ARTICLES

Rasch Analysis of the Fullerton Advanced Balance (FAB) Scale

Penelope J. Klein, Roger C. Fiedler, Debra J. Rose

ABSTRACT

Purpose: This cross-sectional study explores the psychometric properties and dimensionality of the Fullerton Advanced Balance (FAB) Scale, a multi-item balance test for higher-functioning older adults.

Methods: Participants ($n = 480$) were community-dwelling adults able to ambulate independently. Data gathering consisted of survey and balance performance assessment. Psychometric properties were assessed using Rasch analysis.

Results: Mean age of participants was 76.4 (SD = 7.1) years. Mean FAB Scale scores were 24.7/40 (SD = 7.5). Analyses for scale dimensionality showed that 9 of the 10 items fit a unidimensional measure of balance. Item 10 (Reactive Postural Control) did not fit the model. The reliability of the scale to separate persons was 0.81 out of 1.00; the reliability of the scale to separate items in terms of their difficulty was 0.99 out of 1.00. Cronbach's alpha for a 10-item model was 0.805. Items of differing difficulties formed a useful ordinal hierarchy for scaling patterns of expected balance ability scoring for a normative population.

Conclusion: The FAB Scale appears to be a reliable and valid tool to assess balance function in higher-functioning older adults. The test was found to discriminate among participants of varying balance abilities. Further exploration of concurrent validity of Rasch-generated expected item scoring patterns should be undertaken to determine the test's diagnostic and prescriptive utility.

Key Words: aged, balance, fall risk assessment tool, falls, psychometrics, FAB Scale

Klein PJ, Fiedler RC, Rose DJ. Rasch analysis of the Fullerton Advanced Balance (FAB) Scale. Physiother Can. 2010;preprint. doi:10.3138/ptc.2009-51

RÉSUMÉ

Objectif : Cette étude transversale se penche sur les propriétés psychométriques et sur la dimensionnalité de l'échelle d'équilibre FAB (Fullerton Advanced Balance), un test multiple visant à évaluer l'équilibre chez les aînés hautement fonctionnels.

Méthode : Les participants ($n = 480$) étaient des personnes âgées vivant dans la communauté et pouvant se déplacer seules. Les données réunies comprenaient un sondage et une évaluation de l'équilibre. Les propriétés psychométriques ont été évaluées à l'aide d'une analyse de Rasch.

Résultats : L'âge moyen des participants était de 76,4 ans (écart-type = 7.1). Les notes obtenues à l'échelle FAB ont été de 24,7/40 (écart-type = 7,5). Les analyses de dimensionnalité de l'équilibre ont démontré que 9 des 10 points équivalaient à une mesure de l'équilibre unidimensionnelle. Le point 10 (contrôle réactif postural) ne s'accordait pas au modèle. La fiabilité de l'échelle pour séparer les personnes a été de 0,81 sur 1,00; sa fiabilité pour séparer les points en fonction de leur niveau de difficulté était de 0,99 sur 1,00. Le coefficient alpha de Cronbach pour le modèle à 10 points était de 0,805. Les points de difficulté différents formaient une hiérarchie ordinale utile pour les pointages obtenus par une population normative dans des modèles d'échelle de capacité d'équilibre.

Conclusion : L'échelle FAB semble être un outil fiable et valide pour évaluer la fonction d'équilibre chez les aînés hautement fonctionnels. Le test a permis de séparer les participants ayant diverses capacités en termes d'équilibre. Une exploration plus approfondie de la validité concurrente des modèles de pointages attendus pour chaque point à l'aide de Rash devrait être réalisée afin d'établir l'utilité diagnostique et normative de ce test.

Mots clés : aînés, chutes, équilibre, outil d'évaluation du risque de chute, psychométrie

The authors have no conflicts of interest to declare.

Findings of this research were previously reported as a poster presentation at the American Physical Therapy Association, Combined Sections Meeting, Las Vegas, Nevada, 9–12 February 2009.

The authors would like to express their appreciation to those who assisted with the study as site managers, raters, or study participants.

Penelope J. Klein, PT, EdD: Faculty researcher and professor, Department of Physical Therapy, D'Youville College, Buffalo, New York.

Roger C. Fiedler, PhD: Professor of Doctoral Studies, D'Youville College, Buffalo, New York.

Debra J. Rose, PhD: Professor, Division of Kinesiology and Health Science, and Director, Center for Successful Aging, California State University, Fullerton, California.

Address correspondence to *Penelope J. Klein*, D'Youville College, Physical Therapy Department, 320 Porter Ave., Buffalo, NY 14201 USA. Tel. 716-829-7835; E-mail: kleinpj@dyc.edu.

DOI:10.3138/ptc.2009-51

INTRODUCTION

Falls are the leading non-medical cause of serious injury in older adults.[1] Delaying the onset of falls or reducing the severity of a fall injury are goals of both the Canadian and the US national health care agendas.[2,3] Advancing early fall-prevention initiatives that target active older adults prior to fall injury is one possible proactive approach to this known health threat. Distinguishing among varying levels of balance ability and identifying patterns of impairment as part of early intervention require valid screening tools. There are several existing balance assessment tools for use with clinical populations; however, a prospective study of five clinical balance tests, including the Berg Balance Scale (BBS),[4] timed up-and-go (TUG),[5] and Dynamic Gait Index (DGI),[6] concluded that factors contributing to falls may interact differently at different ages and activity levels; current tests are not as successful in predicting fall risk in active older adults as they have been found to be in more frail populations.[7] The authors of this comparative study suggested a need for development and testing of tools that target an increasingly active aging population. The current study responds to this identified need, specifically by adding to existing knowledge of the psychometric properties of a relatively new balance assessment tool that targets higher-functioning older adults.

The Fullerton Advanced Balance (FAB) Scale[8] is a relatively new multi-item balance-assessment test designed specifically to measure balance in higher-functioning active older adults. Content validity is based on theoretical analysis of components of static balance and dynamic balance control, sensory reception and integration, and anticipatory and reactive postural control. The test is composed of 10 items (see Appendix). Previously published research[9,10] has assessed the psychometric properties of the FAB Scale—content and convergent validity, test–retest and intra- and interrater reliability, and internal rater consistency—as well as the test's predictive accuracy. Test–retest reliability has been previously reported as $r = 0.96$, and interrater reliability in the range of $r = 0.91$–0.95, when the test is administered by trained raters.[9] In assessing predictive accuracy, Hernandez and Rose used receiver operating characteristic (ROC) analysis to determine that a cutoff score of 25/40 on the 10-item FAB scale produced the highest sensitivity (74.6%) and specificity (52.6%) in predicting faller status determined by retrospective survey.[10] They further concluded that a practitioner can be confident in more than 7 out of 10 cases that an older adult who scores 25 or lower on the FAB Scale is at high risk for falls and in need of immediate intervention.

Because the FAB Scale is a multi-item test, it is also important to assess the degree to which its individual test items work together to produce a reasonable measure of balance for this higher-functioning population. Knowledge gained from this assessment should provide insights for future use. The primary purpose of the current study was to examine selected measurement properties of the FAB Scale by applying Rasch analysis.[11–14] Specific aims of the research were (a) to define the relationship between patterns of item performance and persons' abilities, (b) to examine dimensionality, (c) to identify item order hierarchy, and (d) to generate a pattern of expected scores for possible use in future research and clinical applications.

METHODS

Design

This cross-sectional study analyzed data from a combined database ($n = 480$). Data were gathered from January 2005 through April 2007.

Participants

Study participants were volunteer community-dwelling adults aged 60 years and older who were able to ambulate independently without an ambulatory aid during test administration.

A standardized research protocol, approved by each respective university's Institutional Review Board, was followed. Participants in both locales were given an opportunity to ask questions about any aspect of the study, after which they signed an informed consent form.

Setting

Data were gathered in upstate New York (NY) and in Orange County, California (CA). Testing sites included community centres for seniors and research-centre facilities.

Instruments

Each participant completed a Health Assessment Questionnaire (HAQ)[15] and a battery of physical balance challenges representing the 10 items of the FAB Scale. Participants tested at the CA locations completed additional physical assessments not part of the current study.

The FAB scale consists of 10 items: "balancing with feet together and eyes closed," "forward reach," "turn 360 degrees," "stepping up, onto and over a 6-inch bench," "tandem walk," "stand on one leg," "standing on foam with eyes closed," "two-footed jump," "walking with head turns," and "postural reaction."[9] The total score range is 0 to 40; higher scores indicate better balance abilities. Test items are scored using a discrete five-point ranked scale. Standardized scoring descriptors are specific to each item scored; item scores range from 0 (unable to perform or not attempted) to 4 (maximum test item score). The FAB scale is relatively easy to administer by trained raters and takes about 10 minutes to complete.[a]

Raters

Raters received equivalent training in test-administration procedures and interpretation of performance for scoring at each locale. A total of six raters, all certified balance and mobility instructors, participated in data gathering at the CA sites; each CA rater completed online coursework related to the correct administration and scoring of the FAB Scale and also completed a practical competency exam to verify his or her accuracy as a test administrator. Two groups of graduate physical therapy students (totalling 13 raters), supervised by a faculty researcher, participated in data gathering at the NY sites; NY raters were trained in two groups (5 raters and 8 raters), and each completed a minimum of 6 hours of training and passed a competency test. Fixed group interrater reliability ranges were assessed for the two groups of NY raters (ICC = 0.955–0.999.)

Procedures

CA participants were recruited both from the community at large and from a population of individuals entering a community-based balance-training program. NY participants were recruited from six community seniors' centres. Volunteers were screened for study eligibility through review of self-report of functional mobility and health status. Once participants entered the study, the FAB Scale was administered.

Statistical Analysis

To provide evidence of homogeneity to support combining of samples, comparability of the two data sets was analyzed using chi-square tests for categorical data and independent-samples *t*-tests for interval-level data. Rating-scale analysis was conducted using the Winsteps computer program (MESA Press, Chicago)

The Rasch measurement model offers several advantages over traditional psychometric approaches and is well suited to meeting the objectives of the current study. Extensive discussion of the theory underlying Rasch analysis is beyond the scope of this paper; however, a brief synopsis of the essential features of Rasch analysis specific to the current study is provided below.

Examination of the measurement properties of test items (how they fit together in forming a balance dimension) and conversion of the raw rating-scale scores to a unidimensional measure of balance are essential elements of the Rasch model advanced by Wright et al.[12,13] Converting a set of raw scores to a Rasch measure requires placing both of them onto the same metric; persons must be expressed in the form of *abilities*, and items in the form of *difficulties*. This application allows the researcher to determine whether items have a range of difficulties sufficient to separate persons in terms of abilities. A basic concept in the Rasch model is that the probability of a person's level on an item is a function of the person's ability and of the difficulty of the item. Unfortunately, raw person or item scores are not typically linear in nature—a rating of 2 on any single item does not represent twice the performance of a person rated 1. The Rasch model converts the raw scores to Rasch measures using a constant metric, called a *logit* or log odds unit. When persons and items can be placed along the same logit metric, comparisons can be made directly, and the results may yield new insights and interpretations. For example, if the range of item difficulties does not match the range of person abilities, the measure is too easy or too hard. The Rasch model suggests that when persons have abilities above the range of the measure's item difficulties, their abilities cannot be determined. Likewise, items with difficulties beyond the abilities of any of the persons tested are not useful, since they do not add any measurement value. Thus, the Rasch model examines how well items and persons work together to form useful measures.

In order to demonstrate this concept, we examined the item difficulties and person abilities together on the Rasch Item Map, with persons mapped with items in the common metric of log odds units (logits). To facilitate interpretation of the mapped logit values, we converted the results to a standardized measure with a mean of 50 and a standard deviation of 10.

Because an important feature of the model is to determine the item difficulties, we examined the items to see how well they fit together (i.e., cooperated) to form a single dimension.[15] Item fit statistics (mean squares and *z* statistics) were used to examine the dimensionality of the FAB Scale. The item-fit statistics are expressed in many ways, two of the more common being "infit" and "outfit" statistics. *Infit* is a standardized measure of unexpected responses to items near the person's measure level, while *outfit* is a standardized, outlier-sensitive measure of unexpected responses to items far from the person's measure level. For example, persons with very low balance ability should have great difficulty on the more difficult of the FAB items; should these persons produce surprising results, such as high ratings on some difficult items, the outfit measure will reflect the erratic behaviour of these items by producing high *z*-score outfit values.

The Rasch model also determines the reliability of the scale using these same concepts. To the extent that items are well separated—very much like inches on a ruler—the measure will be able to detect differences in persons using these markers. Likewise, persons must be well separated along the logit metric in order for the measure to detect individual differences in their abilities. Thus, the results are presented here in the form of person and item separation values, in logit form, and the reliabilities associated with each. For those unfamiliar with these concepts, comparisons are made with more widely recognized reliability values in the form of Cronbach's coefficient alpha measure of internal consistency, with reasons why the Rasch model's results differ slightly from a Cronbach's coefficient alpha analysis.

Item difficulties were also used to determine the item hierarchy and expected score patterns of the FAB Scale. When items are measured along a single dimension, their relative difficulties can be used to form a hierarchy along the dimension, and these difficulties can be used to develop expected scoring patterns that allow identification of persons whose test responses deviate from the population norm. This discriminating psychometric property has applications for diagnosis and for prescription of interventions.

RESULTS

The database ($n = 480$) consisted of 340 (71%) women and 140 (29%) men. Mean age was 76.4 years (SD = 7.1). Twenty percent of the sample reported two or more falls in the past year. The mean FAB Scale score for the entire sample was 24.7/40 (SD = 7.5). Comparability of samples from the two locales is shown in Table 1. We found no significant differences between the two samples with respect to age or gender, and no differences in self-reported health assessment. Differences were observed between the two samples in the number of falls reported in the past year and in mean FAB Scale scores: fewer falls and higher FAB scores were found in the NY sample.

The initial analyses from the Winsteps computer program showed that the 10 items of the FAB scale mapped well to the persons in the combined sample. As Figure 1 shows, person abilities ranged from 29.35 to 78.93 on our logit metric. A close inspection of Figure 1 reveals that the mapped logits for the item difficulties showed a somewhat greater sensitivity for persons of lower abilities but appeared to map across the person abilities with a reasonable spread. Closer examination of the individual FAB scale items in relation to dimensionality is presented in Table 2, in which item-fit statistics and both person and item reliabilities show that the scale and 9 of its 10 items fit very well into a single dimension of balance. The exception was item 10, which showed high misfit, with infit and outfit standardized z-statistics of 5.7 and 6.2 respectively.

Because the focus of the Rasch model is on whether or not the items in a scale work together to measure persons of differing abilities accurately, or reliably, two of the more important indicators are the reliabilities of person and item separation.[13–15] In this case, the person-separation index showed that the FAB Scale was able to detect differences of 2.10 logits, while the item-separation index of 13.85 logits indicated a good separation of the items along the standardized FAB Measure. The reliabilities for these separations are good: the items of the FAB Measure were able to reliably separate persons at $r = 0.81$ and items at $r = 0.99$. For researchers more familiar with the traditional models of internal consistency, comparisons using Cronbach's alpha are also presented for the same participants and the same 10-item data. The alpha value was 0.805 for the 10-item scale and 0.815 for the revised 9-item scale. The differences between the person reliability estimate of the Rasch model and the KR-20, or Cronbach's alpha for internal consistency, have been explained elsewhere.[14] Rasch model results are often found to be somewhat lower than other measures, because extreme scores result in no presumed measurement error and therefore add to the overall estimate of reliability by Cronbach's model.

Initial analysis ranked items from least to most difficult as follows: 1, 2, 4, 7, 10, 3, 9, 8, 5, 6. Removal of the misfitting Item 10 resulted in a final and reliable measure using a 9-item scale, with no change in hierarchical item order or in person and item reliabilities. As a means of further exploring the hierarchy of these items and their usefulness in determining a measure of person ability, the predicted, or expected, FAB Scale item-rating level at each logit value of the measure was determined. A closer review of the item response probabilities of the rating-scale model for the 9-item scale was also conducted.

Figure 2 shows the item step pattern of expected scores for each item, which can be used to provide a diagnostic tool for detecting unexpected results that may indicate unique problems with balance in individuals. The expected pattern of item difficulties mapped against the Rasch-modelled person ability measures for the FAB Scale observed in the study sample are presented in Figure 2. The graph shows the Rasch-converted person-ability scores across the top, from lowest to highest, as well as the mean and 1- and 2-standard-deviation points on the abscissa. Items are presented on the ordinate in increasing order of difficulty, from lowest (Item 1) to highest (Item 6). At each step along the item-response categories (from left to right on each horizontal item line), the difficulties are shown in the order of what increase in person ability was required to advance from the lower response category to the next level up on the item. One interpretation of these results is that it requires a much greater increase in person ability to move from score 0 to score 1 than to move from 2 to 3, as reflected in the widths of the distances between codes of 0 and 3 for each of the items.

As another way of examining item-response scoring, Figure 3 shows the likelihoods of item category responses from 0 to 4. The figure shows that when item difficulty was highest and person ability lowest (shown as a logit

Table 1 Comparison of Demographics between New York State (NY) and California (CA) Participant Sub-groups ($n = 480$)

Category/Variable	*NY* n = *188*	*CA* n = *292*	χ^2 (df)	t-*test* (df)	p
Mean age (SD)	75.864 (6.793)	76.777 (7.269)	–	1.377 (478)	ns
Gender					
Female	128	211	1.804 (1)		ns
Male	60	80			
Perceived Health					
Excellent / Very Good / Good	163	245	2.021 (1)	–	ns
Fair/Poor	21	47			
[else]	4	0			
Daily medications (*n*)					
0	13	30	1.43 (2)	–	ns
1–4	102	158			
>4	65	98			
[else]	8	6			
Engage in regular physical activity					
Yes	136	207	0.3614 (1)	–	ns
No	48	83			
[else]	4	2			
Walk with assistive device					
Yes	14	69	18.51 (1)	–	<0.001
No	166	221			
[else]	8	2			
Falls in past year (*n*)					
0	134	123	48.3 (2)	–	<0.0001
1	23	88			
≥2	20	77			
[else]	11	4			
Mean FAB Scale (SD) score max = 40	27.995 (6.591)	22.586 (7.394)	–	8.151 (478)	<0.0001

[else] = missing or invalid data; FAB Scale = Fullerton Advanced Balance Scale

difference of −30; see Figure 3, far left), the most probable category was 0; when person ability far exceeded item difficulty (logit difference near 30), the most probable category was 4. This finding is consistent with the intent of FAB scoring, which is to show that when persons are expected to do poorly (low ability vs. high item difficulty), scores of 0 are likely, and when persons are expected to do well, scores of 4 are likely. However, the graph also shows that while categories 1 and 2 both emerge as most probable responses as one moves from left to right, in the direction of greater person ability and lower item difficulty, category 3 never emerges as a most probable response at any point on the graph. Examination of response frequencies showed that category 3 added little to participant discrimination, occurring less than 10% of the time (range within items: 4%–14%).

With the information about item difficulties and person abilities shown in Figure 2, an expected FAB score pattern can be mapped against the item difficulty scoring patterns. By generating raw scores from 0 to 36, we can determine response patterns on each of the 9 unidimensional items on the FAB from the expected response patterns in Figure 2. Table 3 shows how this figure can be used to generate expected patterns of responses on each item of the FAB and how unexpected results—that is, individuals who do not show the expected patterns of scores on the items—may be diagnostic in nature. For example, a person with a raw FAB score of 20 would be expected to show the scoring pattern 4, 4, 2, 3, 1, 1, 2, 1, 2 across the 9 items shown in the tables. Should an individual with a score of 20 show a pattern of, say, all 2s, then additional factors may be operating in this individual that should be looked at more closely to determine why his or her scores on items 1 and 2 were lower than expected, why the scores on items 5, 6, and 8 were higher than expected, and why the pattern of scores is so different from the expected pattern.

DISCUSSION

Balance is known to be a complex construct; the control of posture and balance requires integration of neural

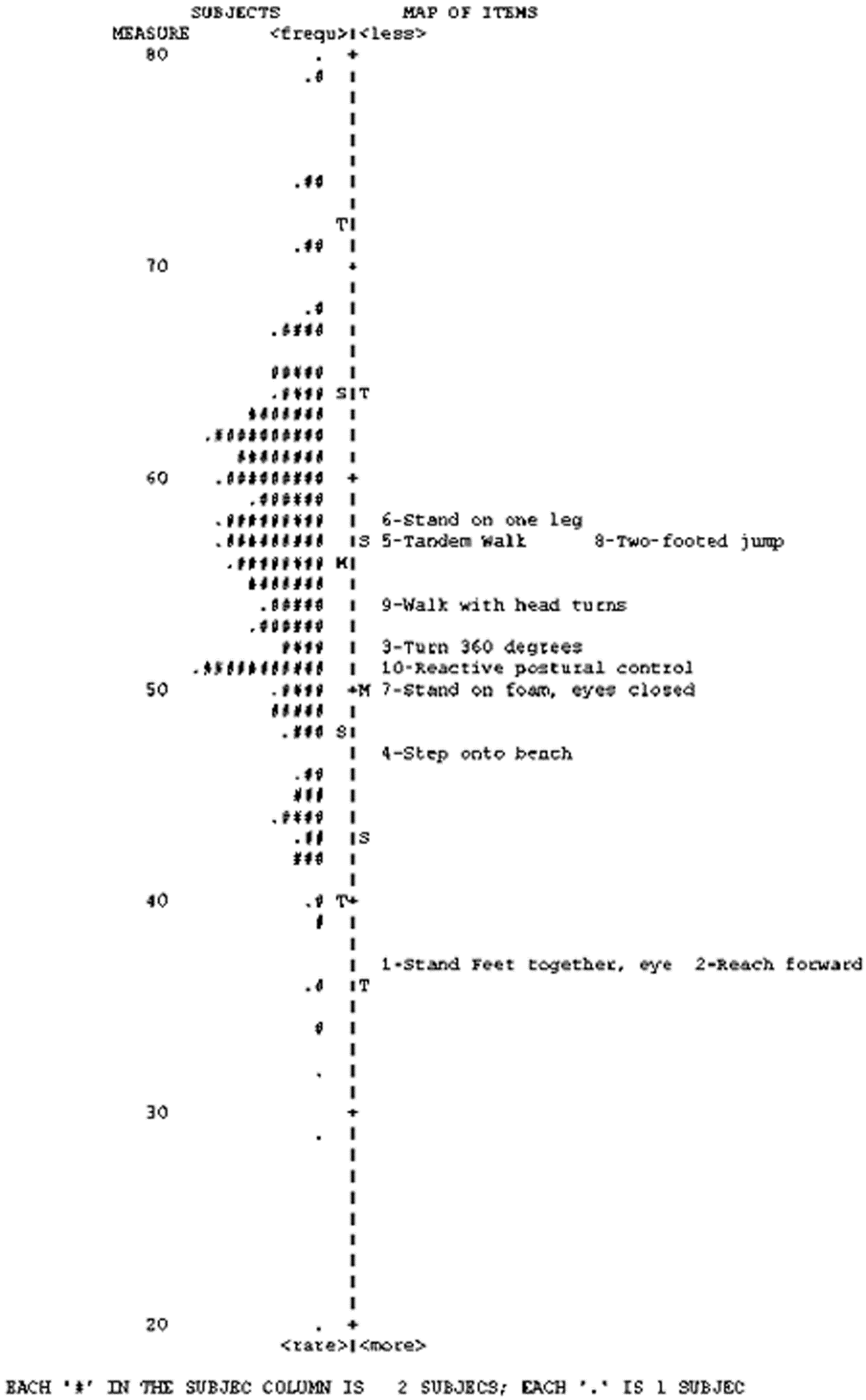

Figure 1 Pattern map of person abilities and item difficulties using the 10-item FAB Scale ($n = 478$)

Table 2 Rasch Item-Fit Statistics and Person and Item Reliabilities for the 10-item FAB Scale in Rasch Measure* Order ($n = 480$)

Item Number and Label	Measure[a]	Error[b]	Infit[c] MNSQ	Infit[c] ZSTD	Outfit[d] MNSQ	Outfit[d] ZSTD	Score Corr.[e]
6. Stand on one leg	58.43	0.44	0.64	−6.7	0.63	−6.5	0.62
5. Tandem walk	56.83	0.43	0.69	−5.9	0.67	−5.8	0.71
8. Two-footed jump	56.69	0.43	1.02	0.4	0.97	−0.5	0.68
9. Walk with head turns	53.70	0.42	1.12	2.0	1.11	1.7	0.56
3. Turn 360 degrees	51.67	0.42	0.82	−3.3	1.10	1.5	0.40
10. Reactive postural control	51.08	0.43	1.37	5.7	1.49	6.2	0.38
7. Stand on foam, eyes closed	50.05	0.43	1.11	1.8	1.04	0.6	0.73
4. Step onto bench	46.96	0.46	1.10	1.5	0.93	−0.9	0.72
2. Reach forward, retrieve object	37.43	0.65	1.19	1.9	1.19	1.2	0.49
1. Stand feet together, eyes closed	37.17	0.65	1.28	2.6	0.79	−1.5	0.55

MNSQ = mean square; ZSTD = standardized *z*-score; FAB Scale = Fullerton Advanced Balance Scale

* The FAB Rasch Measure represents the result of the conversion of FAB raw scores to an interval-level FAB Measure, expressed in standardized logit values, with a mean of 50 and a standard deviation of 10.

a The measure of the item difficulty on the standardized FAB Measure

b Measurement error associated with the item

c Infit is expressed as a mean square and converted to a *z*-score to represent how well the item measures responses at or near the persons' abilities; *z*-scores above 3 standard deviations reflect inaccuracy.

d Outfit is expressed as a mean square and converted to a *z*-score to represent how well the item measures responses at the outer ranges of the persons' abilities; *z*-scores above 3 standard deviations reflect high performances by persons with low ability on high-difficulty items, or vice versa.

e Correlations between the item and the FAB Measure: Person Separation Index = 2.10 logits; Person Reliability = 0.81; Item Separation Index = 13.85 logits; Item Reliability = 0.99.

(sensory and motor) and musculoskeletal systems. From the beginning of FAB Scale test development, it was theorized that individual test items and possible sub-groupings of items within the FAB Scale might involve different balance-control systems to varying degrees.[10] Results of Rasch modelling for the FAB Scale found that 9 of the 10 test items were related within a single domain of balance, which suggests that scores for items 1–9 may be summed into a total score for a meaningful measure of balance ability.

Item 10 (reactive postural control) was found to measure a balance-control mechanism different from that measured by the other nine FAB scale items. Item 10 is intended to measure an individual's ability to respond quickly to an unexpected loss of balance using a protective and involuntarily controlled righting response. To elicit this response, the test administrator places his or her hand in the middle of the participant's back and instructs the participant to lean against the hand in a backward direction. Once the participant reaches a backward lean angle that the test administrator considers to exceed his or her backward limits of stability, the test administrator, without warning, quickly releases the hand, causing the participant to lose his or her balance in a backward direction. The score is based on the number of backward steps required for the participant to regain balance and on whether manual support is required. The assumption that item 10 (reactive postural control) measures something different than the other test items has a theoretical basis: while the act of stepping back to regain balance requires muscle power and proprioception, this test item is primarily viewed as a measure of involuntary postural control that challenges reactive speed and accuracy. Rasch analysis revealed that Item 10 did not fit within a unidimensional model, which suggests that Item 10 should be scored separately from the remaining nine items.

The expected pattern of item difficulties mapped against the Rasch-modelled person-ability measures presented in Figure 2 were used to generate a scoring key of expected normative scores (shown in Table 3). Because Rasch analysis indicated that test items 1–9 were domain related, only scores from these nine items can be included in an expected scoring key (total score range: 0–36). Further exploration of the diagnostic value of the FAB Scale should be addressed in future research by referencing individual person patterns that do not match Rasch-generated expected patterns. Future research should also explore diagnostic and prescriptive test applications, as well as the test's sensitivity to change over time.

LIMITATIONS

The database was purposively limited to community-dwelling older adults aged 60 years and up. The sample was not randomly selected from a representative population; rather, a volunteer sample was recruited, which resulted in non-representative distributions for gender, race, and ethnicity. The diagnostic utility of the measure for diagnostic and prescriptive applications, as well as

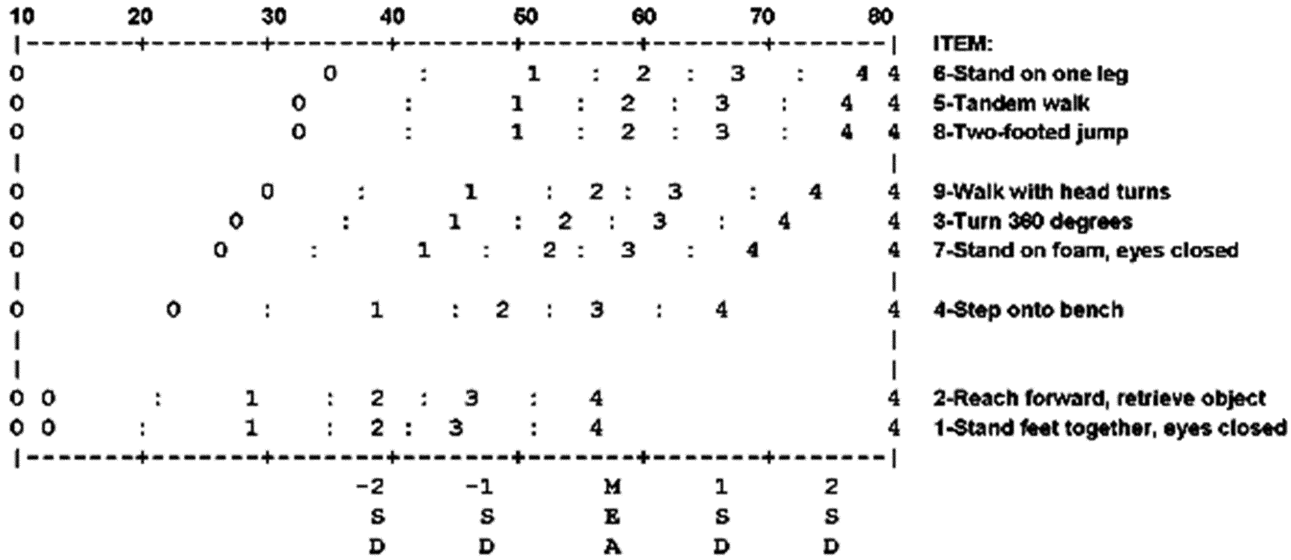

Figure 2 Plot of normative expected score patterns using the FAB Measure (10–80) on the abscissa and the items stacked in difficulty order at the right from lowest (bottom) to highest (top)

Note: MEA and the SDs in the figure represent the mean and standard deviations of the FAB Measure.

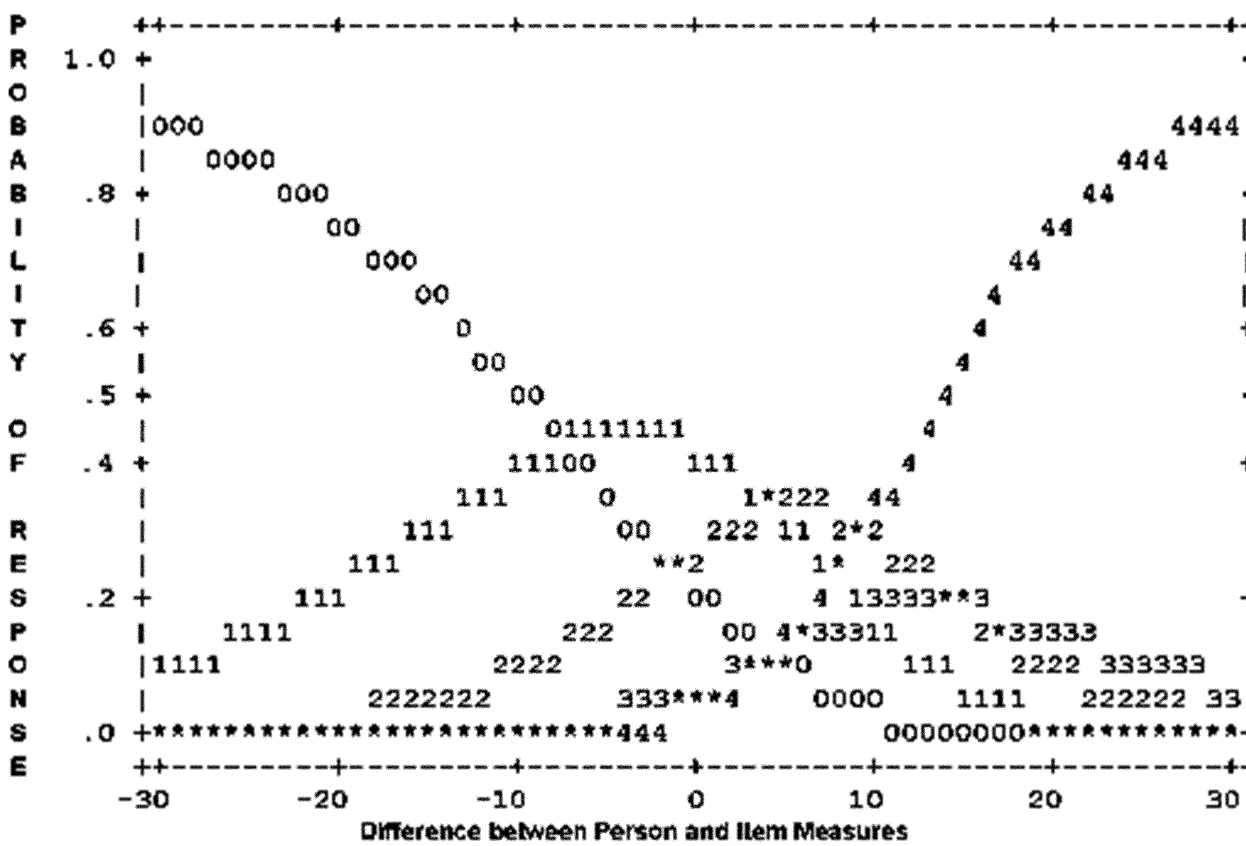

Figure 3 Probabilities of item-response categories plotted against differences between persons and items in logits

Note: When differences between person ability and item difficulty were low (abscissa values near −30), that is, when item difficulty was highest and person ability lowest, the most probable category was 0; when the differences were high (abscissa values near 30), or person ability far exceeded the item difficulty, the most probable category was 4.

Table 3 Expected Scoring Patterns for Individual Items and Total Scores in Item Difficulty Order ($n = 480$)

Raw Score	*Item 1*	*Item 2*	*Item 4*	*Item 7*	*Item 3*	*Item 9*	*Item 8*	*Item 5*	*Item 6*	*Total*
0	0	0	0	0	0	0	0	0	0	N/A*
1	1	0	0	0	0	0	0	0	0	1
2	1	1	0	0	0	0	0	0	0	2
3	2	1	0	0	0	0	0	0	0	3
4	2	2	0	0	0	0	0	0	0	4
5	2	2	1	0	0	0	0	0	0	5
6	2	2	1	1	0	0	0	0	0	6
7	3	2	1	1	0	0	0	0	0	7
8	3	2	1	1	1	0	0	0	0	8
9	3	3	1	1	1	0	0	0	0	9
10	3	3	1	1	1	1	0	0	0	10
11	3	3	2	1	1	1	0	0	0	11
12	3	3	2	1	1	1	1	0	0	12
13	3	3	2	1	1	1	1	1	0	13
14	3	3	2	1	1	1	1	1	1	14
15	3	3	2	2	1	1	1	1	1	15
16	3	3	2	2	2	1	1	1	1	16
17	4	3	2	2	2	1	1	1	1	17
18	4	4	2	2	2	1	1	1	1	18
19	4	4	3	2	2	1	1	1	1	19
20	4	4	3	2	2	2	1	1	1	20
21	4	4	3	3	2	2	1	1	1	21
22	4	4	3	3	2	2	2	1	1	22
23	4	4	3	3	2	2	2	2	1	23
24	4	4	3	3	2	2	2	2	2	24
25	4	4	3	3	3	2	2	2	2	25
26	4	4	3	3	3	3	2	2	2	26
27	4	4	4	3	3	3	2	2	2	27
28	4	4	4	3	3	3	3	2	2	28
29	4	4	4	3	3	3	3	3	2	29
30	4	4	4	3	3	3	3	3	3	30
31	4	4	4	4	3	3	3	3	3	31
32	4	4	4	4	4	3	3	3	3	32
33	4	4	4	4	4	4	3	3	3	33
34	4	4	4	4	4	4	4	3	3	34
35	4	4	4	4	4	4	4	4	3	35
36	4	4	4	4	4	4	4	4	4	N/A*
Totals	120	115	86	75	67	59	52	48	44	

* A Rasch Measure cannot be determined for raw scores that represent all 0's or all 4's.

its level of precision in assessing change over time, were not evaluated. These psychometric properties may be explored in future research.

CONCLUSION

The challenge of identifying valid and reliable assessment tools with practical relevance for assessing balance abilities in higher-functioning community-dwelling older adults is central to the goal of early intervention to reduce falls among active older adults. A relatively new multi-item balance assessment tool, the FAB Scale, may have potential for this intended application. For multi-item tools such as the FAB Scale, knowledge of dimensionality, hierarchy of item difficulty, the relationship between person ability and item difficulty, and expected scoring patterns is essential for clinical interpretation of test results. The current research adds to existing knowledge on the psychometric properties of the FAB Scale. Rasch modelling identified two construct domains. Nine of the 10 test items were found to be construct related, which suggests that these nine items may represent a score total; conversely, Item 10 (reactive postural control) should be interpreted independently of the other nine test items. The FAB Scale was found to have high person and item separation reliability, which suggests that the tool can discriminate among participants of varying balance abilities. A pattern of expected balance-ability scoring for a normative population was also generated that has potential applications in clinical settings. The utility of expected patterns as a reference for making clinical judgments about balance performance may be explored in future research.

KEY MESSAGES

What Is Already Known on This Topic

Preliminary research on psychometric properties of the FAB Scale, addressing test construction, test reliability, and predictive validity, has previously been published.

What This Study Adds

If clinicians are to use test results for any multi-item tool in clinical decision making, knowledge of test dimensionality and hierarchy of items is required. The present study provides this information with respect to the FAB Scale.

NOTE

a. A demonstration of test administration can be viewed at http://www.stopfalls.org/researchers_educators/re_bm.shtml [cited 2009 Sep 20].

REFERENCES

1. Stevens JA, Corso PS, Finkelstein EA, Miller TR. The costs of fatal and nonfatal falls among older adults. Injury Prevention. 2006;12:290–5.
2. Health Canada, Public Health Agency of Canada. Seniors and aging: preventing falls in and around your home [Internet]. It's Your Health. 2006;Jun [cited 2010 Sep 29]. Available from: http://www.hc-sc.gc.ca/hl-vs/alt_formats/pacrb-dgapcr/pdf/iyh-vsv/life-vie/fp-pc-eng.pdf
3. Sleet DA, Moffett DB, Stevens J. CDC's research portfolio in older adult fall prevention: a review of progress, 1985–2005, and future research directions. J Safety Res. 2008;39:259–67. doi:10.1016/j.jsr.2008.05.003
4. Berg KO, Wood-DSL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health. 1992;83(Suppl 2):S7–11.
5. Podsiadlo D, Richardson S. The timed "Up & Go": A test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1989;39:142–8.
6. Chiu YP, Fritz SL, Light KE, Velozo CA. Use of item response analysis to investigate measurement properties and clinical validity of data for the dynamic gait index. Phys Ther. 2006;86(6):778-87.
7. Boulgarides LK, McGinty SM, Willett JA, Barnes CW. Use of clinical and impairment-based tests to predict falls by community-dwelling older adults. Phys Ther. 2003;83:328–39.
8. Rose DJ. Fallproof: a comprehensive balance and mobility program. Champaign, IL: Human Kinetics; 2003.
9. Rose DJ, Lucchese N, Wiersma LD. Development of a multidimensional balance scale for use with functionally independent older adults. Arch Phys Med Rehabil. 2006;87:1478–85. doi:10.1016/j.apmr.2006.07.263
10. Hernandez D, Rose DJ. Predicting which older adults will or will not fall using the Fullerton Advanced Balance (FAB) scale. Arch Phys Med Rehabil. 2008;89:2309–15. doi:10.1016/j.apmr.2008.05.020
11. Wright BD, Stone, MH. Best test design. Chicago: MESA Press; 1979.
12. Wright BD, Masters GN. Rating scale analysis. Chicago: MESA Press; 1982.
13. Linacre JM. True-score reliability or Rasch statistical validity? Rasch Meas. 1996;9:455–6.
14. Tennant A, Pallant, JF. Unidimensionality matters! (a tale of two Smiths?). Rasch Meas Trans. 2006;20:1048–51.
15. Ramey DR. The health assessment questionnaire 1992: status and review. Arthritis Care Res. 1992;5(3):119–29. doi:10.1002/art.1790050303

APPENDIX: INDIVIDUAL TEST ITEMS AND SCORING CATEGORIES ASSOCIATED WITH THE FAB SCALE

Test Item and Verbal Instructions	*Scoring Categories*
1. Standing with feet together and eyes closed "Bring your feet together, fold your arms across your chest, close your eyes when you are ready, and remain as steady as possible until I instruct you to open your eyes."	0 Unable to obtain the correct standing position independently. 1 Able to obtain the correct standing position independently but unable to maintain the position or keep the eyes closed for at least 10 seconds. 2 Able to maintain the correct standing position with eyes closed for more than 10 seconds but less than 30 seconds. 3 Able to maintain the correct standing position with eyes closed for 30 seconds but requires close supervision. 4 Able to maintain the correct standing position safely with eyes closed for 30 seconds.
2. Reaching forward to an object "Try to lean forward to take the pencil from my hand and return to your starting position without moving your feet from their present position." *Equipment: 12" ruler and pencil*	0 Unable to reach the pencil without taking >2 steps. 1 Able to reach the pencil but needs to take 2 steps. 2 Able to reach the pencil but needs to take 1 step. 3 Can reach the pencil without moving the feet but requires supervision. 4 Can reach the pencil safely and independently without moving the feet.
3. Turn in full circle "Turn around in a full circle, pause, and then turn in a second full circle in the opposite direction."	0 Needs manual assistance while turning. 1 Needs close supervision or verbal cueing while turning. 2 Able to turn 360° but takes more than 4 steps in both directions. 3 Able to turn 360° but unable to complete in ≤4 steps in 1 direction. 4 Able to turn 360° safely and takes ≤4 steps in both directions.
4. Step up and over "Step up onto the bench with your right leg, swing your left leg directly up and over the bench, and step off on the other side. Repeat the movement in the opposite direction with your left leg as your leading leg." *Equipment: bench 6" high × 14–18" wide*	0 Unable to step onto the bench without loss of balance or manual assistance. 1 Able to step up onto the bench with leading leg, but trailing leg contacts bench or leg swings around bench during the swing-through phase in both directions. 2 Able to step up onto the bench with leading leg, but trailing leg contacts bench or swings around the bench during the swing-through phase in 1 direction. 3 Able to complete the step up and over in both directions but requires close supervision in 1 or both directions. 4 Able to complete the step up and over in both directions safely and independently.

5. Tandem walk "Walk along the line, placing one foot directly in front of the other such that the heel and toe are in contact on each step forward. I will tell you when to stop." *Equipment: Masking tape (2" wide)*	0 Unable to complete 10 steps independently. 1 Able to complete the 10 steps with >5 interruptions. 2 Able to complete the 10 steps with <5 but >2 interruptions. 3 Able to complete the 10 steps with 2 or fewer interruptions. 4 Able to complete the 10 steps independently and with no interruptions.
6. Stand on one leg "Fold your arms across your chest, lift your preferred leg off the floor, without touching your other leg, and stand with your eyes open as long as you can."	0 Unable to try or needs assistance to prevent falling. 1 Able to lift leg independently but unable to maintain position for >5 seconds. 2 Able to lift leg independently and maintain position for >5 but <12 seconds. 3 Able to lift leg independently and maintain position for >12 but <20 seconds. 4 Able to lift leg independently and maintain position for the full 20 seconds.
7. Stand on foam, eyes closed "Step up onto the foam and stand with your feet shoulder-width apart. Fold your arms over your chest, and close your eyes when you are ready. I will tell you when to open your eyes." *Equipment: Two Airex [Airex AG, Sins, Switzerland] balance pads with 18 × 18" sheet of nonslip material*	0 Unable to step onto foam and/or maintain standing position independently with eyes open. 1 Able to step onto foam independently and maintain standing position but unable or unwilling to close eyes. 2 Able to step onto foam independently and maintain standing position with eyes closed for <10 seconds. 3 Able to step onto foam independently and maintain standing position with eyes closed for >10 seconds but <20 seconds. 4 Able to step onto foam independently and maintain standing position with eyes closed for 20 seconds.
8. Two-footed jump "Try to jump as far but as safely as you can with both feet." *Equipment: Yardstick and masking tape.*	0 Unwilling or unable to attempt or attempts to initiate 2-footed jump but 1 or both feet do not leave the floor. 1 Able to initiate 2-footed jump but one foot leaves the floor or lands before the other. 2 Able to perform 2-footed jump but unable to jump further than the length of their own feet. 3 Able to perform 2-footed jump and achieve a distance greater than the length of their own feet. 4 Able to perform 2-footed jump and achieve a distance greater than twice the length of their own feet.
9. Walk with head turns "Walk forward while turning your head from left to right with each beat of the metronome. I will tell you when to stop." *Equipment: Metronome set to 100 beats per minute*	0 Unable to walk 10 steps independently while performing 30° head turns at an established pace. 1 Able to walk 10 steps independently but unable to perform 30° head turns at an established pace. 2 Able to walk 10 steps but veers from a straight line while performing 30° head turns at an established pace. 3 Able to walk 10 steps in a straight line while performing head turns at an established pace but head turns $<30°$ in one or both directions. 4 Able to walk 10 steps in a straight line while performing 30° head turns at established pacing.
10. Reactive postural control "Slowly lean back into my hand until I ask you to stop."	0 Unable to maintain upright balance; no observable attempt to step; requires manual assistance to restore balance. 1 Unable to maintain upright balance; takes more than 2 steps and requires manual assistance to restore balance. 2 Unable to maintain upright balance; takes more than 2 steps but is able to restore balance independently. 3 Unable to maintain upright balance; takes 2 steps but is able to restore balance independently. 4 Unable to maintain upright balance but able to restore balance independently with only 1 step.

Adapted from Rose (2003).[8] Reprinted with permission.