



ORIGINAL ARTICLE

Application of the Korean version of the Modified Barthel Index: Development of a keyform for use in clinical practice



Ickpyo Hong ^{a,*}, Yoonjeong Lim ^b, HyunSuk Han ^c,
Catherine Cooper Hay ^a, Hee-Soon Woo ^d

^a Division of Rehabilitation Sciences, School of Health Professions, University of Texas Medical Branch, Galveston, TX, USA

^b University of Florida, Rehabilitation Science Program, Department of Occupational Therapy, College of Public Health and Health Professions, Gainesville, FL, USA

^c Program in Research and Evaluation Methodology, College of Education, University of Florida, Gainesville, FL, USA

^d Department of Occupational Therapy, College of Medicine, Wonkwang University, Iksan, South Korea

Received 23 February 2017; received in revised form 20 May 2017; accepted 3 June 2017
Available online 6 July 2017

KEYWORDS

Outcome measure
(health care);
Keyform;
Rasch analysis;
Psychometrics;
Modified Barthel
Index;
Translational
research

Summary Objective: To demonstrate the clinical application of the Korean version of the Modified Barthel Index (K-MBI) using Rasch analysis.

Methods: A total of 276 patients with neurological disorders were assessed with the K-MBI in outpatient clinics. The Rasch partial-credit model was used to generate a keyform based on investigating the psychometric properties of the K-MBI, including dimensionality, precision (person strata and reliability), and hierarchical item difficulty. The Minimal Detectable Change (MDC) in item difficulty was used to establish right-challenging treatment goals and long-term treatment plans.

Results: The findings demonstrated that the Korean version of the MBI satisfied the assumption of unidimensionality. It also showed a hierarchical structure in terms of item difficulty, good reliability (Cronbach alpha, 0.92), and approximately five distinct person strata (4.6). The MDC (raw score, 20.1) of the item difficulty of the test items demonstrated equivalent cut-off scores for targeted short-term treatment goals on the keyform, a Rasch-derived display of patient responses. Long-term treatment goals were identified based on the test items of the keyform.

* Corresponding author. Division of Rehabilitation Sciences, School of Health Professions, University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555, USA. Fax: +1 409 747 1638.
E-mail address: ichong@utmb.edu (I. Hong).

Conclusions: The findings suggest that a Rasch keyform can be applied to clinical practice in Korean settings by identifying clinically and statistically meaningful test items and their step thresholds as short- and long-term goals.

Copyright © 2017, Hong Kong Occupational Therapy Association. Published by Elsevier (Singapore) Pte Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

In the rehabilitation area, approximately 90 activities of daily living (ADL) instruments have been developed (McHorney, 2003). Among those instruments, the Functional Independence Measure (FIM™) is commonly used in inpatient rehabilitation facilities in North America (Ottenbacher, Hsu, Granger, & Fiedler, 1996), and the Barthel Index (BI) and Modified Barthel Index (MBI) are commonly used in Europe and Asia. The MBI is originated in stroke rehabilitation but now its use has been extended to rehabilitation outcome measurement of other people with various kinds of disabilities. In South Korea, the FIM and Korean version of MBI (K-MBI) are most often used in rehabilitation hospitals. However, users of the FIM have to pay to be certified in its administration, and they are required to receive a mandatory training to use the instrument. In contrast with the FIM, the MBI and K-MBI do not require a mandatory training and a certification fee for using the instrument in any facility. For these reasons, the utilization of the MBI and K-MBI has increased. For example, the utilization of the K-MBI was 37.9% in South Korea in 2006 (Yoo, Jung, Park, & Choi, 2006), and the use of the K-MBI is expected to continue to increase in Korea.

Since the 1980s, item response theory (IRT) and the one-parameter model of the IRT, Rasch analysis, have been applied to test psychometric properties of ADL measures, including the BI and MBI (Duncan, Lai, Bode, Perera, & DeRosa, 2003; van Hartingsveld, Lucas, Kwakkel, & Lindeboom, 2006; Kucukdeveci et al., 2000; Liu, Unick, Galik, & Resnick, 2015; de Morton, Keating, & Davidson, 2008). Rasch analysis enables researchers to create a sample- and test-independent measure (Wright, 1968; Wright & Stone, 1979). A measurement construct calibrated by Rasch analysis can estimate person ability and item difficulty regardless of the person ability distributions (norms) of the target samples. In addition, person and item calibrations in a linear scale enable researchers and clinicians to perform mathematical calculations for the calibrated person and item measures (Merbitz, Morris, & Grip, 1989).

Based on a linear scale, hierarchical item difficulties can be created, and the hierarchical structures have been used in establishing logical treatment strategies and goals (Bode, Heinemann, Kozlowski, & Pretz, 2013; Linacre, 1997; Velozo, Warren, Hicks, & Berger, 2013; Velozo & Woodbury, 2011). For instance, a keyform is a visual output of Winsteps® Rasch Measurement software (Linacre, 2016), and the Rasch output provides person-ability and item difficulty on a linear measurement scale. In clinical settings, a keyform has been used to

approximate person-ability and possible treatment tasks based on subjects' responses on the test items (Bode et al., 2013; Linacre, 2016; Velozo et al., 2013; Velozo & Woodbury, 2011).

As the primary purpose of outcome measures was to precisely and accurately measure a patient's current health status and establish achievable treatment goals for the patient, clinicians should be able to easily interpret the measurement outputs. In addition, measurement outputs should suggest evidence-based rationales for clinicians in order for them to establish optimized treatment goals. Therefore, the purpose of this study is to demonstrate clinical applications of a Rasch output, keyform, in accurately measuring patient functional status and logically establishing short- and long-term treatment goals that match a patient's current functional status level as measured by the K-MBI.

Methods

Participants

The participants in this study were extracted from previous research supported by the Wonkwang University in South Korea (Hong et al., 2016). Community-dwelling adults who visited rehabilitation facilities were assessed with the K-MBI. Occupational therapists who had at least 5 years of clinical experience with the instrument assessed the patients. All participants in the study needed to 1) be adults between 18 and 89 years old, and 2) agree to release their de-identified data for research purposes. The data were collected during the 2-month data collection period (April 2015 – June 2015) from three urban rehabilitation hospitals in Gyeonggi-do and Daejeon in South Korea. All participants in the dataset signed the informed consent form for this study. As the extracted participants were de-identified, this study was considered as non-human research by the Institutional Review Board of the Wonkwang University.

Instrument

Korean version of the Modified Barthel Index (K-MBI)

The MBI is one of the ADL outcome measures, and consists of 10 items: feeding, personal hygiene (grooming), bathing, dressing, toilet transfer, bladder control, bowel control, chair/bed transfers, stair climbing, and ambulation (Shah, Vanclay, & Cooper, 1989) (see Fig. 1). In Korea, the MBI was translated into Korean (K-MBI) by bilingual physiatrists using double translation method (back translation). The contents of test items (i.e., eating and grooming) were

revised to reflect the Korean culture and lifestyle (Jung et al., 2007). For instance, in the eating task, “cut meat” was revised to “cut Kimchi” and in the grooming task, “insert the blade” was revised to “trimming nails.” The K-MBI was validated with patients with stroke and demonstrated good intra-rater reliability (0.87~1.00) and good inter-rater reliability (0.93~0.98) among physiatrists, resident physicians of rehabilitation medicine and occupational therapists (Jung et al., 2007). The K-MBI has three different rating scales: a score range of 0–5 (bathing, and personal hygiene), a score range of 0–10 (feeding, dressing, toilet use, bladder control, bowel control, and stair climbing), and a score range of 0–15 (chair/bed transfers and ambulation). A higher score on the K-MBI represents a higher degree of independence in performing basic ADLs.

Study procedure

The keyform development of the K-MBI required four steps: 1) analyzing psychometric properties of the K-MBI, 2) creating the keyform, 3) calculating cut-off scores for short- and long-term treatment goals, and 4) demonstrating the clinical application of the keyform using a subject’s responses.

Step1 Psychometrics

We analyzed the data based on the Rasch partial credit model using Winsteps® Rasch Measurement, Version 3.91.0 (Linacre, 2016). In this study, the Rasch partial credit model was used because it is a flexible model for the various rating scale structure (different step thresholds for each item) of the K-MBI. Equation (1.1) represents the used Rasch model in the study (Wright & Masters, 1982).

$$P_{nix} = \frac{\exp \sum_{j=0}^x (\beta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (\beta_n - \delta_{ij})}, \quad x=0, 1, \dots, m_i \quad (1.1)$$

where the probability (P_{nix}) of person n scoring x on item i is a function of the person’s ability (β_n) and the step thresholds of m_i on item i .

The dimensionality of the K-MBI was investigated with point-biserial correlations, mean square standardized residuals (MNSQ) values, and Rasch principal components analysis of residuals (PCAR). Any negative point-biserial correlations and less than a 0.3 value were considered as potentially problematic test items (Bond & Fox, 2001). The criteria of fit statistics (both infit and outfit) for the clinical observation data were a value of MnSq between 0.5 to 1.7 and a value of standardized fit statistics (ZSTD) between -2.0 and 2.0 (Wright, Linacre, Gustafson, & Martin-Lof, 1994). For PCAR, the first dimension should account for at least 20 percent of the test variance and a variance greater than fifty percent explained by the measure supports unidimensionality (Reckase, 1979). Local independence was considered as less than the absolute value of 0.2 in the residual correlations among the test items (Reeve et al., 2007). We used Winsteps® Rasch Measurement, Version 3.91.0 (Linacre, 2016) to conduct PCAR and Mplus version 7.4 to calculate residual correlations (Muthén & Muthén, 2015).

The precision of the K-MBI was investigated by calculating person strata as $(4 \times G_p + 1)/3$, where G_p is person separation (Wright & Masters, 1982). We considered an acceptable number of person strata as 3.0 which is equivalent to a traditional reliability of 0.8 (Fisher, 1992). Cronbach’s alpha was used to report test reliability, and the acceptable value was considered as at least 0.90 (Bland & Altman, 1997).

Step 2 Creating keyform

A keyform was created based on the constructed hierarchical item difficulty of the K-MBI. Person measures in a

0	10	20	30	40	50	60	70	80	90	100	ITEM	
					0	: 2	: 5	8	:	10	Stairs_Climbing	
		0	:	1	:	3	:	4	:	5	Bathing	
			0	:	3	:	8	:	12	:15	Ambulation	
			0	:	3	:	8	:	12	: 15	Toilet_Transfer	
			0	:	1	:	3	:	4	:	5	Wheelchair_Use
		0	:	2	:	5	:	8	:	10	Dressing	
			0	:	3	:	8	:	12	: 15	Bed_transfer	
	0	:	1	:	3	:	4	:	5		Grooming	
			0	:	2	:	5	:	8	:	10	Feeding
			0	:	2	:	5	:	8	:10	Bladder_Control	
			0	:	2	:	5	:	8	:10	Bowel_Control	
0	10	20	30	40	50	60	70	80	90	100		

Figure 1 Keyform of the Modified Barthel Index. The most difficult item, stairs_climbing, was located on the top and the easiest item, bowel_control, was located on the bottom. The 0–100 point scale on both top and bottom indicates the difficulty levels of each rating category and step thresholds in each item. The numbers in the keyform indicate rating scale categories, and “:” indicates step threshold (half-point units) of each item.

logit scale were converted to a 0–100 point scale (unit) for easy application in clinical areas. The step thresholds (m_i on item i in Equation (1.1)) of each item were located on the keyform according to their hierarchical difficulty calibrations (see Fig. 1). For example, step thresholds of the stair climbing item were the distance among its rating scales: 0, 2, 5, 8, and 10.

Step 3 Calculating cut-off scores for treatment goals

The cut-off score for appropriately challenging tasks was estimated from a previously reported minimal detectable change (MDC) value of the Barthel Index. Hsieh et al. (2007) reported the MDC of the 20-point BI as a raw score of 4.01 (Hsieh et al., 2007). In our study, the MDC of the K-MBI was estimated by multiplying 5 to the MDC of the BI ($5 \times 4.02 = 20.1$) because the BI consists of the same items compared to the K-MBI and the BI's score range (0 to 20) was one fifth of the K-MBI score range (0 to 100). Therefore, the estimated MDC value (20.1 points) was used as a cut-off score for identifying a short-term goal. For long-term goals, a cut-off score is considered as two-step thresholds apart from the cut-off units used for the short-term goals on each test item.

Step 4 Demonstration of clinical application

Three subjects with low, middle, and high abilities were randomly selected, and their responses were marked on the keyform. The three subjects' abilities were estimated based on the actual response patterns on the keyform (Bond & Fox, 2001). Velozo and Woodbury (2011) suggested

that a keyform can be used to approximate person ability by identifying a "transition zone" that indicates an individual's performance pattern fluctuates between two rating scales (p. 1217). Therefore, a vertical line across the middle of the transition zone on a keyform is an approximation of the individuals' functional status level. The three subjects' person ability and MDC scores were presented as vertical lines on the keyform. Based on the two vertical lines, short- and long-term treatment goals were established.

Results

Participants

Table 1 represents the demographics of the participants in this study. A total of 276 participants were extracted as a secondary dataset. In the dataset, males totaled 164 (59.4%), and the average age was 59.4 years old ($SD = 15.8$). The average number of years since onset of chronic diseases was 1.6 year ($SD = 1.9$); the majority of the sample had experienced a stroke, 193 (71.5%), and were unemployed, 265 (96.0%).

Psychometrics

Table 2 presents the item units, infit and outfit fit statistics and point-biserial correlations. All test items demonstrated high positive point-biserial correlations (all > 0.62). According to the fit criteria in this study, the K-MBI demonstrated only a single misfit item, Feeding (MNSQ,

Table 1 Demographic Characteristics of Subjects ($n = 276$).

Characteristics		Number subjects (%)	$M \pm SD$
Age (years)		—	59.4 ± 15.8
Average number of years since onset (years)		—	1.6 ± 1.9
Gender	Male	164 (59.4)	—
	Female	112 (40.6)	—
Marital status	Married	192 (70.0)	—
	Single	32 (11.7)	—
	Widowed	32 (11.7)	—
	Divorced	18 (7.6)	—
Education	Less than high school	108 (39.6)	—
	High school	81 (29.7)	—
	Some college	14 (5.1)	—
	College degree	63 (23.1)	—
	Graduate degree	7 (2.5)	—
Employment status	Unemployed	265 (96.0)	—
	Employed	11 (4.0)	—
Primary impairments	Stroke	193 (71.5)	—
	Brain injury, Non-traumatic	7 (2.6)	—
	Brain injury, Traumatic	19 (7.0)	—
	Neurological conditions, other	16 (5.9)	—
	Guillain–Barre Syndrome	3 (1.1)	—
	Spinal cord injury, Non-traumatic	15 (5.6)	—
	Spinal cord injury, Traumatic	14 (5.2)	—
	Other	3 (1.1)	—

SD: Standard deviation.

Table 2 Fit statistics and Point-biserial Correlation.

Item	Unit	Infit		Outfit		Point-biserial correlation
		MNSQ	ZSTD	MNSQ	ZSTD	
Stair climbing	61.56	1.06	.7	.87	-.7	.62
Bathing	56.10	.72	-3.3	.72	-3.1	.68
Ambulation	52.91	.85	-1.6	.83	-1.8	.74
Toilet use	49.42	.65	-4.4	.63	-4.1	.79
Dressing	48.90	.83	-2.0	1.14	1.3	.76
Transfer bed	47.01	.94	-.6	1.57	4.6	.80
Feeding	43.87	1.45	4.4	1.88	6.9	.72
Grooming	43.80	.59	-3.7	.52	-4.7	.79
Bladder control	36.57	1.68	5.7	1.50	3.6	.75
Bowel control	35.31	1.42	3.6	1.15	1.1	.78

MNSQ: mean square standardized residuals, ZSTD: standardized fit statistics.

1.88; ZSTD, 6.9). However, Feeding was only slightly beyond the 1.7 cutoff, so the item was not removed. The results of PCAR showed that the Rasch measurement model accounted for 75.6% of the variance, which satisfied the unidimensionality assumption. Local independence was met since no residual correlations exceeding 0.2 were found among items. The person separation (Gp) of the K-MBI was 3.25, meaning that the K-MBI can statistically separate the sample into at least four distinct groups (person strata = 4.6). The test reliability of the instrument was 0.92 (Cronbach alpha). The standard error of measurement (SEM) for the Rasch measurement model was 7.09.

The K-MBI demonstrated a hierarchical structure of item difficulty (Table 2). The most difficult items were stair climbing (61.56 unit) and bathing (56.10 unit), while the easiest items were bowel control (35.31 unit) and bladder control (36.57 unit). The person abilities (average of 48.4, SD = 13.04 unit) was well-matched to item difficulty (average of 47.75, SD = 10.45 unit).

Demonstration of clinical application

Keyform

Fig. 1 represents the keyform of the K-MBI. On the right column, the test items are located according to the hierarchy of item average difficulty (bowel control, the easiest, is at the bottom and stair climbing, the most difficult, is at the top). Note that the intervals between step thresholds varied across the test items. Bowel control and bladder control show narrow threshold ranges while grooming, dressing and bathing show wide threshold ranges.

Keyform application

Fig. 2 represents the clinical application of the keyform. We marked the three selected subjects' responses on the keyform with black circles. Based on the subjects' response patterns, the person ability of the subjects who had low (spinal cord injury), middle (stroke), and high ability (stroke) were 30 unit, 34 unit, 65 unit, respectively. However, their accrual person ability levels estimated by the Rasch model were 26.71 unit, 39.87 unit, and 58.80 unit, respectively. The cut-off scores for appropriately challenging tasks for the three subjects were calculated by

adding their person units (estimated on the keyforms) and the MDC of the K-MBI (20.1 units); therefore, their short-term treatment goals are indicated by the solid vertical lines to the far right on the keyform at 50.1 units, 54.1 units, and 85.1 units, respectively. In addition, the step thresholds in the black rectangle shape on the figure would be potential long-term goals for the subject.

For example, the subject having middle person ability (the second panel on Fig. 2b) was able to perform bladder control and bowel control independently; however, he/she had difficulty in conducting the other test items, such as requiring maximum assistance on feeding and being unable to perform dressing. This subject's short-term (the solid vertical lines to the far right on the keyforms) and long-term (squares on the keyforms) treatment goals were estimated based on the cut-off score. Based on the keyform, the subject was able to perform bladder and bowel control independently. However, the subject needed maximal assistance for feeding (score of 2), grooming (score of 1), toilet use (score of 2), and ambulation (score of 3). The short-term goals were considered the step thresholds of the challenging items around the vertical cut-off score. For instance, potential short-term goals would be performing feeding with minimal assistance or supervision (score of 8); grooming (score of 3), bed transfer (score of 8), dressing (score of 5), toilet use (score of 6), ambulation (score of 8), and bathing (score of 3) with moderate assistance; and stair climbing with maximal assistance (score of 2). The long-term goals would be achieving more difficult step thresholds for the challenging items, two step thresholds apart from the cut-off units for the short-term goals. For instance, possible long-term goals would be achieving bed Transfer (score of 15), dressing (score of 10), toilet use (score of 15), ambulation (score of 15), and bathing (score of 5) independently, and performing stair climbing with minimal assistance (score of 8).

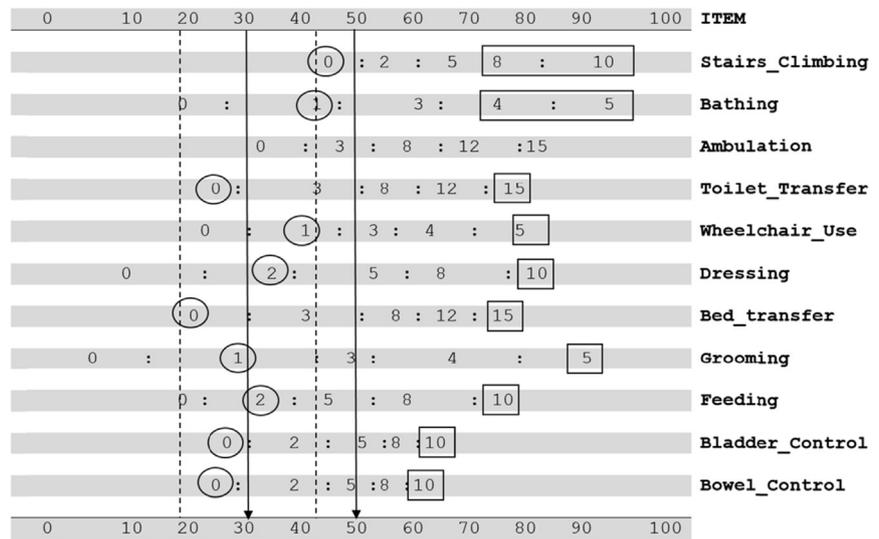
As an individual's ability increases, the challenge of short-term goals and long-term goals also increases. For example, the subject with the lowest ability demonstrated a wide range of step thresholds that could be used for short-term and long-term goals on the keyform (Fig. 2a). However, the individuals with the highest ability needed only short-term goals equivalent to the estimated cut-off unit (85.1 units), and there were no available step thresholds on the keyform for the K-MBI test items that could be considered long-term goals (Fig. 2c).

Discussion

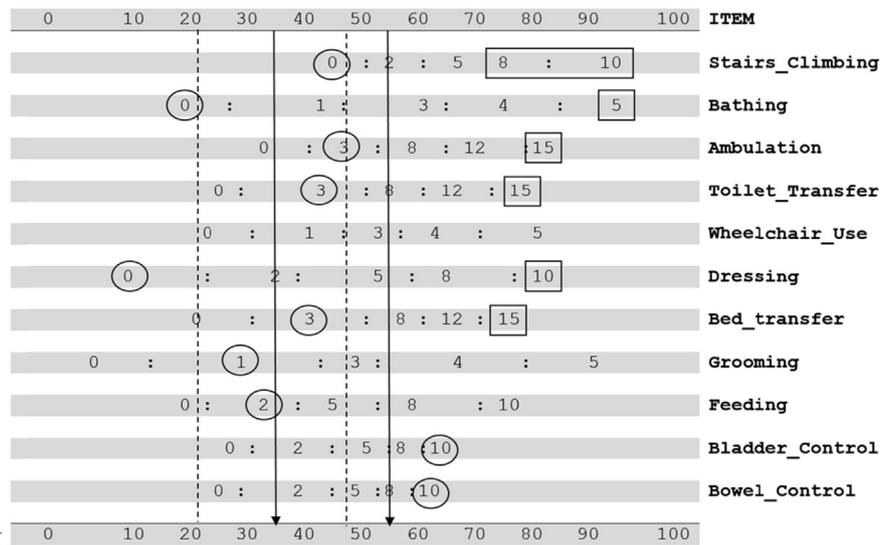
In this study, we investigated the psychometric properties of the K-MBI using IRT and demonstrated the clinical application of the Rasch outputs in establishing evidence-based short- and long-term treatment goals. The instrument demonstrated good precision, good test reliability, and a hierarchical item difficulty.

The identified hierarchical item difficulty of the K-MBI items was supported by previous Rasch analysis studies on the MBI and the BI. de Morton et al. (2008) reported that the most difficult items of the MBI were stair climbing and bathing, and the easiest items were feeding and bowel control (de Morton et al., 2008). Duncan et al. (2003)

a. Low function subject



b. Middle function subject



c. High function subject

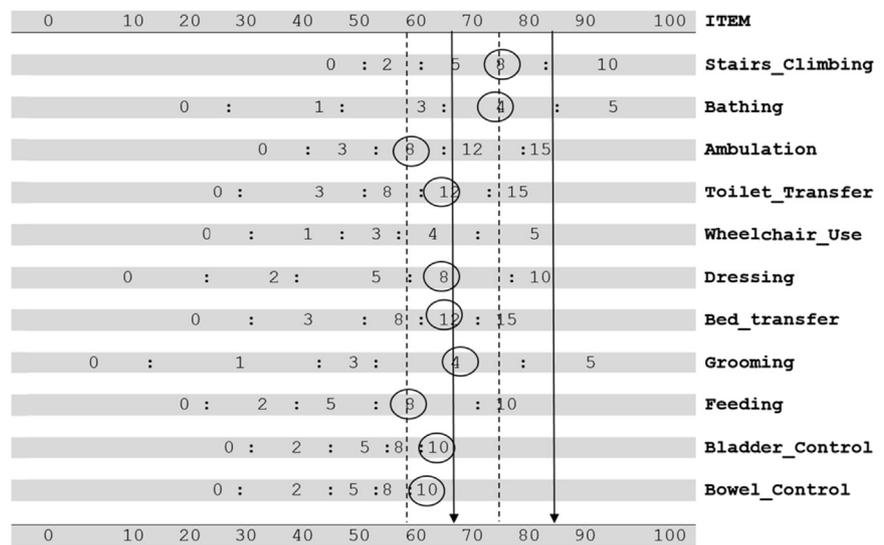


Figure 2 Demonstration of establishing short- and long-term treatment goals using the Keyform. Randomly selected three subjects whose person ability is low, middle, and high (from the top to bottom).

reported the same for most difficult items of the BI but found that bladder control was easier than feeding in their analysis (Duncan et al., 2003). As the keyform of the K-MBI was created based on the hierarchical item difficulty of the K-MBI, the construct validity of the keyform was also supported by previous studies.

We used the MDC score difference (20.1 unit) in order to establish treatment goals, and the SEM of the Rasch measurement model was 7.09 unit. The MDC score difference (20.1 unit) was 2.8 times greater than the value of the SEM (7.09 unit). In other words, when a person demonstrates a functional improvement of more than 20.1 units, the functional gain is a statistically significant improvement. Therefore, the value of MDC score difference (20.1 unit) used in this study can be considered clinically meaningful and a statistically significant improvement. Based on the MDC score difference value, we were able to establish logical short and long-term goals. For clinical applications, we provide a clean copy of the K-MBI keyform in this paper (See Appendix I).

We expect that our research methodologies will be applicable to other cultural versions of the MBI, as well as other rehabilitation outcome measures, such as the Functional Independence Measure and the Berg Balance Scale. Researchers or clinicians who have measurement expertise using Rasch analysis could easily create various keyforms using secondary databases or prospective study databases that contain outcome measure scores recorded at the item level. Once various keyforms are developed, we also expect that clinicians would be able to readily establish evidence-based treatment goals.

We estimated the MDC score of the K-MBI based on the reported MDC score of the 20-point BI because the MDC of the K-MBI has not been reported yet. Although the K-MBI has a score range that is proportional to the score range of the 20-point BI, the actual MDC score of the K-MBI is needed to establish a more accurate cut-off score for short-term treatment goals. In addition, this study was a retrospective design using a secondary dataset. Thus, we did not test whether the established treatment goals were feasible and meaningful for selected subjects. Therefore, a prospective study is needed to verify whether the established treatment goals based on keyform are feasible and meaningful for patients in clinical settings.

Long-term goals were established based on an arbitrary value of two step thresholds apart from the cut-off units used for the short-term goals on each test item. Mathematically, a value of two standard errors of measurement can be considered a significant difference from the measured person ability estimation.

Since a keyform cannot estimate an individual's standard error, we used two step thresholds from the cut-off units for short-term goals. Future studies need to verify this methodology used for establishing long-term goals.

Our study has several limitations. First, using the actual response patterns may inflate the estimation of an individual's actual ability without using actual Rasch statistical software. For example, the person ability estimation differences between the keyform and the Rasch software range from 5.87 unit to 6.20 unit in the three randomly selected subjects. However, these person ability estimation differences were within a half person ability standard

deviation (SD = 13.04 unit) of the sample. While the keyforms may bias person ability estimations, its usability in clinical settings exceeds any generic limitations it suffers in terms of accuracy for individuals. In addition, the population of this study was a convenience sample consisted of patients with various diagnoses which might cause variations in the calculation of dimensionality.

Conclusion

The present study presented a clinical application of the Rasch output keyform by logically establishing treatment goals based on a patient's ability. The keyform of the K-MBI estimated a relatively accurate measure of an individual's ability compared to actual Rasch calibration of Rasch software. In addition, the study demonstrated how to establish evidence-based treatment goals for patients in clinical settings without running sophisticated statistical analyses.

Acknowledgements

We certify that no party having a direct interest in the results of the research supporting this article has or will confer a benefit on us or on any organization with which we are associated. We certify that all financial and material support for this research and work are clearly identified in the title page of the manuscript. We thank Dr. Craig A. Velozo at the Medical University of South Carolina for copyediting the manuscript.

References

- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal*, 314(7080). <http://dx.doi.org/10.1136/bmj.314.7080.572>, 572–572.
- Bode, R. K., Heinemann, A. W., Kozlowski, A., & Pretz, C. (2013). Self-scoring templates for motor and cognitive subscales of the FIM for persons with spinal cord injury. *Archives of Physical Medicine and Rehabilitation*, 4(95), 676–679. <http://dx.doi.org/10.1016/j.apmr.2013.11.009>.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Duncan, P. W., Lai, S. M., Bode, R. K., Perera, S., & DeRosa, J. (2003). Stroke Impact Scale-16: A brief assessment of physical function. *Neurology*, 60(2), 291–296. <http://dx.doi.org/10.1212/01.WNL.0000041493.65665.D6>.
- Fisher, W. P., Jr. (1992). Reliability, separation, strata statistics. *Rasch Measurement Transactions*, 6(3), 238.
- Hong, I., Woo, H.-S., Shim, S., Li, C.-Y., Lim, Y., & Velozo, C. (2016). Equating activities of daily living outcome measures: The Functional Independence Measure and the Korean version of Modified Barthel Index. *Disability and Rehabilitation*. <http://dx.doi.org/10.1080/09638288.2016.1247468> (advance online publication).
- Hsieh, Y. W., Wang, C. H., Wu, S. C., Chen, P. C., Sheu, C. F., & Hsieh, C. L. (2007). Establishing the minimal clinically important difference of the Barthel Index in stroke patients. *Neuro-rehabilitation and Neural Repair*, 21(3), 233–238. <http://dx.doi.org/10.1177/1545968306294729>.
- Jung, H. Y., Park, B. K., Shin, H. S., Kang, Y. K., Pyun, S. B., Paik, N. J., et al. (2007). Development of the Korean version of Modified Barthel Index (K-MBI): multi-center study for subjects

- with stroke. *Journal of Korean Academy of Rehabilitation Medicine*, 31(3), 283–297.
- Kucukdeveci, A., Yavuzer, G., Tennant, A., Suldur, N., Sonel, B., & Arasil, T. (2000). Adaptation of the modified Barthel Index for use in physical medicine and rehabilitation in Turkey. *Scandinavian Journal of Rehabilitation Medicine*, 32(2), 87–92. <http://dx.doi.org/10.1080/003655000750045604>.
- Linacre, J. M. (1997). Instantaneous measurement and diagnosis. *Physical Medicine and Rehabilitation State of the Art Reviews*, 11(2), 315–324. Retrieved from <http://www.rasch.org/pm/pm1-55.pdf>.
- Linacre, J. M. (2016). A user's guide to WINSTEPS® 3.91.2. Retrieved from <http://www.winsteps.com/winsteps.htm>.
- de Morton, N. A., Keating, J. L., & Davidson, M. (2008). Rasch analysis of the Barthel Index in the assessment of hospitalized older patients after admission for an acute medical condition. *Archives of Physical Medicine and Rehabilitation*, 89(4), 641–647. <http://dx.doi.org/10.1016/j.apmr.2007.10.021>.
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus (Version 7.4)*. Los Angeles, CA. Retrieved from <https://www.statmodel.com/download/usersguide/Mplus%20Users%20Guide%20v6.pdf>.
- Liu, W., Unick, J., Galik, E., & Resnick, B. (2015). Barthel Index of activities of daily living: item response theory analysis of ratings for long-term care residents. *Nursing Research*, 64(2), 88–99. <http://dx.doi.org/10.1097/NNR.0000000000000072>.
- McHorney, C. A. (2003). Ten recommendations for advancing patient-centered outcomes measurement for older persons. *Annals of Internal Medicine*, 139(5_Part_2), 403–409. http://dx.doi.org/10.7326/0003-4819-139-5_Part_2-200309021-00005.
- Merbitz, C., Morris, J., & Grip, J. C. (1989). Ordinal scales and foundations of misinference. *Archives of Physical Medicine and Rehabilitation*, 70(4), 308–312.
- Ottenbacher, K. J., Hsu, Y., Granger, C. V., & Fiedler, R. C. (1996). The reliability of the functional independence measure: a quantitative review. *Archives of Physical Medicine and Rehabilitation*, 77(12), 1226–1232.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational and Behavioral Statistics*, 4(3), 207–230.
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*, 45(5 Suppl 1), S22–S31. <http://dx.doi.org/10.1097/01.mlr.0000250483.85507.04>.
- Shah, S., Vanclay, F., & Cooper, B. (1989). Improving the sensitivity of the Barthel Index for stroke rehabilitation. *Journal of Clinical Epidemiology*, 42(8), 703–709.
- van Hartingsveld, F., Lucas, C., Kwakkel, G., & Lindeboom, R. (2006). Improved interpretation of stroke trial results using empirical Barthel item weights. *Stroke*, 37(1), 162–166. <http://dx.doi.org/10.1161/01.STR.0000195176.50830.b6>.
- Veloze, C. A., Warren, M., Hicks, E., & Berger, K. A. (2013). Generating clinical outputs for self-reports of visual functioning. *Optometry and Vision Science*, 90(8), 765–775. <http://dx.doi.org/10.1097/OPX.0000000000000007>.
- Veloze, C. A., & Woodbury, M. L. (2011). Translating measurement findings into rehabilitation practice: an example using Fugl-Meyer Assessment-Upper Extremity with patients following stroke. *Journal of Rehabilitation Research and Development*, 48(10), 1211–1222. <http://dx.doi.org/10.1682/JRRD.2010.10.0203>.
- Wright, B. D. (1968). Sample-free test calibration and person measurement. In *Paper presented at the Proceedings of the 1967 invitational conference on testing problems*.
- Wright, B. D., Linacre, J. M., Gustafson, J., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.
- Yoo, E. Y., Jung, M. Y., Park, S. Y., & Choi, E. H. (2006). Current trends of occupational therapy assessment tool by Korean occupational therapists. *Journal of Korean Society of Occupational Therapy*, 14(3), 27–37.

Appendix I. Keyform of the Korean version of the Modified Barthel Index

0	10	20	30	40	50	60	70	80	90	100	ITEM			
					0	:	2	:	5	:	8	:	10	Stair_Climbing
		0	:	1	:	3	:	4	:	5			Bathing	
			0	:	3	:	8	:	12	:	15		Ambulation	
			0	:	2	:	5	:	8	:	10		Toilet_Use	
			0	:	1	:	3	:	4	:	5		Wheelchair_Use	
			0	:	2	:	5	:	8	:	10		Dressing	
			0	:	3	:	8	:	12	:	15		Bed_transfer	
			0	:	1	:	3	:	4	:	5		Grooming	
			0	:	2	:	5	:	8	:	10		Feeding	
			0	:	2	:	5	:	8	:	10		Bladder_Control	
			0	:	2	:	5	:	8	:	10		Bowel_Control	
0	10	20	30	40	50	60	70	80	90	100				