한국전문물리치료학회지 2015년 22권 3호 61-70 ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
Phys Ther Korea 2015;22(3):61-70 http://dx.doi.org/10.12674/ptk.2015.22.3.061

Selecting Common Items for Linking the Oswestry Low Back Pain Questionnaire and a Short Form of Self-Reported Activity Measure for Low Back Pain

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

Abstract

To develop an effective and efficient measurement system for tracking changes of functional status across two measures, it is essential to integrate information and communicate scores across two measures. The lack of communication between two measures leads to score incompatibility. A potential solution would be the development of a crosswalk table between those measures. Prior to creating a crosswalk table, selecting common items between two measures is critical. By using the Oswestry low back pain disability questionnaire (Oswestry) and a short form measuring disability resulting from low back pain, item level statistics as well as differential item functioning (DIF) using the Rasch measurement were investigated. Eighty-two participants with known group validity were recruited. Based on the application of the Rasch measurement model, item difficulties across the two measures were logically and hierarchically ordered. Ceiling effects for both measures were detected, which were not be able to be effectively measured with the two measures. The DIF analysis across the two measures confirmed that five paired items were found to have DIF and five common items were selected for common items. Although five paired items function differently across the Oswestry and the short form, all items of both measures were well targeted study participants. The common items selected by the Rasch measurement model may be effective when creating a crosswalk table between the Oswestry and the short form.

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

Introduction

A myriad of functional assessments related to low back pain (LBP) has been developed over the past few decades. These assessments typically measure the health domain of disability caused by LBP conditions. Despite the proliferation of such assessments, scores from one instrument cannot be compared to similar scores gathered by another instrument (McHorney, 2002). This lack of "communication" across those instruments may impede clinicians from choosing an efficient instrument that meets the selection criteria for their clinical settings. The incompatibility of assessments prohibits rehabilitation services in monitoring and evaluating the effectiveness of clinical intervention for LBP. Furthermore, hospitals or rehabilitation clinics that use different instruments cannot be compared relative to their outcomes. The most important reason is that being incompatible arise from the different sensitivities of those instruments. That is, most instruments, if not all, are often created to target the average patients, and therefore likely to be more sensitive on the patients with average ability than one with low or high ability (Choi, 2012; White and Velozo, 2002). Other reasons for being incompatible may include that the assessments are specifically developed for the persons they are targeting (Kolen, 2004).

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

Linking, equating in another word, refers to the statistical procedure rendering the need for comparable scores between two assessments. It has traditionally been introduced to compare two different standard assessments in education fields such as the college entrance examinations, the Scholastic Aptitude Test or the American College Test. The admission office of most colleges selects students based on the interchangeable assessment scores. Hence, conversion tables for the scores were developed by researchers to allow communicating between the scores. Similarly, communication between functional assessments in health care settings became a decisive issue determining common items. That is, two assessments measuring the functional status of individuals share common items where the individuals take the assessments that include additional common items to all assessments. For example, researchers suggested to select items that are at the average person's ability level to avoid items that are too easy or too difficult and to spread the linking items across the person ability continuum (Smith and Kramer, 1992). Other researchers used common-item equating a survey of functional status items where a set of items was included on each of the three assessments and a set of common items was administered to all groups (McHorney and Cohen, 2000). Likewise, since the common items are the only link between different assessments, choosing linking items may be a critical issue. In addition, identifying common items and removing misfit items based on individual response pattern were emphasized due to their potential misinterpretations.

Linking two assessments in health care has traditionally been investigated on previously developed functional measures. While those measures are developed to assess the same construct, they are developed differently in many aspects including the number of test items, rating scale categories and item definitions. Since decisions to link the assessments in health care occur following assessment development, researchers or clinicians may have to confront the inherent differences that exist between those assessments. Linking assessments in health care have therefore focused on creating a translation between each item on the assessments. For example, a crosswalk between the functional independence measure (FIM) and the minimum data set (MDS) using an expert panel was created by choosing MDS items, re-scaling and matching to similar FIM items (Velozo et al, 2007). Furthermore, the conversion algorithm was developed based on item-to-item rating scale category comparison (Wang et al, 2008). The reported correlation coefficients between the items later provided an indication of how well scores of one assessment can be predicted from scores of the other assessments, which may indirectly imply how well the algorithm may act in the assessments (Williams et al, 1997). Using similar methods, Barthel index score (i.e., 4-point rating scale) was compared with 7-point rating scale FIM items, which converted to a corresponding 4-point rating scale. The agreement between the converted and actual scores ranged from 75 to 100% and kappa statistics were varied ranging from .53 to 1.00 (Nyein et al, 1999). These equating procedures are based on the rationale that both sets of items are measuring the same construct.

The Oswestry LBP disability questionnaire (Oswestry) has been considered a 'gold standard' instrument measuring the impact on patients' abilities to manage daily life tasks (Fritz and Irrgang, 2001). Its original version was first created by John O'Brien in 1976, then known as Oswestry Disability Index, and published by Fairbanks and colleagues (Fairbanks et al, 1980). An updated version of the Oswestry was later developed by Fritz and Irrgang (2001) which replaced sex item with the employment item. Despite their popularity, several studies indicated a shortcoming of the summated total score yielded. That is, it was unable to provide detailed information on individual test items, only an overall sense of an individual's functional status (Devillis, 2006; Hambleton, 2000). In addition, since the Oswestry was created differently compared to other

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2015.22.3.061

back-related disability instruments, the total score yielded is unable to be compared to other similar instruments. Consequently, by only comparing the total scores, there is no sense in how similar items between two instruments similarly respond across the instruments (Davidson, 2008; Lu et al, 2013). The limitation depends on the set of items selected (test-dependent) and the scores obtained from a sample reflect characteristics of only the sample, which cannot be compared across other samples (sample-dependent) (White and Velozo, 2002). These characteristics often lead to failure to providing item level psychometrics such as how individuals respond on the individual items or how the relative items respond similarly across different assessments. These are due to a function of short-comings in conventional classical test theory (CTT)-based assessments.

Compared to conventional instruments based on CTT-based models, the Rasch model (1-parameter item response theory model) greater emphasizes item level psychometric properties, person and item characteristic, than the test as a whole (Velozo et al, 2006). The item level psychometric properties obtained from the Rasch model are estimated by the probability of an individual's response to test item. Due to the application of the probabilistic mathematical model, the estimated person and item difficulty measures always represent invariant measures over time. These estimated measures are presented as a unit of measurement called a logit (i.e., log-odds unit). The logit scale never changes to whomever one may apply (i.e., sample free measure) and with whichever these measures asses (i.e., scale-free measure). In turn, the invariant property of item difficulty enable one to investigate how well the item difficulties match with the sample person abilities (Taherbhai and Young, 2004).

Differential item functioning (DIF) is a statistical procedure identifying items that appear to be having difficulty levels that are dependent on membership to a particular group after controlling for the ability levels of groups being compared (Finch and Hernandez Finch, 2014; Huang, 2014; Teresi, 2001). The Rasch model has a strong assumption that item discrimination parameters are equal across all items. This assumption makes the Rasch model able to detect DIF, since item characteristics should be consistent across different patient groups, where there is a relative advantage for one group over the other group through the entire ability range. The DIF methodology has become central to the investigation in health-related measurement fields to compare response patterns across group membership (Fleishman and Lawrence, 2003; Haley et al, 2004). This leads to the investigation in the consistency of item performance across two instruments.

The purposes of this study were: 1) to investigate how the Rasch measurement model can be applied to determine the hierarchical order of the Oswestry and the items of a newly created short form, 2) to demonstrate how the items of the measures differently function among the measures, 3) to delineate the common items to be selected for equating the Oswestry and the short form.

Methods

Study participants

The six of 10-item short form were created from a research that created ICFmeasure.com. The web-based program was based on a paper and pencil version of self-reported questionnaire with 255 items as well as 6 domains for 101 individuals with LBP. A newly created short form for walking/moving domain was compared to 10 items of the Oswestry. For a direct comparison of item level psychometric properties, the short form and the Oswestry were applied to eight-four participants for known group validity (42 participants for each back pain and non-back pain groups). Inclusion criteria for participants with LBP were: 1) currently having LBP, 2) having previous treatments for LBP, 3) having ability to read and understand English, and 4) being

age 18 years or older, while the criteria for participants without LBP were: 1) currently having no LBP, 2) able to read and understand English, and 3) 18 years of age or older. The study was approved by the Institutional Review Board at the University of Florida (approval number: #17-2009).

Instrumentation

The Oswestry used in this study includes ten items of pain intensity, personal care, lifting, walking, sitting, standing, sleeping, employment/home-making, and traveling. The rating scale was categorized from a 5 (most disabled) to 0 (least disabled) ordinal scale based on how much difficulty one experiences in daily life. The summated total score yielded is then converted to a percentage score ranging from 0 (no disability) to 100 (most severe disability). These response categories were reversely coded so that high scores on particular items semantically indicated high levels of activity (i.e., less disabled). A 10-item short form of walking/moving which was developed from the paper and pencil version of the ICFmeasure.com measuring activity limitations was administered to 84 individuals. The short form contains 10 items for an underlying construct of activity limitation perceived by individuals with LBP. Participants were asked to respond to one of four categories; '3' (no difficulty), '2' (some difficulty), '1' (a lot of difficulty), and '0' (have not done or never have opportunities to perform). The original logit measure of the short form was later converted to 0~100 scale to be compared with the Oswestry measure.

Winsteps® software program version 3.57.2 (Linacre, Chicago, IL, USA) was used to determine the dimensionality of the Oswestry and the two short forms. Fit statistics from the Rasch measurement model determines dimensionality by scrutinizing mean square standardized residuals (MnSq), which represents observed variance divided by expected variance (Bond and Fox, 2001). Wright and Linacre (1994) suggested that acceptable ranges for the fit statistics would be between .6 and 1.4 for general survey data, while the optimal value of the fit statistics for an item is 1.0. By using UMEAN and USCALE (user-set mean and user-scaled unit respectively) commands of the Winsteps® software program, the logit measures of the short forms were transformed into 0~100 scale and allow meaningful comparisons between the short form and the Oswestry with a same metric system. An item with low or high fit statistics suggest that the item may be redundant or not belong to the underlying construct being measured. Rasch measurement model linearly converts raw scores into logit estimates and places items in the hierarchical order of item difficulty on a linear continuum along with person ability measures (i.e., person-item map). That is, by plotting logit measures of items and persons on the linear continuum, the person-item map can reveal how well the items capture the disability levels and visually inspect any ceiling or floor effects.

The DIF method used in this study was based on the differences between two parameters calibrated on the two relevant items (Wright and Stone, 1979). In general, the DIF findings are similar to those found when including the misfit items. When comparing the Oswestry and the short form in the pairs of item calibrations and associated estimates of the standard error of estimate from the Rasch model, the DIF analysis can determine whether hierarchically ordered test items of both instruments have more difficulty or not. Thus, the analysis allows clinicians or researchers to select the best relevant items measuring disability resulting from LBP.

Results

Dimensionality

By investigating the extent to which items represent the dimensionality of the Oswestry, the fit statistics from the Rasch analysis were inspected. Table 1 presents the Oswestry items in the order of item calibrations with the fit statistics. The item calibrations of all items were ranged from 40.48 (easiest) to 58.70 (most diffi-

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2015.22.3.061

Table 1. Fit statistics of the Oswestry

Items	Difficulty (logits)	SE[a]	Infit MnSq[b]	ZSTD[c]	Outfit MnSq	ZSTD
Lifting	58.70	1.99	.86	-.7	.93	-.3
Standing	55.56	1.93	1.05	.3	.99	0
Pain	52.72	1.94	1.07	.4	.97	-.1
Employment	51.10	1.87	.77	-1.3	.69	-1.7
Sitting	47.30	2.00	1.22	1.2	1.14	.7
Social life	46.54	1.97	1.18	1.0	1.06	.4
Travel	46.11	2.01	.86	-.7	.87	-.6
Walking	45.90	2.03	1.20	1.1	1.11	.6
Self care	44.38	2.01	.92	-.3	.92	-.3
Sleeping	40.48	2.25	.99	0	.89	-.4

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

Table 2. Fit statistics of the walking/moving short form

Items	Difficulty (logits)	SE[a]	Infit MnSq[b]	ZSTD[c]	Outfit MnSq	ZSTD
Running one block	73.33	2.19	1.72	3.1	1.62	2.1
Climbing up or down a 6-foot ladder	72.88	2.19	1.00	.1	.94	-.1
Climbing up or down a 3-step stool	55.59	2.12	.89	-.5	.82	-.8
Walking 4~8 blocks	54.74	2.13	.71	-1.6	.68	-1.5
Climbing down one flight of stairs	53.04	2.14	1.01	.1	1.10	.5
Walking 2~4 blocks	45.31	2.32	.89	-.5	.75	-.8
Walking crowded place	41.98	2.43	.88	-.5	.78	-.6
Stepping up or down a standard curb	35.17	2.66	.99	.0	.92	.0
Walking on carpet	28.44	3.04	1.40	1.5	.81	-.1
Walking within home/living environment	28.42	3.04	.75	-1.0	.63	-.5

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

cult). All items of the Oswestry fit to the Rasch model.

Table 2 presents items of the walking/moving short form with fit statistics representing that an item was (running one block) misfitted. The item calibrations of the most and least challenging items were, 69.36 and 24.45, respectively. In a comparison of the Oswestry versus the short form, the Oswestry showed a slightly less a difficulty range than the short form.

Table 3 represents the hierarchically paired items of the Oswestry and the walking/moving short form in logit scale. Items difficulties for 5 items (pain, employment, sitting, social life, and travel items) were similarly calibrated those of the short form (climbing up or down a 3-step stool, walking 4~8 blocks, climbing down one flight of stairs, walking 2~4 blocks, and walking crowded place). These

Table 3. Item difficulty of the Oswestry and the short form

Oswestry items	Difficulty (logits)	Difficulty (logits)	Short form items
Lifting	58.70	73.33	Running one block
Standing	55.56	72.88	Climbing up or down a 6-foot ladder
Pain	52.72	55.59	Climbing up or down a 3-step stool
Employment	51.10	54.74	Walking 4~8 blocks
Sitting	47.30	53.04	Climbing down one flight of stairs
Social life	46.54	45.31	Walking 2~4 blocks
Travel	46.11	41.98	Walking crowded place
Walking	45.90	35.17	Stepping up or down a standard curb
Self care	44.38	28.44	Walking on carpet
Sleeping	40.48	28.42	Walking within home/living environment

paired items were considered as common items for two measures. Excluding the 5 items, however, the other 5 items were differently calibrated.

DIF analysis for selecting common items

DIF analysis revealed that 5 items exhibited significant DIF (Figure 1). When comparing the paired items of the short form to the Oswestry, five items demonstrated significant DIF. Of the items showing significant DIF, three easy paired items plotted in the third quadrant were "stepping up or down a standard curb/walking", "walking on carpet/self care", and "walking within home/living". In addition, the two most difficult items located in the first quadrant were "running one block/lifting" and "climbing up or down a 6-foot ladder/standing". The other five items which

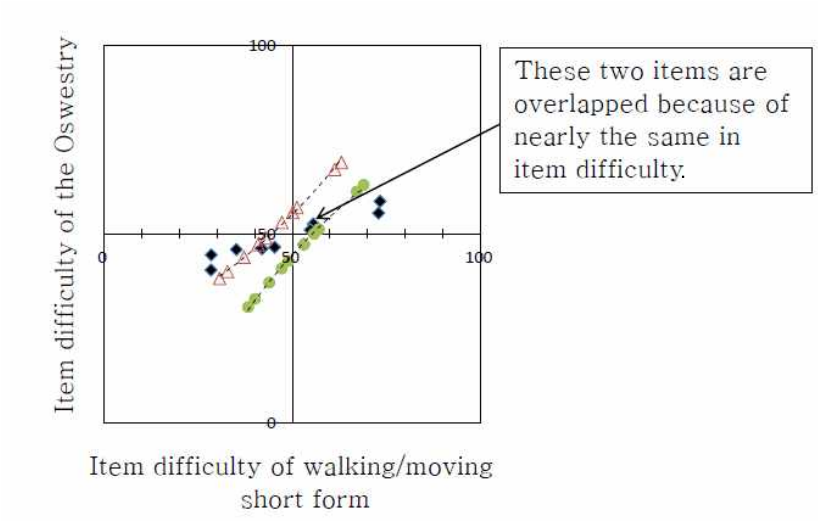

Figure 1. Differential item functioning (DIF) plots for the Oswestry items versus the walking/moving short form (The dashed lines connecting the triangles represent the upper and lower 95% confidence intervals. The circled 5 items were selected for common items, which were climbing up or down a 3-step stool/pain, walking 4~8 blocks/employment, climbing down one flight of stairs/sitting, walking 2~4 blocks/social life, walking crowded place/travel. The other 5 items represent DIF responding differently across the two measures. The measures were converted to 0~100 score from logits following the Rasch analysis.).

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2015.22.3.061

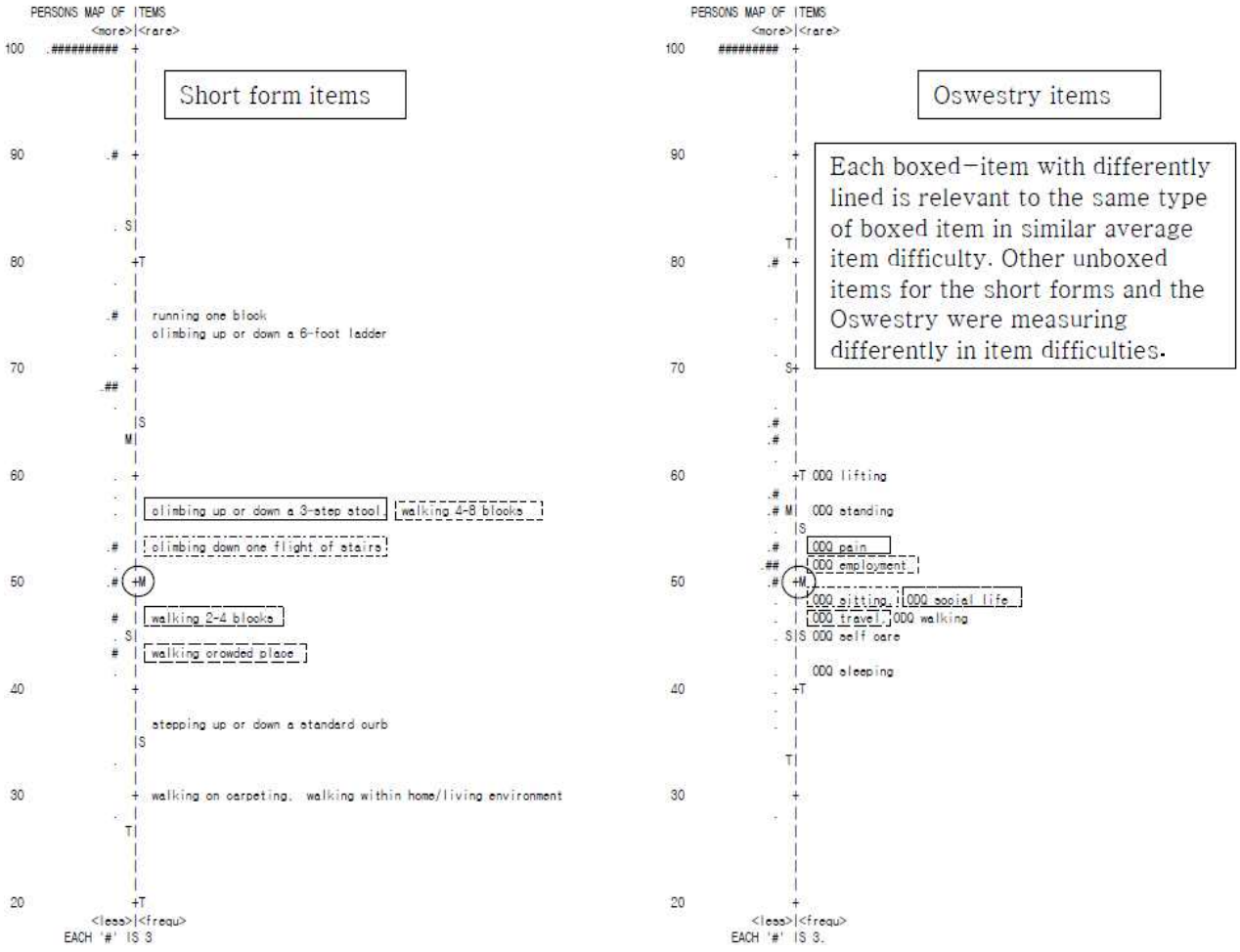

Figure 2. Item-person map of the walking/moving short form (left) versus the Oswestry (right) (The graph represents item difficulty measures on the right side of each map with the 0~100 converted score in logit and the person ability measures on the left side following the Rasch analysis. Each analysis is anchored on the average item difficulty measure to 50 for comparisons. The circled 'M' represents the average item calibrations for both measures.).

were plotted within the upper and lower 95% confidence intervals were selected for common items to equate the Oswestry and the short form. These items were hierarchically presented as well as persons in Figure 2.

Hierarchical order of the common items for the Oswestry versus the short form

Figure 2 presents the hierarchical order of item difficulty of the Oswestry and the walking/moving short form. The Rasch analysis placed persons and items onto the same linear scale with the same local origin. The average item difficulty was anchored to "0" for direct comparisons between two measures. Ceiling effects were noted for both measures which were unable to measure persons with high ability. However, all items of both measures targeted persons with middle and low ability (i.e., middle and high disability). The five common paired items previously stated from Figure 1 measured persons with persons with slightly low ability.

Discussion

A direct comparison of a single measure to another measure may be necessary to determine the score

compatibility across the measures. The score incompatibility generally lead to the development of new instruments that meet the criteria of rehabilitation clinicians and their rehabilitation programs. That is, these inconsistent instruments developed had later been proven to be cumbersome measurement tools. For maximizing the effectiveness of a single measure, linking or equating across the measures may ultimately be a promising method. A newly created short form of an activity measure for LBP (Choi, 2014) was attempted to link the Oswestry measuring disability resulting from LBP.

The present study demonstrated that acceptable fit statistics were exhibited for the Oswestry and the short form except "running one block" item of the short form. This indicates that the participants' responses to the item were not predictable. Otherwise, the intent of the item was to measure a latent trait other than the walking/moving construct. This finding would prompt further investigations to clarify if the item of "running one block" would be appropriate in the short form. Despite the unacceptable value obtained, the item was included for further analyses due to the necessity of the most challenging item within the short form. The criterion for the acceptable fit statistics in the study was ≥1.4 or ≤.6 for infit or outfit square (MnSq), which was suggested by Bond and Fox (2001). Additionally, although the "pain" and "employment" items of the Oswestry have traditionally been reported within the instrument (Lu et al, 2013; White and Velozo, 2002), all items except the item of "running one block" were fit to the Rasch model.

The hierarchically ordered item difficulties of the Oswestry and the short form were supported in a logical fashion. That is, participants in the present study had a tendency to rate their disability on the difficult items (i.e., lifting and running one block items) more severly relative to the easy items (i.e., sleeping and walking within home/living environment). The lifting activity is typically one of the leading causes of back related injury compared to the injury from during sleeping or walking activities. This empirical evidences also supported a postulation that item difficulties across two measures may possibly be calibrated at the same time. In fact, item difficulties of the five items across these two measures were similarly calibrated (boxed items in Table 3). In empirical evidences, participants would have more difficulty with "running one block" activity than lifting activities or appear to have more difficulty with "climbing up/down a 6-foot ladder" than standing activities. In addition, participants appeared to be somewhat slightly less influenced by the Oswestry items than the short form items ranged from 58.70 to 40.48 versus from 73.33 to 28.42. That is, the Oswestry items were targeting more on average persons than the shot form, while the short form items were measuring slightly more on the persons with high and low ability.

As previously demonstrated in tables and figures, five items were selected for common items that showed similar item calibrations and further analyzed to determine if any paired items across the Oswestry and the short form would differently function as their ability levels are controlled. The DIF analysis showed that participants differently perceived those items of "stepping up or down a standard curb/walking", "walking on carpet/self care", "walking within home or living environment/sleeping", "running one block/lifting", and "climbing up or down a 6-foot ladder/standing". That is, these participants perceived "lifting" function item of the Oswestry to be easier than the "running one block" item of the short form (58.70 versus 73.33 logits), while they rated the "sleeping" item of the Oswestry to be more difficult than the "walking within home or living environment" item of the short form (40.48 versus 28.42 logits). In addition, the walking item of the Oswestry showed DIF in the present study. The finding was somewhat expected because the walking item has traditionally been reported as a DIF item within the Oswestry (Davidson, 2008). Except for the DIF items, five paired items were identified for common items. These common items are later to link or equate the Oswestry and the short form. While sample dis-

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2015.22.3.061

tributions were similar, however, individual item comparisons fell short of expectations. Only half of the paired items of the two measures were within a 95% confidence interval. This lack of "connection" between two measures was probably due to severe ceiling effects on the both measures (34 persons for the short form and 27 persons for the Oswestry). That is, if individuals are at the ceiling of both measures, we would expect a better estimate despite the subtle difference in disability levels resulting from LBP. However, item difficulty range of the short form pushed study participants into wider range of the measure than did the Oswestry. This may mean that the Oswestry items converge into average individuals, while the short form items spread out slightly wider ranges of the disability level.

A limitation of the current study is the application of the Oswestry and the short form to participants with no back pain despite the use of known group validity. Those two measures were originally intended to detect a potential disability status resulting from LBP. This method had led participants with low disability level (or high ability) to high ceiling effects on both measures. Another limitation would be on the issue of local independence assumption of the short form items. This must have violated the assumption of the item response theory model. That is, test items should be locally independent from which test items in the short form should not be related each other. In turn, the response to a test item should not affect the response to another item. However items in the short form were similarly worded such as walking 2~4 blocks and walking 4~8 blocks. Further research in the future should consider the assumption for validating the application of item response theory models.

Conclusion

The goals of this study were to demonstrate how to apply the Rasch measurement model to generate item difficulty, the hierarchical order and DIF analysis of the Oswestry and the short form items with the use of known group validity. The items of both measures showed acceptable fit statistics except for "running one block" item of the short form. The hierarchical order of item difficulty was logically supported by the empirical evidences. Five paired items across the two measure were detected as a significant DIF indicating the items were differently functioned across the two measures. Despite the DIF items across the two measures, five paired items were selected for common items and well targeted participants with LBP.

References

Bond TG, Fox CM. Applying the Rasch Model: Fundamental measurement in the human sciences. 2nd ed. Mahwah, NJ, Lawrence Erlbaum Associates Publishers, 2001:23-28.

Choi BS. Computer adaptive testing method for measuring disability in patients with back pain. Phys Ther Korea. 2012;19(3):124-131. http://dx.doi.org/10.12674/ptk.2012.19.3.124

Choi BS. The use of rasch model in developing a short form based on self-reported activity measure for low back pain. Phys Ther Korea. 2014;21(4):56-66. http://dx.doi.org/10.12674/ptk.2014.21.4.056

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

DeVellis RF. Classical test theory. Med Care. 2006;44(11 suppl 3):S50-S59.

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

Finch WH, Hernandez Finch ME. Differential item functioning analysis using a multilevel rasch mixture model: Investigating the impact of disability status and receipt of testing accommodations. J Appl Meas. 2014;15(2):133-151.

Fleishman JA, Lawrence WF. Demographic variation in SF-12 scores: True differences or differential item functioning? Med Care. 2003;41(7 suppl):III75-III86.

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.

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.

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

Huang HY. Effects of the common scale setting in the assessment of differential item functioning. Psychol Rep. 2014;114(1):104-125.

Kolen MJ. Linking assessments: Concept and history. Appl Psychol Meas. 2004;28(4):219-226.

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. http://dx.doi.org/10.1186/1477-7525-11-119

Mallinson T, Stelmack JA, Velozo CA. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. Med Care. 2004;42(1 suppl):I17-I24.

McHorney CA. Use of item response theory to link 3 modules of functional status items from the asset and health dynamics among the oldest old study. Arch Phys Med Rehabil. 2002;83(3):383-394.

McHorney CA, Cohen AS. Equating health status measures with item response theory: Illustrations with functional status items. Med Care. 2000;38(9 suppl):II43-II59.

Nyein K, McMichael L, Turner-Stokes L. Can a barthel score be derived from the FIM? Clin Rehabil. 1999;13(1):56-63.

Smith RM, Kramer GA. A comparison of two methods of test equating in the rasch model. Educ Psychol Meas. 1992;52(4):835-846.

Taherbhai HM, Young MJ. Pre-equating: A simulation study based on a large scale assessment model. J Appl Measure. 2004;5(3):301-318.

Teresi JA. Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical and mental health. J Mental Health Aging. 2001;7(1):31-40.

Velozo CA, Byers KL, Wang YC, et al. Translating measures across the continuum of care: Using rasch analysis to create a crosswalk between the functional independence measure and the minimum data set. J Rehabil Res Dev. 2007;44(3):467-478.

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, 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.

Wang YC, Byers KL, Velozo CA. Validation of FIM-MDS crosswalk conversion algorythm. J Rehabil Res Dev. 2008;45(7):1065-76.

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

Williams BC, Li Y, Fries BE, et al. Predicting patient scores between the functional independence measure and the minimum data set: Development and performance of FIM-MDS “crosswalk”. Arch Phys Med Rehabil. 1997;78(1):48-54.

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

Wright BD, Stone MH. Best test design. Rasch measurement. Chicago, IL, MESA press, 1979:93-95.

This article was received June 22, 2015, was reviewed June 22, 2015, and was accepted July 23, 2015.