted culturally to maintain the content validity of the instrument at a conceptual level across different cultures.[6,11–13,15,24] Attention to this level of detail allows increased confidence that the impact of a disease or its treatment is described in a similar manner in multinational trials or outcome evaluations. The term "cross-cultural adaptation" is used to encompass a process that looks at both language (translation) and cultural adaptation issues in the process of preparing a questionnaire for use in another setting.

Cross-cultural adaptations should be considered for several different scenarios. In some cases, this is more obvious than in others. Guillemin et al[11] suggest five different examples of when attention should be paid to this adaptation by comparing the target (where it is going to be used) and source (where it was developed) language and culture. The first scenario is that it is to be used in the same language and culture in which it was developed. No adaptation is necessary. The last scenario is the opposite extreme, the application of a questionnaire in a different culture, language and country—moving the Short Form 36-item questionnaire from the United States (source) to Japan (target)[7] which would necessitate translation and cultural adaptation. The other scenarios are summarized in Table 1 and reflect situations when some translation and/or adaptation is needed.

The guidelines described in this document are based on a review of cross-cultural adaptation in the medical, sociological, and psychological literature. This review led to the description of a thorough adaptation process designed to maximize the attainment of semantic, idiomatic, experiential, and conceptual equivalence between the source and target questionnaires.[13]. Further experience in cross-cultural adaptation of generic and disease-specific instruments and alternative strategies driven by different research groups[18] have led to some refinements in methodology since the 1993 publication.[11].

These guidelines serve as a template for the translation and cultural adaptation process. The process involves the adaptation of individual items, the instructions for the questionnaire, and the response options. The text in the next section outlines the methodology suggested (Stages I–V). The subsequent section (Stage VI) presents a suggested appraisal process whereby an advisory committee or the developers review the process and determine whether this is an acceptable translation. Although such a committee or the developers may not be engaged in tracking translated versions of the instrument, this stage has been included in case there is a tracking system. Records of translated versions not only can save considerable time and effort (by using already available questionnaires) but also avoid erroneous comparisons of results across different translated versions.

The process of cross-cultural adaptation tries to produce equivalency between source and target based on content. The assumption that is sometimes made is that this process will ensure retention of psychometric properties such as validity and reliability at an item and/or a scale level. However, this is not necessarily the case: For instance, if the new culture has a different way of approaching a task that makes it inherently more or less difficult compared with other items, it would change the validity, certainly in terms of item-level analyses (such as item response theory, similar to Rasch). Further tests should be conducted on the psychometric properties of the adapted questionnaire after the translation is complete.[3,10,26,20] This will be discussed briefly at the end of the guidelines. In fact, the translation process outlined in this article is the first step in the three-step process adopted by the International Society for Quality of Life Assessment (IQOLA) project.[8,25,26] The other two steps

From the *Institute for Work and Health; †St. Michael's Hospital; Departments of ‡Occupational Therapy and ÷Medicine and the §Clinical Epidemiology and Health Care Research Program, University of Toronto ¶The University Health Network, Toronto General Hospital; and #Mt. Sinai Hospital, Toronto, Ontario, Canada; **Ecole de Sant, Publique, Facult, de M,decine,Vandoeuvre-les-Nancy, France; and the ††Division of Rheumatology, Department of Medicine, Escola Paulista de Medicina, Universidade Federal de Sôo Paulo, Sôo Paulo, Brazil.
Supported in part by the American Academy of Orthopaedic Surgeons and the Institute for Work & Health who co-sponsored the development of these guidelines. DB was supported by a Medical Research Council of Canada PhD Fellowship during the preparation of this work.

Table 1. Possible Scenarios Where Some Form of Cross-Cultural Adaptation is Required

Wanting to use a questionnaire in a new population described as follows:		Results in a Change in . . .			Adaptation Required	
		Culture	Language	Country of Use	Translation	Cultural Adaptation
A	Use in same population. No change in culture, language, or country from source	—	—	—	—	—
B	Use in established immigrants in source country	✔	—	—	—	✔
C	Use in other country, same language	✔	—	✔	—	✔
D	Use in new immigrants, not English-speaking, but in same source country	✔	✔	—	✔	✔
E	Use in another country and another language	✔	✔	✔	✔	✔

Adapted from Guillemin et al.[4]

are, first, verification of the scaling requirements (item performance, item weights) and, second, the validation of and establishing normative values for the new version.

■ Guidelines for the Cross-Cultural Adaptation Process

Figure 1 outlines the cross-cultural adaptation process being recommended. It is the method currently used by the American Association of Orthopaedic Surgeons (AAOS) Outcomes Committee as they coordinate the translation of the different components of their outcomes battery. The written documentation of each step helps to record that it was performed but can also serve as a memory aid at later stages. For instance, if an item is not working in the field testing, there will be a record showing whether the translators had difficulty with that item, and how they resolved it. Sample forms have been designed for one questionnaire[19] so that the worksheets used for the translation can formulate the written report

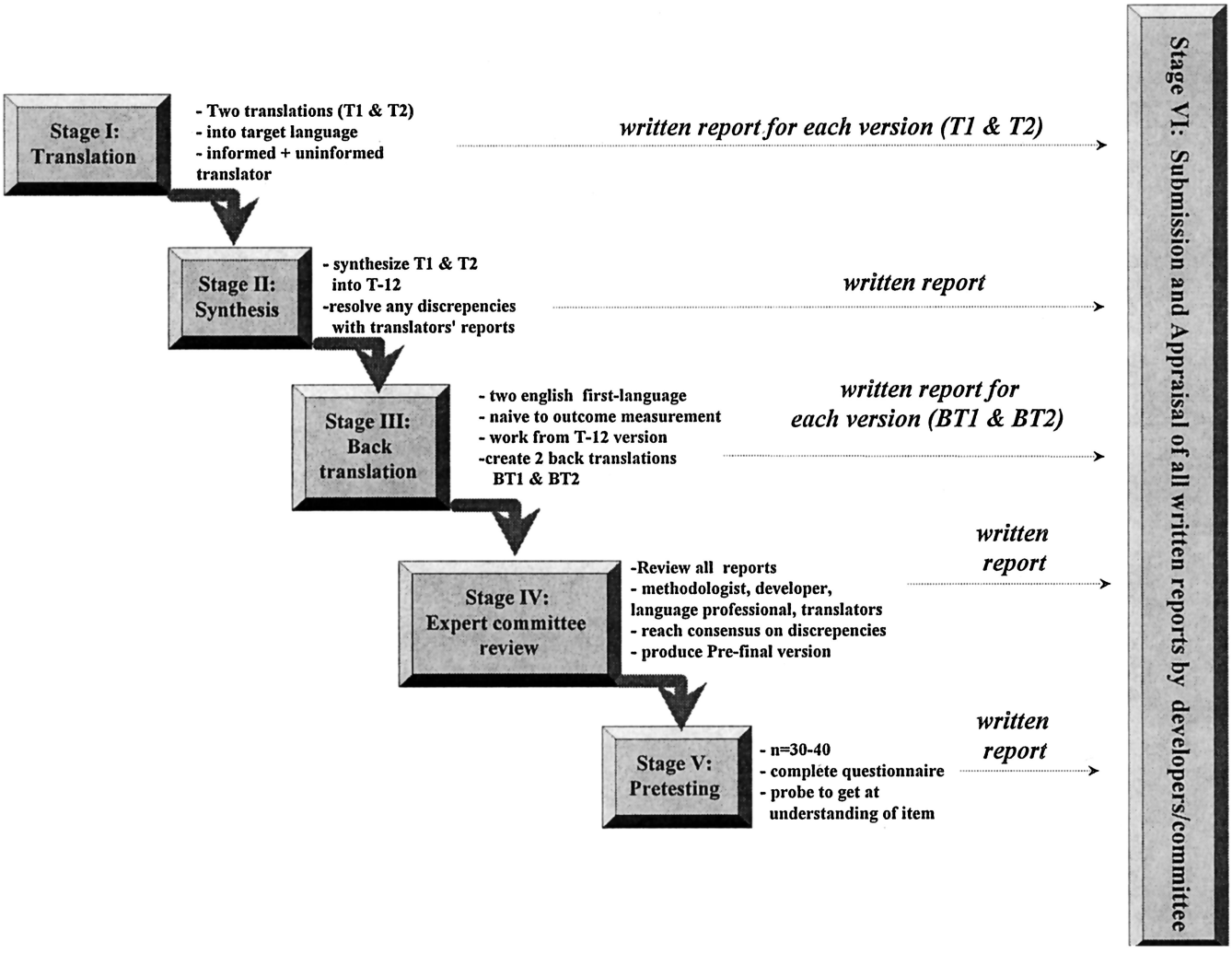

Figure 1. Graphic representation of the stages of cross-cultural adaptation recommended.

as well. The forms are available through the authors, or through the AAOS. Each stage in the recommended protocol is described in detail in the following sections.

Stage I: Initial Translation

The first stage in adaptation is the forward translation. Many recommend that at least two forward translations be made of the instrument from the original language (source language) to the target language. In this way, the translations can be compared and discrepancies that may reflect more ambiguous wording in the original or discrepancies in the translation process noted. Poorer wording choices are identified and resolved in a discussion between the translators.

Bilingual translators whose mother tongue is the target language produce the two independent translations. Translations into the mother tongue, or first language, more accurately reflect the nuances of the language.[13] The translators each produce a written report of the translation that they complete. Additional comments are made to highlight challenging phrases or uncertainties. Their rationale for their choices is also summarized in the written report. Item content, response options, and instructions are all translated in this way.

The two translators should have different profiles, or backgrounds.

Translator 1. One of the translators should be aware of the concepts being examined in the questionnaire being translated (functional disability or neck and shoulder disorders). Their adaptations are intended to provided equivalency from a more clinical perspective and may produce a translation providing a more reliable equivalence from a measurement perspective.

Translator 2. The other translator should neither be aware nor informed of the concepts being quantified and preferably should have no medical or clinical background. This is called a naive translator, and he or she is more likely to detect different meaning of the original than the first translator. This translator will be less influenced by an academic goal and will offer a translation that reflects the language used by that population, often highlighting ambiguous meanings in the original questionnaire.[11]

Stage II: Synthesis of The Translations

The two translators and a recording observer sit down to synthesize the results of the translations. Working from the original questionnaire as well as the first translator's (T1) and the second translator's (T2) versions, a synthesis of these translations is first conducted (producing one common translation T-12), with a written report carefully documenting the synthesis process, each of the issues addressed, and how they were resolved. It is important that consensus rather than one person's compromising her or his feelings resolve issues. The next stage is completed with this T-12 version of the questionnaire.

Stage III: Back Translation

Working from the T-12 version of the questionnaire and totally blind to the original version, a translator then translates the questionnaire back into the original language. This is a process of validity checking to make sure that the translated version is reflecting the same item content as the original versions. This step often magnifies unclear wording in the translations. However, agreement between the back translation and the original source version does not guarantee a satisfactory forward translation, because it could be incorrect; it simply assures a consistent translation.[18] Back translation is only one type of validity check, highlighting gross inconsistencies or conceptual errors in the translation.

Once again, two of these back-translations are considered a minimum. The back-translations (BT1 and BT2) are produced by two persons with the source language (English) as their mother tongue. The two translators should neither be aware nor be informed of the concepts explored, and should preferably be without medical background. The main reasons are to avoid information bias and to elicit unexpected meanings of the items in the translated questionnaire (T-12),[11,18] thus increasing the likelihood of "highlighting the imperfections."[18]

Stage IV: Expert Committee

The composition of this committee is crucial to achievement of cross-cultural equivalence. The minimum composition comprises methodologists, health professionals, language professionals, and the translators (forward and back translators) involved in the process up to this point. The original developers of the questionnaire are in close contact with the expert committee during this part of the process.

The expert committee's role is to consolidate all the versions of the questionnaire and develop what would be considered the prefinal version of the questionnaire for field testing. The committee will therefore review all the translations and reach a consensus on any discrepancy. The material at the disposal of the committee includes the original questionnaire, and each translation (T1, T2, T12, BT1, BT2) together with corresponding written reports (which explain the rationale of each decision at earlier stages).

The expert committee is making critical decisions so, again, full written documentation should be made of the issues and the rationale for coming to a decision about them.

Decisions will need to be made by this committee to achieve equivalence between the source and target version in four areas[11]:

Semantic Equivalence. Do the words mean the same thing? Are their multiple meanings to a given item? Are there grammatical difficulties in the translation?

Idiomatic Equivalence. Colloquialisms, or idioms, are difficult to translate. The committee may have to formu-

late an equivalent expression in the target version. For example the term "feeling downhearted and blue" from the SF-36 has often been difficult to translate, and an item with similar meaning would have to be found by the committee.

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. An example might be in an item worded: Do you have difficulty eating with a fork? when that was not the utensil used for eating in the target country.

Conceptual Equivalence. Often words hold different conceptual meaning between cultures (for instance the meaning of "seeing your family as much as you would like" would differ between cultures with different concepts of what defines "family"—nuclear versus extended family).

The committee must examine the source and back-translated questionnaires for all such equivalences. Consensus should be reached on the items, and if necessary, the translation and back-translation processes should be repeated to clarify how another wording of an item would work. The advantage of having all translators present on the committee is obvious, because tasks such as that could be undertaken immediately. Items, instructions, and response options[7,16] must be considered. The translators should also make sure that the final questionnaire would be understood by the equivalent of a 12-year-old (roughly a Grade 6 level of reading), as is the general recommendation for questionnaires.

Stage V: Test of the Prefinal Version

The final stage of adaptation process is the pretest. This field test of the new questionnaire seeks to use the prefinal version in subjects or patients from the target setting. Ideally, between 30 and 40 persons should be tested.

Each subject completes the questionnaire, and is interviewed to probe about what he or she thought was meant by each questionnaire item and the chosen response. Both the meaning of the items and responses would be explored. This ensures that the adapted version is still retaining its equivalence in an applied situation. The distribution of responses is examined to look for a high proportion of missing items or single responses.

It should be noted that although this stage provides some useful insight into how the person interprets the items on the questionnaire, it does not address the construct validity, reliability, or item response patterns that are also critical to describing a successful cross-cultural adaptation. The described process provides for some measure of quality in the content validity. Additional testing for the retention of the psychometric properties of the questionnaire is highly recommended and will be discussed briefly later.

Stage VI: Submission of Documentation to the Developers or Coordinating Committee for Appraisal of the Adaptation Process

The final stage in the adaptation process is a submission of all the reports and forms to the developer of the instrument or the committee keeping track of the translated version. They in turn probably have a means to verify that the recommended stages were followed, and the reports seem to be reflecting this process well. In effect it is a process audit, with all the steps followed and necessary reports followed. It is not up to this body or committee to alter the content, it is assumed that by following this process a reasonable translation has been achieved.

Further Testing of the Adapted Version The goal of this article was to outline the process of translation and adaptation of self-report measures of health. Cross-cultural adaptation tries to ensure a consistency in the content and face validity between source and target versions of a questionnaire. It should therefore follow that the resultant version has sound reliability and validity if the original version did. However, this is not always the case, perhaps because of subtle differences in the living habits in different cultures that render that item more or less difficult than other items in the questionnaire.[3,20] Such changes could alter the statistical or psychometric properties of an instrument.

It is highly recommended that, after the translation and adaptation process, the investigators ensure that the new version has demonstrated the measurement properties needed for the intended application.[2,8,25] The new instrument should retain both the item-level characteristics such as item-to-scale correlations and internal consistency; and the score-level characteristics of reliability, construct validity, and responsiveness. It is possible to work some of these tests of reliability and validity into the pretesting process (stage V of the adaptation), although often they need larger sample sizes.

There are many examples of ways in which translated questionnaires have been tested for their psychometric comparability with the source version. Many are published in a special issue of the *Journal of Clinical Epidemiology* (1998, volume 51 number 11) dealing with the IQOLA project.[4,26] Items are checked for the distribution of responses to them, and the correlation of each with its scale and not with other scale (if there is more than one dimension in the instrument). Ware[25,26] suggests item response theory could play a role in verifying the calibration or location of each item on the underlying attribute of health. It would be anticipated that similar calibrations, and item–total correlations would be found in a well-translated item. The final step is a full assessment of the score level attributes: construct validity, reliability, and responsiveness. Comparisons of these tests are made against similar tests preformed in the original setting using the original instrument. It is expected that the adapted version would perform in a similar manner.

For instance, correlations with other measures of overall health or comparisons between groups known to differ in their health should result in similar values in each culture (with the use of the appropriate version of the instrument). In this way, there is more confidence that the adapted instrument is measuring a construct comparable to the original.[8]

Several examples from the IQOLA project demonstrate the different types of adaptation that are needed—for instance the translation of the SF-36 for use in China[17]—and then a retest of the translated version to see whether it is suited for Chinese-speaking persons living in the United States[21] (a situation in which cultural adaptation could have been necessary). Another large project was the testing of the human immunodeficiency virus (HIV) version of the Medical Outcome Study short form so that it could be used in five different cultures and languages.[23] Other examples show the need for adaptation even between English-language versions of the questionnaires used in the United Kingdom and the United States.[5]

Two examples of adaptations were found for self-report measures of low back pain. First, the Roland–Morris questionnaire was translated into German, using techniques similar to those described in this article.[27] Second, Schoppink et al[22] described the psychometric testing of the Dutch version of the Quebec Back Pain Questionnaire.

Any of these papers could serve as examples of the types of testing that must be carried out in addition to the translation process. The final step would be determining normative data on relevant populations using the new instrument.

■ Discussion

The authors' best understanding to date is that a poor translation process may lead to an instrument that is not equivalent[11,14] to the original questionnaire. The lack of equivalence limits the comparability of responses across populations divided by language or by culture. In this article, a guideline for the process of adapting a questionnaire for use in a different setting has been presented (Table 1). The need has also been acknowledged for psychometric testing and normative data collection using the new instrument. The authors' choice was to separate the adaptation from the testing, because the need for additional testing is the same as would have to be undertaken after any adaptation of any existing questionnaire whether it be shortening it or performing a cross-cultural adaptation. The authors concur with the IQOLA group recommendations for formal testing of the final instrument.[8,25]

The process described in this article is a process of translating and, if necessary, replacing items or scaling to make it relevant and valid in a new culture. Herdmann et al[14] remind researchers to be aware that item-level translations can often work on the assumption that the same items, once translated, will be meaningful reflections of health in a new culture. The authors' method allows for that adaptability, but it is useful to remember that attention must be paid not only to item equivalence, but to the others described as well. The key learning point is that the translation does not automatically provide a valid measure of another culture's health, and this should be verified carefully throughout the process, and in the final testing.[12]

In this article, reference is made to submitting reports to a body such as the AAOS who are tracking the adaptation of Modems instruments. The same procedure should be followed whether or not a given instrument has a formal repository for translated versions. If there is such a gathering place, than submission of a report of the adaptation process and the resultant questionnaire will help ensure multiple translations are not in use and more importantly that the extensive amount of work entailed is not needlessly repeated. If there is no repository, efforts should be made to publish the adaptation so that other researchers can be made aware of the available version.

Adaptation of a questionnaire for use in a new setting is time consuming and costly. However, to date the authors believe it is the best way to get an equivalent metric for whatever self-report attribute is being considered. It allows data collection efforts to be the same in cross-national studies or to avoid the selection bias that may be associated with studies that must exclude all patients who were unable to complete a form in English, for example, because there are no translated versions of the questionnaire.

References

1. Anderson RT, Aaronson N, Wilkin D. Critical review of the international assessments of health-related quality of life generic instruments. In: The International Assessment of Health-Related Quality of Life: Theory, Translation, Measurement and Analysis. Oxford, UK: Rapid Communication of Oxford; 1995: 11–37.
2. Anderson RT, Aaronson NK, Wilkin D. Critical review of the international assessments of health-related quality of life. Qual Life Res 1993;2:369–95.
3. Bjorner JB, Kreiner S, Ware JE, et al. Differential item functioning in the Danish translation of the SF-36. J Clin Epidemiol 1998;51:1189–202.
4. 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.
5. Bushnell DM, Martin ML. Quality of life and Parkinson's disease: translation and validation of the US Parkinson's Disease Questionnaire (PDQ-39). Qual Life Res 1999;8:345–50.
6. Ferraz MB. Cross cultural adaptation of questionnaires: what is it and when should it be performed [editorial; comment]? J Rheumatol 1997;24:2066–68.
7. Fukuhara S, Bito S, Green J, et al. Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. J Clin Epidemiol 1998;51:1037–44.
8. Gandek B, Ware JE Jr, IQOLA Group. Methods for validating and norming translations of health status questionnaires: the IQOLA project approach. J Clin Epidemiol 1998;51:953–59.
9. Gonzalez-Calvo J, Gonzalez VM, Lorig K. Cultural diversity issues in the development of valid and reliable measures of health status. Arthritis Care Res 1997;10:448–56.
10. Goycochea-Robles MV, Garduno-Espinosa J, Vilchis-Guizar E, et al. Validation of a Spanish version of the childhood health assessment questionnaire [see comments]. J Rheumatol 1997;24:2242–5.
11. 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.
12. Guyatt GH. The philosophy of health-related quality of life translation. Qual Life Res 1993;2:461–5.
13. Hendricson WD, Russell IJ, Prihoda TJ, et al. Development and initial vali-

dation of a dual-language English-Spanish format for the arthritis impact measurement scales. Arthritis Rheum 1989;32:1153.
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;7:323–335.
15. Herdman M, Fox-Rushby J, Badia X. "Equivalence" and the translation and adaptation of health-related quality of life questionnaires. Qual Life Res. 1997; 6:237–47.
16. Keller S, Ware JE, Gandek B, et al. Testing the equivalence of translations of widely used response choice labels: results from the IQOLA project. J Clin Epidemiol 1998;51:933–44.
17. Lam CL, Gandek B, Ren XS, et al. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. J Clin Epidemiol 1998;51:1139–47.
18. Leplege A, Verdier A. The adaptation of health status measures: a discussion of certain methodological aspects of the translation procedure. In: The International Assessment of Health-Related Quality of Life: Theory, Translation, Measurement and Analysis. Oxford, UK: Rapid Communication of Oxford; 1995: 93–101.
19. McConnell S, Beaton DE, Bombardier C. The DASH Outcome Measure: A User's Manual. Toronto, Ontario: Institute for Work & Health, 1999.
20. Raczek AE, Ware JE, Bjorner JB, et al. Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:1203–14.
21. Ren XS, Amick B III, Zhou L, et al. Translation and psychometric evaluation of a Chinese version of the SF- 36 Health Survey in the United States. J Clin Epidemiol 1998;51:1129–38.
22. Schoppink LE, van Tulder MW, Koes BW, et al. Reliability and validity of the Dutch adaptation of the Quebec Back Pain Disability Scale. Phys Ther 1996; 76:268–75.
23. Scott-Lennox JA, Wu AW, Boyer JG, et al. Reliability and validity of French, German, Italian, Dutch, and UK English translations of the Medical Outcomes Study HIV Health Survey. Med Care 1999;37:908–25.
24. Wagner AK, Gandek B, Aaronson NK, et al. Cross-cultural comparisons of the content of SF-36 translations across 10 countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:925–32.
25. Ware JE Jr, Gandek B, Keller S, IQOLA Group. Evaluating instruments used cross-nationally: Methods from the IQOLA project. In: Spilker B, editor. Quality of Life and Pharmacoeconomics in Clinical Trials. 2nd ed. Philadelphia: Lippincott–Raven Publishiers, 1996:681–692.
26. Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol 1998;51:945–52.
27. Wiesinger GF, Nuhr M, Quittan M, et al. Cross-cultural adaptation of the Roland-Morris questionnaire for German- speaking patients with low back pain. Spine 1999;24:1099–103.

Address reprint requests to

Dorcas Beaton
Institute for Work & Health
250 Bloor St, Ste. 702
Toronto, ON, Canada, M4W 1E6
E-mail: dbeaton@iwh.on.ca