Cross-Cultural Adaptation, Reliability, Internal Consistency and Validation of the Spinal Function Sort (SFS) for French- and German-Speaking Patients with Back Complaints

S. Borloz · M. A. Trippolini · P. Ballabeni ·
F. Luthi · O. Deriaz

Abstract *Introduction* Functional subjective evaluation through questionnaire is fundamental, but not often realized in patients with back complaints, lacking validated tools. The Spinal Function Sort (SFS) was only validated in English. We aimed to translate, adapt and validate the French (SFS-F) and German (SFS-G) versions of the SFS. *Methods* Three hundred and forty-four patients, experiencing various back complaints, were recruited in a French ($n = 87$) and a German-speaking ($n = 257$) center. Construct validity was estimated via correlations with SF-36 physical and mental scales, pain intensity and hospital anxiety and depression scales (HADS). Scale homogeneities were assessed by Cronbach's α. Test–retest reliability was assessed on 65 additional patients using intraclass correlation (IC). *Results* For the French and German translations, respectively, α were 0.98 and 0.98; IC 0.98 (95% CI: [0.97; 1.00]) and 0.94 (0.90; 0.98). Correlations with physical functioning were 0.63 (0.48; 0.74) and 0.67 (0.59; 0.73); with physical summary 0.60 (0.44; 0.72) and 0.52 (0.43; 0.61); with pain −0.33 (−0.51; −0.13) and −0.51 (−0.60; −0.42); with mental health −0.08 (−0.29; 0.14) and 0.25 (0.13; 0.36); with mental summary 0.01 (−0.21; 0.23) and 0.28 (0.16; 0.39); with depression −0.26 (−0.45; −0.05) and −0.42 (−0.52; −0.32); with anxiety −0.17 (−0.37; −0.04) and −0.45 (−0.54; −0.35). *Conclusions* Reliability was excellent for both languages. Convergent validity was good with SF-36 physical scales, moderate with VAS pain. Divergent validity was low with SF-36 mental scales in both translated versions and with HADS for the SFS-F (moderate in SFS-G). Both versions seem to be valid and reliable for evaluating perceived functional capacity in patients with back complaints.

Keywords Spinal function sort (SFS) · Self-administered questionnaire · Outcome assessment · Translation and validation

S. Borloz (✉) · F. Luthi
Département de l'appareil locomoteur, Clinique Romande de réadaptation SUVA Care, Sion, Switzerland
e-mail: stephane.borloz@hcuge.ch

M. A. Trippolini
Department of Work Rehabilitation, Rehaklinik Bellikon, SUVA Care, Bellikon, Switzerland

M. A. Trippolini
Department of Rehabilitation Medicine, Center for Rehabilitation, University Medical Center, Groningen, The Netherlands

P. Ballabeni · F. Luthi · O. Deriaz
Institut de recherche en réadaptation and Service de recherche médicale, Clinique Romande de réadaptation SUVA Care, Sion, Switzerland

P. Ballabeni
Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland

F. Luthi
Département de l'appareil locomoteur, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland

Introduction

The follow-up of patients with musculoskeletal disorders in clinical and research settings is not only based on clinical exams or radiography but also on self-administered

questionnaires which are inexpensive and give insight into the patient's perspective.

In occupational rehabilitation, one important activity is the functional capacity evaluation (FCE) of patients in order to determine readiness or ability for safe return to work following musculoskeletal injury [1]. The patient's self efficacy (SE) level was proposed as a relevant psychosocial factor that may influence FCE. Perceived SE refers to the individual's beliefs about their own competence or ability [2]. SE beliefs may influence the patient's behavior, e.g. the ability to overcome negative experiences. It has been suggested that SE is more closely related to work disability than actual physical abilities [3]. Assessment of SE by self-report therefore plays an important role in predicting health outcome [4–5]. It has also been recommended that patients with low back pain should be assessed with both instruments (i.e. self-report and performance tests) because these strategies may lead to different results [6–8].

Self-administered questionnaires should be developed with accurate and rigorous instruments to ensure that they are specific to the studied concept as well as reliable and responsive (clinimetric qualities) [9–10]. A great variety of questionnaires have been developed to assess the perceived function of patients with back pain. Some of them such as the Oswestry disability index, the Roland Morris disability questionnaire and the Quebec back pain disability scale have been recommended for clinical purposes by an expert panel [11]. The utility of questionnaires in rehabilitation settings is often limited by the literacy level of the patients [12]. One approach to improving the comprehension of the questionnaire by patients with low literacy levels is to inform the patient through pictorial activities and task sorts (PATS) designed for self-assessment of functional ability in occupational rehabilitation such as were developed in the 70s [13].

Recently, efforts have focused on the creation of questionnaires more oriented towards functional limitations and occupational perspectives [14]. However, those picture-based questionnaires are often validated in English only and not for use by non-English speaking patients.

The Spinal Function Sort (SFS), published in English in 1989 [15], has proven to be of advantage in work-related rehabilitation settings [16–20]. It is often used in addition to FCEs to assess the self-perceived functional capacity of patients with back complaints [21]. It is a picture-based generic tool that is useful for all kinds of back disorders. The reliability and validity of the SFS have been reported [22–24] but, to the best of our knowledge, no German or French versions have been properly cross-culturally adapted and translated. The aim of this study was to do a cross-cultural adaptation and validation of the SFS in French and German.

Methods

Spinal Function Sort (SFS)

The French and German translations of the SFS consists, as the original SFS, of a booklet containing drawings (Picture 1 in Electronic supplementary material) with a brief a description of 50 tasks. These tasks are performed by men and women and reflect a wide range of daily living or vocational activities that involve the spine. The pictured activities are graded from light to heavy material handling, so that scores can be compared to the physical demand characteristics from the United States Department of Labor's Dictionary of Occupational Titles [25]. Subjects are asked to answer quickly without spending too much time on any one drawing. They are told that their "first impression is usually the best". There is no time limit to fill out the questionnaire. Subjects rate their ability to perform the task on a 5-point Lickert scale (from "able" to "restricted" to "unable"). An additional category depicted as "?" means "I don't know", for example, for an unfamiliar task. Items are scored from 4 (able) to 0 (unable or "?"). The SFS is scored manually by the assessor and yields a total score, which can range from 0 to 200. This total score corresponds to the level of perceived physical work, ranging from sedentary to very heavy, and can be compared to the Dictionary of Occupational Titles.

Following the scoring instruction of the original SFS, questionnaires with 4 or more "I don't know" responses, were excluded from the present study because of potential bias. Moreover, the SFS has 2 internal validity check drawings with the same questions but different images to test the reliability of subjects (questions #6 and #50; #17 and #49). Subjects who showed inconsistencies greater than 3 points on the 5 point scale were also excluded.

Cross-Cultural Adaptation

The cross-cultural adaptation of the SFS was performed according to the guidelines of the American academy of orthopaedic surgeons (AAOS) outcomes committee [26] and as recommended by others in the literature [27–28]. The following five steps were documented in a written report: (1) Forward translation from English to French and to German by two translators whose native language was French, or German, and fluent in English (T1 and T2). One of the translators was informed about the aims of the study, and the other received only limited information (so-called naïve translator). Moreover, none of the translators were physicians. (2) Synthesis of T1 and T2 were amalgamated to form the unique translated version T12 by resolving any discrepancies under supervision of a methodologist who was not involved in the translation process. (3) Back

translation of the T12 version from French or German into English by two translators whose native language was English, and who were fluent in French, or German (BT1 and BT2). These two translators were naïve to the study and not directly linked with the medical domain. (4) Consensus meeting with all the involved subjects (translators, methodologist, specialist physicians in occupational rehabilitation) in order to resolve any discrepancies and doubts met during the translation, and to establish the pre-final French and German versions of the SFS. (5) Pre-testing of the French and German versions for the accuracy of the words and ease of understanding of the SFS was conducted with 20 consecutive patients with back complaints. Patients were asked to mention any difficulties encountered during a phone call. The last steps were realized by submitting the final version of the French SFS (SFS-F) and German SFS (SFS-G) and all reports and forms to a committee keeping track of the translated version in order to verify that the recommended stages were followed.

Participants

For each language, two sets of participants were recruited: one for the assessment of construct validity and scale homogeneity, and a second set for test–retest reliability.

For construct validity of the French version, 17 women and 70 men were recruited. These 87 subjects were consecutive inpatients hospitalized because of persistent back pain between 2004 and 2005 at the Clinique romande de réadaptation at Sion, Switzerland. The mean age was 44 years (SD: 10; range: 19–61). Test–retest reliability of the French version was assessed on a sample of 21 patients (9 women, 12 men; mean age 43 years, SD: 14, range 19–65) recruited in 2009. In addition to a history of back pain, subjects in both samples had diagnoses such as fracture (operated or treated conservatively), discal prolapse or hernia, degenerative disorders, discopathies, status after discal hernia operation, contusion(s), tight canal, o-listhesis, spina bifida occulta, isthmic lysis, non specific lumbalgia, whiplash, cervical strain, transitional anomaly, Scheuermann's disease.

Construct validity of the German version was assessed on 257 consecutive inpatients hospitalized between November 2003 and February 2006 (53 women, 204 men; mean age 40 years, SD 11, range 18–64). These subjects were recruited at the Rehaklinik Bellikon in Bellikon, Switzerland, because of persistent back pain. Test–retest reliability of the German version was assessed on a convenience sample of 51 patients (9 women; 41 men, mean age 43.6 years, SD 13 years, range 21–65) recruited in 2009. Diagnoses for both samples were similar to the French cohort.

Patients with upper and/or lower limb complaints were excluded because of the risk of influencing the SFS scores. Patients with psychopathology in which pain is the central element (such as somatoform trouble) were also excluded. Patients who had other non-disabling psychopathologies were included.

The study was approved by the ethical committees of the canton Valais and the canton Aargau, where the two clinics are located. All patients signed a written informed consent form.

Validation

All patients completed the French or German version of the SFS, the medical outcomes short form (SF-36) [29], the Hospital Anxiety and Depression Scale (HADS) [30], and the Visual Analogue Scale for Pain Intensity (VAS) [31]. Construct validity of the SFS translated versions was assessed by estimating Pearson's correlation coefficients between the French (resp. German) versions of the SFS and the HADS, VAS, and relevant subscales of the SF-36. The physical functioning subscale (PF), the physical summary scale (PCS) and the VAS were used to assess convergent validity (high correlations expected); the mental health scale (MH), the mental summary scale (MCS) and HADS for divergent validity (low correlations expected). Ninety-five percent confidence intervals for the correlation coefficients were calculated by means of Fisher's transformation.

Ceiling and floor effects were defined as present if at least 15% of results reached the maximum or the minimum value [32].

Internal consistency was determined by Cronbach's α [33–34], which is a general coefficient of homogeneity between items. Values for α can range from 0 (no internal consistency) to 1 (perfect internal consistency), where a value above 0.8 is considered acceptable [35].

The reliability of the translated versions was assessed by test–retest reliability and quantified by the intraclass correlation coefficient (ICC) [36]. Patients who were not expected to have a significant health status change between tests were asked to fill out the SFS-F (resp. SFS-G) on two occasions separated by 2 days. Values for ICC can range from 0 (no agreement) to 1 (perfect agreement). Bland–Altman plots were used to assess the disagreement between test and retest values [37]. Such plots show the individual score differences between tests as a function of the individual mean scores of the two tests. 95% limits of agreement were calculated as the mean difference ± 1.96 SD of the difference. The narrower the limits of agreement, the smaller the disagreement between the repeated tests.

All calculations were performed using the statistical package Stata 11.0 for Windows [38].

Results

Cross-Cultural Adaptation

The translations and back-translations of the SFS items were carried out in both French and German without any relevant difficulties. The back-translations of the T12 versions to English were very similar to the original versions. Only some typically US expressions or words were different as our back-translators were native from the United Kingdom and India. Moreover, patients did not mention any difficulties in understanding the items.
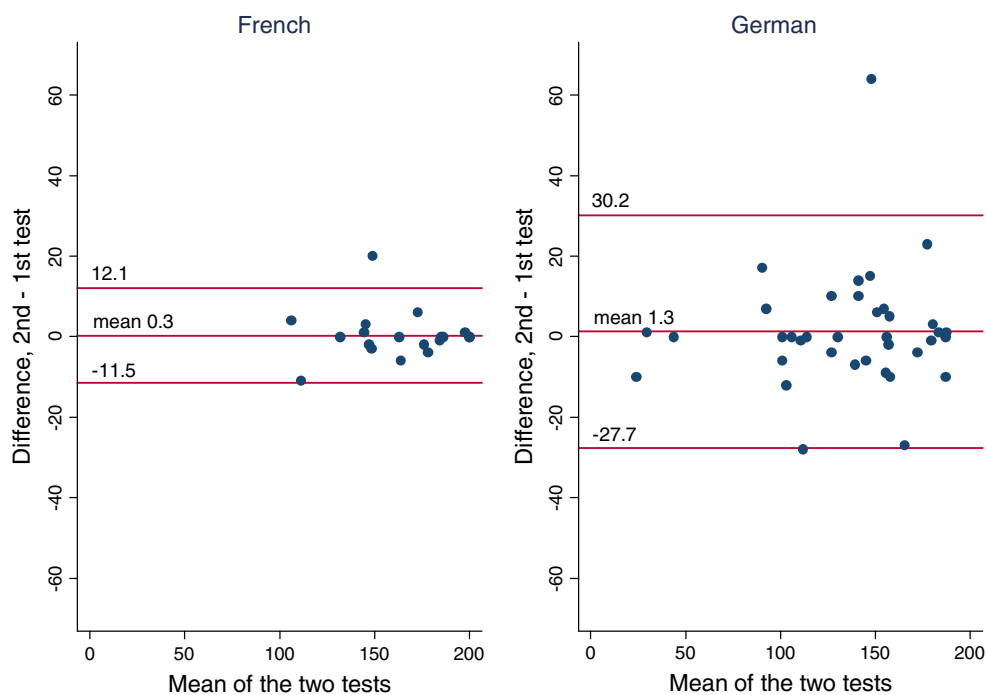
Validation

SFS-French Version

Eighty-seven subjects were eligible for the validation of the SFS-F. The excluded patients had inconsistency in the internal validity check or more than 4 "I don't know" answers.

For convergent validity, we found a correlation coefficient of 0.63 (95% CI: 0.48–0.74) between SFS-F and PF, 0.60 (95% CI: 0.44–0.72) between SFS-S and PSC, and −0.33 (95% CI: −0.51 to −0.13) between SFS-F and VAS. The assessment of divergent validity resulted in an SFS-F-MH correlation of −0.08 (95% CI: −0.29 to 0.14), an SFS-F-MCS correlation of 0.01 (95% CI: −0.21 to 0.23), an SFS-F-HADS depression correlation of −0.26 (95% CI: −0.45 to −0.05), and an SFS-F–HADS anxiety correlation of −0.17 (95% CI: −0.37 to −0.04).

No evidence for floor or ceiling effects was found for the total score since no patient reached the minimum or maximum possible score. A floor effect was found in the items 45–48 with more than 99% of the participants rating their ability to perform the task as "restricted" (14%) or "unable" (85%) on a 5-point Lickert scale from "able" to "restricted" to "unable"). For internal consistency, Cronbach's α was 0.98 for the SFS-F. The reliability, assessed by test–retest in 21 patients, resulted in ICC values of 0.98 (95% CI: 0.97–1.00). The mean difference between test and retest was 0.3, with 95% upper and lower limits of agreement at −11.5 and 12.1 (Fig. 1).

SFS-German Version

Three hundred and nine subjects were eligible for validation of the SFS-G. The excluded patients had inconsistency in the internal validity check or more than 4 "I don't know" answers.

For convergent validity, we found a correlation coefficient of 0.67 (95% CI: 0.59–0.73) between SFS-G and PF, 0.52 (95% CI: 0.43–0.61) between SFS-G and PSC, and −0.51 (95% CI: −0.60 to −0.42) between SFS-G and VAS. The assessment of divergent validity resulted in an SFS-G-MH correlation of 0.25 (95% CI: 0.13–0.36), an SFS-G–MCS correlation of 0.28 (95% CI: 0.16–0.39), an SFS-G–HADS depression correlation of −0.42 (95% CI: −0.52 to −0.32), and an SFS-G–HADS anxiety correlation of −0.45 (95% CI: −0.54 to −0.35).

No evidence for floor or ceiling effects was found for the total score since no patient reached the minimum possible

Fig. 1 Bland–Altman plot for the two languages. The *middle line* represents the mean difference between the two tests. The *upper* and *lower lines* represent the upper and lower limits of agreement, i.e. mean difference + 1.96 SD of the differences and mean difference − 1.96 SD of the differences, respectively

score and only one scored the maximum possible value. A floor effect was found in items 45–48 with more than 97% of the participants rating their ability to perform the task as "restricted" (12%) to "unable" (85%). For internal consistency, Cronbach's α was 0.98 for the SFS-G. Reliability, assessed by test–retest in 44 patients, resulted in ICC values of 0.94 (95% CI: 0.90–0.98). Mean difference between test and retest was 1.3, with 95% lower and upper limits of agreement at −27.7 and 30.2 (Fig. 1). A look at Fig. 1 shows a highly influential patient with a difference of over 60 units between tests. Limits of agreement calculated without that patient were −20.9 and 19.9 for a mean difference of −0.5.

Discussion

The original English version of the SFS was translated and adapted into French and German, respectively, to create the SFS-F and SFS-G versions. Evidence for reliability and validity was shown, supporting the use of the SFS-F and SFS-G as a self-report instrument for individuals with a wide range of chronic back disorders. Specifically, evidence for convergent validity, divergent validity, internal structure, and score stability were provided for the SFS-F and SFS-G.

Typical activities, especially those regarding gardening with specific tools, had to be adapted for the French and German culture. For example, a "spade-shovel" is not commonly used by patients in our countries and was modified as "shovel" ("pelle" in French and "Schaufel" in German). Thus, and although the SFS is a pictorial questionnaire, cross-cultural adaptation shows the importance of following the complete AAOS guidelines for a valuable final version.

As hypothesized for convergent validity, the correlation coefficients between the translated SFS versions and the SF-36 physical scales were fairly high, i.e. 0.63 and 0.67 for the SFS-F and SFS-G, respectively. Moreover, they were similar to values found by Gibson et al. [22] with other scales such as the pain disability index (−0.64), the work re-entry questionnaire (0.67), the SE questionnaire (0.55) and the pain self-efficacy questionnaire (0.78) [22]. Those questionnaires could not have been used in the present study because of the lack of French and German validated versions. The pain scale also showed a significant correlation with the SFS-G (−0.51), but only a low correlation with the SFS-F (−0.33). This last estimation is rather imprecise (95% CI −0.51 to −0.13), probably due to a smaller sample compared to the German version. For divergent validity, we found no correlation (−0.08) between the SF-36 mental scales and the SFS-F, and a low correlation with the SFS-G (0.25) as hypothesized. The small differences between the French and German versions

might be explained either by some cultural differences regarding the implication of back problems in daily living and, consequently, the interaction with MH of the SF-36 (which has questions regarding irritability, sadness, motivation), or by sampling. The correlation between HADS and the SFS was low (0.26) for the French version and moderate (0.42) for the German version. These correlations are possibly due to the chronicity of back problems in our patients, who were recruited in tertiary centers. Patient populations with chronic occupational back pain are known to exhibit higher prevalence of psychological disorders compared to the general population [39]. Moreover, the difference between the French and German versions may, as for MH, be explained by cultural differences related to either the patients or medical practice, but also to the difference in the timing of hospitalization in the two centres after back problems were diagnosed. Furthermore, it must be kept in mind that our study samples were not randomly drawn from a population but were convenience samples. A previous study performed at the Clinique Romande de réadaptation Suva care (Switzerland) has shown that questionnaire responders differed from non-responders in some sociodemographic and biophsychosocial aspects [40]. Thus, some degree of selection bias, which may differ between clinics, may well have occurred in the present study.

A floor effect was found in items 45–48. Those items describe activities where weights of 50 kg are lifted either from floor, waist or overhead height or down again. Most participants felt they could not carry out such strenuous activities. It may be questioned whether these items are of great value for the clinical purpose of the questionnaire. Furthermore, lifting tasks involving weights over 25 kg are nowadays prohibited in most occupations in Switzerland, France and Germany.

According to the literature, a Cronbach's α over 0.80 (over 0.90 for clinical applications) represents a good internal consistency. We found excellent α value far above these thresholds with 0.98 for both SFS-F and SFS-G. This high internal consistency may be partly influenced by the high number of items since α has the property of becoming larger with increasing item number, given equal between-item correlations [41]. However, our values are similar to those of the English versions (0.98), suggesting that the French and German translations bear the same level of internal consistency as the original version.

The reliability of both the SFS-F and the SFS-G was excellent with regard of an ICC of 0.98 and 0.94, respectively. These coefficients are higher than the values reported in the original version (0.89) [23]. Moreover, the confidence intervals (0.97–1.00 for the French version, and 0.90–0.98 for the German version) were narrow for both translations, indicating rather precise estimates.

The limits of agreement were calculated to determine the magnitude of disagreement between the two measurement occasions. With all patients included, the interval between the limits of agreement of the German version was over twice that of the French version (57.9 and 23.6 units, respectively). After exclusion of a highly influential patient, the German version's interval was reduced to 41.8 units. However, further studies should be done to evaluate the minimal clinical important change [42] to establish whether the difference in score is clinically relevant.

Some limitations of the present study have to be recognised. First, only patients hospitalized in tertiary centers for chronic back problems were included. Thus, results of SFS-F and SFS-G questionnaires have to be interpreted with caution in other clinical settings. The use of convenience samples instead of random samples was discussed above.

Although the SFS has been successfully used for the last 20 years, some recommendations may be given here to improve its clinical utility. First, old fashioned drawings (i.e. old type of vacuum cleaner) should be replaced by new pictures of tools used nowadays. Second, reduction in the number of items would lead to important time saving and therefore further improve its clinical utility. In this context, we calculated PACT scores using either the 25 even or the 25 uneven items and ranked the patients on the full score and the two half-scores. Correlations between the full score rank and each of the half-score ranks were 0.99 for both languages, showing item redundancy. A reduction in the number of items is also supported by the high internal consistency. Third, relevant items which include posture of spinal load, such as sitting, should be included in the SFS. The development of a brief version of the SFS is clearly needed. Further studies exploring these measurement properties in different settings and with other validation tools are therefore needed.

In conclusion, the French and the German versions of the SFS, seem to be valid and reliable, and it is a tool that is easy to administer to evaluate perceived functional capacity for native French-speaking and German-speaking patients with back disorders, both for clinical purposes and research.

References

1. Genovese E, Galper JS. Guide to the evaluation of functional ability: how to request, interpret, and apply functional capacity evaluations. 1st ed. American Medical Association; 2009.
2. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev. 1977;84:191–215.
3. Reneman MF, Jorritsma W, Schellekens JM, Goeken LN. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. J Occup Rehabil. 2002;12:119–29.
4. Holden G. The relationship of self-efficacy appraisals to subsequent health related outcomes: a meta-analysis. Soc Work Health Care. 1991;16:53–93.
5. Lackner JM, Carosella AM. The relative influence of perceived pain control, anxiety, and functional self efficacy on spinal function among patients with chronic low back pain. Spine (Phila Pa 1976). 1999;24:2254–60.; Discussion 60–61.
6. Lee CE, Simmonds MJ, Novy DM, Jones S. Self-reports and clinician-measured physical function among patients with low back pain: a comparison. Arch Phys Med Rehabil. 2001;82:227–31.
7. Schiphorst Preuper HR, Reneman MF, Boonstra AM, Dijkstra PU, Versteegen GJ, Geertzen JH, et al. Relationship between psychological factors and performance-based and self-reported disability in chronic low back pain. Eur Spine J. 2008;17:1448–56.
8. Smeets RJ, van Geel AC, Kester AD, Knottnerus JA. Physical capacity tasks in chronic low back pain: what is the contributing role of cardiovascular capacity, pain and psychological factors? Disabil Rehabil. 2007;29:577–86.
9. de Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. J Clin Epidemiol. 2003;56:1137–41.
10. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737–45.
11. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research: a proposal for standardized use. Spine (Phila Pa 1976). 1998;23:2003–13.
12. 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.
13. Matheson LN. History, design characteristics, and uses of the pictorial activity and task sorts. J Occup Rehabil. 2004;14:175–95.
14. Mayer J, Mooney V, Matheson L, Leggett S, Verna J, Balourdas G, et al. Reliability and validity of a new computer-administered pictorial activity and task sort. J Occup Rehabil. 2005;15:203–13.
15. Matheson LN, Matheson ML. Spinal function sort. Rating of perceived capacity. Test booklet and examiners manual. Performance Assessment and Capacity Testing PACT; 1989/1991.

16. Henchoz Y, de Goumoens P, So AK, Paillex R. Functional multidisciplinary rehabilitation versus outpatient physiotherapy for non specific low back pain: randomized controlled trial. Swiss Med Wkly. 2010;140:131–3.

17. Innes E, Hardwick M. Actual versus perceived lifting ability in healthy young men (18–25 years). Work. 2010;36:157–66.

18. Kool JP, Oesch PR, Bachmann S, Knuesel O, Dierkes JG, Russo M, et al. Increasing days at work using function-centered rehabilitation in nonacute nonspecific low back pain: a randomized controlled trial. Arch Phys Med Rehabil. 2005;86:857–64.

19. Robinson RC, Kishino N, Matheson L, Woods S, Hoffman K, Unterberg J, et al. Improvement in postoperative and nonoperative spinal patients on a self-report measure of disability: the spinal function sort (SFS). J Occup Rehabil. 2003;13:107–13.

20. Sufka A, Hauger B, Trenary M, Bishop B, Hagen A, Lozon R, et al. Centralization of low back pain and perceived functional outcome. J Orthop Sports Phys Ther. 1998;27:205–12.

21. Oliveri M. Functional capacity evaluation. In: Gobelet C, Franchignoni F, editors. Vocational rehabilitation. Paris: Springer; 2005.

22. Gibson L, Strong J. Assessment of psychosocial factors in functional capacity evaluation of clients with chronic back pain. Br J Occup Ther. 1998;61:399–404.

23. Matheson LN, Matheson ML, Grant J. Development of a measure of perceived functional ability. J Occup Rehabil. 1993;3:15–30.

24. Oesch PR, Hilfiker R, Kool JP, Bachmann S, Hagen KB. Perceived functional ability assessed with the spinal function sort: is it valid for European rehabilitation settings in patients with nonspecific non-acute low back pain? Eur Spine J. 2010;19:1527–33.

25. U.S. Department of Labor. The revised handbook for analyzing jobs. 4th ed. Indianapolis: JIST Works, inc.; 1991.

26. 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.

27. 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.

28. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, 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.

29. Ware JE Jr, Gandek B. Overview of the SF-36 health survey and the international quality of life assessment (IQOLA) project. J Clin Epidemiol. 1998;51:903–12.

30. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the hospital anxiety and depression scale: an updated literature review. J Psychosom Res. 2002;52:69–77.

31. Boonstra AM, Schiphorst Preuper HR, Reneman MF, Posthumus JB, Stewart RE. Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain. Int J Rehabil Res. 2008;31:165–9.

32. Bent NP, Wright CC, Rushton AB, Batt ME. Selecting outcome measures in sports medicine: a guide for practitioners using the example of anterior cruciate ligament rehabilitation. Br J Sports Med. 2009;43:1006–12.

33. Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.

34. Osburn HG. Coefficient alpha and related internal consistency reliability coefficients. Psychol Methods. 2000;5:343–55.

35. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Upper Saddle River, NJ: Prentice-Hall Health; 2007.

36. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med. 1990;20:337–40.

37. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307–10.

38. StataCorp LP, 4905 Lakeway Drive, College Station, TX 77845, USA

39. Dersh J, Gatchel RJ, Mayer T, Polatin P, Temple OR. Prevalence of psychiatric disorders in patients with chronic disabling occupational spinal disorders. Spine (Phila Pa 1976). 2006;31:1156–62.

40. Burrus C, Ballabeni P, Deriaz O, Gobelet C, Luthi F. Predictors of nonresponse in a questionnaire-based outcome study of vocational rehabilitation patients. Arch Phys Med Rehabil. 2009;90:1499–505.

41. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 3rd ed. Oxford: Oxford University Press; 2003.

42. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976). 2008;33:90–4.