한국전문물리치료학회지 2014년 21권 4호 56-66 ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
Phys Ther Korea 2014;21(4):56-66 http://dx.doi.org/10.12674/ptk.2014.21.4.056

The Use of Rasch Model in Developing a Short Form Based on 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

For maintaining adequate psychometric properties when reducing the number of items from an instrument, item level psychometrics is crucial. Strategies such as low item correlation or factor loadings, using classical test theory, have traditionally been advocated. The purpose of this study is to describe the development of a new short form assessing the impact of low back pain on physical activity. Rasch measurement model has been applied to the International Classification of Functioning, Disability and Health Activity Measure (ICF-AM). One hundred and one individuals with low back pain aged 19-89 years (mean age: 48.1±17.3) who live in the community were participated in the study. Twenty-seven items of lifting/carrying construct of the ICF-AM were analyzed. Ten items were selected from the construct to create a short form. Item elimination criteria include: 1) high or low mean square (out of the range: .6-1.4 for the fit statistics), 2) similar item calibrations to adjacent items, 3) person separation value, and item-person map for potential gap in person ability continuum. All 10 items of the short form fit to the Rasch model except one item (i.e., carrying toddler on back). Despite its high infit and outfit statistics (1.90/2.17), the item had to be reinstated due to potential gaps at the upper extreme of person ability level. The short form had a slightly better spread of person ability continuum compared to the entire set of item. The created short form separated individuals with low back pain into nearly 4 groups, while the entire set of items separated the individuals into 6 groups. The findings prompted multidimensional models for better explanation of the lifting/carrying domain. The item level psychometrics based on the Rasch model can be useful in developing short forms with rationally retained items.

Key Words: Activity; Assessment; Item response theory; Low back pain; Rasch analysis; Short form.

Introduction

Creating fixed short forms have primarily been used for decades to achieve "psychometric efficiency" in health assessments (Caronni et al, 2014; Haley et al, 2004; Jette, 2003; Velozo et al, 2000-2001). Shortened instruments have developed in response to growing demands for reducing test administration time, respondent burden and study costs. In creating the short form of an assessment, a goal would be selecting the least number of items necessary while maintaining adequate precision in measuring the latent trait (Mallinson et al, 2004). That is, the major challenge of shortening an existing well-developed instrument is to achieve psychometric efficiency with fewer items without sacrificing measurement precision (Haley et al, 2004; Jette and Haley, 2005; Lee and Kang, 2013; Lerdal et al, 2013). Creating fixed short form has largely been driven by the necessity of establishing comprehensiveness and breadth of prior assessments. However, when the number of items are substantially reduced, as it is often the case, the partial loss of measurement precision is inevitable (Ware and Sherbourne, 1992). Several studies indicated that balance between comprehensiveness and precision of measurement should be taken into

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

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2014.21.4.056

consideration when developing short forms (Gum et al, 2013; Haley et al, 2004; Johnsen et al, 2013). The loss of precision may appear regardless of which items investigators eliminate because fewer items would leave more "gaps" in measurement across the ranges of person ability (i.e., disability level). In general, deficits in measurement precision often occur when items do not closely match particular ability levels. Thus, items should be chosen to match ability in order to enhance measurement precision (Haley et al, 2004).

Traditionally, classical test theory (CTT)-based methodologies often include the deletion of items focusing low item-total correlations that would be the least impact on overall internal consistency and factor loadings (Nunnally, 1994). Of these methods, Cronbach's α is one of the most commonly used methods for selecting and eliminating items that have the least impact on internal consistency. However, copious studies indicated that Cronbach's α is depending on the particular sample used (i.e., sample-dependent) and is not reflecting stable property of the test (Raykov, 2008). The estimated Cronbach's α that is a property of observed responses of a sample cannot be generalized to different samples. A study indicated that Cronbach's α could be influenced by many factors such as test length (i.e., longer tests are more reliable than shorter ones) and missing data (Nunnally, 1994). Consequently, the test items may not well matched to the individuals.

In addition, the use of the separation ratio (SR) based on the Rasch measurement model has also been advocated (Davidson, 2009; Mallinson et al, 2004). The SR indicates the impact that removing an item or items has on measurement precision. The previous studies recommended deleting items with high/low mean square residuals, similar item difficulty calibrations, and substantial influence on person separation. In other studies using item response theory (IRT) methods, items were selected based on: 1) frequency of administration in computer adaptive testing, 2) high test information, and 3) broad item difficulty coverage (Mallinson et al, 2004; Velozo et al, 2000-2001).

In the current study, we attempted to develop a short form using the Rasch measurement model, one-parameter IRT model, for an underlying construct of an instrument measuring activity that were mostly relevant to individuals with low back pain. The goal was to create an efficient short form while maintaining adequate precision. The purpose of the present study is two-fold. First, we removed items from the lifting/carrying construct of an activity measure to create a 10-item short form which is psychometrically comparable to the entire set of items. Second, we investigated the item level psychometrics as well as precision of the created short form.

Methods

Participants and design

The International Classification of Functioning, Disability and Health (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) ICF model. The ICF model provides 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 subcategories of the ICF activity dimension consulted in the development of items. Funding for the development of ICF-AM was obtained from the National Institute of Disability and Rehabilitation Research (NIDRR, #H133G000227). The study was approved by the Institutional Review Board of the University of Florida (Approved by IRB # 568-2000). Data from the 101 individuals with back pain who completed the paper-pencil version was retrieved and analyzed for the current study.

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2014.21.4.056

Instrumentation

The lifting/carrying construct was selected from six constructs of the ICF-AM due to its relevance to activity limitations resulting from low back pain. In an effort to overcome limitations of the CTT-based short form construction procedure, the Rasch model (one-parameter IRT model) was employed. The Rasch model is the most robust of the IRT models with respect to sample size with a recent simulation study proving invariant item difficulty (Wang and Chen, 2005). That is, stable item calibrations can be obtained with a relatively small sample size.

A series of Rasch analysis was performed to identify items that could be eliminated based on the following four criteria: 1) high mean square, 2) low mean square, 3) similar calibrations to other items, and 4) person separation value (i.e., item was retained if analysis with the item removed substantially decreased person separation). High or low mean square values indicate that the item may measure a different construct or need further clarification. Similar item calibrations may indicate redundant items. Removal of redundant items was considered to be appropriate if the range of ability level and intervals between items were maintained on the item-person map. To reduce the impact of local item dependence, deleting similar items was attempted.

Data analysis

Using Winsteps® software program ver. 3.57.2 (Linacre, Chicago, IL, USA), the Rasch rating scale model was employed to determine model fit as well as item level psychometrics of the ICF-AM. The Winsteps® program produces goodness of fit statistics for each item and person. These fit statistics are used to identify items that did not fit the unidimensional Rasch model. Infit and outfit mean square (MnSq) values greater than 1.4 and smaller than .6 indicate misfitting. That is, the item might have been responded to erratic or overly predictive (Bond and Fox, 2001; Wright and Linacre, 1994). High MnSq values may indicate that the item is measuring a different construct or that the item was poorly understood and needs clarification for that. Infit is inlier-sensitive or information-weighted fit. This fit is more sensitive to the pattern of responses to items at a person's ability level. Outfit is outlier sensitive fit. In contrast to infit, outfit is more sensitive to the pattern of responses to items with difficulty far from a person (Linacre, 2002).

Rasch analysis provides point measure correlation coefficients as an immediate evaluation of response-level scoring. If the item-level scoring accords with the latent variable, these correlations will be positive. A negative correlations coefficient may indicate a reverse scored item. The point measure correlations are acceptable if they are greater than .3. The Rasch analysis also produces estimates of person ability and item difficulty. These estimates are on a log-odds unit (i.e., logit) scale. The average item difficulty is arbitrarily set at zero logits with positive logits indicating higher than average probabilities and negative logits indicating lower than average probabilities (Bond and Fox, 2001).

Rasch analysis also provides person separation, which is an index of the sample standard deviation in terms of standard error units and person reliability (analogous to Cronbach's α), and the proportion of observed sample variance that is not attributable to measurement error (Wright and Masters, 1982). The SR values, which allows determining whether items are effective in separating individuals into statistically distinct ability levels. The SR provides an indication of the number of statistically significant strata into meaningful categories (e.g., low, medium, and high ability back pain groups). The formula used to calculate is SR=(4Gp+1)/3, where "Gp" represents person separation (Wright and Masters, 2002).

Confirmatory factor analysis (CFA) and exploratory factor analysis (EFA) were used to test the unidimensionality of the short form by using Mplus™ software program ver. 4.21 (Muthén & Muthén, Los Angeles, CA, USA). The software was used to de-

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2014.21.4.056

termine the goodness of fit to one-factor model of the short form. The following criteria were used to determine the goodness of fit to the one factor model: 1) p-value of chi square >.05, 2) Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) close to 1.0, 3) root mean square error of approximations (RMSEA) <.06, and 4) weighted root mean square residual (WRMR) <.01 (Brown, 2003). Since the one-factor model was not sufficient, EFA was performed to further investigate the potential factor structure. We applied the unweighted least squares method for estimators, varimax rotation following the initial factor extraction, and replaced missing values with mean values. Criteria to determine the number of factors to retain was: 1) Kaiser's eigenvalues greater than 1, 2) factors accounting for greater than 5% of total variance, and 3) scree test where the slope changes substantially in the factor versus eigenvalue graph was performed (Cattell, 1966). A criterion of greater than .5 was used to indicate a significant loading on a factor.

Test information function reports the "statistical information" in the data corresponding to the complete test. In general, the precision with which a parameter is estimated is measured by the variability of the estimates around the value of the parameter. The amount of information is the reciprocal of variance. Statistically, when the standard deviation of person ability estimates about the examinee's ability is squared, the term represents the variance and is a measure of the precision with which a given ability level can be estimated. From the above explanation, the amount of information at a given level is the reciprocal of this variance. If the amount of information is large, it means that the person ability may be estimated with high precision at a given ability level and the estimates will be close to the true value of ability. If the amount of information is small, it means that the person ability may be estimated with low precision and the estimates will be widely scattered around the true value of ability. In order to determine how precisely the items on each of the short forms estimate person ability across the full range of the construct, the test information function was examined.

Results

The average age of the sample was 48.1±17.3 years and nearly 80% of participants reported having back pain more than a year, which indicated a chronic condition. 64% (65/101) of the sample in the study were females and nearly 31% (31/101) were males. Five subjects did not fill out this section.

After an initial Rasch analysis run with 27 items, items with high infit/outfit statistics (boxed items) were removed (Table 1). With several iterations of Rasch analysis, attempting to maintain adequate person separation and confirm if there is any changes in the fit statistics, ten items were selected from the entire set of 27 items (Table 2). The 10 items retained to create the short form all conformed to the Rasch model except one item. The item, carrying toddler on back, had a problematic fit statistic (outfit 1.47). However, despite the high fit statistic (1.90/2.17), this item had to be reinstated into the short form due to a potential gap. All 10 items of the short form exhibited moderate to high point measure correlations ranging from .42 to .83, compared to the range of the entire set of items (.41 to .78). The 10 items of the short form had a slightly better spread of person ability (-3.10 to 4.80 logits) than the entire set of items (-2.72 to 4.30 logits). Item calibrations of the 10-item short form remained relatively stable after the 17 items were deleted. However, person separation decreased from 3.67 to 2.49 (SR decreased from 5.23 to 3.65). That is, the created 10-item short form separated the sample into nearly 4 groups, while the entire item separated the sample into 6 groups. After the short form creation, person reliability (analogous to Cronbach's α) decreased from .93 to .86.

A CFA was conducted to test for dimensionality of the created short form. The one factor model

Table 1. Fit statistics for the lifting/carrying construct of the ICF-AM

Items	Measure (logits)	Error	Infit MnSq[a]	ZSTD[b]	Outfit MnSq	ZSTD	Corr[c]
Carrying toddler on shoulders	2.84	.19	1.45	2.0	1.73	1.7	.43
Carrying toddler on back	2.69	.19	1.35	1.7	1.47	1.3	.51
Lifting 25 pounds shoulder to above head	1.77	.15	.86	-.9	.76	-1.1	.70
Carrying toddler on hip	1.55	.14	1.43	2.4	1.26	.9	.56
Carrying infant in arms	1.39	.13	1.82	4.3	2.14	3.0	.46
Lifting 25 pounds waist to shoulder	1.25	.14	.77	-1.7	.77	-1.3	.72
Lifting 25 pounds floor to waist	1.18	.14	.79	-1.6	.81	-1.2	.73
Carrying 10 pounds down one flight stairs	1.15	.14	1.43	2.8	1.63	2.8	.56
Carrying 10 pounds up one flight stairs	.99	.13	1.27	1.9	1.48	2.4	.58
Carrying 25 pounds 25 feet	.89	.14	.80	-1.5	.76	-1.6	.74
Lifting 10 pounds shoulder to above head	.67	.13	.63	-3.1	.58	-2.8	.78
Lifting 10 pounds waist to shoulder	.35	.13	.63	-3.2	.60	-3.1	.78
Lifting 10 pounds floor to waist	.07	.13	.93	-.5	.89	-.7	.69
Lifting 5 pounds shoulder above head	-.31	.13	.87	-.9	.77	-1.2	.70
Pulling open a heavy door	-.53	.17	.98	-.1	.91	-.5	.58
Carrying 10 pounds 25 feet	-.55	.14	.76	-1.8	.67	-2.0	.72
Lifting 5 pounds floor to waist	-.71	.14	.83	-1.2	.87	-.7	.68
Lifting 5 pounds waist to shoulder	-.85	.14	.73	-2.1	.66	-2.1	.71
Pushing open a heavy door	-.86	.17	.86	-1.0	.74	-1.1	.61
Pulling wet laundry out washing machine	-.94	.15	.97	-.1	.93	-.3	.62
Lifting 1 pound shoulder to above head	-1.08	.14	1.16	1.0	1.05	.3	.57
Lifting 1 pound floor to waist	-1.42	.15	1.00	.1	1.38	1.2	.55
Carrying 5 pounds 25 feet	-1.45	.17	.74	-1.5	.80	-.8	.63
Pushing a shopping cart	-1.73	.18	1.60	2.4	1.42	1.2	.41
Carrying 1 pound 25 feet	-1.94	.19	1.29	1.2	1.19	.5	.45
Lifting 1 pound waist to shoulder	-2.00	.18	.89	-.5	.70	-1.0	.57
Pulling open refrigerator door	-2.43	.27	1.00	.1	.71	-.4	.40

[a]mean square standardized residuals, [b]Z-score standardized, [c]point measure correlation coefficient.

proved to be inadequate because none of criteria were met to determine the goodness of fit. The p-value was significant at .001, the CFI and TLI were far less than 1.0 (.016 and .016 respectively), the RMSEA was larger than .06 (.579), and the WRMR was larger than .1 (5.728) (Table 3). For the two factor model, both CFI and TLI were approximate to 1.0, yet other criteria were not met.

An EFA to further investigate the factor structure suggested that a two factor solution would be more appropriate (Table 4). We retained two factors based on the Kaiser criterion of eigenvalue greater than one. These two factors accounted for 64% of total variance (the first factor accounting for 48% and the second factor accounting for 16%). The table presents factor loadings of 10 items (factor loadings greater than .5 are in bold). Five items loaded onto factor 1 and another 5 items loaded onto factor 2. The factor loadings of item 5 and 6 involving "lifting 10 pounds waist to shoulder and lifting 5 pounds

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2014.21.4.056

Table 2. Fit statistics for lifting/carrying construct following 17 items removal

Items	Measure (logits)	Error	Infit MnSq[a]	ZSTD[b]	Outfit MnSq	ZSTD	Corr[c]
Carrying toddler on back	3.58	.21	1.90	3.7	2.17	2.2	.49
Lifting 25 pounds shoulder to above head	2.25	.17	.80	-1.3	.75	-1.1	.79
Lifting 25 pounds floor to waist	1.51	.16	.77	-1.7	.77	-1.3	.80
Carrying 25 pounds 25 feet	1.20	.15	.78	-1.6	.71	-1.8	.82
Lifting 10 pounds waist to shoulder	.50	.15	.69	-2.5	.65	-2.5	.83
Lifting 5 pounds shoulder above head	-.33	.15	1.15	1.0	.99	0	.72
Pulling wet laundry out washing machine	-1.02	.16	1.07	.5	1.05	.3	.65
Pulling open a heavy door	-1.45	.17	.96	-.2	1.13	.5	.61
Lifting 1 pound waist to shoulder	-2.45	.20	1.06	.4	.74	-.5	.57
Pulling open refrigerator door	-3.78	.29	1.05	.3	.62	-.2	.42

[a]mean square standardized residuals, [b]Z-score standardized, [c]point measure correlation coefficient.

Table 3. Results of confirmatory factor analysis for the short form of the ICF-AM

Indices criteria	1-factor model	2-factor model
Chi-square	1380.940	810.136
Degree of freedom	39	32
p-value (>.001)	.000	.000
CFI[a] (1.0)	.016	.897
TLI[b] (1.0)	.016	.945
RMSEA[c] (<.06)	.579	.294
WRMR[d] (<.1)	5.728	3.907

[a]comparative fit index, [b]Tucker-Lewis index, [c]root mean square error of approximation, [d]weighted root mean square residual.

shoulder above head" were nearly the criterion used. Items had a tendency to load onto factors based on item difficulty. That is, items with activities involving lifting heavy objects loading onto factor 1 and those involving lifting light objects onto the other factor2.

Figure 1 presents how 10 items of the short form are spreaded before and after the item reduction and carrying a toddler on back item measures individuals at the upper extreme in ability level. The item reduction has led the ICF-AM instrument to an improvement in terms of comprehensiveness. That is, the short form showed slightly better capability to measure a wide range of ability levels compared to 27 items of its original instrument.

Figure 2 presents that the removal of 17 items from entire 27 items resulted in considerable loss of test information from 12.09 to 4.85. That is, the entire 27 items of the construct of the ICF-AM estimated person ability with greater precision than did the 10 item short form, particularly near the center of ability range.

Discussion

This study demonstrated how the Rasch model can be applied to achieve measurement efficiency and reduce items while maintaining adequate precision. Extensive attempts to create short forms for the individuals with low back pain have previously focused on classical test theory methodologies such as mostly

Table 4. Factor structure of short form for the lifting/carrying contruct of the ICF-AM

Items (difficulty order)	Factor 1	Factor 2
1. Carrying a toddler on your back (for example, piggyback)?	.636	-.151
2. Lifting 25 pounds (for example, large bag of dog food or cat litter) from shoulder height to above your head with your hand(s) and arm(s)?	.859	.223
3. Lifting 25 pounds (for example, large bag of dog food or cat litter) from floor to waist height with your hand(s) and arm(s)?	.806	.289
4. Carrying 25 pounds (for example, a large bag of dog food or cat litter) in your hand(s) and arm(s) 25 feet?	.814	.320
5. Lifting 10 pounds (for example, bag of groceries or 12-pack of soft drinks) from waist height to shoulder height with your hand(s) and arm(s)?	.696	.488
6. Lifting 5 pounds (for example, bag of sugar or large telephone book) from shoulder height to above your head with your hand(s)?	.429	.631
7. Pulling wet laundry out of a washing machine?	.323	.647
8. Pulling open a heavy door (for example, department/convenience store door)?	.225	.727
9. Lifting 1 pound (for example, a can of soup) from waist height to shoulder height with your hand(s)?	.128	.747
10. Pulling open a full-size refrigerator door?	-.092	.790
Percent of total variance accounted for by factors	48%	16%

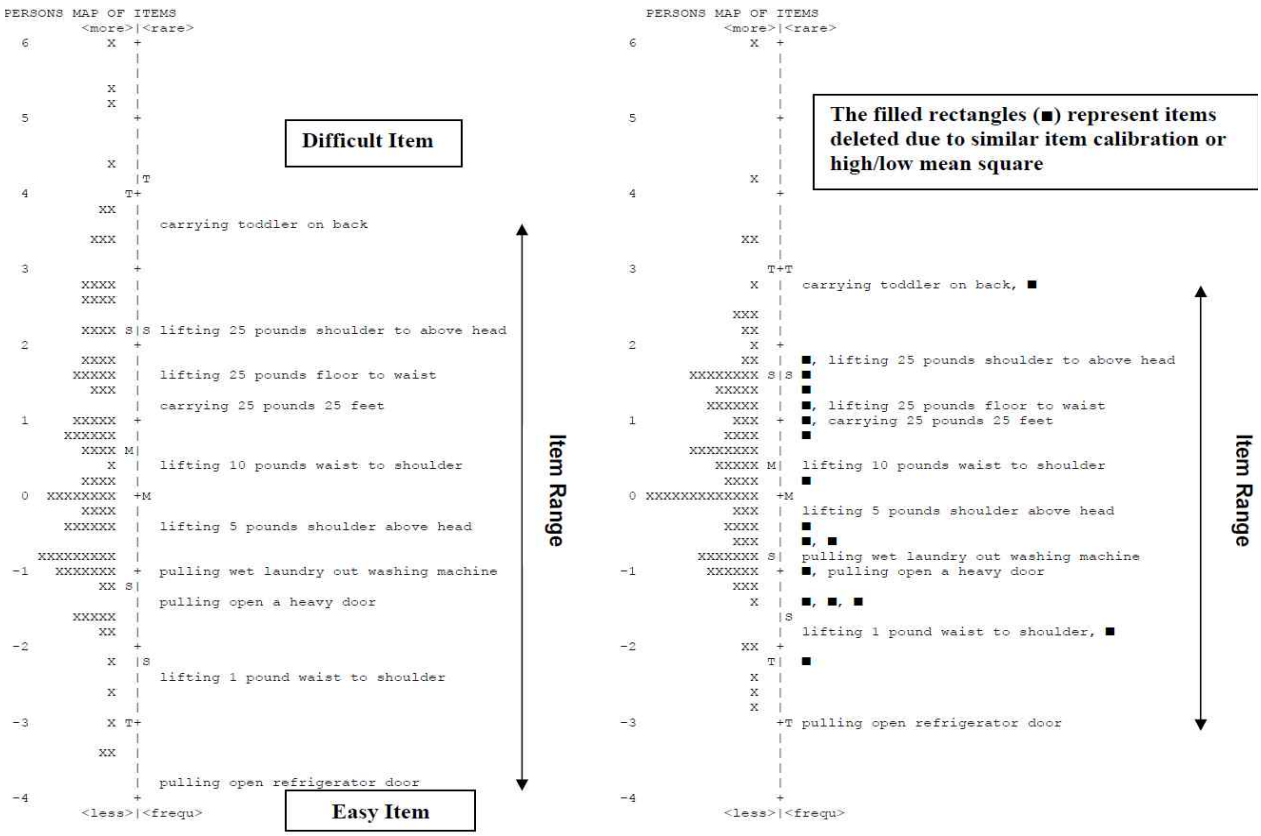

Figure 1. Item-person map of lifting/carrying construct of the ICF-AM before/after 17 items removal (Note that item calibrations are provided from easy to difficult items.).

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2014.21.4.056

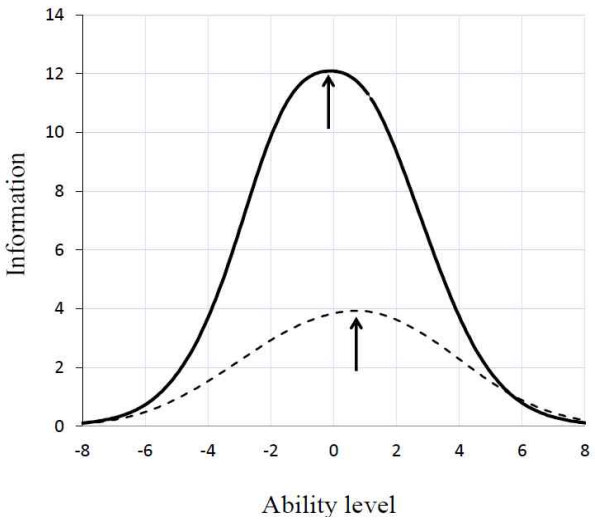

Figure 2. Test information function of short form versus entire item set of lifting/carrying construct of the ICF-AM. [The graph presents to what extent item reduction lost test information (i.e., precision) with the short form (dotted line) and the entire 27 items (solid line). The arrow presents what level of ability was targeted with each measure.].

internal consistency and test-retest reliability (Kopec et al, 1996; Müller et al, 2004; Müller et al, 2006). In this study, we used the Rasch model to provide item level psychometrical properties on a construct of the ICF-AM. Criteria for the item removal were focused on infit/oufit MnSq, person separation, item-person map, and hierarchical order of item difficulty.

Ten items of the short form fit to the Rasch model except one item (i.e., carrying toddler on back). The individuals with low disability may have tendency to select low ratings or individuals with high disability may have tendency to select unexpected high ratings on the item. These response patterns might have been the result of a lack of observations for the item. Rasch analysis also aided item selection by identifying items that best capture the range of person ability to be estimated and identifying gaps where item difficulty calibrations did not match person-ability measures. These gaps provided directions in selecting items along with item statistics. Thus, in determining whether or not items are equally distributed across the full ranges of ability continuum, items are selected based on the person location on the map. That is, we placed items at or near the middle of the scale where average individuals aggregate even though candidate items distributed toward both extremes. For example, in the initial modification phase, "carrying toddler on back" item was identified due to high fit statistics. By inspecting the item-person map (Figure 1) revealed that the item was needed to reduce possible ceiling effects as no other items remained on the short forms that were as difficult as these items. The item was later reinstated to the short form because of a lack of the most difficult item to match individuals at the extremes of the scale.

It should be noted that we treated a response category 'have not done' as the lowest rating based on the rationale that the most likely explanation for an activity not occurring was that the item could not be performed (Jette et al, 2003). Thus, we determined that treating the category 'have not done' as the lowest rating would have been more appropriate. In fact, nearly half of individuals with above average person ability scored the rating on "carrying toddler on back" item. A plausible explanation is that these individuals might have responded to the absence of opportunity on these items (i.e., you can do the activity but have not done so for any reason in the last 30 days). In addition, other individuals might have responded to other instructions indicating the lowest score.

The dimensionality of the 10-item short form was examined by CFA and EFA. The CFA failed to support the proposed unidimensional structure of the short form. Since CFA failed to support one factor model, EFA was performed. One factor model for the short form accounted for a moderate percentage of the variance (>48%). Based on the Kaiser rule, we retained the two factors for the short form. The finding may implicate that the theoretically generated the construct of the ICF-AM instrument might already have more than one dimension as well. The EFA showed that the items grouped by the hier-

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2014.21.4.056

archical order of item difficulty. The difficult items had a tendency to load on factor 1, while the moderate/easy items had a tendency to load on factor 2 (Table 4). The findings may indicate that dividing the short form into more than one subscale would be preferred.

The SR for the short form was adequate, separating the samples nearly 3 to 4 statistically meaningful strata. The SR of the short form considerably decreased compared to its full 27 item set. This decrement was unavoidable because nearly 63% of items for the lifting/carrying construct were removed. However the person reliability (analogous to Cronbach's α) only slightly decreased from .93 to .86. Perhaps the reason for this is that the removal of redundant items allowed the item deletion without loss of internal consistency. Despite the reduction of person reliability, the values were still in acceptable ranges suggested by George and Mallery (2002).

Test information function (TIF) showed that the entire set of items estimated the person ability with greater precision than did the short forms near the center of the ability range. The statistical meaning of information is defined as the reciprocal of the precision with which a parameter could be estimated (Fisher, 1925). Thus, when we estimate person ability with precision, we would know more about the values of the person ability than if we estimated it with less precision. The precision with which person ability is estimated is measured by the variability of the estimates around the value of person ability. Therefore, a measure of precision is the variance of the estimators (i.e., σ^2) and the amount of information at a given ability level is the reciprocal of this variance. That is, if the amount of information is large, person ability at a particular level can be estimated with precision. Similarly, if the amount of information is small, person ability at a particular level cannot be estimated with precision. In this study, TIF showed a considerable loss of information as the number of items was reduced. As items were eliminated to create the short form, the information decreased about 60%. The peak of the TIF for the positioning/transfer short form slightly moved to the left side of the center, while the peak of the TIF the short form slightly moved to the right side of the center. This may suggest that we should have selected items with lower item calibrations (i.e., easier items) when we deleted items. In fact, we should have selected items with higher item calibrations (i.e., more difficult items). However, the total number of individuals in the ceiling did not differ before and after item reduction.

These evidences implicate that the newly created short form could be improved in future research addressing by: 1) replacing problematic item, 2) developing items that more adequately fill the gaps in the person ability to cover the wider range of ability. In addition, the results of the present study suggest that the short form was multidimensional. However it is unrealistic to use a multidimensional model with the sample size of the current study, which leads to a limitation of this study. These findings may prompt the use of multidimensional models with adequate sample sizes to better explain physical activity domains. In addition, it is apparent that relative to the entire item banks, the short form showed decrement in measurement precision despite the use of an IRT methodology (McHorney, 1999). One way to avoid this decrement in measurement precision would be to combine the IRT and computer adaptive testing methodology. By selectively presenting items that are matched to the ability levels of respondents, these methodologies may accomplish both measurement efficiency and precision (McHorney, 1997; Velozo et al, 1999).

Conclusion

This study aimed to describe how to apply Rasch measurement model to create a short form from its original instrument. The short form separated individuals with low back pain into nearly 4 groups,

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2014.21.4.056

while the full item set separated the same individuals into 6 groups. Cronbach's α slightly decreased from .93 to .86. A considerable loss of test information was inevitable because such a large number of items deleted from its original instrument. However the created short form had a slightly better spread of test items covering the ability continuum. Despite few disadvantages, the use of the Rasch model would be useful in developing short forms with rationally retained items.

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, 183-184.

Brown TA. Confirmatory factor analysis of the penn state worry questionnaire: Multiple factors or method effects? Behav Res Ther. 2003;41(12): 1411-1426.

Caronni A, Zaina F, Negrini S. Improving the measurement of health-related quality of life in adolescent with idiopathic scoliosis: The SRS-7, a rasch-developed short form of the SRS-22 questionnaire. Res Dev Disabil. 2014;35(4):784-799. http://dx.doi.org/10.1016/j.ridd.2014.01.020

Cattell RB. The scree test for the number of factors. Multivar Behav Res. 1966;1:245-276.

Davidson M. Rasch analysis of 24-, 18- and 11-item versions of the roland-morris disability questionnaire. Qual Life Res. 2009;18(4):473-481. http://dx.doi.org/10.1007/s11136-009-9456-4

Fisher RA. Theory of statistical estimation. Math Proc Camb Phil Soc. 1925;22(5):700-725.

George D, Mallery P. SPSS for Windows Step by Step: A simple guide and reference, 11.0 update. 4th ed. Boston, Allyn and Bacon, 2002:123-124.

Gum JL, Glassman SD, Carreon LY. Clinically important deterioration in patients undergoing lumbar spine surgery: A choice of evaluation methods using the oswestry disability index, 36-item short form health survey, and pain scales: Clinical article. J Neurosurg Spine. 2013;19(5): 564-568. http://dx.doi.org/10.3171/2013.8.SPINE12804

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.

Jette AM. Assessing disability in studies on physical activity. Am J Prev Med. 2003;25(3 suppl 2): 122-128.

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. Comparison of functional status tools used in post-acute care. Health Care Financ Rev. 2003;24(3):13-24.

Johnsen LG, Hellum C, Nygaard OP, et al. Comparison of the SF6D, the EQ5D, and the oswestry disability index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskelet Disord. 2013;14:148. http://dx.doi.org/10.1186/1471-2474-14-148

Kopec JA, Esdaile JM, Abrahamowicz M, et al. The quebec back pain disability scale: Conceptualization and development. J Clin Epidemiol. 1996;49(2): 151-161.

Lee TW, Kang SJ. Development of the short form of the korean health literacy scale for the elderly. Res Nurs Health. 2013;36(5):524-534. http://dx.doi.org/10.1002/nur.21556

Lerdal A, Kottorp A, Gay CL, et al. Development of a short version of the lee visual analogue fatigue scale in a sample of women with HIV/AIDS: A rasch analysis application. Qual Life Res. 2013;22(6):1467-1472. http://dx.doi.org/10.1007/s11136-012-0279-3

Linacre JM. What do infit and outfit, mean-square and standardized mean? Rasch Meas Trans. 2002;16(2):878.

Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the

ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
http://dx.doi.org/10.12674/ptk.2014.21.4.056

creation of minimum item sets. Med Care. 2004;42(1 suppl):I17-I24.

McHorney CA. Generic health measurement: Past accomplishments and a measurement paradigm for the 21st century. Ann Intern Med. 1997;127(8 Pt 2):743-750.

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

Müller U, Duetz MS, Roeder C, et al. Condition-specific outcome measures for low back pain. Part I: Validation. Eur Spine J. 2004;13(4):301-313.

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

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

Raykov T. Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha. Br J Math Stat Psychol. 2008;61(Pt2):275-285.

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. 2000-2001;4(3):667-680.

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

Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473-483.

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

Wright BD, Masters GN. Rating Scale Analysis: Rasch measurement. Chicago, MESA Press, 1982:8-9.

Wright BD, Masters GN. Number of person or item strata. Rasch Meas Trans. 2002;16(3):888.

This article was received September 10, 2014, was reviewed September 10, 2014, and was accepted November 3, 2014.