한국전문물리치료학회지 2013년 20권 4호 22-31 ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
Phys Ther Kor 2013;20(4):22-31 http://dx.doi.org/10.12674/ptk.2013.20.4.022

A Comparative Study of Oswestry Back Pain Disability Questionnaire Versus Computer Adaptive Testing for Measuring Back Pain

Bong-sam Choi, PhD, MPH, PT
Dept. of Physical Therapy, College of Health and Welfare, Woosong University

Abstract

The aim of the present study was to compare measurement precisions of the Oswestry Back Pain Disability Questionnaire (ODQ) and a computer adaptive testing (CAT) method. The ODQ has been regarded as one of the most reliable condition-specific measure for back pain for decades. Cross-sectional study was carried out with two independent convenient samples from two out-patient rehabilitation clinics for back pain (n_1=42) and non-back pain group (n_2=42). Participants were asked to fill out the ODQ and CAT of International Classification of Functioning, Disability and Health-Activity Measure (ICF-AM). A series of Rasch analyses were performed to calculate person ability measures. The CAT measures had greater relative precision in discriminating the groups than did the ODQ measure in comparisons of the relative precision. The CAT measure appears to be more effective than did the ODQ measure in terms of measurement precision. By administering test items calibrated in a way, CAT measures using item response theory may promise a means with measurement precision as well as efficiency.

Key Words: Back pain; Computer adaptive testing; Item response theory; Measurement; Rasch model; Relative precision.

Introduction

Many self-report measures have been developed specifically for the back pain population due to their several advantages. These advantages include decreasing administration costs, reducing respondent burden, and potentially accessing scattered sample. Many studies suggest that self-report disability measures for back pain are as reliable as performance measures and appear to be sensitive indicators of long-term outcome (Kopec, 2000). In general, these self-report disability measures are commonly classified into generic and condition specific measures (McHorney, 1997; McHorney, 1999). To date, nearly 82 condition specific disability measures for back pain have been developed and have been shown to have adequate psychometrics. Of these widely accepted disability instruments, the Oswestry Back Pain Disability Questionnaire (ODQ) is regarded as one of the most reliable back pain instruments.

However studies have shown that the ODQ may lack sensitivity to discriminate between individuals at the high extreme of ability range (Fairbank and Pynsent, 2000; White and Velozo, 2002), only occasionally being responsive to individuals with severe back pain (Baker et al, 1989; Taylor et al, 1999). Several studies also indicate that the ODQ is more sensitive for patients who have improved but less sensitive for patients whose condition remained unchanged (Muller et al, 2006; Taylor et al, 1999). Thus, despite its adequate psychometrics, the ODQ may not precisely measure the disability of back pain across the full range of ability.

Deficits in measurement precision may be the result of using items that do not closely match the ability of the sample of interest (McHorney, 1999). That is, when easy items are administered to individuals with high ability (i.e., low disability) and/or

Corresponding author: Bong-sam Choi bchoi@wsu.ac.kr

difficult items are administered to individuals with low ability, there is a lack of measurement precision with a resulting inability to discriminate among individuals (Fairbank and Pynsent, 2000; White and Velozo, 2002). Problems with measurement precision occur many times in conventional instruments with fixed number of item, because it is unrealistic for one instrument to include enough items to precisely measure individuals across a wide range of ability. Even instruments with excellent breadth may still have inadequate depth of measurement (Liang et al, 2002). Additionally problematic issues are the fact that long assessments covering a wider range of ability level contain items that appear unnecessary and induce a concern over respondent burden and administration costs (Jette and Haley, 2005). Thus creating the "ideal" measure consisting of enough items to cover the full range of the trait with adequate precision is challenging when using short forms. Despite the popularity and widespread use of the existing instruments developed using classical test theory (CTT), the limitations of these instruments still remain in criticism (Hambleton, 2000).

In contrast to CTT, item response theory (IRT) focuses on the psychometric properties of the items making up the instrument instead of the instrument as a whole (Velozo et al, 2006; Velozo et al, 2008). By estimating the probability that a respondent will select a particular rating for an item, item difficulty and person ability can be placed on the same linear continuum. Thus, IRT model allows "connecting" individuals' responses to items with their ability level (Velozo and Peterson, 2001; Velozo et al, 2006). Estimates of person ability on an underlying construct obtained using IRT methods are invariant regardless of the items used, whereas under the CTT paradigm, person scores vary depending on the difficulty of the instrument (Velozo et al, 2006). Furthermore, item difficulty estimates derived from the IRT analyses remain the same regardless of the ability of the sample, while test statistics in CTT are dependent on the sample taking the test. In addition, the IRT models linearly transform raw scores into equal interval measures (McHorney et al, 1997). These advantages of IRT allow for the creation of invariantly calibrated large item banks that can more precisely discriminate individuals' ability levels and thus, capture smaller increments of change.

While IRT methodologies provide the means for generating and linking person ability and item difficulty calibrations, Computer Adaptive Testing (CAT) methods promise a means for administrating items in a way that is both efficient and precise (Bjorner and Ware, 1998; Elhan et al, 2008; Jette and Haley, 2005; Haley et al, 2004a; Haley et al, 2004b; McHorney, 1997; Velozo et al, 1999; Velozo et al, 2000). Studies have shown that CAT improves test efficiency maintaining adequate precision with fewer items than the full test (Elhan et al, 2008; Flynn et al, 2008; Haley et al, 2008; Hart et al, 2006; Hol et al, 2007; Jette et al, 2008; Velozo et al, 2006; Ware et al, 2003; Weiss, 1982). CAT measures are highly correlated with other assessments intending to measure the same construct and require fewer items. The CAT is based on a testing algorithm which defines iterative processes with a set of rules specifying the test questions to be administered to respondents. This includes procedures for item selection, ability estimation, and termination criteria. By selectively administering items that are matched to the ability level of the individuals, measurement efficiency can be accomplished without the loss of precision provided by the full item bank. With this technology, a small number of items can be selected from the item bank which are most relevant for a person of a particular ability level (Velozo et al, 1999). The IRT in combination with CAT method has recently become an alternative to conventional fixed-format disability measurement (Jette and Haley, 2005; Kopec, 2000).

The ICF Activity Measure (ICF-AM) has recently been developed to create an efficient and precise measurement system based on the activity dimension of World Health Organization's (WHO) International Classification of Functioning, Disability and Health

(ICF). The ICF provided the conceptual framework and classification system for generating the items on the ICF-AM. Activities involving movement, moving around and daily life activities were the sub-categories of the ICF activity dimension consulted in the development of items. Items were developed with the intent to represent the entire range of ability on each construct, thus, creating an equiprecise measurement which means precise measurement across the entire range of the underlying construct. Using IRT and CAT methods, Velozo and colleagues (2008) created ICF-AM, a web based computer adaptive survey system. The administrative core of the instrument allows adjusts to be made to various settings making it possible to change, the initial theta value (i.e., difficulty of question first given to respondent) and stopping rule (i.e., guidelines for terminating the test). Because questions are targeted to individuals at their ability level requiring 5~10 questions per construct are required to reach at a final measure of person ability with acceptable error. In addition, immediate feedback is provided to the respondents/clinicians in the form of a graph and summary statistics.

We hypothesized that the CAT measures will discriminate more precisely than do the ODQ measures. The purpose of this study was to compare relative precision of the person measures generated from the CAT measures of the ICF-AM and the ODQ.

Methods

Participants

Forty-two individuals with back pain were recruited from rehabilitation clinics in Gainesville, Florida including the University of Florida and Shands Orthopaedics and Sports Medicine Institute and Shands Rehabilitation Hospital. Forty-two individuals without back pain were recruited from multiple public sites in Gainesville, Florida, USA. Criteria for participants with back pain included: 1) currently experiencing back pain, 2) having previously received treatment for back pain, 3) ability to read and understand English, and 4) 18 years of age or older. The criteria for non back pain participants included: 1) currently experiencing no back pain, 2) able to read and understand English, and 3) 18 years of age or older. All appropriate participants presenting to the recruiting sites between November 3, 2009 and June 30, 2010 were recruited for the back pain group. This study was approved by the Institutional Review Board at the University of Florida (IRB #17-2009).

Instrumentation

The ODQ, a conventional back pain disability instrument developed under CTT, was used in this study. The ODQ is among the most popular self-report condition specific instruments assessing how back pain affects patients' ability to manage daily life tasks (Fairbank, 2000). The ODQ and its revised versions provide an index of the perceived disability experienced by individuals with back pain. It consists of ten items including pain intensity, personal care, lifting, walking, sitting, standing, sleeping, employment/home making, and traveling. Participants respond on a 5 point ordinal scale (5; pain does not interfere with activities, 0; pain so severe that activities cannot be performed). The summed total score is converted to a percentage score meaning no disability (zero) to most severe disability (one hundred). Thus, a higher score is indicative of a higher level of disability.

In an effort to achieve both psychometric efficiency and precision, the ICF-AM was developed using item response theory. The original ICF-AM consists of 6 activity constructs: positioning/transfers, lifting/carrying, fine hand, walking/climbing, wheelchair/scooters, and self care activities. Constructs for use in this study were selected based on the following two criteria: 1) tasks represented by items within the construct frequently cited as problematic for individuals with back pain and 2) tasks within the construct represent a potential activity limitation for individuals with back pain. Based on these criteria, three relevant constructs were chosen for this study: 1) positioning/transfers, 2) lifting/carrying, and 3) walking/moving. For each of the questions on the CAT, respondents are asked to select one of four response categories with a lower score representing a lower level of ability; "3" (no difficulty), "2" (some difficulty), "1" (a lot of difficulty), and "0" (have not done).

CAT technology was used to administer items of the ICF-AM instrument for each construct. Figure 1 presents the CAT algorithm used for the ICF-AM instrument. First, the CAT begins with an initial person ability estimate (Bn) for a particular construct (i.e., positioning/transfer). The initial person ability measure is set at the mean person ability of the sample used in the preliminary paper and pencil field test (during ICF-AM development phase). The CAT presents an item with a difficulty measure (Di) that is identical or closes to this initial person ability measure. After the initial item is presented and responded to, a new person ability estimate and standard error (SE) is generated. The stopping rule for

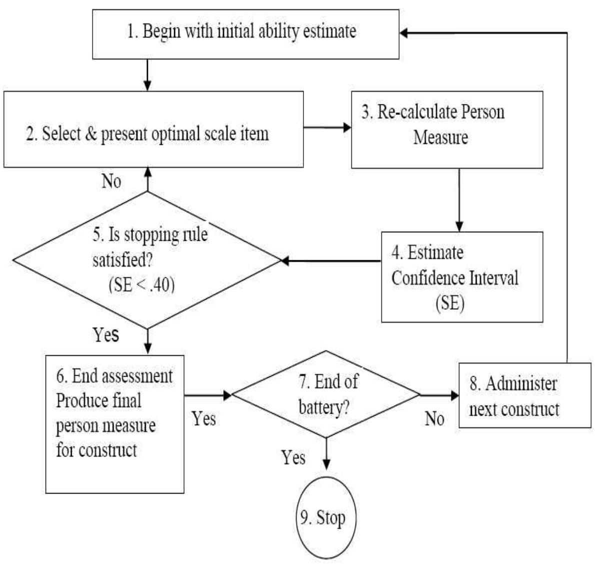

Figure 1. Computer adaptive testing algorithm. Adapted from Wainer, Dorans, Eignor, Flaugher, Green, Mislevy, Steinberg, and Thissen (2000) (SE: standard error).

the CAT is preset based on the standard error associated with a person ability estimate (i.e., SE <.40) and the maximum number of items administered (i.e.,<10 items). That is, the test finishes when an individual's ability is estimated with a standard error less than .40 or 10 items have been administered to the individual. Since the stopping rule is unlikely to be reached with the presentation of a single item, a second item is presented to the respondent. Based on the response, the person ability estimate is recalculated. This procedure continues until the SE associated with the person ability estimate is less than the preset SE, which defines the stopping rule. Once the stopping rule is satisfied, the respondent's final ability measure for that construct is formulated. After completed positioning/transfer construct, the next construct (i.e., lifting/carrying and walking/moving construct) is presented until the CAT reaches the final ability measure.

Data analysis

A series of Rasch analyses were performed using Winsteps software program version 3.57.2 (Winsteps.com, Chicago, USA) to calculate person measures for back pain and non-back pain groups (Linacre, 2005). The Rasch model transforms total raw scores into estimate of person ability in logits. To maximize the comparability of summative scores the ODQ instrument, the CAT measures were linearly transformed from the original logit estimates to a 0~100 metric scale.

All descriptive statistical and correlation analysis were calculated using SPSS software ver. 21.0 (SPSS Inc., Chicago, IL, USA). Pearson product moment correlations were obtained to compare the measurement properties of CAT (i.e., 10-item stopping rule and standard errorless than .40) and ODQ. Scatter plots of ability estimates for the CAT versus the ODQ measure were used to further examine these relationships.

To examine potential differences in precision across the three measures, the method of known groups validity to test relative precision (RP) in discriminating back pain and non-back pain groups was used. Methods included under the general linear model were used to test for hypothesized differences in group mean estimates. The magnitude of the F value from the analysis of variance represents a measure of precision. F statistics associated with chance probabilities $p<.05$ were considered significant. If the RP ratio is equal to 1, both methods of estimating function are equally discrimi. natory. If the RP>1 the measurement method in the numerator is superior in differentiating function compared to method in denominatorThe greater the F value, the greater the amount of systematic variance a measurement method accounts for and, therefore, the greater its ability to discriminate groups of subjects.

Results

Participant characteristics

Sample demographic characteristics and clinical information are presented in Table 1. The average age was 53.7 years for the back pain group and 48.7 years for the non-back pain group. Nearly 60% of participants reported having back pain more than a year indicating it was a chronic condition. Five percent of the non-back pain participants reported having another pain related condition.

Participants in the back pain group answered slightly more questions than those in the non-back pain group. For the back pain group, the average respondent answered 5.6 questions in the positioning/transfer construct, 6.3 questions in the lifting/carrying construct, and 6.2 questions in the

Table 1. Demographic characteristic of study participants (N=84)

Characteristics		Back pain group n_1=42 (%)	Non-back pain group n_2=42 (%)
Age			
	< 20	1(.4)	3(7.0)
	21～30	3(.71)	6(14.4)
	31～40	10(23.8)	9(21.4)
	41～50	8(19.0)	6(14.4)
	51～65	8(19.0)	8(19.0)
	> 65	12(28.6)	10(23.8)
	Mean±SD	53.7±20.1	48.7±19.7
Gender			
	Female	29(69.0)	27(64.3)
	Male	13(31.0)	15(35.7)
Education			
	Middle/Junior high	2(4.7)	0(0)
	High school	19(45.3)	14(33.3)
	College	12(28.5)	23(54.8)
	Graduate	9(21.5)	5(11.9)
Race/Ethnic			
	African American	7(16.6)	5(11.9)
	Hispanic American	1(2.3)	2(4.8)
	American Indian	1(2.3)	0(0)
	White, not Hispanic origin	32(76.2)	25(59.5)
	Asian/Pacific Islander	2(4.6)	10(23.8)
Years that has had related problems			
	Less than a year	14(33.3)	0(0)
	1 through < 4 years	5(12.0)	0(0)
	More than 4 years	20(47.6)	2(4.7)
Missing		3(7.1)	40(95.3)

Table 2. Correlation coefficients for CAT and ODQ measure for back pain group

	CAT[a] P/T[b]	CAT L/C[c]	CAT W/M[d]
CAT L/C	.837*		
CAT W/M	.614*	.647*	
ODQ[e]	.605*	.530*	.594*

[a]computer adaptive testing measure, [b]positioning/transfer construct, [c]lifting/carrying construct, [d]walking/moving construct, [e]Oswestry Back Pain Disability Questionnaire measure. *Pearson's correlation is significant at the .01 level (2-tailed).

Table 3. Difference between means for back pain and non-back pain groups

Measure	Mean±SE		F statistic	Relative precision
	Back pain	Non-back pain		
CAT[a] P/T[b]	49.83±.61	55.55±.61	41.76*	2.16
ODQ[c]	53.69±2.69	85.38±2.69	19.26*	1.00
CAT L/C[d]	50.24±.77	77.00±.77	27.36*	1.42
ODQ	53.69±2.69	85.38±2.69	19.26*	1.00
CAT W/M[e]	53.14±.89	58.33±.89	16.34*	.84
ODQ	53.69±2.69	85.38±2.69	19.26*	1.00

[a]computer adaptive testing, [b]positioning/transfer measure, [c]Oswestry Back Pain Disability Questionnaire measure, [d]lifting/carrying measure, [e]walking/moving measure. *F statistics is significant at the .001 level.

walking/moving construct. For non-back pain group, the average respondent answered 4.6 questions in the positioning/transfer construct, 5.1 questions in the lifting/carrying construct, and 5.4 questions in the walking/moving construct. The CAT administered more questions for back pain group than non-back pain group.

Correlations among the measures

In order to inspect the linear association between the measures, Pearson product moment correlations were calculated. Table 2 provides Pearson correlation coefficients between the CAT measures and the ODQ measures. Overall, the CAT measures had moderate correlations with the ODQ measures. All correlations were statistically significant at the $p<.01$ level.

In an auxiliary investigation of the linear relationships between the CAT and the ODQ measures, each pair of measures was plotted against each other (Figure 2). Scatter plots of the CAT and the ODQ measures clustered around the center of graph. In addition, the ODQ measures were more dispersed in the y-coordinate direction than other measures, while the CAT measures clustered into the center of the graph. As noted in Figure 2, the CAT measures had 22～36% less variance than the ODQ measure. Scatter plots of all relationships showed linear relationships.

Relative precision

Comparisons of the relative precision (RP) of the two measures to discriminate groups differing in back pain are presented in Table 3. As was hypothesized, the CAT measure for positioning/transfer construct achieved almost 2 times greater RP than did the ODQ measure. In addition, the CAT for the lifting/carrying construct had 42% greater RP in discriminating the groups, while the CAT measure for walking/moving construct had 16% less RP in discriminating the groups than did the ODQ. That is,

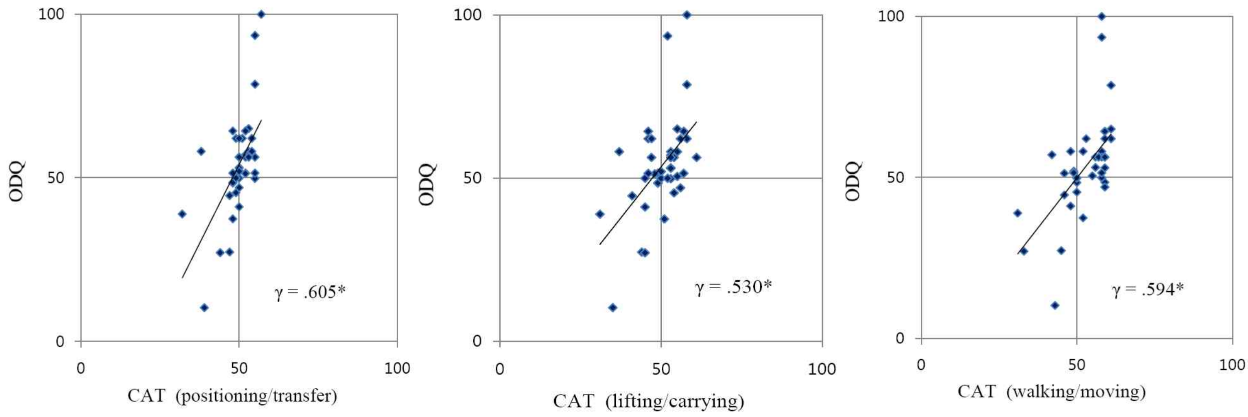

Figure 2. Scatter plot of ability measures from the CAT versus the ODQ measure. Figures represent the plot of ability measures for the CAT and the ODQ measure (CAT: computer adaptive testing, ODQ: Oswestry Back Pain Disability Questionnaire measure. *Pearson's correlation is significant at the .01 level.

The RP ratio for discriminating the groups did not favor the CAT measure for the walking/moving construct showing 16% less precision than the ODQ measure.

Discussion

The ODQ and its versions are widely used as outcome measures for disability resulting from back pain. They have been extensively cited more than 200 times in the Science Citation Index (Fairbank, 2000). Despite the popularity of the ODQ, numerous studies reveal substantial concerns regarding its measurement precision as well as measurement breadth (Davidson, 2008; Fairbank, 2007; White and Velozo, 2002). That is, the ODQ is often recommended to use for the assessment of a high severity group (Muller et al, 2006). This commonly leads to deficits in measurement precision.

Relative precision is based on the ratio of pairwise F statistics (an index of between subject variability to within subject variability) of two different measures. The magnitude of the F statistics from the analysis of variance represents a measure of precision. Thus, the relative precision estimates indicate how much more or less precise a measure is relative to another measure (McHorney et al, 1997). In this study, relative precision comparisons were conducted using known group validity in discriminating back pain and non-back pain groups. This known group validity addresses the extent to which a measure differs as predicted between groups who should score low and high on an ability trait. Supportive evidence of known-group validity typically is provided by significant differences in mean score across independent samples (Netemeyer et al, 2003). As was hypothesized, the results showed that the CAT measures achieved greater relative precision in discriminating back pain and non-back pain groups than did the ODQ measure except for with the walking/moving construct. This may indicate that CAT measures for positioning/transfer and lifting/carrying construct outperform the ODQ measure in terms of measurement precision. On the other hand, for the walking/moving construct, the CAT measure achieved less relative precision than did the ODQ measure in discriminating the groups. This indicates that CAT's ability for the positioning/transfer construct to discriminate between individuals in the back pain and non-back pain groups was twice as effective as the ODQ's ability. Likewise, the CAT measure for the walking/moving also achieved less relative precision than did the ODQ for this construct. This may indicate that individuals appear to report with higher rating rather than lower rating on the construct.

The present study has several limitations. Computer adaptive testing methods shorten test length by 62.5%, or require only an estimated nine items (Hol et al, 2007). In the present study, our CAT used much fewer items than the preset ten items and average respondents answered slightly more than 6 items for each construct. Since the standard error of CAT measures was not included in analysis, it is unknown whether all criteria were met to reach the person measure. Future research is needed to investigate the effects of adjusting the stopping rules to make them more rigorous, thus allowing more information to be obtained about respondents.

Conclusion

We attempted comparisons between the CAT measure as a modern measurement tool and the ODQ measure as a conventional instrument. Excluding the walking/moving construct, the CAT measure appeared to be more effective than did the ODQ measure in terms of measurement precision.

Acknowledgement

I am most grateful to Dr Craig A. Velozo, Medical University of South Carolina for the contribution in multiple comments and guides to this study.

References

Baker D, Pynsent PB, Fairbank JCT. The oswestry disability index revisited: Its reliability, repeatability, and validity, and a comparison with the st. thomas's disability index. In: Roland MO, Jenner JR eds. Back Pain: New approaches to rehabilitation and education. New York, Manchester University Press, 1989:174-186.

Bjorner J, Ware Jr JE. Using modern psychometric methods to measure health outcomes. Med Outcome Trust Monitor. 1998;3:12-16.

Davidson M. Rasch analysis of three versions of the oswestry disability questionnaire. Man Ther. 2008;13(3):222-231.

Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine (Phila Pa 1976). 1998;23(18):2003-2013.

Elhan AH, Oztuna D, Kutlay S, et al. An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. BMC Musculoskelet Disord. 2008;9:166.

Flynn KE, Dombeck CB, DeWitt EM, et al. Using item banks to construct measures of patient reported outcomes in clinical trials: Investigator perceptions. Clin Trials. 2008;5(6):575-586.

Haley SM, Gandek B, Siebens H, et al. Computerized adaptive testing for follow-up after discharge from inpatient rehabilitation: II. Participation outcomes. Arch Phys Med Rehabil. 2008;89(2):275-283.

Hart DL, Cook KF, Mioduski JE, et al. Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. J Clin Epidemiol. 2006;59(3):290-298.

Hol A, Vorst HCM, Mellenbergh GJ. Computerized adaptive testing for polytomous motivation items: Administration mode effects and a comparison with short forms. Appl Psychol Meas. 2007;31(5):412-429.

Fairbank JC. Revised oswestry disability questionnaire. Spine (Phila Pa 1976). 2000;25(19):2552.

Fairbank JC. The use of revised oswestry disability questionnaire. Spine (Phila Pa 1976). 2000;25(21):2846-2847.

Fairbank JC, Pynsent PB. The oswestry disability index. Spine (Phila Pa 1976). 2000;25 (22):2940-2952.

Fairbank JC. Use and abuse of oswestry disability index. Spine (Phila Pa 1976). 2007;32(25):2787-2789.

Hambleton RK. Emergence of item response modeling in instrument development and data analysis. Med Care. 2000;38(9 Suppl):II60-II65.

Haley SM, Andres PL, Coster WJ, et al. Short-form activity measure for post-acute care. Arch Phys Med Rehabil. 2004;85(4):649-660.

Haley SM, Coster WJ, Andres PL, et al. Score comparability of short forms and computerized adaptive testing: Simulation study with the activity measure for post-acute care. Arch Phys Med Rehabil. 2004;85(4):661-666.

Jette AM, Haley SM. Contemporary measurement techniques for rehabilitation outcomes assessment. J Rehabil Med. 2005;37(6):339-345.

Jette AM, Haley SM, Ni P, et al. Creating a computer adaptive test version of the late-life function and disability instrument. J Gerontol A Biol Sci Med Sci. 2008;63(11):1246-1256.

Kopec JA. Measuring functional outcomes in persons with back pain: A review of back-specific questionnaires. Spine (Phila Pa 1976). 2000;25 (24):3110-3114.

Liang MH, Lew RA, Stucki G, et al. Measuring clinically important changes with patient-oriented questionnaires. Med Care. 2002;40(4 Suppl):II45-II51.

Linacre JM. Winsteps rasch measurement computer program. Chicago, IL, winsteps.com, 2005:13-14.

McHorney CA. Generic health measurement: Past accomplishments and a measurement paradigm for the 21st century. Ann Intern Med. 1997;127:743-750.

McHorney CA, Haley SM, Ware JE Jr. Evaluation of the mos sf-36 physical functioning scale (pf-10): II. Comparison of relative precision using likert and rasch scoring methods. J Clin Epidemiol. 1997;50(4):451-461.

McHorney CA. Health status assessment methods for adults: Past accomplishments and future challenges. Annu Rev Public Health. 1999;20:309-335.

Müller U, Röder C, Greenough CG. Back related outcome assessment instruments. Eur Spine J 2006;15 Suppl 1:S25-S31.

Netemeyer RG, Bearden WO, Sharma S. Scaling Procedures, Issues and Applications. 1st ed. Thousand Oaks, California, Sage Publications Inc., 2003:78-93.

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, Kielhofner G, Lai JS. The use of rasch analysis to produce scale-free measurement of functional ability. Am J Occup Ther. 1999;53(1):83-90.

Velozo CA, Lai JS, Mallinson T, et al. Maintaining instrument quality while reducing items: Application of rasch analysis to a self-report of visual function. J Outcome Meas. 2000;4(3):667-680.

Velozo CA, Peterson EW. Developing meaningful fear of falling measures for community dwelling elderly. Am J Phys Med Rehabil. 2001;80(9):662-673.

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, Wang Y, Lehman L, et al. Utilizing rasch measurement models to develop a computer adaptive self-report of walking, climbing, and running. Disabil Rehabil. 2008;30(6):458-467.

Ware JE Jr, Kosinski M, Bjorner JB, et al.

Applications of computerized adaptive testing (CAT) to the assessment of headache impact. Qual Life Res. 2003;12(8):935-952.

Weiss D. Improving measurement quality and efficiency with adaptive testing. Appl Psychol Meas. 1982;6(4):473-492.

White LJ, Velozo CA. The use of rasch measurement to improve the oswestry classification scheme. Arch Phys Med Rehabil. 2002;83(6):822-831.

This article was received September 31, 2013, was reviewed September 31, 2013, and was accepted November 6, 2013.