A Comparative Study of Item Difficulty Hierarchy of Self-Reported Activity Measure Versus Metabolic Equivalent of Tasks

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

Abstract

The purposes of this study were: 1) to show the item difficulty hierarchy of walking/moving construct of the International Classification of Functioning, Disability and Health-Activity Measure (ICF-AM), 2) to evaluate the item-level psychometrics for model fit, 3) to describe the relevant physical activity defined by level of activity intensity expressed as Metabolic Equivalent of Tasks (MET), and 4) to explore what extent the empirical activity hierarchy of the ICF-AM is linked to the conceptual model based on the level of energy expenditure described as MET. One hundred and eight participants with lower extremity impairments were examined for the present study. A newly created activity measure, the ICF-AM using an item response theory (IRT) model and computer adaptive testing (CAT) method, has a construct on walking/moving construct. Based on the ICF category of walking and moving, the instrument comprised items corresponding to: walking short distances, walking long distances, walking on different surfaces, walking around objects, climbing, and running. The item difficulty hierarchy was created using Winstep software for 20 items. The Rasch analyses (1-parameter IRT model) were performed on participants with lower extremity injuries who completed the paper and pencil version of walking/moving construct of the ICF-AM. The classification of physical activity can also be performed by the use of METs that is often preferred to determine the level of physical activity. The empirical item hierarchy of walking, climbing, running activities of the ICF-AM instrument was similar to the conceptual activity hierarchy based on the METs. The empirically derived item difficulty hierarchy of the ICF-AM may be useful in developing MET-based activity measure questionnaires. In addition to convenience of applying items to questionnaires, implications of the finding could lead to the use of CAT method without sacrificing the objectivity of physiologic measures.

Key Words: Computer adaptive testing; Item response theory; Metabolic equivalent of tasks; Rasch analysis.

Introduction

To improve public health and chronic disease control, valid and appropriate measurement of physical activity is an essential process in health care (Pereira et al, 1997). Assessing physical activity often classified into direct method using motion sensors such as accelerometers, pedometer, or monitoring heart rate and indirect method using self-reported questionnaires. The direct method often converts the measures into metabolic equivalent of tasks (METs) that are widely accepted notion in energy expenditure of physical activity. It is commonly viewed as a measure that could provide a common descriptor of workload levels across populations (Byrne et al, 2005). Due to its optimal accuracy, the use of METs is commonly preferred to determine the level of physical activity (Ainsworth, 2009; Ainsworth et al, 2011). However measuring METs is often cumbersome because it requires the use of in-

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

struments or impractical complex system in clinical settings. Consequently, self-reported questionnaires as an indirect measure are used most often in health surveillance system by identifying activity patterns and quantifying activity performed because of its easiness or efficiency (Freedson et al, 2008).

Several self-reported physical activity questionnaires were developed to determine the brief classification of activity level such as active or inactive status (Taylor-Piliae et al, 2007) and the estimated energy expenditure of various activities expressed as kilocalories per day or MET-minutes per day (Kriska et al, 1988; Paffenbarger et al, 1978). A study found 27 self-reported physical activity questionnaires in which energy cost accounted for many activities computed scores and converted them into the forms of energy expenditure (Kriska and Caspersen, 1997). Most questionnaires, if not all, contain a great number of various types of physical activities in their test questions. From a myriad of physical activities that one can incorporate into questionnaire, walking is considered as the one of the most common dimensions of physical activity. Hence, walking activity has been incorporated in activity questionnaires (Craig et al, 2003). Additionally, many questionnaires have been concurrently validated by comparing the estimated MET level and relevant activities of the domain of walking/climbing/running (Bassett et al, 2000; Berlin et al, 2006). MET is an energy cost of physical activities that is described as a multiple of resting metabolic rate (RMR) (Blair et al, 1985). One MET is defined as the oxygen uptake when a person is at rest, which is equivalent to 3.5 milliliter of oxygen per kilogram of body weight per minute (Balady, 2002). Based on the rate of energy expenditure expressed as METs, the activity intensity of walking is classified as multiples of 1 MET or as the ratio of the associated metabolic rate for the specific activity divided by a standard RMR. For example, the MET of 'walking less than 2.0 mile per hour (mph)' is 2.0, while 'walking the dog' and 'walking 5.0 mph' are 3.0 and 8.0 respectively.

In contrast with the activity classification of MET level, the activity classification of the International Classification of Functioning, Disability and Health-Activity Measure (ICF-AM) instrument was based on empirically derived item difficulty hierarchy, which was created by applying Rasch model (1-parameter Item Response Theory model) to activities involving movement, moving around and daily life tasks (Velozo et al, 2008). The ICF-AM is accessible worldwide through the web and information obtained from ICFmeasure.com (available at http://www.ICFmeasure.com; accessed 23 July 2013). The ICF-AM allows us to measure physical activities in efficient and precise manner by selectively directing individualized questions to a respondent from a large item bank. The Rasch model orders activities being measured based on how difficult the activity is to perform. That is, activities easier to perform will receive lower mean calibrations, while activities harder to perform will receive higher mean calibrations (Sabari et al, 2005). The empirically derived item difficulty hierarchy of the ICF-AM may challenges to the hierarchy of existing physiological measures such as the MET level.

The purpose of this study are to determine: 1) dimensionality, 2) item-level psychometrics, 3) empirical item hierarchy of walking/moving domain of the ICF-AM, and 4) to what extent the empirical hierarchical structure of the ICF-AM is linked to the conceptual model based on the level of energy expenditure described as MET.

Methods

Instrumentation

The data used in the present study was partially retrieved from a research that developed the ICFmeasure.com funded by the National Institute of Disability and Rehabilitation Research (NIDRR). The study approved by the Institutional Review Board of the University of Florida (approved by IRB #

568-2000).

The ICF-AM was developed to create an efficient and precise measurement system based on the activity dimension of International Classification of Functioning, Disability and Health (ICF). The ICF by World Health Organization (WHO) provided the conceptual framework and classification system for developing items used in the study. Equiprecise measurement (i.e., measurement across the entire range of a construct) was applied to activities involving movement, moving around and daily life tasks as defined by the activity dimension of the ICF (Velozo et al, 2008). The ICF-AM is a web-based computer adaptive survey system. The administrative core of the instrument allows setting a wide range of functions, including initial theta value (i.e., directing the initial question that most closely matches the ability level of the respondent) and standard error (i.e., for terminating the test). The questions are targeted to individuals at their ability level requiring only 5-10 questions per construct to reach at a final measure of person ability. In addition, immediate results are provided to the respondents/clinician in the form of graphs and summary statistics.

The present study retrieved partial data from the paper and pencil version of self-reported instrument with 255 items (6 domains of activity dimension) prior to developing the ICFmeasure.com in effort to capture limitations in activities resulting from lower extremity impairments. The walking/moving domain of the ICF-AM instrument constructed with 20 items is selected for the present study. The 20 items are listed in the Table 1. Each item consists of a closed-ended question with 4 response categories: 1) a lot of difficulty, 2) some difficulty, 3) no difficulty, and 4) have not done yet. Participants were instructed to answer 'have not done yet', if participants have not performed the activity for the past 30 days, unable to perform the activity, require any help/assistance of persons, or medical doctor suggested participants not to do the activity. This category was regarded as missing value in the analysis.

Participants

Through 1) focus group presentation with test items, 2) professional panel consultations, 3) cognitive interviewing, and 4) paper-pencil version filed test with 255 items for different diagnostic groups, the study was conducted to develop the ICF-AM with 264 items measuring activity limitation. Of the total 413 participants who were in a larger project to develop an activity measure, the 108 participants who reported impairments of lower extremity on paper-pencil version measuring activity limitations were selected for this study.

Statistical Analysis

The Rasch rating scale model was applied to analyze to determine the model fit as well as item level psychometrics (item calibrations) of the ICF-AM. The Rasch model transforms raw scores into estimates of item difficulty (calibrations) in logits. These calibrations in logits are empirically generated estimates of item difficulty hierarchy of physical activity. In order to determinate the relevance of the empirical item hierarchy of the ICF-AM and conceptual hierarchy of activity physiological-based MET levels, the METs of varied physical activities were compared to the empirical hierarchy based on the item calibrations of the ICF-AM. The specific levels of METs were adopted from the Compendium of Physical Activities, which is a coding scheme that classified specific physical activity determined by the energy expenditure on varied physical activities (Ainsworth et al, 2011).

Rasch analysis with rating scale model (Wright and Masters, 1982) using Winsteps[®] computer program (Linacre, Chicago, IL, version 3.57.2) was conducted to determine the model fit as well as the item level psychometrics of the ICF-AM for back pain patients. Rasch model (i.e., one-parameter item response theory model (IRT)) is the most robust of the IRT models in which stable and accurate item parameters such as fit statistics could be obtained with relatively small sample size (Wang and Chen,

Table 1. Fit statistics for walking/moving construct following Rasch modelling

Items	Measure (logits)	Error	Infit MnSq[a]	ZSTD[b]	Outfit MnSq	ZSTD	Correlation
Jogging one mile	2.85	.23	1.68	3.1	1.71	1.8	.39
Running one block	2.51	.18	1.77	3.2	2.11	2.5	.57
Climbing up or down a 6-foot ladder	1.28	.14	1.45	2.4	1.28	1.0	.69
Climbing up two flights of stairs	.97	.14	.83	−1.2	.73	−1.6	.77
Walking one mile	.84	.14	.69	−2.3	.71	−1.5	.79
Climbing down two flights of stairs	.72	.13	.86	−1.0	.77	−1.2	.76
Climbing up or down a 3-step stool	.65	.13	1.03	.3	1.02	.2	.72
Stepping onto or off a bus	.63	.13	1.32	2.1	1.75	2.9	.65
Walking 4-8 blocks	.17	.13	.74	−2.0	.73	−1.5	.75
Climbing up one flight of stairs	.11	.14	.69	−2.2	.69	−2.0	.76
Climbing down one flight of stairs	−.02	.14	.70	−3.1	.69	−1.9	.74
Walking 2-4 blocks	−.43	.14	.83	−1.2	.70	−1.7	.71
Walking on gravel	−.66	.14	1.33	1.9	1.44	1.9	.57
Walking crowded place	−.77	.15	.81	−1.3	.69	−1.7	.69
Walking small obstacles on floor	−.89	.15	1.17	1.0	1.23	1.0	.56
Stepping into or out of an elevator	−1.12	.15	1.20	.9	1.88	1.5	.50
Walking within home /living environment	−1.31	.20	.95	−.3	.78	−1.1	.57
Stepping up or down a standard curb	−1.55	.17	.97	−.2	.77	−1.1	.61
Walking on grass	−1.98	.19	1.04	.3	.79	−.5	.50
Walking on carpeting	−2.0	.23	1.00	.1	.75	−.5	.45

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

2005). The Winsteps program produces goodness of fit statistics for each item and person, which were used to identify items that did not fit the unidimensional Rasch model. Items with infit and outfit mean square (MnSq) presented greater than 1.4 and smaller than .6 indicate misfit, which means that the items were responded erratically relative to other items (Bond and Fox, 2001; Wright and Linacre,

1994). The erratic pattern of response may indicate that the item might be measuring a different construct or the item needs further clarification. Infit means inlier-sensitive or information-weighted fit, which is more sensitive to the pattern of responses to items targeted on the person, while outfit means outlier sensitive fit, which is more sensitive to the pattern of responses to items with difficulty far from

a person (Linacre, 2002). Rasch analysis also provides point measure correlation coefficients as an immediate check that the item-level scoring accords with the latent variable. A negative correlations coefficient may indicate reversed survey item. The point measure correlations should be >.3 or better (available at http://www.winsteps.com/winman/index.htm/correlations.htm; accessed 23 July 2013).

Confirmatory factor analysis (CFA) was conducted to determine a goodness of fit of the items to one factor model for the ICF-AM and one factor model for each construct of the ICF-AM. The following criteria were used to determine goodness of fit to the one and multi factor model: 1) the p-value of chi square >.05 indicating a significant fit, 2) comparative fit index (CFI) and Tucker-Lewis Index (TLI) 1.0 indicating the closer to 1.0, the better the fit, 3) root mean square error of approximations (RMSEA) <.06, and 4) weighted root mean square residual (WRMR) <.01 (Brown, 2003; Brown, 2008).

We conducted exploratory factor analysis (EFA) on the construct of the ICF-AM when the CFA failed to confirm the unidimensionality of each construct to further investigate the potential factor structure. EFA was performed using MplusTM (Muthén and Muthén, LosAngeles, CA, version 4.21). We used the unweighted least squares method for estimators, varimax rotation following the initial factor extraction, and replaced missing data with mean values. Criteria to determine the number of retaining factors were: 1) Kiser's eigen values greater than 1, 2) factor s accounting for greater than 5% of the variance, and 3) scree test where the slope changes substantially in the factor versus eigen value graph (Cattell, 1966). A criterion of greater than .46 was used as a significant factor loading (Norman and Steiner, 1994).

Results

Table 1 presents item measures, errors, infit/outfit statistics, and point measure correlations for the 20 items of the walking/moving construct. Fifteen items showed an acceptable infit/outfit, while 5 items showed high infit/outfit (i.e., jogging one mile, running one block, climbing up or down a 6-foot ladder, walking on gravel and stepping into or out of an elevator). Of these items, 'jogging one mile' and 'running one block' items significantly misfit on both infit/outfit criteria (presented in bold, 1.68/1.71 and 1.77/2.11, respectively). In addition, all 20 items showed adequate point measure correlations distributing from .39 to .79. Items of walking/moving construct were effective in differentiating individuals with back pain into almost 5 statistically distinct levels of person ability. Person separation index was 3.44, defining 4.92 statistically meaningful levels of disability. These items also showed good person separation reliability (Cronbach's α), which was .92.

The table also presents the item difficulty hierarchy of walking/moving construct. Items least likely to be endorsed with high rating (i.e., the most difficult item) were 'jogging one mile' and 'running one block' with similar item difficulty calibrations (2.85 and 2.51 logits, respectively). In addition, items most likely to be endorsed with low rating (i.e., the easiest items) were 'walking on grass' and 'walking on carpet' -1.98 and -2.0 logits, respectively). That is, individuals with impairments on lower extremities are more likely to have difficulties with jogging/running related activities than walking on grass/carpet.

The result of the CFA only partially confirmed the one factor model for walking/moving construct (Table 2). The CFI and TLI of the indices were marginally adequate (.960 and .978, respectively) as well as other indices (RMSEA: .22, WRMR: 2.842).

To further explore the factor structure of walking/moving construct, EFA was conducted. We retained three factors based on a criterion of eigenvalue greater than 1, three factors based on a criterion of variance greater than 5%, and two factors based on a criterion of scree test. These factors based on each criterion accounted for 70%, 65%, and

Table 2. Confirmatory Factor Analysis for walking/moving construct

Indicis (criterion)	Walking/moving (1-factor model)
Chi-square	359.304
DF[a]	24
P-value (>.05)	<.001
CFI[b] (>.95)	.960
TLI[c] (>.95)	.978
RMSEA[d] (<.06)	.220
WRMR[e] (<.10)	2.842

[a]degree of freedom, [b]comparative fit index, [c]Tucker-Lewis Index, [d]root mean square error of approximations, [e]weighted root mean square residual.

Table 3. Factor structure of walking/moving domain following Exploratory Factor Analysis

Walking/moving domain	F1	F2	F3
Jogging one mile	.064	.139	.838
Running one block	.088	.214	.840
Climbing up or down a 6-foot ladder	.284	.387	.551
Climbing up two flights of stairs	.259	.776	.396
Walking one mile	.411	.470	.474
Climbing down two flights of stairs	.291	.789	.353
Climbing up or down a 3-step stool	.371	.442	.503
Stepping onto or off a bus	.320	.376	.463
Walking 4-8 blocks	.570	.462	.360
Climbing up one flight of stairs	.388	.787	.231
Climbing down one flight of stairs	.412	.802	.193
Walking 2-4 blocks	.662	.416	.271
Walking on gravel	.649	.262	.223
Walking crowded place	.690	.382	.178
Walking small obstacles on floor	.662	.282	.165
Stepping into or out of an elevator	.499	.250	.165
Walking within home/living environment	.802	.236	.124
Stepping up or down a standard curb	.672	.370	.138
Walking on grass	.832	.178	.130

54% of total variance, respectively. Based on these results, we extracted three factors to further investigate the interpretability of the factor loadings.

Most items loaded onto factors that contained items which appeared to be activities with walking/stepping, climbing/walking, and climbing/running/jogging (Table 3). Ten items loaded onto factor 1, 6 items had tendency to load onto factor 2 and 5 items loaded onto factor 3. Walking related items except 'walking one mile' item and stepping related items had tendency to load onto factor 1. Four climbing related items and two walking-related items had tendency to load onto factor 2, while two climbing-related items, running and jogging items had

tendency to load onto factor 3. Two items (walking 4-8 blocks without stopping and walking one mile without stopping) loaded onto more than one factor (factorial complex).

Item difficulty calibrations matched person ability measures fairly well on walking/moving construct (Figure 1). The item-person map shows a relative normal distribution of individual abilities ranging from -5.0 and 6.0 logits. The person ability distribution also showed 8 individuals in the floor but no ceiling effect. The average item difficulty was 1.31±2.41 logits lower than average person ability.

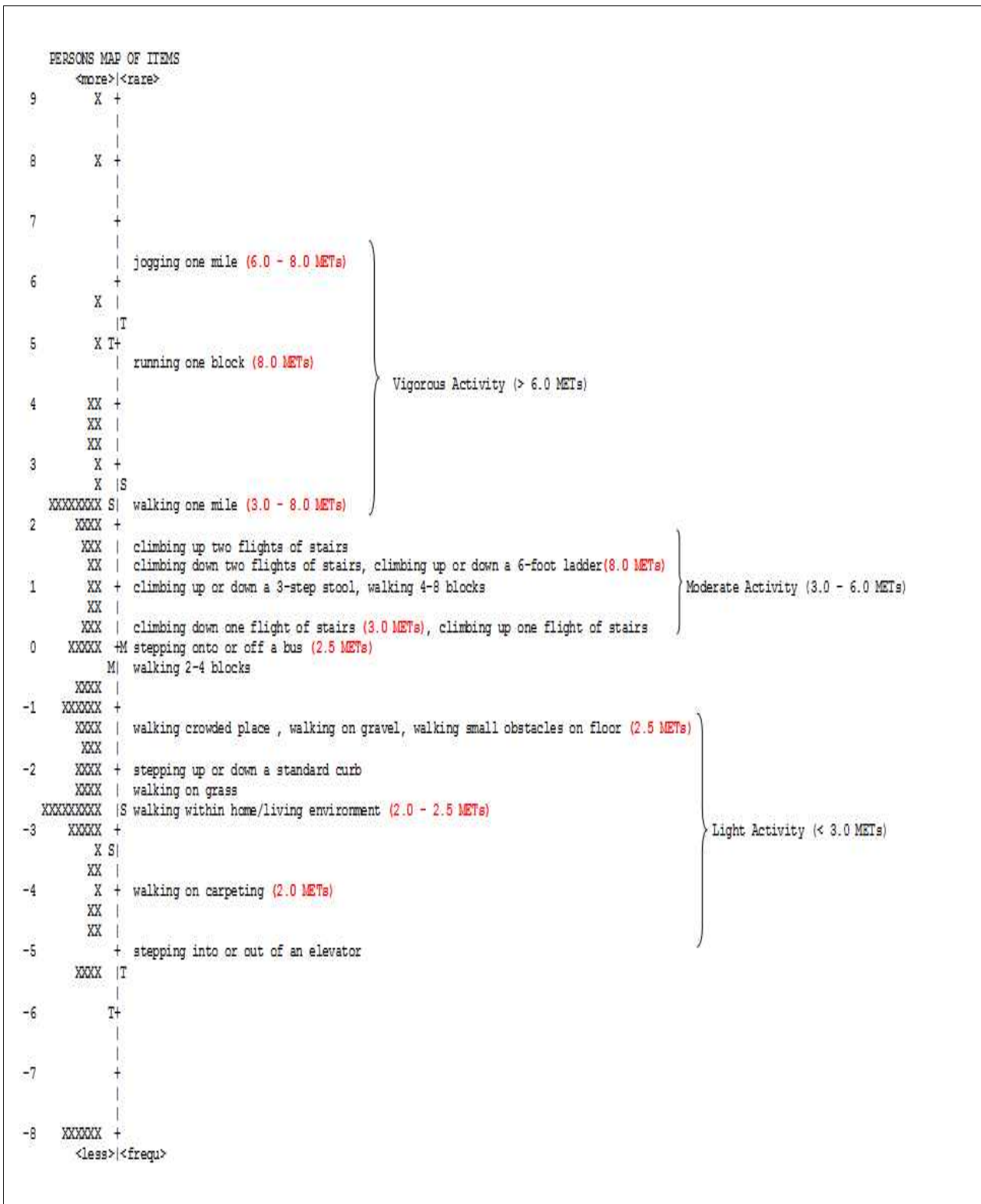Empirical evidence generated by Rasch analysis supports the hypothetically derived hierarchy based

Figure 1. Item-person map of walking/moving construct of the ICF-AM. Each 'X' on the left side of map represents 1 subject, with Xs and at the top of map representing individuals with high ability and at the bottom of map representing individuals with low ability.

on MET levels. The hierarchy of walking/running construct for MET is primarily determined by the speed pertinent to particular activities (Hambleton, 2000; Liang et al, 2002; Wright and Linacre, 1989), while our empirical hierarchy of item difficulty is determined by its conceptual difficulties such as challenges of distance or environments. The most challenging items in our empirical hierarchy, such as "jogging one mile" and "running one block", match with the vigorous activity (>6.0 METs) in MET. In addition, moderately challenging items in our empirical item difficulty hierarchy, such as most climbing related items, matched with moderate activity in MET ranging from 3.0 METs to 6.0 METs. Moreover, the least challenging items in our empirical item difficulty hierarchy clearly match the light activity (<3.0 METs) in MET. That is, empirically derived item difficulty hierarchy of walking/moving construct of the ICF-AM is similar to the activity hierarchy of associated MET levels.

Discussion

Traditionally, EFA has been used to explore the possible underlying structure of a set of interrelated variable without any preconceived structure on the outcome (Child, 2006). In this study, we conducted EFA on walking/moving construct of the ICF-AM since the CFA failed to confirm the dimensionality of the construct to further investigate the potential factor structure. EFA suggested three factor solution for the walking/moving construct. Several possible reasons might accounted for the failure to support dimensionality. First, although the instrument was theoretically generated, it might differ from their practical dimensionality. In addition, there is tendency to separate the construct into three latent traits, which could be labeled as 1) walking/stepping, 2) climbing/walking, and 3) climbing/running/jogging. These findings may suggest that the theoretically generated construct of the ICF-AM instrument have

multidimensional structures. Further investigation would be necessary to ascertain the dimensionality. Second, although dimensionality is a requisite assumption for IRT approaches the concept of dimensionality remains obscure. Researchers indicated that no measures are purely unidimensional construct (Reckase, 1985) and emphasized that there is no single test available for checking the dimensionality (Ware and Sherbourne, 1992). However, in many cases, studies can be justified by applying the concept of essentially unidimentional construct (Ware, 1996). Thus, dimensionality may not a definitive term but a quantitative attribute that may only be approximated (Wright and Linacre, 1989). Future research should take into account the influence multidimensionality on measuring the activity levels of individuals.

In regards to the fit statistics obtained from Rasch analysis, none of items misfit significantly except two items (i.e., jogging one mile and running one block). These fit statistics are a measure of observed variance over expected variance. Infit/oufit statistics of these two items were 1.68/1.71 and 1.77/2.11, respectively. Since these two items are among the most difficult items, individuals with low disability (i.e., high ability) were likely to score either low or high ratings. The bimodal distribution of responses might have resulted due to a lack of observations (the middle categories) and lead to the large observed variances for this item.

An empirical evidence of hierarchical order of item difficulty based on Rasch analysis did support the hierarchical order of physical activity based on MET values. For example, the relevant MET value of 'walking on carpeting' was 2.0 METs, while 'climbing down one flight of stairs' was 3.0 METs. That is, 'climbing down one flight stairs' would be a more challenging activity than 'walking on carpeting' in terms of the MET value because the challenging climbing stairs activity would require more energy to perform than the walking activity. Similarly, the MET value of 'climbing up or down 6-foot ladder'

was 8.0 METs, while 'stepping onto or off a bus' was 2.5 METs. This indicates that the climbing activity would require more energy to perform than stepping onto or off activity in terms of MET value. That is, the empirical hierarchical order of item difficulty corresponds with a physiological measure, which is determined by caloric energy expenditure.

The empirically created item difficulty hierarchy of the walking/moving construct reflected the hypothesized hierarchy of activities based on MET values. In our empirical hierarchy generated through Rasch analysis, an individual with low extremity impairments who is having difficulty on average difficulty item such as 'climbing down one flight of stairs' would be expected to have more difficulty on 'climbing up or down a 6-foot ladder' (more difficult than 'climbing down one flight of stairs'). Similarly, an individual with low extremity impairments who is capable of 'climbing down one flight of stairs' would be expected to be more capable of 'step up or down a standard curb' (easier than 'climbing down one flight of stairs'). This logical pattern is similar to that of hypothetical activity hierarchy based on the MET values. That is, the empirical hierarchy of walking/moving related activities may correspond to the hypothesized activity hierarchy based on physiological energy expenditure using the METs. The similarity of these two hierarchies may allude to areas of unexplored research. Future research investigating associations between self report measures and physiological measures could be indicated.

Conclusion

The purpose of this study are to determine dimensionality, item-level psychometrics, empirical item hierarchy of walking/moving domain of the ICF-AM, and what extent the empirical hierarchical structure of the ICF-AM is linked to the conceptual model based on the level of energy expenditure described as MET. Exploratory and confirmatory factor analysis were applied to determine the dimensionality and the Rasch rating scale model (1-parameter item response theory model) was applied to determine the hierarchical order of item difficulty of the walking/moving construct of the ICF-AM. Comparisons were drawn between the physiologically based MET value as a conceptual hierarchy and the empirically based item difficulty hierarchy of activity. The logical patterns of two hierarchies of walking/moving activity may imply correspondence between empirical and physiological measures. Thus, the advantage of comparable self-reported measures may be achieved by scrutinizing item difficulty hierarchy using IRT methodology.

References

Ainsworth BE. How do I measure physical activity in my patients? Questionnaires and objective methods. Br J Sports Med. 2009;43(1):6-9.

Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 compendium of physical activities: A second update of codes and met values. Med Sci Sports Exerc. 2011;43(8):1575-1581.

Balady GJ. Survival of the fittest--more evidence. N Engl J Med. 2002;346(11):852-854.

Bassett DR, Cureton AL, Ainsworth BE. Measurement of daily walking distance-questionnaire versus pedometer. Med Sci Sports Exerc. 2000;32(5):1018-1023.

Berlin JE, Storti KL, Brach JS. Using activity monitors to measure physical activity in free-living conditions. Phys Ther. 2006;86(8):1137-1145.

Blair SN, Haskell WL, Ho P, et al. Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. Am J Epidemiol. 1985;122(5):794-804.

Bond TG, Fox CM. Applying the Rasch Model, Fundamental measurement in the human sciences. 2nd ed. Mahwah, NJ, Lawrence

Erlbaum Associates Publishers, 2001:158-172.

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

Brown TA. Confirmatory factor analysis for applied research. 2008:65-73.

Byrne NM, Hills AP, Hunter GR, et al. Metabolic equivalent: One size does not fit all. J Appl Physiol. 2005;99(3):1112-1119.

Cattell RB. The scree test for the number of factors. Mutivariate Behav Res. 1996;1(2):245-276.

Child D. The Essentials of Factor Analysis. 3rd ed. New York, Continuum, 2006:45-47.

Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35(8):1381-1395.

Freedson PS, Brendley K, Ainsworth BE, et al. New techniques and issues in assessing walking behavior and its contexts. Med Sci Sports Exerc. 2008;40(7 Suppl):S574-S583.

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

Institute for Objective Measurement Inc.. Correlations: point-biserial, point-measure, residual [Internet]. Chicago, 2005 [cited 2013 July 30]. Available from: http://www.winsteps.com/ winman/index.htm? correlations.htm. Correlations: Point-biserial, point-measure, residual.

Pereira MA, FitzerGerald SJ, Gregg EW, et al. A collection of physical activity questionnaires for health-related research. Med Sci Sports Exerc. 1997;29(6 suppl):S1-S205.

Kriska AM, Caspersen CJ. Introduction to a collection of physical activity questionnaires. Med Sci Sports Exerc. 1997;29(6):5-9.

Kriska AM, Sandler RB, Cauley JA, et al. The assessment of historical physical activity and its relation to adult bone parameters. Am J Epidemiol. 1988;127(5):1053-1063.

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

Institute for Objective Measurement Inc.. What do infit and outfit, mean-square and standardized mean? [Internet]. Chicago, 2002 [cited 2013 July 30]. Available from: http://www.rasch.org/ rmt/rmt162f.htm. What do infit and outfit, mean-square and standardized mean?.

Linacre JM. Winsteps Rasch Measurement Computer Program. Chicago, IL, winsteps.com, 2005:13-14.

Norman GR, Steiner, D.L. Biostatistics: The bare essentials. St. Louise, MO, Mosby Yearbook Inc., 1994:125-128.

Paffenbarger RS, Jr., Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. Am J Epidemiol. 1978;108(3): 161-175.

Reckase MD. The difficulty of test items that measure more than one ability. Appl Psychol Meas. 1985;9(4):401-412.

Sabari JS, Lim AL, Velozo CA, et al. Assessing arm and hand function after stroke: A validity test of the hierarchical scoring system used in the motor assessment scale for stroke. Arch Phys Med Rehabil. 2005;86(8):1609-1615.

Taylor-Piliae RE, Haskell WL, Iribarren C, et al. Clinical utility of the stanford brief activity survey in men and women with early-onset coronary artery disease. J Cardiopulm Rehabil. 2007;27(4):227-232.

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

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

Ware JE Jr. A 12-item short-form health survey: Construction of scales and preliminary tests of

reliability and validity. Med Care. 1996;34(3): 220-233.

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. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil. 1989;70(12):857-860.

Wright BD, Linacre JM. Reasonable mean-square fit values. Rasch Meas Transac. 1994;8:370.

Wright BD, Masters GN. Rating Scale Analysis. Chicago, IL, MESA Press, 1982:94-101.