한국전문물리치료학회지 2012년 19권 3호
Phys Ther Kor. Vol. 19 No. 3 2012

Responsiveness Comparisons of Self-Report Versus Therapist-Scored Functional Capacity for Workers With Low Back Pain

Bongsam Choi[1], PhD, MPH, PT, So-yeon Park[2], PhD, PT
[1]Dept. of Physical Therapy, College of Health and Welfare, Woosong University,
[2]Dept. of Physical Therapy, College of Alternative Medicine, Jeonju University

Abstract

The primary aim of this study was to compare responsiveness of self-report by worker and therapist-scored functional capacity instrument. Self-report and therapist-scored interval-level person measures and item difficulties were compared at admission and discharge. Therapist and worker ratings were collected on 230 clients from 27 rehabilitation sites using the newly developed Occupational Rehabilitation Data Base (ORDB) functional capacity instrument. ORDB comprises several subscales measuring relevant variables of "a return-to-work model" in work-related rehabilitation clinics. The functional capacity scale deals with 10 DOT job factors. The rating scale categories were 1-severely impaired, 2-moderately impaired, 3-mildly impaired, and 4-not impaired. Only data from clients with low back pain (n=98) with complete data (both admission and discharge scores) were used for the present study. Therapists and workers completed the functional capacity instrument at admission and discharge. Rasch analysis [1-parameter item response theory model (IRT)] was applied to calibrate item difficulty and person ability measure of therapist and workers ratings. Effect sizes for therapist and self-report ratings were slightly different, .69 and .30, respectively. Therapist and worker ratings were more consistent at discharge (r=.54) than at admission (r=.32). Workers have a tendency to be more severe in their ratings (show higher item difficulties) than therapists at admission and discharge. Therapists and workers report similar magnitudes of improvement following treatment program. These findings challenge the belief that injured workers may unreliable source for monitoring therapeutic outcomes. Self-report measures have the advantage of conserving therapist time for treatment (versus evaluation). While the therapist and self-report ratings are comparable at discharge, there is less consistency at admission. Comparable therapist-worker ratings may be achieved by controlling for rating severity using IRT methodologies.

[Bongsam Choi, So-yeon Park. Responsiveness Comparisons of Self-Report Versus Therapist-Scored Functional Capacity for Workers With Low Back Pain. Phys Ther Kor. 2012;19(3):91-97.]

Key Words: Functional capacity; Low back pain; Responsiveness.

Introduction

Low back pain is one of the most common cause of rapidly growing work related loss, health care expenditure and disability benefits. It is the most costly symptom of injured worker with economic loss exceeding $50 billion per year in the United States (Frymoyer and Cats-Baril, 1991). Every year the particular concern shows that more than 2% of all workers sustain low back injuries (Fathallah et al, 1998).

Functional Capacity Evaluation (FCE) is a tool that evaluates the capacity of injured worker to perform the physical demands of their job or the demands postulated for their work situation (Gibson et al, 2002). The therapist administering the FCE evaluates the person's performance of physical demands,

Corresponding author: So-yeon Park soyeonparkpt@gmail.com

including any major limiting factors (Gibson et al, 2002). Physical demands included in FCE are often based on those listed in the Dictionary of Occupational Title (DOT), which include sitting, standing, walking, lifting, carrying, pushing, pulling, crouching, kneeling, and stooping (U.S. Department of Labor, 1981). The classification based on DOT is obvious to professional working in the vocational rehabilitation. The validity of the DOT classification is considered to be adequate to measure functional capacity (Innes and Straker, 1999; King et al, 1998; Reneman et al, 2002a; Reneman et al, 2002b) though DOT classification recently has been replaced by the United States Department of Labor with the O*Net (http://www.oalj.dol.gov/libdot.htm).

Patient views about their health status are becoming increasingly important in the evaluation of the outcome of treatment (Taylor et al, 1999). Self-reported and performance-based FCE performed by clinician are among the common methods of assessment of musculoskeletal disorders. FCE instruments have been widely used to determine the physical work abilities of individuals who have sustained musculoskeletal injury such as low back pain (Abdel-Moty et al, 1993; Gross, 2004; King et al, 1998; Mayer et al, 1985; Simmonds et al, 1998; Velozo, 1993). Although the performance measures are commonly used to measure physical functions and are thought of as objective biomechanical measures, in reality, they are frequently influenced by rater's subjectivity (Cox et al, 2000). In addition while rater's judgments of maximum performance during FCE appear reliable, FCE does not appear to be purely tests of physical capacity as performance during assessment is influenced by multiple factors (Gross, 2004). Compared to performance-based measure with FCE, self-report measure can be influenced by psychological distress, correlating with emotional, cognitive function and pain behavior rather than true functional capacities (Gross, 2004; Kaplan et al, 1996).

While reliability and validity have been traditionally considered as the important psychometric criteria for evaluating outcome measures, the sensitivity of a measure to detect clinical change, responsiveness, may be the one of the most critical criteria for evaluating the efficacy of a therapeutic intervention (Guyatt et al, 1989; Guyatt et al, 1992). Responsiveness is defined as the ability of an instrument to detect clinically important changes over time, even if those changes are small (Kirshner and Guyatt, 1985). Cohen (1977) proposed that the effect size statistic relates the magnitude of score change on a particular scale to the variability within the score change. It represents a standardized measure of change that allows direct comparison between different instruments that use different units or scales of measurement.

While FCE has been widely used as an outcome measure for the injured worker's capacity to perform the physical demands, the comparison of self-reports versus therapist scored ratings has not been investigated with respect to patient view in Korea. The purpose of this study aims to determine responsiveness of self-report by injured worker and therapist-scored functional capacity evaluation, and compare those measures at admission and discharge.

Methods

Instruments and Procedures

Velozo and colleagues (1994) developed a DOT-based functional capacity instrument as a part of an Occupational Rehabilitation Data Base (ORDB) instrument. The instrument comprises several subscales measuring relevant variables of "a return to work model" in work-related rehabilitation clinics. The functional capacity scale deals with 10 DOT job factors: standing, walking/running, sitting, lifting, carrying, pushing/pulling, climbing, stooping/crouching/kneeling, reaching, handling/fingering. These factors were derived from the list of the DOT job factors (Fishbain et al, 1994; U.S. Department of Labor,

1981; U.S. Department of Labor, 1986). While the DOT is being replaced by the O*Net, the ORDB physical functioning items reflect the content of 7/9 of the physical demand questions on the O*Net Questionnaire (#34~42) (www.onetcenter.org/questionnaires.html). In this study, all DOT classifications were not used and some of the classifications were combined (i.e., pushing/pulling, stooping/crouching/kneeling, handling/fingering). The rating scales were rated four-points: 1-severely impaired, 2-moderately impaired, 3-mildly impaired and 4-not impaired.

The ORDB instrument used therapist and worker rater forms. Both forms defined key aspects the 10 DOT job factors and described the four rating scale categories in detail. Both therapists and workers instructed to rate item based on the demands of their previous job and items as an anticipated job if the worker did not intend to return to their previous work.

Participants

Workers incorporated in 28 rehabilitation facilities in Illionois, United States of America due to return-to work issues were recruited as a part of a larger project that intended to develop the ORDB functional capacity instrument. The facilities incorporate work hardening and conditioning treatment program into a return-to-work model as a specialized rehabilitation program. Representatives from the facilities were educated on the theoretical concepts of the instrument before the field testing. Workers who were accepted to participate in the project were selected for the study. For 3 months (between August 22, 1994 and November 11, 1994), a total of 230 workers were selected for actual data collection, however only valid data were included for the present study (n=98). Sixty-seven percent (66/98) of the workers participating in the study were male and 33 percent (32/98) were female with an average age of 38.76 (ranges from 21 through 65 years of age).

Statistical Analysis

Application of the Rasch model (1-parameter item response theory model) can produce a stable item calibrations and item fit statistics with as small as 100 subjects for a 10-item test. (Linacre, 2005; Wang and Chen, 2005; Wright and Masters, 1982). Admission and discharge scores rated by therapists and workers were analyzed with Winsteps Rasch analysis computer program (Winsteps, Chicago, IL, U.S.A.), using the rating scale model (Linacre, 2005). Rasch analysis generates a log odds unit (logit) unit scale presenting person ability and item difficulty measures. Logits are logarithmic transformation of item and person data into an interval scale. For a dichotomous item, item difficulty is the ratio of the probability of success relative to the probability of failure on an item. If items comprise a rating scale, item difficulty is the probability of success relative to failure for an item at a particular rating.

The measurement property used to quantify responsiveness is the standardized effect size. Effect sizes for therapist and worker ratings were calculated using the method proposed by Guyatt et al (1987) in the present study, which was calculated as the ratio of a minimal clinically important difference to the standard deviation in subjects who do not change clinically. Correlation analysis between therapist and worker ratings at admission and discharge was performed to determine how the two ratings related.

Results

Using the ORDB functional capacity instrument, an effect size was examined for therapist and worker ratings and is shown in Table 1. The effect size for therapist ratings was “moderate” at .69 and the effect size for worker ratings was at .30. Using the criterion of effect size, an effect size of .20 is considered to be small, .50 to be moderate, and .80 or greater to be large. Moderate effect size was ob-

Table 1. Effect size (ES) calculations based on person measure changes

	Admission score	Discharge score	Change score	Effect size*
Therapist ratings	.11±1.68	1.27±3.04	1.68	.69
Client ratings	.01±1.46	.45±2.60	1.46	.30

*Effect size (ES) is defined as mean change score in Functional Capacity Evaluation (FCE) instrument divided by the standard deviation of the initial mean score.

tained for therapist rating who rated less "impaired" than worker rating, while relatively small effect size was obtained for worker ratings who rated their functions "more impaired" than therapist rating following the treatment program.

By plotting Rasch-modeled estimates of person ability for therapist and self-reported worker ratings, Figure 1 delineates how therapist and worker ratings were differently measured at admission/discharge. The x and y coordinates are derived based on person ability measures. After obtaining the person measures from both ratings are cross-plotted. A pair of 95% confidence interval lines based on the conjoint error estimates is constructed. Points outside the 95% confidence interval lines are flagged as potential disagreements between therapist and worker ratings.

While the plot at admission appears to converge within the confidence interval line and be less linear than one at discharge. The plot at admission could be described as being more clustered than one at discharge. In order to confirm whether the associations between the two ratings are linear, the correlation coefficients are calculated. The correlation coefficients show that therapist and worker ratings are more consistent at discharge ($r=.54$, $p<.001$) than at admission ($r=.32$, $p=.002$).

Discussion

A primary issue of this study was to examine

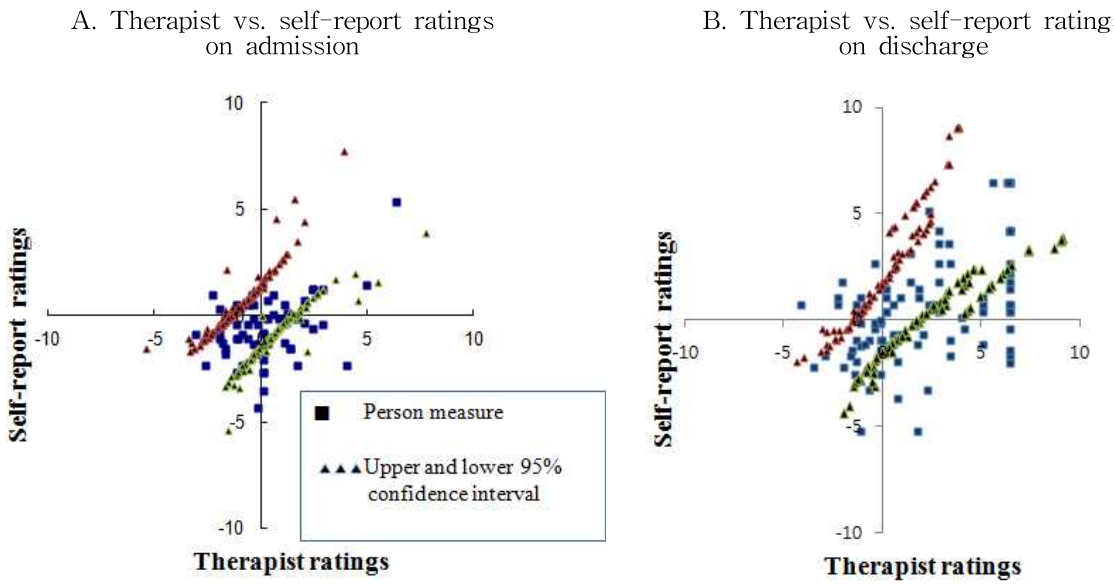

Figure 1. Comparisons of therapist versus self-report ratings by workers on admission and discharge. The graph represents the relationship between therapist scored and self-report ratings in person ability measures. Each "■" on the graph represent plot between the two ratings, with the "▲" creating two diagonal lines representing upper and lower limits to 95% confidence interval.

whether the functional capacity instrument of the ORDB would show a similar responsiveness between therapists scored and self-reported workers ratings in workers with low back pain. The functional capacity instruments have been criticized because little is known about their psychometric properties (King et al, 1998). The psychometrics of responsiveness has recently gained considerable attention in outcome measure. By using one of Item Response Theory (IRT) methodologies, Rasch analysis, the functional capacity scale of the ORDB showed acceptable psychometric responsiveness for both ratings, therapist scored and self-report ratings by workers. Whereas a moderate effect size was obtained for the therapist (.69), a small size was obtained for the worker ratings (.30) across admission and discharge outcome. This finding indicates that therapist scored rating is more sensitive to the changes than in worker's rating in response to ORDB functional capacity scale.

Responsiveness may be the most important property of evaluative outcome measures of health status in terms of 'sensitivity to change'. While Kirshner and Guyatt's definition of responsiveness prevails, several authors have defined responsiveness or sensitivity to change as the ability of an instrument to detect minimal clinically important changes in patient's function as a result of treatment (Kirshner and Guyatt, 1985). There seems to be no objection to use the term "responsiveness" to detect the clinical changes following rehabilitation interventions. The measurement property used to quantify responsiveness is the standardized effect size that represents a standardized measure of change that allows direct comparison between different instruments that use different units or scales of measurement (Guyatt et al, 1989; Guyatt et al, 1992). The effect size of therapist scored ratings, the magnitude of responsiveness measures, was (.69) slightly better than self-reported worker ratings (.30), though both ratings showed good item level psychometrics as determined by the IRT.

In the present study, the correlation between therapist rating and self-report rating by worker was r=.54 at discharge and r=.32 at admission. Low correlation coefficient between admission and discharge for worker rating is of our concern. This evidence implicates that self-report might be less objective resource due to "sick role" and the potential influence on secondary reward (Velozo et al, 2006). By contrast, Sadosky and colleagues (2011) provided a convincing argument for the usefulness of self-reported severity and other outcome measures as appropriate indicators. The result is comparable to a similar study on the comparison of self-report and performance-based ratings of functional capacity evaluation (Reneman et al, 2002c). The correlations between those two measures were found to be poor to moderate relative to the performance-based functional capacity measure. These findings suggest that multifactorial approach should consider the possibility of disagreement of treatment outcome. In reality, it is a well known phenomenon that patients with chronic pain may overrate disability (Hazard et al, 1991). In general, self-report is known for assessing not necessarily indicative of actual functional activities but the perception of patient. The self-report instrument does not give an accurate reflection of functional capacity, hindering determination of how reported physical activity is associated with low back injuries (Simmonds et al, 1998). Our study result also showed that workers had a tendency to be more severe than therapists in their ratings. That is, the instrument is probably influenced by nonphysical factors. However, self-report measure has been recently accepted as an useful tool due to its potential usage for as a screening tool (Hart and Wright, 2002).

Conclusion

The purpose of this study was to compare rater's severity and responsiveness of self-report by worker and therapist-scored ORDB functional capacity in-

strument for low back pain. The comparison of therapist and self-report ratings in responsiveness showed slightly different magnitudes of improvement following the treatment. Both workers and therapist had a tendency to score more consistently at discharge. Effect sizes for therapist and self-report ratings were slightly different, .69 and .30, respectively. Also, the consistency of therapist and self-report ratings showed r=.54 at discharge and r=.32 at admission. This results showed that workers have a tendency to be more severe in their ratings than therapists at admission and discharge.

In summary, the comparison of therapist and self-report ratings by workers in responsiveness shows acceptable measurement qualities, reporting slightly different magnitudes of improvement following the treatment. Although these findings challenge the belief that injured workers may be unreliable source for monitoring therapeutic outcomes, self-report measures have many advantages of conserving therapist time for treatment versus evaluation. Physical therapist and other rehabilitation professionalists are encouraged to use self-report as an alternative of health outcome measure.

References

Abdel-Moty E, Fishbain DA, Khalil TM, et al. Functional capacity and residual functional capacity and their utility in measuring work capacity. Clin J Pain. 1993;9(3):168-173.

Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ, Lawrence Erlbaum Associates Inc., 1977:1-27.

Cox ME, Asselin S, Gracovetsky SA, et al. Relationship between functional evaluation measures and self-assessment in nonacute low back pain. Spine (Phila Pa 1976). 2000;25(14):1817-1826.

Fathallah FA, Marras WS, Parnianpour M. The role of complex, simultaneous trunk motions in the risk of occupation-related low back pain disorders. Spine (Phila Pa 1976). 1998;23:1035-1042.

Fishbain DA, Abdel-Moty E, Cutler R, et al. Measuring residual functional capacity in chronic low back pain patients based on the Dictionary of Occupational Titles. Spine (Phila Pa 1976). 1994;19(8):872-880.

Frymoyer JW, Cats-Baril WL. An overview of the incidence and costs of low back pain. Orthop Clin North Am. 1991;22(2):263-271.

Gibson L, Allen M, Strong J. Re-integration into work. In: Strong J, Unruh A, Wright A, eds, Pain: A textbook for therapists. Edlinburgh, Churchill Livingston, 2002:17.

Gross DP. Measurement properties of performance-based assessment of functional capacity. J Occup Rehabil. 2004;14(3):165-174.

Guyatt GH, Deyo RA, Charlson M, et al. Responsiveness and validity in health status measurement: A clarification. J Clin Epidemiol. 1989;42(5):403-408.

Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: What are the necessary measurement properties? J Clin Epidemiol. 1992;45:1341-1345.

Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40(2):171-178.

Hart DL, Wright BD. Development of an index of physical functional health status in rehabilitation. Arch Phys Med Rehabil. 2002;83(5):655-665.

Hazard RG, Bendix A, Fenwick JW. Disability exaggeration as a predictor of functional restoration outcomes for patients with chronic low back pain. Spine (Phila Pa 1976). 1991;16(9):1062-1067.

Innes E, Straker L. Validity of work-related assessments. Work. 1999;13(2):125-152.

Kaplan GM, Wurtele SK, Gillis D. Maximal effort during functional capacity evaluations: An ex-

amination of psychological factors. Arch Phys Med Rehabil. 1996;77(2):161-164.

King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. Phys Ther. 1998;78(8):852-866.

Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis. 1985;38(1):27-36.

Linacre JM. A User's Guide to Winsteps Rasch-Model Computer Programs. Chicago, Winsteps.com, 2005:351.

Mayer TG, Gatchel RJ, Kishino N, et al. Objective assessment of spine function following industrial injury: A prospective study with comparison group and one-year follow-up. Spine (Phila Pa 1976). 1985;10(6):482-493.

Reneman MF, Dijkstra PU, Westmaas M, et al. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. J Occup Rehabil. 2002a;12(4):269-275.

Reneman MF, Jaegers SM, Westmaas M, et al. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. Work. 2002b;18(1):23-27.

Reneman MF, Jorristsma W, Schellekens JMH, et al. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. J Occup Rehabil. 2002c;12(3):119-129.

Sadosky AB, Taylor-Stokes G, Lobosco S, et al. Relationship between self-report low-back pain severity and other patient-reported outcomes: Results from an observational study. J Spinal Disord Tech. 2011;24(5):1-7

Simmonds MJ, Olson SL, Jones S, et al. Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. Spine (Phila Pa 1976). 1998;23(22):2412-2421.

Taylor SJ, Taylor AE, Foy MA, et al. Responsiveness of common outcome measures for patients with low back pain. Spine (Phila Pa 1976). 1999;24(17):1805-1812.

Velozo CA. Work evaluations: Critique of the state of the art of functional assessment of work. Am J Occup Ther. 1993;47(3):203-209.

Velozo CA, Choi B, Zylstra SE, et al. Measurement qualities of a self-report and therapist-scored functional capacity instrument based on the dictionary of occupational titles. J Occup Rehabil. 2006;16(1):109-122.

Velozo CA, Santopoalo R. Training Manual: Occupational rehabilitation data base manual. Chicago, University of Illinois at Chicago, 1994.

Wang WC, Chen CT. Item parameter recovery, standard error estimates, and fit statistics of the Winsteps program for the family of Rasch models. Ed Psych Meas. 2005;65(3):376-404.

Wright BD, Masters GN. Rating Scale Analysis. Chicago, MESA Press, 1982.

U.S. Department of Labor, Employment and Training Administration. Selected characteristics of occupations defined in the Dictionary of Occupational Titles. Washington, DC, U.S. Government Printing Office, 1981.

U.S. Department of Labor, Employment and Training Administration. Dictionary of Occupational Titles, 4th eds., Supplement. Washington, DC, U.S. Government Printing Office, 1986.

This article was received July 7, 2012, was reviewed July 7, 2012, and was accepted August 29, 2012.