한국전문물리치료학회지 2017년 24권 4호 20-28
Phys Ther Korea 2017;24(4):20-28
ISSN (Print) 1225-8962, ISSN (Online) 2287-982X
https://doi.org/10.12674/ptk.2017.24.4.020

Rasch Analysis of the Korean Version of the Fullerton Advanced Balance Scale

Yong-jin Jeon[1], PhD, PT, Gyoung-mo Kim[2], PhD, PT
[1]Dept. of Physical Therapy, School of Medical and Public Health, Kyungdong University
[2]Dept. of Physical Therapy, Daejeon Health Sciences College

Abstract

Background: Rasch analysis has the advantage of placing both the items and the person along a single ratio scale and calibrates person ability and item difficulty onto an interval scale by logits. Therefore, Rasch analysis has been recommended as a better method for evaluating functional outcome questionnaires than traditional analyses.

Objects: The aim of current study was to investigate item fit, item difficulty, rating scale, and separation index of the Korean version of the Fullerton Advanced Balance (KFAB) scale using Rasch analysis.

Methods: In total, 93 patients with stroke (male=58, female=35) participated in this study. To investigate the item fit, difficulty, rating scale, and separation index of the KFAB scale, Rasch analysis was completed by the Winsteps software program.

Results: In this study, all items of the KFAB scale were included in the Rasch model. The most difficult item was 'standing with feet together and eyes closed', and the easiest item was 'two-footed jump'. The rating scale was a 4-point scale instead of the original 5-point scale. Person and item separation indices showed high values that can identify a person with a wide range of balance ability.

Conclusion: The KFAB scale appears to be a reliable and valid tool to assess balance function in patients with stroke. Furthermore, the scale was found to discriminate among stroke patients of varying balance abilities.

Key Words: Fullerton advanced balance scale; Rasch analysis; Stroke.

Introduction

The recovery of activities of daily living (ADL) performance is one of the most important goals in patients with stroke (van de Port et al, 2006). In patients with stroke, balance function is an important predictor for functional movement and walking as well as ability to perform independent ADL (Weerdesteyn et al, 2008). However, many patients with stroke have poor balance ability due to lack of integration and control of various sensory input such as vision, vestibular, and somatosensory, which contribute to maintaining balance and postural control (Bonan et al, 2004). Although approximately 75% of stroke patients recover independent standing balance, asymmetric weight bearing and unstable postural sway decreased ability to maintain balance and increased risk of falling (Geurts et al, 2005).

Falls and fall-related injuries are among the most common complications after stroke (Langhorne et al, 2000). Many hemiplegic stroke patients shift their center of gravity (COG) to the unaffected side when maintaining balance, and show asymmetric weight bearing in performing motor function. These factors lead to falls, and consequently decreased ADL and quality of life (Suzuki et al, 2013). Improvement of balance is often a major goal of rehabilitation. Therefore measurements of balance that are reliable and valid are necessary for evaluating the treatment outcome and for monitoring the balance function of stroke patients

Corresponding author: Gyoung-mo Kim kgm6240@naver.com

over time (Hellström and Lindmark, 1999).

Clinical balance scales are useful assessment tools to evaluate various dimensions of postural control and balance (Schlenstedt et al, 2016). In particular, multi-dimensional clinical balance scales have the advantage of providing a detailed explanation of symptoms and predictive information about potential risk factors for falls (Li et al, 2012). The Berg Balance Scale (BBS) has strong psychometric properties and is useful in evaluating balance function after stroke. However, caution should be taken when electing to use the BBS to measure change in patients assigned to either severe or mild impairment groups, due to floor and ceiling effects (Blum and Korner-Bitensky, 2008).

Recently, the Fullerton Advanced Balance (FAB) scale was used to evaluate balance function and to predict falls with neurologic disorders such as Parkinson's disease and cerebral palsy (Schlenstedt et al). The FAB scale is a performance-based measure that was developed by Rose et al (2006) to avoid ceiling effect and to evaluate higher balance functioning. The FAB scale items included that static and dynamic postural control, sensory reception and integration, and reactive postural control in response to a perturbation. The FAB scale items also included a secondary task during walking. These features might reflect subtle balance challenges during ADL, possibly allowing the detection of small differences in balance performance, and imporved sensitivity to the change of intervention (Rose et al, 2006; Schlenstedt et al, 2015).

The Rasch analysis estimates the probability that a respondent will endorse an item and select a particular rating for that item, and focuses on the psychometric properties of the individual items of a measurement tool instead of the tool as a whole (Velozo and Peterson 2001). For these reasons, the Rasch analysis has been used increasingly to validate functional outcome questionnaires in various fields, such as neurology, orthopedics, and physical medicine (Duncan et al. 2003).

Although the FAB scale was translated Korean and provided reliability and validity in older adults and individuals with cerebral palsy (Jeon and Kim, 2016; Kim, 2016a), it remains unclear whether this scale is valid to evaluate balance function and identify psychometric properties in patients with stroke.

The primary purpose of the current study was to examine measurement properties of the Korean version of the FAB scale (KFAB) by applying Rasch analysis, and to identify misfit items from the domains to improve the measurement of the KFAB scale in patients with stroke.

Methods

Participants

Ninety-three patients (male=58, female=35) with stroke who had been receiving rehabilitation at several hospitals participated in this study. Inclusion criteria were: first-ever stroke, at least six months since onset, ability to walk indoors without any walking aid, and a Mini-Mental State Examination-Korean score of 23 or higher. Exclusion criteria were: other neurological diseases, fractures, and other obvious symptoms that might affect balance. The general characteristics of the participants are presented in Table 1.

Table 1. Demographic characteristics of the participants (N=93)

Parameters	Males (n_1=58)	Females (n_2=35)	Total
Height (㎝)	169.5±6.6[a]	156.5±5.4	164.6±8.8
Weight (㎏)	69.0±10.7	58.4±8.5	65.1±11.2
Age (year)	60.1±10.8	63.2±12.2	61.7±11.4
Onset duration	22.2±24.2	24.8±24.9	23.2±24.2
Total FAB[b] score	20.1±11.4	16.3±10.2	18.7±11.1

[a]mean±standard deviation, [b]Fullerton advanced balance scale.

Procedures

Participants were sufficiently informed of the study procedure, submitted written consent forms, and completed a form requesting general characteristics (gender, age, and medical information). After completing the information form, each participant was assessed to determine their degree of balance function using the KFAB scale. To minimize measurement error, evaluations were performed in a comfortable and familiar physical therapy room where patients had been treated. To demonstrate that the patients understood the KFAB scale, verbal and physical cues were allowed, and only one practice trial per item was provided.

Instruments

The FAB scale was developed to evaluate balance function in higher functioning older adults. It consists of 10 items that require static and dynamic activities in multiple environments, including: standing with feet together and eyes closed, reaching forward to an object, turn in full circle, step up and over, tandem walk, stand on one leg, stand on foam with eyes closed, two-footed jump, walk with head turns, and reactive postural control. Each item is scored on a 5-point ordinal scale (0 to 4), and the total score range is 0 to 40. Standardized scoring descriptors are specific to each item; item score range from 0 (unable to perform or not attempted) to 4 (maximum test item score). The higher score indicate better balance ability, and a score of 25 or lower indicates a high risk of falling. It has high test - retest reliability ($r=.96$) and interrater reliability ($r=.92$), and a good correlation with the BBS ($r=.75$, $p<.001$) (Hernandez and Rose, 2008; Klein et al, 2011; Rose et al, 2006). In this study, the KFAB scale was used. The KFAB scale demonstrated a high reliability (intraclass correlation coefficients=.99) and validity ($r=.89$) in elderly people study (Kim, 2016b).

Statistical analysis

The collected KFAB scale data were analyzed using Rasch analysis with Winsteps program version 3.71.0 (Linacare, Chicago, IL, USA). Rasch analysis is based on the probability that a person's level on an item is a function of the person's ability and the difficulty of the item. Raw scores are converted to equal-interval measures using a constant metric, called a logit or log odds unit (Klein et al, 2011). Rasch analysis uses these equal-interval measures to assess multiple psychometric characteristics: item fit (the extent to which items measure a single construct); item difficulty (the ordering of items from least to most difficult to perform), and separation (the extent to which the items distinguish distinct levels of functioning within the domain) (Duncan et al, 2003). We verified the item fit of the scale for 10 items of the KFAB scale using the infit and outfit values. These fit statistics are represented as mean square residual (MNSQ). The MNSQ represents the observed response value divided by expected response, and the desired MNSQ value is 1 (White and Velozo, 2002). Both infit and outfit MNSQ values in the range .6 to 1.4 are deemed acceptable in case of Likert scale (Wright and Linacare, 1989). The standardized residual is considered adequate if the value does not exceed ±2, which provides a significance test for the MNSQ (Davidson, 2009). Furthermore, the Rasch analysis converts ordinal-scale scores to interval-scale scores, which are calibrated on a single linear measurement continuum divided into equal intervals, or logits, for each item (Duncan et al, 2008). The measurement properties of the rating scale were assessed based on the following criteria: the number of observed counts per category is at greater than 10, the average measure considered acceptable if the score ordered from low to high, category fit meets an outfit MNSQ value of less than 2, threshold calibration should be difference of at least 1.4 logits but no more than 5 logits between categories, and each category should have a distinct peak in the probability curve graph (White and Velozo, 2002). Based on these results, we analyzed item fit, item difficulty, and separation index.

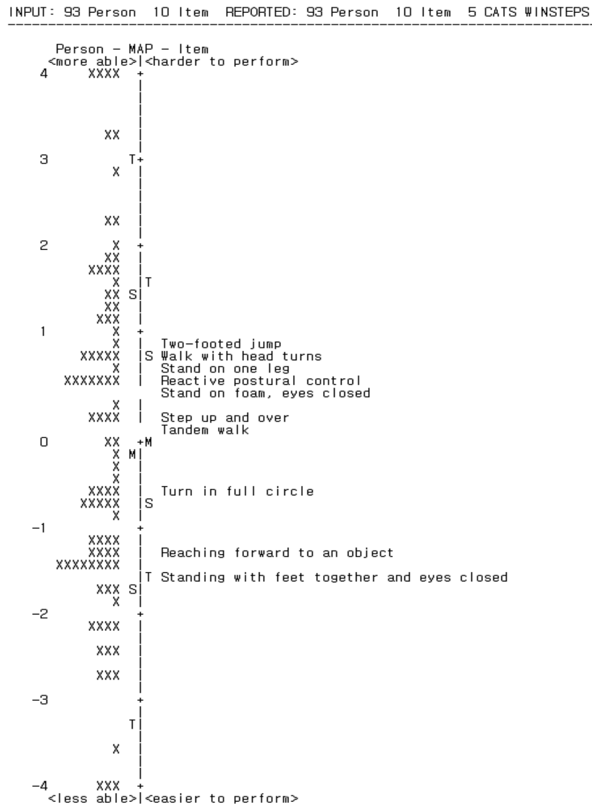

Figure 1. Item and person map for the 10 items of the Korean version of Fullerton advanced balance scale: Mean (M), 1 standard deviation (S), and 2 standard deviation (T).

Results

Item fit and difficulty of the KFAB

None of the 10 items of the KFAB scale showed misfit. The most difficult item was 'two-footed jump' (logit value, .82), whereas the easiest item was ' standing with feet together and eyes closed' (logit value, -1.51) (Table 2). Table 2 demonstrate the fit statistical of the KFAB scale items, arranged by order of item difficulty. Furthermore, the item and person map (Figure 1) illustrates that the item difficulty was plotted to the right of the vertical line and more difficulty items were located at the top of the map. The ability of the participants was to the left, and higher ability participants were located at the top of the map.

Rating scale of the KFAB

The observed count met the minimum criterion for each category, and the average measures were arranged from low to high scores sequentially. In the case of the 5-point rating scale of the KFAB scale, the category fit values acceptable, but the threshold calibration showed insufficient distinction between categories 2 (-.12) and 3 (.52) (Table 3). After converting, 4-point rating scale results showed that the threshold calibration demonstrated distinct difference between two category (Table 4). Furthermore, the probability curve graph illustrated that each category had a distinct peak in the 4-point compared as compared with the 5-point rating scale (Figure 2).

Item and person index and reliability of the KFAB

The item separation index of the KFAB scale was 5.37, and item reliability was .97. The person separation index of the KFAB scale was 2.95, and person reliability was .90 (Table 5).

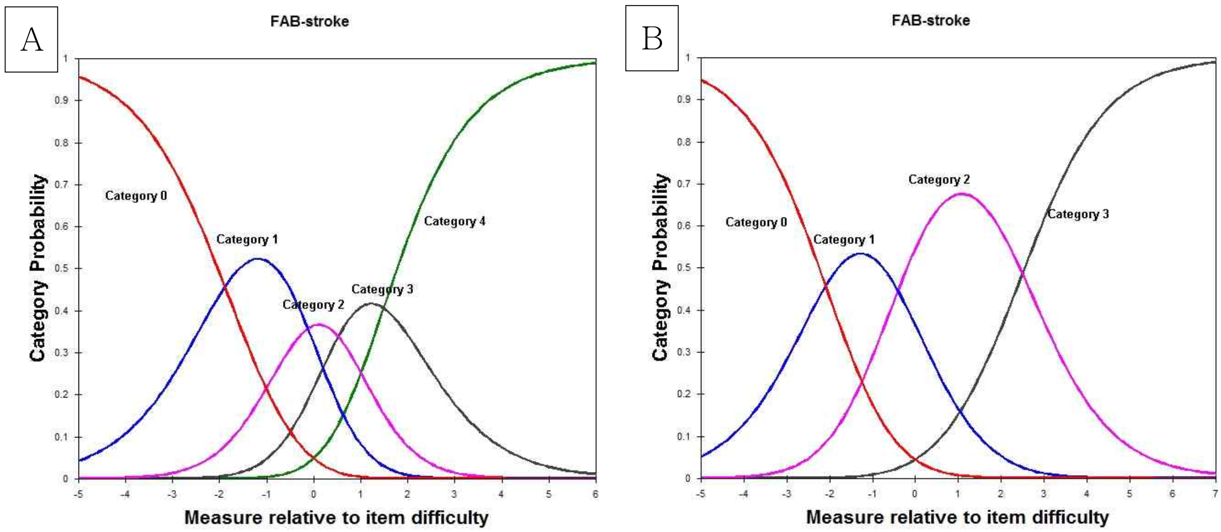

Figure 2. Category probability curve of the Korean version of Fullerton advanced balance scale (A) 5-point rating scale, (B) 4-point rating scale.

Table 2. Item fit and difficulty of the KFAB

No	Item	Logit	Error	Infit		Outfit	
				MNSQ[a]	ZSTD[b]	MNSQ	ZSTD
1	Standing with feet together and eyes closed	-1.51	.14	1.08	.5	1.10	.5
2	Reaching forward to an object	-1.22	.14	1.22	1.3	1.16	.8
3	Turn in full circle	-.56	.13	.97	-.1	.90	-.6
4	Step up and over	.13	.13	.97	-.1	1.02	.2
5	Tandem walk	.17	.13	1.05	.4	1.10	.6
6	Stand on one leg	.63	.14	.90	-.6	.90	-.6
7	Stand on foam, eyes closed	.39	.14	.97	-.8	.85	-.9
8	Two-footed jump	.82	.14	.71	-2.0	.73	-1.7
9	Walk with head turns	.65	.14	.72	-2.0	.73	-1.8
10	Reactive postural control	.50	.14	1.39	2.3	1.23	1.4
	Mean	.00	.14	.99	-.1	.97	-.2

[a]mean square residual, [b]standardized residual.

Table 3. The 5-point rating scale of the KFAB

Category score	Observed counts (%)	Average measure	Infit MNSQ[a]	Outfit MNSQ	Threshold calibration
0	198(21)	-2.00	1.00	1.02	None
1	239(26)	-1.15	.94	1.04	-1.89
2	162(17)	.13	.80	.74	-.12
3	151(16)	1.08	.83	.80	.52
4	180(19)	1.97	1.30	1.28	1.49

[a]mean square residual.

Table 4. The 4-point rating scale of the KFAB

Category score	Observed count (%)	Average measure	Infit MNSQ[a]	Outfit MNSQ	Threshold calibration
0	198(21)	−2.36	1.02	1.01	None
1	239(26)	−1.15	.95	.98	−2.09
2	312(34)	.94	.75	.78	−.41
3	181(19)	2.71	1.26	1.19	2.50

[a]mean square residual.

Table 5. Item and person separation index and reliability of the KFAB

		Standard error	Separation index	Separation reliability
KFAB[a]	Item	.14	5.37	.97
	Person	.66	2.95	.90

[a]Korean version of the Fullerton advanced balance scale.

Discussion

The process of measuring functional ability is important to predict the prognosis and confirm the results of interventions clinical setting. Therefore, it is imperative to verify the psychometric properties of the assessment scales developed to evaluate functional status (Darr et al, 2015; Verbecque et al, 2015). The aim of the present study was to investigate the item fit, item difficulty, and separation index of the KFAB scale using Rasch analysis in patients with stroke.

Rasch analysis has recently been used to evaluate the construction and validation of functional assessment questionnaires in the various medical field (Duncan et al, 2003). Rasch analysis compares the response patterns of individuals to the entire sample to estimate person ability and item difficulty. It is a probability model that converts the ordinal scores obtained by summing item scores into interval measures. In traditional analysis, the ordinal scores are typically used as if they were interval scores in nature, but the measures produced by Rasch analysis are on an equal interval scale that is common to both persons and items. Therefore, it provides psychometric information that is not provided in traditional analyses (Wright and Linacre, 1989).

We used the Rasch analysis to investigate item fit, item difficulty, rating scale, and person and item separation index and reliability. The item fit residual statistic indicates how well a set of empirical data meets the requirements of the model. The two fit statistics, infit and outfit, are used to verify the dimensionality of the scale. Infit represents the information-weighted MNSQ difference between the observed and expected responses, and the outfit is the unweighted MNSQ and is more sensitive to outlier (Gothwal et al. 2009; White, and Velozo 2002). Acceptable MNSQ values range between .6 and 1.4 when used with a Likert scale (Wright and Linacre, 1989). There are no items of msifit in the KFAB in this study. Although, all items of the KFAB scale fit the Rasch model, item 10 (reactive postural control) was nearest to misfit criteria (infit 1.39, MNSQ, 2.3). It might indicate that item 10 was inconsistent responses, poorly defined, or measured different ability. The result was similar to previous research indicating that item 10 was found to measure a balance-control mechanism different from that measured by the other 9 items of the FAB scale (Klein et al, 2011).

Item difficulty is expressed by logits in Rasch analysis, with a higher value representing increasing item difficulty. It serves as evidence of construct validity and helps in understanding the progress of functional recovery (Duncan et al. 2003). Our results showed that the easiest item was item 1 (standing with feet together and eyes closed), and the most difficult item was item 8 (two-footed jump). This result was also similar to Klein et al (2011) of study. The person-item map demonstrated an even spread of items across the entire range of person ability. It indicates that there was appropriate targeting of item difficulty to patient ability. Most of the items lined up with persons, representing that these items were targeting the patients, and patients could be differentiated well by the items (Gothwal et al, 2009). As Figure 1 shows, person abilities ranged from -4 to 4 on our logit metric. A close inspection of Figure 1 demonstrates that the mapped logits for the item difficulties showed a somewhat inappropriate for persons of lower and higher abilities but appeared to map across the person abilities with a reasonable spread.

Investigating rating scale diagnostics is useful, as it demonstrates the extent to which respondents utilized each rating scale option. Total observed counts and percentages are provided to illustrate these points, and infit and outfit mean square statistics provide information about how well each category fits the rating scale (Kornetti et al, 2004). Most respondents utilized all ratings in this study. However, threshold calibration showed insufficient distinction between categories 2 (-.12) and 3 (.52). Where response categories are disordered, we recommended combining or collapsing adjacent response categories. Collapsing categories can often improve the performance of the overall scale, as well as reduce burden on the respondent and save time (Comins et al, 2008). To improve the rating scale, we converted it from a 5-point scale to a 4-point scale by combining categories 2 and 3. After converting, 4-point rating scale results showed that respondents utilized all rating, and the threshold calibration demonstrated distinct difference between two category (Table 4). Also, the probability curve graph illustrated that each category had a distinct peak in the 4-point (Figure 2).

Rasch analysis provides an index that indicates the number of distinct strata of persons within each domain to represent the reliability of the assessment tools. In a separation index, the numerator is the variance in the person measures for the group and the denominator is the average error in estimating these measures. Higher indices indicated that more distinct levels of functioning can be distinguished in the measure (Davidson, 2009). A person separation index of 1.50 indicates an acceptable level of separation, an index of 2.00 indicates a good level of separation, and index of 3.00 indicates an excellent level of separation (Duncan et al, 2008). The number of distinct strata that can be distinguished in the sample is calculated by the [4(separation index)+1)/3] formula (Mallinson et al, 2004). Person separation index was 2.95 in this study, it means that can discern 4 strata.

A limitation of the study is that it did not considered other psychometric properties that may affect balance function in patients with stroke, such as, age, cognition, and fear of falling. Further research is needed to investigate and incorporate these factors.

Conclusion

The present study investigated measurement properties such as item fit, item difficulty, rating scale, and separation index of the KFAB scale by Rasch analysis in patients with stroke. All items of the KFAB scale were suitable, and rating scale was recommended to 4-point scale for assessing the balance ability of stroke patients. The KFAB scale was found to have high person and item separation reliability, which indicated that the scale can distinguish among participants of varying balance abilities. Furthermore, the easiest item was 'standing with feet together and eyes closed', and most difficult item was 'two-footed jump'. Based on the above results, we will expect that more accurate measurement will be completed, when evaluating the balance ability by considering the difficulty of items.

References

Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: A systematic review. Phys Ther. 2008;88(5):559-566. https://doi.org/10.2522/ptj.20070205

Bonan IV, Colle FM, Guichard JP, et al. Reliance on visual information after stroke. Part I: Balance on dynamic posturography. Arch Phys Med Rehabil. 2004;85(2):268-273.

Comins J, Brodersen J, Krogsgaard M, et al. Rasch analysis of the knee injury and osteoarthritis outcome score (KOOS): A statistical re-evaluation. Scand J Med Sci Sports. 2008;18(3):336-345.

Darr N, Franjoine MR, Campbell SK, et al. Psychometric properties of the pediatric balance scale using Rasch analysis. Pediatr Phys Ther. 2015;27(4):337-348. https://doi.org/10.1097/PEP.0000000000000178

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. https://doi.org/10.1007/s11136-009-9456-4

Duncan PW, Bode RK, Min Lai S, et al. Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale. Arch Phys Med Rehabil. 2003;84(7):950-963.

Geurts AC, de Haart M, van Nes IJ, et al. A review of standing balance recovery from stroke. Gait Posture. 2005;22(3):267-281.

Gothwal VK, Wright TA, Lamoureux EL, et al. Rasch analysis of the quality of life and vision function questionnaire. Optom Vis Sci. 2009;86(7):E836-844. https://doi.org/10.1097/OPX.0b013e3181ae1ec7

Hellström K, Lindmark B. Fear of falling in patients with stroke: A reliability study. Clin Rehabil. 1999;13(6):509-517.

Hernandez D, Rose DJ. Predicting which older adults will or will not fall using the Fullerton Advanced Balance scale. Arch Phys Med Rehabil. 2008;89(12):2309-2315. https://doi.org/10.1016/j.apmr.2008.05.020

Jeon YJ, Kim GM. Comparison of the psychometirc properties of two balance scales in children with cerebral palsy. J Phys Ther Sci. 2016;28(12):3432-3434. https://doi.org/10.1589/jpts.28.3432

Kim GM. Comparison of the Pediatric Balance Scale and Fullerton Advanced Balance Scale for predicting falls in children with cerebral palsy. Phys Ther Korea. 2016a;223(4):63-70.

Kim GM. Reliability and validity on the Korean version of the Fullerton Advanced Balance scale. Phy Ther Korea. 2016b;23(1):31-37.

Klein PJ, Fiedler RC, Rose DJ. Rasch analysis of the Fullerton Advanced Balance (FAB) Scale. Physiother Can. 2011;63(1):115-125. https://doi.org/10.3138/ptc.2009-51

Kornetti DL, Fritz SL, Chiu YP, et al. Rating scale analysis of the Berg Balance Scale. Arch Phys Med Rehabil. 2004;85(7):1128-1135.

Langhorne P, Stott DJ, Robertson L, et al. Medical complications after stroke: A multicenter study. Stroke. 2000;31(6):1223-1229.

Li F, Harmer P, Fitzgerald K, et al. Tai chi and postural stability in patients with Parkinson's disease. N Engl J Med. 2012;366(6):511-519. https://doi.org/10.1056/NEJMoa1107911

Mallinson T, Stelmack J, Velozo C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. Med Care. 2004;42(1 Suppl):17-24.

Rose DJ, Lucchese N, Wiersma LD. Development of a multidimensional balance scale for use with functionally independent older adults. Arch Phys Med Rehabil. 2006;87(11):1478-1485.

Schlenstedt C, Brombacher S, Hartwigsen G, et al. Comparing the Fullerton Advanced Balance Scale with the Mini-BESTest and Berg Balance Scale to assess postural control in patients with Parkinson disease. Arch Phys Med Rehabil. 2015;96(2):218-225. https://doi.org/10.1016/j.apmr.2014.09.002

Schlenstedt C, Brombacher S, Hartwigsen G, et al. Comparison of the Fullerton Advanced Balance Scale, Mini-BESTest, and Berg Balance Scale to predict falls in Parkinson disease. Phys Ther. 2016;96(4):494-501. https://doi.org/10.2522/ptj.20150249

Suzuki M, Fujisawa H, Machida Y, et al. Relationship between the Berg Balance Scale and Static Balance Test in Hemiplegic Patients with Stroke. J Phys Ther Sci. 2013;25(8):1043-1049. https://doi.org/10.1589/jpts.25.1043

van de Port IG, Kwakkel G, Schepers VP, et al. Predicting mobility outcome one year after stroke: A prospective cohort study. J Rehabil Med. 2006;38(4):218-223.

Velozo CA, Peterson EW. Using Rasch analysis to develop meaningful fear of falling measures for community dwelling elderly. Am J Phys Med Rahabil. 2001;80(9):662-673.

Verbecque E, Lobo Da Costa PH, Vereeck L, et al. Psychometric properties of functional balance tests in children: A literature review. Dev Med Child Neurol. 2015;57(6):521-529. https://doi.org/10.1111/dmcn.12657

Weerdesteyn V, de Niet M, van Duijnhoven HJ, et al. Falls in individuals with stroke. J Rehabil Res Dev. 2008;45(8):1195-1213.

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

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

This article was received October 10, 2017, was reviewed October 10, 2017, and was accepted November 13, 2017.