*Journal of Occupational Rehabilitation, Vol. 12, No. 3, September 2002 (© 2002)*

# Concurrent Validity of Questionnaire and Performance-Based Disability Measurements in Patients With Chronic Nonspecific Low Back Pain

**Michiel F. Reneman,**[1,4] **Wim Jorritsma,**[1] **Jan M.H. Schellekens,**[2] **and Ludwig N.H. Göeken**[1,3]

*This study aimed to investigate the concurrent validity of two approaches to disability measurement in patients with chronic nonspecific low back pain (CLBP). It was hypothesized that if both are measuring the same construct, the instruments would lead to similar disability results and would correlate strongly* ($r > 0.75$). *The study compared the results of self-reported and performance-based measures of disability in 64 consecutive patients with CLBP. Participants mean age was 38.0 years, the mean duration of the current episode of back pain 9.9 months, and 90% were off work due to CLBP. The self-report measures used were: the Roland Disability Questionnaire (Roland); the Oswestry Disability Questionnaire (Oswestry); and the Quebec Back Pain Disability Questionnaire (Quebec). Performance was measured using the Isernhagen Work Systems Functional Capacity Evaluation (FCE). The mean scores from the self-report measure are as follows: Roland 13.5 (scale 0–24), Oswestry 28.2 (scale 0–100), Quebec 37.8 (scale 0–100) consistent with moderate to severe disability. In contrast the results from the performance-based measures suggested that the subjects should be able to work at a physical intensity level of moderate to heavy. Little to moderate correlation was observed between the self-report and performance-based measures (Spearman rank correlations: Roland-FCE (−0.20),* $p > 0.05$*; Oswestry-FCE (*$-0.52$*),* $p < 0.01$*; Quebec-FCE (*$-0.50$*),* $p < 0.01$). *Results are interpreted to suggest that both performance-based and self-report measures of disability should be used in order to obtain a comprehensive picture of the disability in patients with CLBP.*

**KEY WORDS:** concurrent validity; disability; functional status; functional capacity evaluation.

[1]Center for Rehabilitation, University Hospital Groningen, Haren, The Netherlands.
[2]Department of Occupational Psychology, University of Groningen, Groningen, The Netherlands.
[3]Institute for Human Movement Sciences, University of Groningen, Groningen, The Netherlands.
[4]Correspondence should be directed to Michiel F. Reneman, Center for Rehabilitation, University Hospital Groningen, PO Box 30002, 9750 RA Haren, The Netherlands; e-mail: m.reneman@beatrixoord.nl.



1053-0487/02/0900-0119/0 © 2002 Plenum Publishing Corporation

## INTRODUCTION

Chronic nonspecific low back pain (CLBP) accounts for the consumption of major portions of health care funds and financial compensation for temporary or permanent inability to work. Rehabilitation programs aim to reduce the disability of the CLBP patient. Disability can be measured by questionnaires and performance-based instruments. Questionnaires measure self-reported disability and contain perceived or subjective information. A person's "objective" disability can be assessed by means of direct performance measurement in a laboratory setting, often referred to as Functional Capacity Evaluations (FCEs). FCEs are batteries of standardized tests designed to assess a person's functional capacity related to work.

All self-report and performance-based instruments operationalize the construct across several different activities, which combined should reflect disability as defined. Without a universally accepted definition of disability (1–3), however, it is not surprising that different terms are used. Besides "disability" (4–6), terms that are also used are functional disability (7), functional status (8–10), and functional capacity (11–14). The content of the terms is very similar and appears to be used interchangeably. Even though the wordings used are not equal, they appear to measure the same construct, defined as "difficulty of a person to perform an activity" (15).

The advantages of using questionnaires over performance-based testing include less expense, less time consumed, and proven psychometric qualities (1,7–10,16). FCEs have been criticized because little is known about their psychometric properties (17,18), the results can be influenced by a patient's pain behavior (19–21) and the time needed to collect data is relatively long (3–5 h), thus making it expensive (17,22).

Thus, the question can be asked, whether the extra time and expense of the FCE is warranted. This question cannot be answered, as no head-to-head comparison between the outcome of self-reported and performance-based disability measurements has been performed (9). The objective of this investigation was to study the concurrent validity of two approaches of measuring disability: self-report and performance based. It was hypothesized that, if questionnaires and FCEs are in fact measuring the same construct, these two different measurement approaches should lead to similar disability outcome and correlate strongly; patients with high levels of self-reported disability are expected to perform poorly in an FCE, and vice versa.

## METHODS

### Setting

This study was conducted in an outpatient university rehabilitation and occupational assessment center in the northern part of The Netherlands.

### Patients

Patients were referred to the rehabilitation department by their general physicians. The patients were assessed by a staff physiatrist. Included were patients diagnosed with nonspecific CLBP who appeared willing to improve their functional status through participation in

a rehabilitation program. Excluded from the program were: patients with significant comorbidity, specific pathology related to LBP (i.e., disc herniations, tumors, spondylolisthesis grade 3 or 4, etc.) and patients with extensive psychological or behavioural problems (as diagnosed by a staff psychologist). A total of 67 consecutive patients were identified as eligible for inclusion in this study. Of these, three patients were immigrants who had not mastered the Dutch language sufficiently to complete the questionnaires by themselves, leaving a total of 64 patients who make up the population of this study.

## Disability Measures

Three questionnaires were used to measure self-reported disability related to CLBP. These were:

1. The Roland Morris Disability Questionnaire (Roland) (4), developed to measure self-rated disability due to low back pain, without defining the term "disability." The questionnaire consists of 24 items, each one a statement that the patient is asked to indicate whether this statement applies to him/her. The total score is the sum of all questions answered in the affirmative and ranges from 0 to 24.
2. The Oswestry Back Pain Disability Scale (Oswestry) (6) which defines disability as "the limitations of a patient's performance compared with that of a fit person." The questionnaire consists of 10 items, each item consisting of six statements. The patient is asked to check the statement that most closely resembles his/her status. The total score ranges from 0 to 100.
3. The Quebec Back Pain Disability Scale (Quebec) (1) which adopts the World Health Organization's 1980 definition of disability as "any restriction or lack of ability to perform an activity in a manner or within the range considered normal for human beings." The questionnaire consists of 20 items; each item consists of a description of an activity. The patient is asked to what extend he/she experiences difficulties in performing that activity. Answering is possible on a 6-point scale ranging from "no difficulty" to "unable to perform." The total score ranges from 0 to 100.

A detailed review of the literature on the psychometric properties of the questionnaires lies beyond the scope of this article. In summary, the reliability, sensitivity, internal consistency, and the validity of the questionnaires have been reported to be high in populations with subacute and chronic LBP (1,7–10,16).

The FCE is a performance-based evaluation designed to measure a person's capacities in multiple activities. In this study, the standardized FCE protocol of Isernhagen Work Systems (IWS; 23) was used. Disability is not defined, but inferred to be the ability to perform (maximally) purposeful activities in the context of work (14). Activity selection is based on the Dictionary of Occupational Titles (DOT (5)). The DOT is used to express work demands and the capacities needed to work for specific occupations. Face and construct validity of the DOT classifications is considered to be adequate (17,18,24–27). The DOT categorizes the physical characteristics into 20 activities (e.g., lifting and carrying). In this study, patients were asked to perform to their maximum abilities on 14 different activities (lifting, carrying, pushing, pulling, overhead work, stooping, crouching, kneeling, standing, walking, sitting, and stair climbing). Testing of an activity was terminated when one of the following occurred: 1) the patient wished to terminate (e.g., due to pain), 2) the heart rate

**Table I.** Dictionary of Occupational Titles (5) Physical Demands Strength Rating. Limits of Weights Lifted/Carried or Forces Exerted During the Workday

| Category | Occasionally | Frequently | Constantly |
|---|---|---|---|
| 1. Sedentary | [a]–4.5 | [a] | [a] |
| 2. Light | [a]–9.0 | [a]–4.5 | [a] |
| 3. Medium | 9.0–22.5 | 4.5–11.4 | [a]–4.5 |
| 4. Heavy | 22.5–45.0 | 11.4–22.5 | 4.5–9.0 |
| 5. Very heavy | >45.0 | >22.5 | >9.0 |

[a] Negligible weight.

reached 85% of the age-related maximum, indicated by a heart rate monitor, 3) the evaluator deemed further testing to be unsafe, or 4) a preset performance criterion was reached (e.g., 30 min standing). The inter- and intra-rater reliability of the material handling procedures is fair to excellent (28–32). Test-retest reliability of lifting and carrying is good in patients with CLBP (28,32). Test-retest reliability of static overhead work and static forward bend is good in healthy young adults (33). Pilot testing of test-retest reliability of the other FCE-subtests in 10 healthy young adults resulted in good correlations ($r$ ranges from 0.82 to 1.00, $p < 0.00$) (34).

The DOT-classification was used to transform the FCE scores into a single measure. Five strength levels express the physical demands of work: very light (category 1), light (category 2), moderate (category 3), heavy (category 4), and very heavy (category 5). The characteristics for each category are given in Table I. Besides lifting, carrying, and pushing/pulling, the duration of sitting, standing, and walking is also taken into account. The performance of each subject was classified according to the DOT scales and entered as a separate variable. Based on the information from the patient, the same classification was used to classify the physical characteristics of the work.

## Procedures

Prior to the first visit, patients completed the Roland scale (5–10 min), a pain intensity rating on a numeric rating scale (0–10) (1 min), a pain drawing, out of which the distribution of the symptoms were classified according to the Quebec Task Force (35), and a study consent form. A standardized medical intake procedure lasting approximately 45 min followed, consisting of a review of the medical history and a physical examination, including the determination of Waddell's nonorganic signs (36). The FCE was performed 1–2 weeks after the medical intake procedure. Prior to FCE, the patients filled out the Quebec and Oswestry questionnaires. They were also asked if their pain and/or functional status were different compared to the medical intake.

## Data Analysis

Levels of disability were computed using descriptive statistical analysis. The association between the scores on the questionnaires and the FCE outcome (DOT 1–5), and associations between the questionnaire scores was tested with Spearman rank correlations.

Some of the questionnaire items relate directly or closely to activities tested on the FCE. The operational definitions of all items are different. For example, "lifting low" on the

FCE is operationally defined as the maximal ability (in kilograms) to safely lift a receptacle from table height to the floor and vice versa for 5 times in no longer than 90 s (23). Lifting is not an item specifically asked about in the Roland questionnaire. In Question 3 of the Oswestry (6), statements do include lifting activities, for example: "I can lift a heavy object but this causes extra pain" or, or "I can only lift light objects." The client was asked to check the statement that most closely resembles their status. On Question 20 of the Quebec scale (1), the client is asked to rate the difficulty experienced from lifting 20 kg. It was presumed that the more difficulty the client experienced, the lesser amount of weight can be lifted maximally and vice versa. This presumption was analyzed statistically by using a Spearman rank correlation test. Somers' $d$ index of symmetric association was used with ordinal items with no more than two categories (dichotomous) (37).

No strict criteria are available to establish concurrent validity. In order for instruments to be concurrently valid, the results of the measurements should be similar and they should correlate strongly, that is, have a considerable amount of shared variation ($r^2$). In this study, correlations are interpreted as follows (27): correlation coefficients ($r$) of 0.50 or less are considered poor (little similarity; $r^2 \leq 25\%$), correlations of 0.51–0.75 are considered moderate (some similarity; $r^2 = 26$–$56\%$), and correlations of 0.76 or more are considered good (substantial similarity; $r^2 \geq 75\%$).

## RESULTS

### Patients

Characteristics of the patients are presented in Table II. The characteristics indicate that this represents a population of chronic LBP sufferers, including accompanying characteristics such as a high rate of medical consumption and work absence. In line with the Dutch

**Table II.** Characteristics of the Patients ($N = 64$; Male $n = 54$, Female $n = 10$)

| Population characteristics | % | Mean | SD | Min–Max |
|---|---|---|---|---|
| Age (years) | | 38.0 | 8.9 | 23–58 |
| Current episode (months) | | 9.8 | 11.3 | 2–72 |
| Pain intensity (0–10) | | 5.1 | 2.1 | 0–10 |
| Waddell signs (0–5) | | 1.2 | 1.3 | 0–4 |
| DOT work (0–5) | | 2.4 | 1.0 | 1–5 |
| Married (or common-law) | 86 | | | |
| Smoker | 49 | | | |
| Using NAISDs | 19 | | | |
| Sx location back-gluteal | 42 | | | |
| Sx location back-thigh | 31 | | | |
| Sx location back-leg | 27 | | | |
| Previous Rx: 1 | 100 | | | |
| Previous Rx: >1 | 63 | | | |
| Visited >1 medical specialists | 41 | | | |
| Status postlumbar surgery | 16 | | | |
| Employed | 95 | | | |
| % Of which out of work due to CLBP | 90 | | | |
| Receiving final compensation | 87 | | | |

*Note.* DOT = Classification physical intensity of work according to the Dictionary of Occupational Titles; Sx = Symptoms; Rx = Physical/rehabilitation Therapy.

**Table III.** Results of Self-Reported and Performance-Based Disability Assessments

| | Scale | Mean | SD |
|---|---|---|---|
| Roland | 0–24 | 13.5 | 3.7 |
| Oswestry | 0–100 | 28.2 | 15.8 |
| Quebec | 0–100 | 37.8 | 15.7 |
| FCE-DOT | 1–5 | 3.6 | 0.6 |

worker's compensation system, the majority of the patients were financially compensated while absent due to CLBP. Most patients (71%) were off work for more than 3 months, 14% were off work between 6 weeks and 3 months, and 5% were off work for less than 6 weeks. All patients stated to the investigators that their functional and pain status prior to the FCE had not changed relevantly since their medical intake.

## Disability Measures Descriptives

The scores on the different disability measures are presented in Table III. The average scores on the Oswestry reflect a "moderate" level of disability (6). The average Quebec scores are interpreted as "somewhat disabled," analogous to the wording in the questionnaire (1). Interpretation of the Roland scores is not as straightforward. The subjects in this study reported difficulty on more than half of the questionnaire items (mean score 13.5). A mean score of 14.3 can be interpreted as "significant disability" (38). The results of the performance-based FCE testing indicate that the subjects were able to perform work that is (according to DOT categories) physically moderate to heavy. Results of self-reported disability were not significantly ($p < 0.01$) different between males and females. Performances differed significantly between the genders (males FCE-DOT mean 3.8, SD = 0.5; females mean 2.9, SD = 0.7, $p < 0.01$).

## Correlations Between Disability Measures

Correlations between the questionnaires and FCE performance are presented in Table IV. Correlations among the questionnaires are generally higher than correlations between the questionnaires and the FCE. The correlation of the FCE with the Roland Disability questionnaire is poor, and poor to moderate with the Oswestry Back Pain Disability Scale and the Quebec Back Pain Disability Scale.

**Table IV.** Associations Between Self-Reported and Performance-Based Disability Measures

| | Roland | Oswestry | Quebec |
|---|---|---|---|
| Roland | — | — | — |
| Oswestry | 0.50* | — | — |
| Quebec | 0.60* | 0.74* | — |
| FCE | −0.20 | −0.52* | −0.50* |

* $p < 0.01$.

**Table V.** Summary of Performance on Selected FCE Activities and Correlations With Corresponding Questionnaire Items ($n = 64$)

| FCE activities | Criterion | Performance | Correlation with questionnaire item |
|---|---|---|---|
| Lifting | 5 lifts in <90 s, max | Mean 29.5 kg, SD 11.3 (range 4–56 kg) | Oswestry 3, $r = -0.20$; Quebec 20, $r = -0.51^{**}$ |
| Overhead lift | 5 lifts in <90 s, max | Mean 16.1 kg, SD 6.0 (4–40 kg) | — |
| Short carry two handed | 5 carries 1.5 m in <90 s, max | Mean 36.3 kg, SD 15.6 (4–70 kg) | — |
| Long carry two handed | 1 carry 20 m in <90 s, max | Mean 37.6 kg, SD 12.4 (12–75 kg) | Quebec 19, $r = -0.44^{**}$ |
| Pushing | Static full body push, average of 3 repetitions | Mean 43.0 kg, SD 12.2 (12–72 kg) | Quebec 18, $r = -0.47^{**}$ |
| Pulling | Static full body pull, average of 3 repetitions | Mean 51.9 kg, SD 16.1 (17–80 kg) | Quebec 18, $r = -0.52^{**}$ |
| Walking[a] | Shuttle walk, max speed | Mean speed 5.6 kph, SD 0.9 (1.8–7.2) | Roland 3, $d = 0.13$ |
| Walking[a] | Shuttle walk, distance | 22% > 300 m, 32% > 400 m, 38% > 500 m | Roland 17, $d = 0.03$; Oswestry 4, $r = -0.17$; Quebec 8, $r = -0.27$; Quebec 9, $r = -0.32^{*}$ |
| Sitting tolerance | 30 min | Met by 94% (6% 20–29 min) | Roland 20, $d = 0.06$; Oswestry 5, $d = 0.03$; Quebec 6, $d = -0.09$ |
| Standing tolerance | 30 min | Met by all subjects | Oswestry 6, $d = C$; Quebec 5, $d = C$ |
| Kneeling[a] | 1 min | Met by 97% | Roland 11, $d = -0.08$ |
| Overhead work[a] | Static 1 min, dynamic 20 reps in 1 min | Static met by 95%, dynamic met by 81% | Quebec 10, $d = 0.05$ (dyn); $d = 0.05$ (stat) |
| Standing forward bend[a] | Static 1 min, dynamic 20 reps in 1 min | Static met by 98%, dynamic met by 66% | Roland 11, $d = 0.01$ (dyn); $d = -0.05$ (stat); Quebec 16, $d = 0.01$ (dyn); $d = 0.08$ (stat) |
| Stairclimbing[a] | 100 steps total, <35 s per 25 steps, use handrail (y/n) | Speed met by 90%, no handrail 98% | Roland 23, $d = 0.23$ (speed); Roland 5, $d = -0.05$ (handrail); Quebec 7, $d = -0.15$ (speed) |

*Note.* $r$ = Spearman's rho, $d$ = Somers' $d$, $C$ = statistics cannot be computed because one of the variables is a constant (FCE item).
[a] For comparison, performance criteria listed differ from standardized IWS FCE (manual).
$^{*}p < 0.05$; $^{**}p < 0.01$.

### Correlations Between Activities and Questionnaire Items

Results of the FCE performances and correlations with corresponding questionnaire items are presented in Table V. Most of the results are not significant at $p < 0.05$.

## DISCUSSION

This study was performed to investigate the concurrent validity between two approaches of measuring disability in patients with CLBP. Although subjects performed at a level that resulted in them being defined as being able to perform work at a physical

intensity level of medium to heavy, they reported their disability to be moderate or severe on the questionnaires. Correlations between self reported and performance-based disability measures were found to be poor to moderate: −0.20 (Roland), −0.50 (Quebec), and −0.52 (Oswestry). Based on these figures, 4–27% of the total variance ($r^2$) between the two measurement approaches can be explained. The strengths of the correlations among the questionnaires are moderate (0.50–0.74). At item level, statistically significant correlations between the questionnaire items and performance-based items were poor to moderate. Based on the differences in disability outcome and strengths of the correlations, it is concluded that the two measurement approaches do not meet the criteria of concurrent validity.

A strength of this study is that well-known clinical instruments were used. This means that the results of this study contribute to the existing body of knowledge with regards to the psychometric properties and clinical utility of commonly used instruments. As a result of choosing existing instruments, some weaknesses have occurred. First, the construct of disability is operationalized differently in each of the instruments, likely because a universally accepted definition of disability does not exist. Second, there is a difference in context. The questionnaires ask for the ability to perform activities in a context of daily living, whereas FCEs measure disabilities in a laboratory context. Third, there is a difference in measurement strategy. The questionnaires assess loss of ability or difficulty in performance (negative approach, "glass half empty"), whereas the FCE assesses the ability to perform (positive approach, "glass half full"). Fourth, there are differences in item scaling. The Roland questionnaire uses dichotomous items, whereas the other two use a 6-point scale for each item. The relevance of this is that even the slightest amount of limitation will contribute a full point on the Roland disability score (1/24), where the same amount of difficulty would contribute only 1/50th of the total Oswestry score, and 1/100th of the Quebec score. This may contribute to the observation that the differences between the FCE and the Roland are the most significant. In line with this, a wide variety of individual performances are reduced to only five categories in the DOT classification. Effectively, the majority of the patients in this study were classified into only three of these categories. To promote variation in results, future studies should aim to measure activities at ratio level, and prevent reduction of data to a small number of ordinal levels (e.g., test and measure maximum capacity for overhead work rather than cutting off at 5 min, based on the FCE protocol).

Each of these weaknesses may be a source of variance, and may thus explain part of the difference in results. The relative contribution of each towards the results is unclear and should be studied separately. Differences between the questionnaires were not studied. As such, based on the data presented here, no recommendation can be made regarding the value or utility of one questionnaire over another.

Both self-report and performance based measurement approaches involve effort-related performance. As such, they cannot be defined independently of the person's behavior. Pain-related disability is a matter of human performance, whether it is observed/measured or reported (39). This is likely to apply in an admission phase of a rehabilitation program as well. In this stage patients not only reveal their perception of their disability, but also present an image to the practitioner (40), justifying the need for treatment (41,42) and perhaps also justifying the fact that they are off work due to CLBP. The scores on the questionnaires may thus be an overrating of the "real" disability. The FCE is a measure of demonstrated ability. A known phenomenon in FCEs is the existence of pain behaviors (19–21). There were indications that pain behaviors may have interfered with the FCE results in this study.

Thus, without these pain behaviors, the FCE results would probably have been higher. On average, the performance-based scores presented in this study may thus be considered an underrating of the abilities of the patients.

The IWS FCE has similarities with other FCEs with regards to construct validity (17,18,27). Because of differences in measurement approaches and the lack of studies that have compared FCEs head-to-head, it is not known whether the results of this study can be generalized to other FCEs. Compared with the questionnaires, a drawback of FCEs is the time needed to collect data (3–6 h) which makes them less practical in their use, and the expenses that accompany FCEs are higher (17). Research is warranted to identify approaches to shorten the FCE without compromising its validity. Additionally, further research on the psychometric properties of FCEs, including the IWS FCE, is warranted to establish test-retest reliability of some of the nonmaterial handling test items.

A source of variance is the use of different measurement approaches: self-report and performance based. This finding is particularly clear at item level, where differences in results are evident and the correlations are poor. Given the fact that the activities measured are the same or very similar, the differences should be attributed to differences in measurement approaches. The results of this study confirm the findings of others (43–46) that self-report of ability to perform certain activities cannot be interchanged with the actual ability to perform that same activity. Given that 90% of the patients in this study were off work, though considered physically able to work, it appears that work disability here is more closely related to self-report than to actual physical abilities. Because the patient's physical capacities exceed physical demands of certain jobs, it can be debated whether treatment modalities designed to improve physical capacities (e.g., physical reconditioning and work simulation) are appropriate for many patients in CLBP return to work programs.

## CONCLUSION

The measurement objectives of users of the different instruments are similar, but results of questionnaire and performance-based disability measurements differ substantially and correlations between the approaches suggest poor concurrent validity. Instruments based on self-report or based on performance appear to measure different dimensions of the disability construct. A performance measure should be used to measure "a person's ability to perform an activity," whereas a questionnaire should be used to measure "a person's self-reported ability to perform an activity." Replication studies of similar design are recommended with larger patient samples, and different questionnaires and performance measures, to provide broader generalization. Based on the results of this study, both a performance measure and a questionnaire are recommended in order to obtain a more comprehensive picture of disability with CLBP.

## ACKNOWLEDGMENTS

The authors thank Roy Stewart, research methodologist, and Pieter Dijkstra, clinical epidemiologist, for their statistical assistance and critical remarks on previous versions of this manuscript.

## REFERENCES


1. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI. The Quebec back pain disability scale. Measurement properties. *Spine* 1995b; 20: 341–352.
2. Pfeiffer D. The problem of disability definition: Again. *Disab Rehabil* 1999; 21: 392–395.
3. Wunderlich GS (ed.). *Measuring functional capacity and work requirements; summary of a workshop*. Washington, DC: National Academic Press, 1999.
4. Roland M, Morris R. A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983; 8: 141–144.
5. *Dictionary of Occupational Titles*, 4th edn., Rev. Washington, DC: U.S. Department of Labor, 1991.
6. Fairbank JCT, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980; 66: 271–273.
7. Kopec JA, Esdaile JM. Spine update: Functional disability scales for back pain. *Spine* 1995a; 20: 1943–1949.
8. Beurskens AJ, de Vet HC, Köke AJ, van der Heijden GJ, Knipschild PG. Measuring the functional status of patients with low back pain: Assessment of the quality of four disease-specific questionnaires. *Spin*e 1995; 20: 1017–1028.
9. Deyo RA. Measuring the functional status of patients with low back pain. *Arch Phys Med Rehabil* 1988;69: 1044–1053.
10. Deyo RA, Battie M, Beurskens AJHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Korff Mvon, Waddell G. Outcome measures for low back pain research. *Spine* 1998; 23: 2003–2013.
11. Abdel-Moty E, Fishbain DA, Khalil TM, Sadek S, Cutler R, Rosomoff RS, Rosomoff HL. Functional capacity and residual functional capacity and their utility in measuring work capacity. *Clin J Pain* 1993; 9: 168–173.
12. Fishbain DA, Abdel-Moty E, Cutler R, Khalil TM, Saedek S, Rosomoff RS, Rosomoff HL. Measuring residual functional capacity in chronic low back pain patients based on the dictionary of occupational titles. *Spine* 1994; 19: 872–880.
13. Fishbain DA, Khalil TM, Abdel-Moty E, Cutler R, Sadek S, Steele-Rosomoff R, Rosomoff HL. Physician limitations when assessing work capacity: A review. *J Back Musculoskeletal Rehabil* 1995; 5: 107–113.
14. Isernhagen SJ. Functional capacity evaluation: Rationale, procedure, utility of the kinesiophysical approach. *J Occup Rehabil* 1992; 2: 157–168.
15. *ICIDH-2: International Classification of Functioning and Disability*, Beta-2 draft, Full Version, Geneva, World Health Organisation, 1999.
16. Nachemson A, Vingard E. Assessment of patients with back and neck pain: A best-evidence synthesis. In: Nachemson A, Jonsson E, eds. *Neck and back pain: The scientific evidence of causes, diagnostics, and treatment.* Philadelphia: Lippincott Williams & Wilkins, 2000, pp. 189–236.
17. King PM, Tuckwell N, Barrett TE. A critical review of Functional Capacity Evaluations. *Phys Ther* 1998; 78: 852–866.
18. Reneman MF, Jaegers SMHJ, Muskee C, Schröer HTh, Göeken LNH. Functional Capacity Evaluation: Toepassing in Nederland? *TBV* 1997; 5: 139–146.
19. Cooke C, Dusik LA, Menard MR, Fairburn SM, Beach GN. Relationship of performance on the ERGOS work simulator to illness behavior in a workers' compensation population with low back versus limb injury. *JOM* 1994; 36: 757–762.
20. Kaplan GM, Wurtele SK, Gillis D. Maximal effort during Functional Capacity Evaluations: An examination of psychological factors. *Arch Phys Med Rehabil* 1996; 77: 161–164.
21. Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: A summary of methods and approaches. *Phys Ther* 1998; 8: 867–888.
22. Reneman MF, Dijkstra SJ, Jorritsma W, Muskee C, Schiphorst Preuper HR, Göeken LNH. Assessment and treatment of chronic work-related pain disorders in an outpatient university rehabilitation setting in the Netherlands. *Work* 2001; 16: 23–30.
23. Isernhagen Work Systems. *Manual Functional Capacity Evaluation.* Duluth, MN: Isernhagen & Associates, 1989.
24. Dusik AL, Menard MR, Cooke C, Fairburn SM, Beach GN. Concurrent validity of the ERGOS work simulator versus conventional functional capacity evaluation techniques in a workers' compensation population. *JOM* 1993; 35: 759–767.
25. Lechner DE, Jackson RJ, Roth DL, Straaton KV. Reliability and validity of a newly developed test of physical work performance. *JOM* 1994; 36: 997–1004.
26. Mul CAM, Douwes M, Hazelet AM, Wevers CWJ. *Schadebeoordeling en FCE-methoden.* Hoofddorp: TNO Arbeid, 1999.
27. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13: 125–152.
28. Gross DG. Reliability of safe maximum lifting determinations of a functional capacity evaluation. *Phys Ther* 2002; 82: 364–371.
29. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgements of level of lift effort in a kinesiophysical Functional Capacity Evaluation. *Work* 1999; 12: 145–150.

30. Smith RL. Therapists' ability to identify safe maximum lifting in low back pain patients during functional capacity evaluation. *JOSPT* 1994; 19: 277–281.
31. Reneman MF, Jaegers SMHJ, Westmaas M, Göeken LNH. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *WORK* 2002; 18: 23–28.
32. Reneman MF, Dijkstra PU, Westmaas M, Göeken LNH. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil*.
33. Reneman MF, Bults MMWE, Engbers LH, Mulders KKG, Goeken LNH. Maximum holding times and perception of static elevated work and forward bending in healthy young adults. *JOR* 2001; 11: 87–97.
34. Dijkink A, Kuis M.*Onderzoek naar de betrouwbaarheid van de Functional Capacity Evaluation van Isernhagen* [research report]. Groningen, The Netherlands: University of Groningen, Institute for Movement Sciences, 2000.
35. Spitzer WO, (ed.). Scientific approach to the assessment and management of activity related spinal disorders: A monograph for clinicians. Report of the Quebec Task Force on spinal disorders. *Spine* 1987; 12: 1–59.
36. Waddell G, McCulloch JA, Kummel E, Venner RM. Nonorganic physical signs in low-back pain. *Spine* 1980; 5: 117–125.
37. Siegel S, Castellan NJ. *Nonparametric systems for the behavioral sciences*, 2nd edn. Singapore: McGraw-Hill, 1988.
38. Stratford PW, Binkley JM. Measurement properties of the RM-18. *Spine* 1998; 22: 2416–2421.
39. Fordyce WE (ed.). Back pain in the work place. Management of disability in nonspecific conditions. *A report of the task force on pain in the work place of the International Association of the Study of Pain*. Seattle, WA: IASP press, 1995.
40. Simmons M. Physical function and physical performance in patients with pain: What are the measures and what do they mean. In: *Syllabus IASP refresher courses on pain management held in conjunction with the 9th world congress on pain*, Vienna, Austria, Aug. 22–27, 1999, pp. 127–136.
41. Watson PJ. Non-physiological determinants of physical performance in musculoskeletal pain. In: *Syllabus IASP refresher courses on pain management held in conjunction with the 9th world congress on pain*, Vienna, Austria, Aug. 22–27, 1999.
42. Hadler NM. If you have to prove you are ill, you can't get well. *Spine* 1996; 21: 2397–2400.
43. Piela CR, Hallenberg KK, Geoghegan AE, Monsein MR, Lindgren BR. Prediction of functional capacities. *Work* 1996; 6: 107–113.
44. Wijlhuizen GJ, Ooijenduik W. Measuring disability, the agreement between self evaluation and observation of performance. *Disab Rehabil* 1999; 21: 61–67.
45. Verbunt JA, Westerterp KR, van der Heyden GJ, Seelen HA, Vlaeyen JW, Knottnerus JA. Physical activity in daily life in patients with chronic low back pain. *Arch Phys Med Rehabil* 2001; 82: 726–730.
46. Lee CE, Simmonds MJ, Novy DM, Jones S. Self-reports and clinician measured function among patients with low back pain. *Arch Phys Med Rehabil* 2001; 82: 227–231.