한국전문물리치료학회지 2014년 21권 2호 74-81 ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
Phys Ther Korea 2014;21(2):74-81 http://dx.doi.org/10.12674/ptk.2014.21.2.074

Keyform Analysis of Rasch Measurement Accessible to Clinicians in Rehabilitation Clinics

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

Abstract

The versions of the Oswestry disability questionnaire (ODQ) is regarded as one of the most extensively used condition-specific instruments measuring disability resulting from low back pain. It has been shown to have adequate psychometrics, reliability, validity, and responsiveness as a whole, yet the summated total score of the instrument often provide little information to rehabilitation clinicians. A keyform analysis based on Rasch measurement model is an innovative way of illustrating the specific test items that an individual may or may not perform. By applying the keyform of the Rasch measurement model to the ODQ, rehabilitation clinicians may able to select more challenging ODQ items matching an individual's ability and document them as attainable treatment goals. The results demonstrated how a keyform analysis assist to setting possible goals for the treatment of individuals with low back pain. Forty-two individuals with low back pain were recruited from rehabilitation clinics in Gainesville, Florida. A series of Rasch analyses on the 10 items of the ODQ were performed using Winsteps software. The performance of two individuals on those 10 items was illustrated on the keyform. The keyform analysis of the Rasch measurement model may be translated into a useful tool for making clinical judgements.

Key Words: Assessment; Item response theory; Low back pain; Measurement; Rasch analysis; Rehabilitation.

Introduction

Most, if not all, existing assessments measuring disability resulting from back pain yield a summated score obtained by adding individual item scores. The score only provides a general sense of a person's disability or ability level (or disability) offering an understanding of the underlying construct as a whole. In order to attain specific information from the existing standardized assessments, item level psychometrics as well as how respondents perform on test items should be taken into account. Therefore rehabilitation clinicians would not be reluctant to use many standardized instruments for low back pain (Bossons et al, 1996; Holmqvist et al, 2009; Müller et al, 2006).

A study reported that less than half of the physical therapists incorporated newly created assessments into their practice despite concerted efforts to motivate the therapist to use the assessments. A reason for being that is because there is an inherent dissatisfaction with standardized assessments that have widespread use (Kay et al, 2001). That is, although the standardized assessments have a critical role in clinical rehabilitation practice, they may not be used unless there is any immediate benefit to rehabilitation clinicians. The restricted use of assessments in clinical setting may be influenced by several limitations such as test and sample dependent properties of classical test theory (CTT)-based measurement (White and Velozo, 2002).

In contrast to CTT, item response theory (IRT) models focus on item level statistics rather than the test as a whole. By applying the Rasch model (1-parameter IRT model), the item level statistics always represent the same thing and the same way

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

over time. Additionally they are invariant and consistent measures assessing the latent traits over time. It never changes like 'centimeters on a ruler' by using a unit of measurement called a logit (i.e., log-odds unit). The Rasch model allows the estimations based on the probability of selecting a particular rating of test items in relation to the relationship of item difficulty and person ability. Individuals with high ability will have higher probability of getting higher ratings on difficult items and individuals of low ability will have lower probability of getting lower ratings on easy items. The Rasch model places item difficulty relative to person ability on the same linear continuum. This leads to connect a respondent's rating on a particular item to the person's level of ability (Velozo and Peterson, 2001) and translates ordinal data into equal interval data. Consequently, the Rasch model forms the basis of a keyform that is available to rehabilitation professionals for instantaneous use.

The keyform is a term used by Michael Linacre who first proposed the idea (Institute for objective measurement, 1997). It orders the test items according to item difficulty relative to each other, which displays most difficult items at the top and easiest item at the bottom for each individual. Then it allows one to examine which items an individual is having challenges with and which they are not. For example, on an assessment, an individual with a lower person measure (i.e., an estimate for person ability) would have less ability than one with a higher person measure by selecting lower ratings.

Of the many condition specific measures for low back pain, only a few measures have commonly been accepted such as the Oswestry disability questionnaire (ODQ). The original version of the ODQ was first developed by John O'Brien in 1976 and originally called the Oswestry disability index (ODI). The ODI further updated multiple times for newer versions. The updated ODI includes 10 items of the level of pain, interference with personal care and physical functions (i.e., lifting, walking, sitting, and standing, sleeping, sex life, social life, and traveling) (Fairbanks and Pynsent, 2000). Fritz and Irrgang (2001) developed the ODQ by replacing 'sex life' item with 'employment/home making' item and indicated that the instrument is being much more sensitive to individuals with severe low back pain, while it appears to be less responsive to those with less severe low back pain (Fritz and Irrgang, 2001). However the ODQ has demonstrated ability to detect clinical change (responsiveness) at the level of the assessment as a whole (Fairbank, 2000; Fairbank et al, 1980; White and Velozo, 2002). Although it is presently considered as a "gold standard" instrument due to its popularity, it often fails to provide detail information on test items other than the summated score.

The purposes of this study are: 1) to demonstrate, using data on the ODQ, how Rasch measurement can be used to generate a hierarchy of item difficulty of the ODQ, 2) to visually inspect the hierarchical order as well as optimal rating scale responses of the ODQ items, 3) to show how two individuals with various ability responded on the keyform and how the keyform can be useful in clinical setting in terms of determining treatment goals.

Methods

Participants

Forty-two individuals with low back pain participated in this study. The participants were asked to complete the ODQ after screening for inclusion criteria. The criteria for participants with back pain included if: 1) currently experiencing low back pain, 2) having previously received treatment for back pain, 3) having ability to read and understand English, and 4) being age 18 years or older. The mean age of the study participants was 53.7 years ranging from 18 to 74 years. The participants included 29 females (69.0%) and 13 males (31.0%). The participants with more than a year of back related

problem were nearly 60.0% of our sample. Data collection sites included 3 rehabilitation clinics in Gainesville, Florida: University of Florida, Shands Orthopaedics and Sports Medicine Institute and Shands Rehabilitation Hospital. Prior to actual appointment for data collection, screening procedures for the inclusion criteria were performed. All selected participants presenting to the recruiting sites between November 3, 2009 and June 30, 2010 were recruited and scheduled for the data collection. This study was approved by the Institutional Review Board at the University of Florida (Approved by IRB #17-2009).

Instruments and measurements

The ODQ, a most widely accepted conventional back pain disability instrument, was the instrument used in this study. The ODQ is one of the most popular self-report condition specific instruments assessing how back pain impacts on patients' ability to manage daily life tasks (Fritz and Irrgang, 2001). The ODQ generally provides an indication of perceived disability resulting from back pain. The latest version of the ODQ contains ten items of pain intensity, personal care, lifting, walking, sitting, standing, sleeping, employment/home-making, and traveling. Participants respond on a 6 point ordinal scale according to how much difficulty individuals experience in their daily life from 5 to 0. For each test item, the possible score would be 5 (more disabled) if the last statement is marked and 0 (least disabled) if the first statement is marked. The total score (i.e., sum of all item responses) is converted to a percentage score ranging from 0 (no disability) to 100 (most severe disability).

To determine the dimensionality of ODQ, goodness of fit statistics were obtained using Winsteps® software program (Linacre, Chicago, IL, ver. 3.57.2). Given the small number of study participants, factor analysis was not possible. Dimensionality can be determined by scrutinizing mean square standardized residuals (MnSq) produced for each item. The residuals represents observed variance divided by expected variance. The optimal value of the residuals for an item is 1.0. The acceptable criterion can be determined by the intended purpose of the measure and the degree of rigor desired. For general survey study, the residuals between .6 and 1.4 is suggested by Wright and Linacre (1994). An item with low or high residual values suggests that the item may be redundant or not be belonging to the trait being measured.

In addition, a series of Rasch analyses were performed to calibrate the person measure and item difficulty. Rasch scores were linearly transformed from the original logit estimates. Rasch analysis places test items and persons on the same linear continuum along with the item calibrations and person measures. The general keyform output from the Winsteps program was generated as the basis for the creation of a data form displaying items on the right of the response categories. Participants' responses on the keyform were to illustrate how the completed form would look based on the two different disability levels.

Results

To provide an evidence that confirms the extent to which items represent a unidimensional construct, the fit statistics (i.e., mean square standardized residuals) were inspected. Table 1 presents the ODQ items in order of difficulty with fit statistics. All of the items on the ODQ fit to the Rasch model except for the employment and sleeping item. Of these two misfitting items, the sleeping item was nearly the acceptable range. However the employment item remains problematic with low fit statistics indicating that the item had more difficulty fitting the Rasch model.

Rasch item-person map depicting person ability to the hierarchical order of item difficulty is presented in Figure 1. The person abilities are displayed on the left side of the figure from the lowest at the bottom

Table 1. Fit statistics for the ODQ

Items	Difficulty (Logits)	Error	Infit MnSq[a]	ZSTD[b]	Outfit MnSq	ZSTD
Lifting	.80	.15	.92	-.4	.84	-.7
Standing	.49	.15	.95	-.2	.92	-.3
Pain	.30	.16	1.30	1.4	1.26	1.2
Employment	-.03	.16	.46	-3.0	.51	-2.5
Sitting	-.05	.17	1.23	1.0	1.23	1.0
Social Life	-.08	.17	.88	-.5	.93	-.2
Traveling	-.22	.17	.78	-.9	.82	-.7
Walking	-.25	.17	1.36	1.5	1.30	1.2
Sleeping	-.28	.17	1.42	1.7	1.19	.8
Personal care	-.68	.19	.67	-1.4	.70	-1.2

[a]mean square standardized residuals, [b]Z-score standardized.

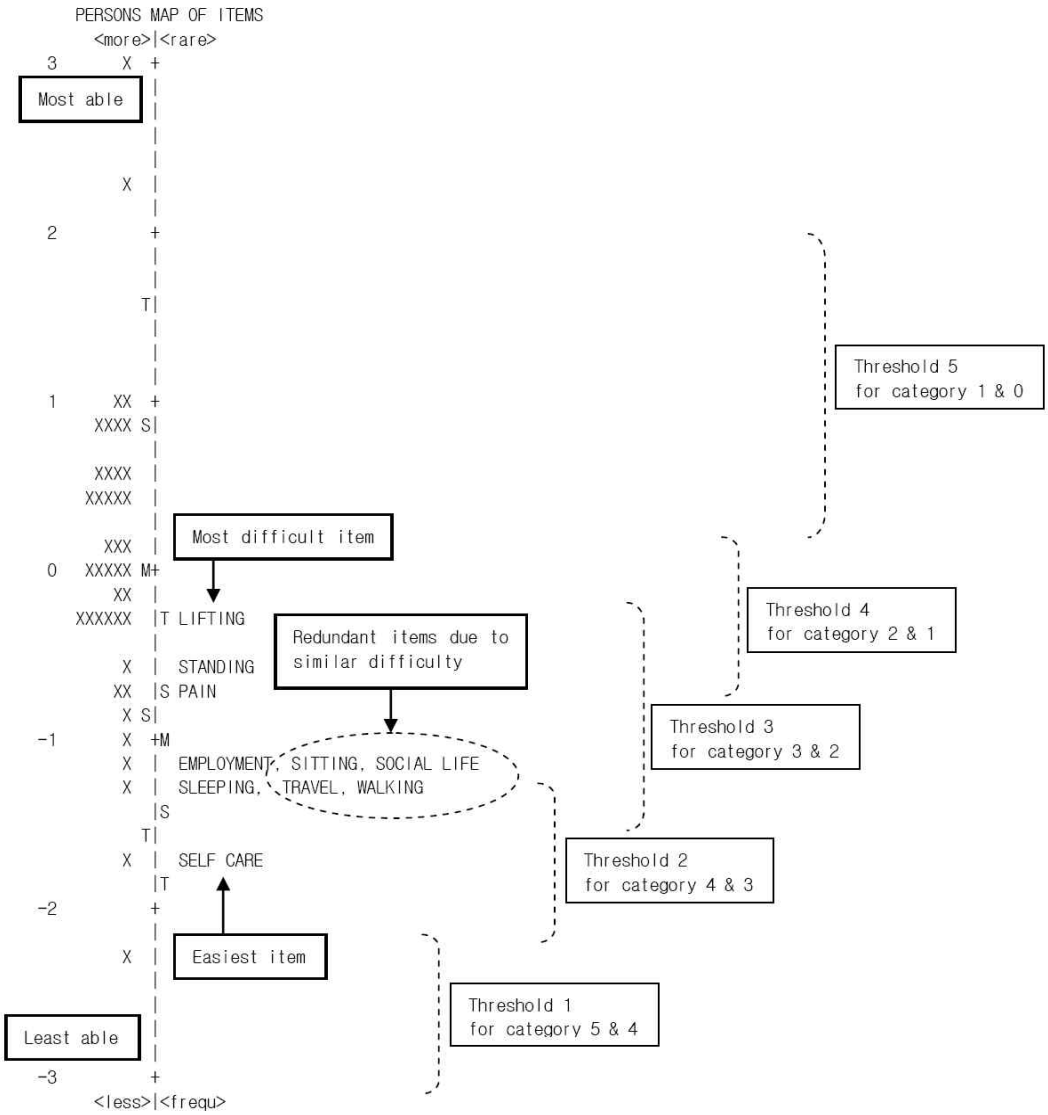

Figure 1. Item-person map for the ODQ with 6 response categories.

to the highest at the top, while the ODQ items are presented on the right side of the figure. As shown in the figure, lifting item was the most difficult, while self care item was the easiest task. This logical fashion of the item hierarchy was supported by the empirical evidences. These items are successfully measuring person ability indicating that the ODQ is sensitive to the wide range of disability level. Additionally 6 response categories of all items of the ODQ were measuring nearly the whole range of ability level for our sample.

The key form resulting from the Rasch analysis illustrated in Figure 2. The figure showed how logits and category response were linked to the test items in hierarchical order. Items are located at the right in order of decreasing difficulty. Response categories (0 to 5) are presented to the left of each item. These categories are placed at a location relative to the person ability scale at the bottom. Two respondents with different levels of disability (i.e., moderately and mildly impaired in our sample) resulting from back pain were presented. The person B with a person measure of +3.34 logits, mildly impaired, would be most likely to rate the pain item with 0 category (I have no pain at the moment) and the lifting/standing items with more difficulty than 0 category. Thus the patient had no difficulty with all but two items, that are lifting and standing items. These two items would be appropriate goals for treatment plan in rehabilitation clinics. Similarly, the person A with a person measure of +1.12 logits, moderately impaired, would most likely to rate the travel/walking/sleeping

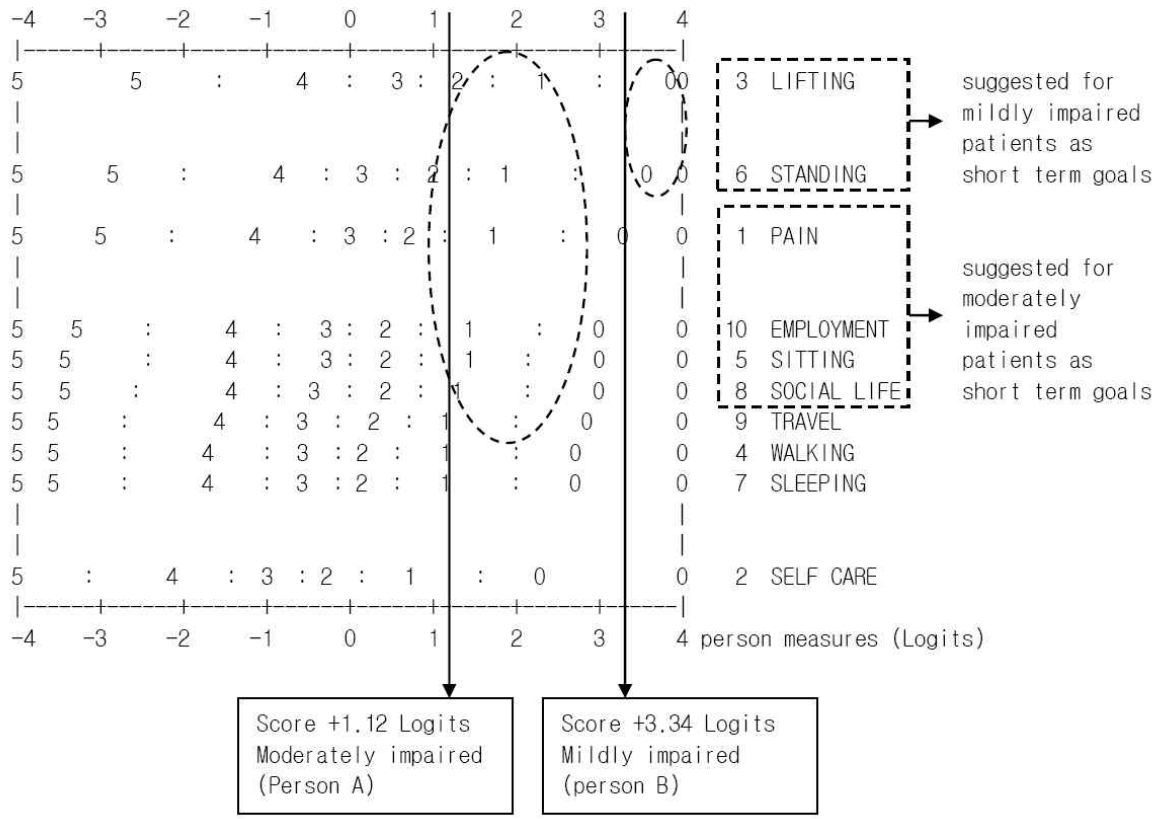

Figure 2. The general keyform analysis for 2 different disability levels (i.e., person measure +1.12 and +3.34 logits). The keyform demonstrates that most difficult items (i.e., lifting and standing) would be considered for person B and moderately difficult items (i.e., pain, employment, sitting and social life) could be rendered for person A as potential short term goals of a rehabilitation treatment plan.

items with category 1 (slightly more challenge than category 0) and more difficult items than those three items with category 2 or 1, that would be more challenging categories. The person A had slightly more challenge with travel/walking/sleeping items (i.e., rated with category 1) and would have more challenge with more difficult item than travel item.

Discussion

In summary, the goals of this study were to demonstrate how the keyform of Rasch analysis can be translated into a useful tool in establishing treatment goals. The keyform illustrated how study participant's responses to test items would look on the form in terms of different ability levels and how the form can be used in setting patient goals. Prior to addressing these goals, empirical evidences were presented to support the use of Rasch analysis for the creation of the keyform. Two evidences which were fit statistics for model fit and item hierarchy were tested, These were model fit testing if all items contribute to a single underlying trait and logically present the hierarchy of item difficulty.

Rasch analysis revealed that all items of the ODQ fit to the Rasch model for measuring disability resulting from low back pain except employment item (infit/outfit: .46/.54). The ODQ, a modified version of ODI, in which sex life was replaced by employment item showed inadequate fit to the Rasch model. This finding suggest that the item do not perform as intend to measure within the instrument. The sleeping item misfit to the model (infit: 1.42) as well, however it is a nearly acceptable range. Compared with previous studies, pain item in the ODQ and versions of it has been rendered problematic due to the issue of taking painkillers for pain (Page et al, 2002; White and Velozo, 2002). That is, respondent's response pattern on the pain item was erratic and unpredictable since pain intensity depends upon the use of pain medications. Therefore respondents were uncertain how to respond to the pain item.

The item difficulty hierarchy provides clinicians more meaningful information than does ODQ instrument. By presenting test items in order of difficulty, one can suggest that difficult items require more challenge due to dealing with the use of multiple joints, environmental factors and complex functional activities in terms of motor control perspectives (Velozo and Woodbury, 2011). In this study, the item difficulty hierarchy included personal care item as an easiest item, followed in order of difficulty by sleeping, walking, traveling, social life, sitting, employment, pain, standing and lifting. Therefore, logically, it would be expected that for individuals with low back pain, lifting and standing would be more difficult than personal care and sleeping activities. This pattern of item difficulty was consistent with the hierarchy of several previous studies (Lu et al, 2013; White and Velozo, 2002).

In regards to the response pattern of the keyform, respondents do not rate with higher ratings on easy items and lower ratings on difficult items. As can be seen in the figure, responses to particular items do not fit ideal response pattern such as a deterministic Guttman's scale model (Nunnally and Bernstein, 1994). The Guttman's scale, previously known as cumulative scaling, constructs a matrix that includes the responses of all the respondents for all of the items (Kline RB, 1999). The respondents who earn highest score are listed on the top and those who earn lowest score are at the bottom. Then the person with highest score would demonstrate a pattern showing highest ratings on easy items, while getting lowest ratings on difficult items (i.e., the response of 'I can lift heavy weights without extra pain' to lifting item and the response of 'I am unable to do any washing without help' to self care item). This table would allow one to perfectly predict item responses. However this perfectly matched matrix does not exist in real world but similar patterns of the matrix do.

The use of probabilistic Rasch model versus the Guttman's scale model is supported by the findings

presented. By using the keyform, it is possible to get a general idea of an respondent's overall capability to perform on particular items based on the probabilistic Rasch model. On the keyform figure presented, a line be drawn by person ability measure could be estimated. Such a line could be estimated even with missing responses to some items (Kielhofner et al, 2005). The line may allow clinicians to visually predict respondent's next levels of challenging items as setting goals as depicted in Figure 2. In realty, some respondents' patterns may differ from what would be expected based on the hierarchy obtained. However such a keyform may give rehabilitation professionals a general sense of what functional activities or physical activities an individual is having difficulty with completing. By visually scrutinizing the form along with a respondent's ability level estimated (i.e., the ability line with 3.34 logits), one could easily see that the respondent appears to have more difficulty with lifting than standing. Additionally, a moderately impaired respondent with 1.12 logits appears to have less difficulty on self care than sleeping, walking, travel items. Therefore more challenging items such as social life, sitting, employment and pain item would be suggested for the individual for short term goals. Once these goals are achieved later, sitting or lifting item could be considered for next short term goals or long term goal (s) in a treatment plan.

Future directions for this study would include investigations on: 1) the psychometric properties of the ODQ based on a larger data set of persons with low back pain; 2) the feasibility of performing factor or principle component analysis to determine the unidimensional construct of the ODQ.

Conclusion

This study attempt to apply the Rasch model on the ODQ and create a clinically useful form. The item level psychometrics as well as item difficulty hierarchy should be taken into consideration in order to make the ODQ valuable in clinical settings. Additionally, the keyform analysis would be a useful tool for making clinical judgement in terms of goal setting in a rehabilitation treatment plan.

References

Bossons CR, Levy J, Sutterlin CE 3rd. Reconstructive spinal surgery: Assessment of outcome. South Med J. 1996;89(11):1045-1052.

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

Fairbank JC, Couper J, Davies JB, et al. The oswestry low back pain disability questionnaire. Physiotherapy. 1980;66(8):271-273.

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

Fritz JM, Irrgang JJ. A comparison of a modified oswestry low back pain disability questionnaire and the quebec back pain disability scale. Phys Ther. 2001;81(2):776-788.

Holmqvist K, Kamwendo K, Ivarsson AB. Occupational therapists' descriptions of their work with persons suffering from cognitive impairment following acquired brain injury. Scand J Occup Ther. 2009;16(1):13-24. http://dx.doi.org/10.1080/11038120802123520

Institute for objective measurement. Instantaneous measurement and diagnosis [Internet]. Durham, 1997 June [Cited 2014 May 21]. Available from: http://www.rasch.org/memo60.htm

Kay TM, Myers AM, Huijbregts MPG. How far have we come since 1992? A comparative survey of physiotherapists' use of outcome measures. Physiother Can. 2001;53:268-275.

Kielhofner G, Dobria L, Forsyth K, et al. The construction of keyforms for obtaining instantaneous measures from the occupational performance history interview rating scales. OTJR (Thorofare NJ). 2005;25(1):23-32.

Kline RB. Book review: Psychometric theory (3rd ed.). J Psychoeduc Assess. 1999;17(3):275-280.

Lu YM, Wu YY, Hsieh CL, et al. Measurement precision of the disability for back pain scale-by applying Rasch analysis. Health Qual Life Outcomes. 2013;11(119):1-10. http://dx.doi.org/10.1186/1477-7525-11-119

Müller U, Röder C, Greenough CG. Back related outcome assessment instruments. Eur Spine J. 2006;15(Suppl1):S25-S31.

Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York, McGraw-Hill, 1994:275-280.

Page SJ, Shawaryn MA, Cernich AN, et al. Scaling of the revised oswestry low back pain questionnaire. Arch Phys Med Rehabil. 2002;83(11):1579-1584.

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

Velozo CA, Woodbury ML. Translating measurement findings into rehabilitation practice: An example using Fugl-Meyer assessment-upper extremity with patients following stroke. J Rehabil Res Dev. 2011;48(10):1211-1222.

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

Wright BD, Linacre JM. Reasonable mean-square fit values. Rasch Measurement Transactions. 1994;8(3):370.

This article was received April 7, 2014, was reviewed April 7, 2014, and was accepted May 7, 2014.