6-1-2006

Use of Item Response Analysis to Investigate Measurement Properties and Clinical Validity of Data for the Dynamic Gait Index

Yi-Po Chiu
Marshall University, chiuy@marshall.edu

Stacy L. Fritz

Kathye E. Light

Craig A. Velozo

Use of Item Response Analysis to Investigate Measurement Properties and Clinical Validity of Data for the Dynamic Gait Index

Background and Purpose. The Dynamic Gait Index (DGI) is a standardized clinical assessment that aids in evaluating a subject's ability to modify gait in response to changing demands. The purpose of this study was to use Rasch measurement theory to examine whether the DGI rating scale meets suggested psychometric guidelines, whether the hierarchical order of DGI tasks is consistent with a clinically logical testing procedure, and whether the DGI represents a unidimensional construct. **Subjects.** Subjects were 84 community-dwelling male veterans (age range=64–88 years; mean±SD=75±6.47 years). **Methods.** Data were retrieved retrospectively from the participants' clinical records. The Rasch measurement model with the WINSTEPS program was used in this study because it offers distinct advantages over traditional psychometric approaches. **Results.** Overall, the DGI showed sound item psychometric properties. Each of the original 4 rating scale categories appeared to distinctly identify subjects at different ability levels. The analysis revealed a clear item difficulty hierarchical order that is generally consistent with clinical expectations. In addition, fit statistics and principal components analysis indicated that the 8 items of the DGI appear to represent a single construct. **Discussion and Conclusion.** The results suggest that the rating scale of the DGI is used appropriately for community-dwelling older subjects with balance problems. The findings support the continued use of this well-constructed scale for clinical and research assessment in a community-dwelling population of older subjects. [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;86:778–787.]

Key Words: *Balance, Gait, Psychometric measurement*

Yi-Po Chiu, Stacy L Fritz, Kathye E Light, Craig A Velozo

espite preventive interventions, the incidence of falls and fall-related injuries is expected to rise with the aging population. Approximately 33% of people over the age of 65 years fall at least 1 time per year, and approximately 50% over the age of 85 years fall each year.[1,2] Falls continue to be the leading cause of injury and a frequent cause of morbidity in elderly people.[3,4] Falls produce a threat to quality of life and independence secondary to impaired mobility and loss of function.[5] Emphasis on early identification, prevention, and intervention for elderly people identified to be "at risk" for falls is becoming increasingly important in the fields of physical therapy and rehabilitation.[6–12] Standardized clinical assessments are used widely in both research and clinical settings for the identification of people who are at risk for falls. In general, these tests have been adopted readily because of their simplicity and low cost.

The Dynamic Gait Index (DGI)[9(pp405–406),13] is a standardized clinical assessment that aids in evaluating a person's ability to modify gait in response to changing gait task demands. The DGI is a performance-based test developed as part of a profile of tests and measurements that are effective in predicting likelihood for falls in community-dwelling older adults.[14] The DGI has been shown to yield ratios of subject variability to total variability with excellent interrater reliability (.96) and test-retest reliability (.98) when rated by physical therapists.[15] The DGI correctly classifies 59% of people with a history of falls (sensitivity) while correctly classifying 64% of those without a history of falls (specificity).[16] The DGI rates performance from 0 (severe impairment) to 3 (normal) on 8 different gait tasks. The 8 tasks, administered from item 1 to item 8, consist of gait on even surfaces, gait when changing speeds, gait and head turns in a horizontal direction, gait and head turns in a vertical direction, gait with pivot turns, stepping over obstacles, stepping around obstacles, and ascending and descending steps.[9(pp405–406)] Scores on the DGI range from 0 to 24. Although a recent study by Boulgarides et al[14] showed that the DGI (along with 4 other commonly used balance assessments) cannot predict falls in a sample of community-dwelling, active, independent older adults, the DGI has been shown to be correlated with falls in other populations.[9(p401),15] Shumway-Cook et al[15] showed that a score of 19 or less, out of 24, indicates an increased risk of falling in older adults.

Many rehabilitation specialists believe that balance assessment under multitask conditions (ie, performing more than one activity at the same time, such as walking forward and simultaneously looking up and down) may be a more sensitive indicator of balance problems and falls than balance assessment in a single-task context.[17–20] This belief is attributable to the fact that elderly people often fall when they try to perform 2 activities at once.[21] Given that the DGI has many tasks that allow for testing under multitask conditions (eg, walking with head turns or stepping over obstacles), it should be a more sensitive indicator of balance problems than other commonly used balance assessments that do not incorporate multiple tasks into the evaluation. Resnick[22] reported that 63% of falls occurred while walking, which is the key factor used across items in the DGI. Furthermore, the DGI has been shown to be a sensitive assessment tool for identifying people who are at risk for falls because of vestibular disorders.[23,24] However, one component of the DGI scale not addressed so far in the literature is the hierarchy of item difficulty. No explicit hierarchy was

YP Chiu, PT, MHS, is a doctoral candidate in rehabilitation science, Department of Physical Therapy, College of Public Health and Health Professions, University of Florida, Box 100154, UFHSC, Gainesville, FL 32611-0154 (USA) (yipaoc@ufl.edu). Address all correspondence to Mr Chiu.

SL Fritz, PT, PhD, is Clinical Assistant Professor, Department of Exercise Science, University of South Carolina, Columbia, SC; Graduate Student, Department of Physical Therapy, College of Public Health and Health Professions, University of Florida, Gainesville, Fla; and Pre-Doctoral Fellow, VA Brain Rehabilitation Research Center, Malcom Randall VA Medical Center, Gainesville, Fla.

KE Light, PT, PhD, is Associate Professor, Department of Physical Therapy, College of Public Health and Health Professions, University of Florida.

CA Velozo, OTR/L, PhD, is Associate Professor, Department of Occupational Therapy, College of Public Health and Health Professions, University of Florida, and Research Health Scientist, Rehabilitation Outcome Research Center, North Florida/South Georgia Veterans Health System, Gainesville, Fla.

intended when the DGI was developed, although it is usually administered in a standardized order (Anne Shumway-Cook, PT, PhD; verbal communication with YP Chiu; February 2003). Nevertheless, knowing the hierarchy of item difficulty can be an asset to both the researcher and the clinician.

Administering items by starting with the easiest and moving to the most difficult may be a logical progression in testing clients. Furthermore, if the hierarchical structure of the DGI is validated, the selective administration of items depending on an individual's ability level may prove to be efficient. For example, if a client is functionally ambulatory, instead of testing "gait on level surface," a more challenging item, such as "gait and pivot turn," could be administered initially. On the basis of the importance of the DGI as a clinical tool and research instrument in the assessment of balance and in the identification of people who are at risk for falls, it is worthwhile to evaluate further the item characteristics of the instrument by use of the Rasch measurement model. A number of articles recently published in the physical therapy literature support the use of Rasch analysis to clinically validate functional assessments.[25–28]

Although traditional psychometric approaches focus on the total score of a given instrument, the Rasch measurement model allows analysis of instruments at the item and rating scale levels. First, Rasch analysis converts ordinal raw-score data, such as the scale from 0 to 3 on the DGI, into an interval-based measure, the log-odd metric, or logit. Second, the analysis allows the determination of whether the rating scale is used in the expected manner (eg, people with lower balance ability would be expected to use lower item ratings, whereas people with higher balance ability would be expected to use higher item ratings). Third, the Rasch measurement model provides a connection between a person's total score and the items of the instrument by placing the person's ability (person measure) and item difficulty (item measure) on the same linear continuum. Ceiling and floor effects are revealed when person's ability and item difficulty fail to match at the extremes of the continuum. Item goodness-of-fit statistics provided by the analysis determine the extent to which each item fits the construct it is intended to measure. High fit statistics may indicate that the item is mismarked, poorly worded, or misinterpreted. In combination with principal components analysis (PCA), high fit statistics may identify a subset of items that measure a unique construct.[29]

The purpose of this study was to use Rasch measurement theory to examine: (1) whether the DGI rating scale meets suggested psychometric guidelines, (2) whether the hierarchical order of DGI tasks is consistent with a clinically logical testing procedure (ie, moving from easy

items to more difficult items), and (3) whether the DGI represents a unidimensional construct (ie, all items reflect a single latent trait [balance] rather than multiple constructs [both balance and endurance]).

Method

Participants

Data were retrieved retrospectively from 84 community-dwelling male veterans (mean age=75 years, SD=6.47, range=64–88) who were receiving care at the North Florida/South Georgia Veterans Affairs Malcom Randall Medical Center. These veterans were participating in a gait and balance rehabilitation program upon referral by their primary care physicians. The DGI was a component of a 90-minute comprehensive initial physical therapy assessment. The inclusion criteria for this study were: 65 years of age or older, 1 or more falls or numerous "near falls" in the preceding year, disequilibrium, persistent complaints of dizziness or balance problems, Mini-Mental State Examination score of at least 24 out of 30, Geriatric Depression Scale score of less than 5 out of 15, positive Romberg test, and inability to maintain single-leg stance. All subjects had multiple comorbidities. In addition to the DGI, a standardized falls history interview revealed that these subjects had experienced 0 to 12 or more falls over the preceding 12-month period ($\overline{X}$=5.5, median=4). The Timed "Up & Go" Test scores[12,30] ranged from 7.58 seconds to 43.66 seconds ($\overline{X}$=18.47, median=16.53), and the Berg Balance Scale scores ranged from 29 to 56 ($\overline{X}$=42.65, median=43).[31]

Analyses

The Rasch measurement model with the WINSTEPS program[32] was used in this study because it offers distinct advantages over traditional psychometric approaches. As stated above, Rasch analysis focuses on the psychometric properties of the item, person, and rating scale categories. Two values are used throughout the analysis: logit measures and fit statistics. Logits, or log-odd units, convert ordinal raw scores into linear interval measures.[33(pp17–19)] The logit is the natural logarithm of the odds of a person being successful at a specific task or an item being successfully carried out.[34] For the person category, logit measures indicate whether one person is more able than another (eg, Does one person have better balance ability than another?); for items, logit measures indicate whether one item is more difficult than another (eg, Is stepping over an obstacle more difficult than walking on a level surface?); and for rating scale categories, logit measures indicate whether one rating scale category is greater or less than another in degree (eg, Does a rating of 2 [mild impairment] represent less impairment than a rating of 3 [moderate impairment] in the DGI?).

Fit statistics[33(p208)] monitor the compatibility of the raw data with the Rasch measurement model. Fit to the Rasch measurement model requires that high ratings on more difficult items are accomplished by people with higher ability and that people have a greater probability of attaining higher scores on easier items than on more difficult ones.[35] In general, mean square (MnSq) fit statistics, which are used to identify item and person ratings that deviate from expectations, range from 0 to positive infinity. The MnSq fit statistics value is the ratio of observed variance (variance attributable to the data) to expected variance (variance estimated by the Rasch measurement model). Ideally, the ratio will be 1.0, so that observed variance equals expected variance. When the MnSq fit statistics value is greater than 1.0, for example, 1.70, there is 70% more variation in the observed data than the Rasch model predicted. When the fit statistics value is less than 1.0, there is less variation in the observed data than the Rasch model predicted (ie, overfit).[33(p177)] Two types of fit statistics are provided in this study: outfit and infit statistics.[33(p208)] Both are the average of standardized residual variance. Standardized residual variance is the difference between the observed score and the Rasch estimated score divided by the square root of the Rasch model variance.[36] Outfit statistics are unweighted, being affected more by unexpected responses far from the person, item, or rating scale category measure (eg, a person of low ability unexpectedly having a normal score on a difficult item). Infit statistics are weighted, being affected more by unexpected responses close to the person, item, or rating scale category measure (eg, a person of low ability unexpectedly having a score indicating severe impairment on an easy item).

Rating scale analysis was accomplished by determining whether the DGI 4-point rating scale met Linacre's 3 essential criteria for optimizing rating scale category effectiveness.[37] The criteria are as follows: 10 observations are obtained per rating scale category, category logit measures advance (eg, the average logit measure for the rating scale category "mild impairment" is greater than the average logit measure for the rating scale category "moderate impairment"), and the outfit MnSq value for each rating scale category is less than 2.0. In the present study, the frequency of each of the 4 rating scale categories in the DGI was computed. Average logit measures for 4 rating scale categories were used to determine whether the rating scale categories of the DGI advance monotonically. As proposed by Linacre,[37] the outfit MnSq value for each rating scale category was compared with the threshold value of 2.0. Values of greater than 2.0 suggest that more unexplained variance than explained variance is found.

The hierarchical order of the DGI items also was determined with the WINSTEPS program. The items of the DGI were arranged from the least difficult to the most difficult according to their corresponding logit measures. Item hierarchy can be used to investigate construct validity (ie, support or refute the expectation that "stepping over obstacles" is more challenging than "walking on a level surface" in the DGI). Furthermore, the comparison of item difficulty with person ability (ie, item-person map) can be used to determine whether the items of an instrument cover the range of person abilities in the sample (ie, reveal ceiling or floor effects).

Next, Rasch fit statistics in combination with PCA were used to test the unidimensionality of the DGI.[38] Reasonable ranges of MnSq fit values are between 0.6 and 1.4 and are with standardized z values of less than 2.0.[39] Recent studies[40–44] suggested that fit statistics alone are inadequate for determining unidimensionality. Therefore, to test further for unidimensionality, a PCA based on residuals[45] was conducted.[41–43] The PCA transforms correlated items into principal components. In the determination of unidimensionality, it is expected that after the removal of the Rasch dimension (eg, the trait that the DGI intends to measure), the residuals for pairs of items should be uncorrelated and normally distributed.[42] That is, there should be no principal components. When the first principal component has an eigenvalue of less than 1.4, then the measure is considered unidimensional.[44]

Finally, the WINSTEPS program provides several summary statistics for person ability and item difficulty logit measures. Person separation and person separation reliability are indicators of how well the items of the instrument separate or spread out the subjects in the sample. Person separation is an index of the sample standard deviation in terms of standard error units.[46(p106)] Person separation reliability is the proportion of observed sample variance that is not attributable to measurement error.[46(p106)] This value is analogous to the Cronbach alpha.[33(p207)] Similarly, item separation and item separation reliability are indicators of how well the subjects in the sample separate or spread out the items of the instrument. Item separation is an index of the item standard deviation in terms of calibration error units.[46(p92)] Item separation reliability is the proportion of observed item variance that is not attributable to estimation error.[46(p92)]

Results

Rating Scale Analysis

The results for the rating scale analysis of the DGI are shown in Table 1. Category frequency counts for the 4-point rating scale (ie, severe impairment, moderate

Table 1.
Rating Scale Analysis for the 4 Categories of Dynamic Gait

Rating Scale Category (Score)	Category Count	Category Percentage	Average Measure	Expected Measure	Outfit Mean Square
Severe impairment (0)	55	8	−1.32	−1.33	1.01
Moderate impairment (1)	229	35	−0.37	−0.39	1.01
Mild impairment (2)	269	41	0.66	0.7	0.99
Normal (3)	104	16	2.16	2.08	0.94

impairment, mild impairment, and normal) of the DGI were large, and all categories had more than 10 observations. The use of the rating scale categories was approximately normally distributed, with the middle categories "mild impairment" and "moderate impairment" representing 35% to 41% of the ratings and the extreme categories "normal" and "severe impairment" representing 8% to 16% of the ratings. The average logit measures for the 4 rating scale categories increased monotonically with rating scale category from −1.32 to −0.37 logits, from −0.37 to 0.66 logits, and then from 0.66 to 2.16 logits. As expected, lower rating scale categories (ie, "severe impairment" and "moderate impairment") were associated with lower average logit measures, whereas higher rating scale categories (ie, "normal" and "mild impairment") were associated with higher average logit measures. In addition, across all rating scale categories, the average logit measures derived from observations were all close to the expected logit measures predicted by the Rasch measurement model. Regarding outfit MnSq values, all 4 categories had values between 0.94 and 1.01, clearly meeting Linacre's requirement of an outfit MnSq value of less than 2.0.[37]

Item-by-item analysis of each rating scale category showed similar results as well. Seventy-five percent of rating scale categories (24/32) had more than 10 observations. All of the average logit measures for each rating scale category of each item increased monotonically. Their outfit MnSq values were all less than 2.0 (range=0.64–1.78).

Hierarchical Order of the DGI Tasks

Table 2 shows the DGI item administration order compared to the Rasch analysis-derived item difficulty order. The left-most column of Table 2 shows the original DGI item administration order (1–8). Item difficulty order was determined by use of the Rasch logit measures in the second column. "Level surface," "speed change," and "around obstacles" appeared to be the easiest items (lowest logit measures), whereas "vertical head turns,"

"steps," and "horizontal head turns" appeared to be the most difficult tasks (highest logit measures). Although the first 2 items routinely administered represent the easiest items (ie, "level surface" and "speed change"), the order of administration of the remainder of the items does not match the item difficulty order. For example, the item "around obstacles" is the next-to-last item administered, although it represents a fairly easy item, and the item "horizontal head turns" is the third item administered, even though it represents the most difficult item of the DGI.

Item difficulty order can be described further in relation to person ability logit measures, as graphically shown in the Figure.[47] To fully reflect the item difficulty spread, items are presented (at the left of the Figure) 3 times, or at 3 "step" calibrations. Step calibrations represent the increments in difficulty as the scoring criteria progress from a rating of 0 to 1 (the lowest presentation of the items) to a rating of 2 to 3 (the highest presentation of the items). The middle presentation of items reflects the average item difficulties (item mean logit measure values shown in Tab. 2). Therefore, the difficulty spread of the items is between −3.2 and 3.0 logits. Item separation is 1.98, and item separation reliability is .80. The range of person ability logit measures is represented as the bars at the right of the Figure. Although the person ability logit measures spread beyond the item difficulty range (−2.6 to 5.0 logits), only 4% of the sample (4/84), that is, subjects with logits above 3.0, is not covered by the item difficulty range. Technically, only 1 of these subjects obtained a perfect score on the DGI. Person separation is 1.98, and person separation reliability is .80.

Unidimensional Construct

Fit statistics and PCA were used to determine the unidimensionality of the DGI. Table 2 shows that all items had infit and outfit statistics within the reasonable range for observations (ie, between 0.6–1.4 and associated with standardized z values of <2.0[39]), except for the task "vertical head turns" (MnSq infit value=1.41; z value=2.5).

The PCA of the DGI showed that the residual component (ie, the component beyond the single latent trait) has an eigenvalue of 1.8, representing only 22.5% (1.8/8) of the residual variance. In simulation studies, Smith and Miao[44] reported that eigenvalues of less than 1.4 are at the random level. Therefore, the DGI items are essentially unidimensional. Table 3 shows the factor loadings of 8 items for the secondary dimension in the DGI. Three items (items 3, 4, and 5) with head turns

Table 2.
Item Difficulty and Fit Statistics

| | | Infit | | Outfit | | |
| | **X̄ (SD)** | **Mean** | **Standardized** | **Mean** | **Standardized** | |
Entry	**(Logits)**	**Square**	**z Value**	**Square**	**z Value**	**Item**[a]
3	0.52 (0.17)	1.14	0.9	1.19	1.2	Horizontal head turns
8	0.41 (0.18)	1.26	1.6	1.24	1.5	Steps
4	0.32 (0.18)	1.41	2.5	1.38	2.3	Vertical head turns
5	0.15 (0.18)	1.05	0.4	1.04	0.3	Pivot turn
6	−0.02 (0.18)	0.66	−2.6	0.67	−2.5	Over obstacle
7	−0.16 (0.18)	0.88	−0.8	0.86	−1.0	Around obstacles
2	−0.55 (0.18)	0.89	−0.8	0.89	−0.8	Speed change
1	−0.68 (0.18)	0.68	−2.4	0.67	−2.4	Level surface

[a] Arranged in order of difficulty from most difficult (top) to least difficult (bottom).

("horizontal head turns," "vertical head turns," and "pivot turn") load in the direction opposite that of the remaining 5 items.

Discussion

The purpose of this study was to investigate the rating scale structure, task hierarchical order, and unidimensionality of the DGI by use of item response analysis. Overall, the DGI showed sound item psychometric properties. Each of the original 4 rating scale categories appeared to distinctly identify people who are at different ability levels. Fit statistics and PCA indicated that the 8 items of the DGI appear to represent a single construct. The analysis revealed a clear item difficulty hierarchical order that is generally consistent with clinical expectations.

Rating Scale

The sound psychometric properties of the rating scale of the DGI may reflect the consistency of the scale with typical clinical observations and language. The use of each of the rating scale categories was distributed normally. The middle categories "mild impairment" and "moderate impairment" were the most frequently used responses, and the 2 extreme categories "normal" and "severe impairment" were used the least. Furthermore, the use of each rating scale category was connected to person ability level. That is, as subject ability increased, there was a clear tendency for higher ratings to be used.

These findings challenge the suggestions of Krishnan et al[48] that the DGI would be improved by expanding its rating scale categories by adding either extra timing components or time for completing tasks. They claimed that without mutually exclusive and exhaustive rating scales, an evaluator would have difficulty accurately assigning scores of 2 (mild impairment) and 1 (moderate impairment) because some people may demonstrate certain characteristics from more than one category. The DGI rating scale categories met or exceeded Linacre's guidelines for optimizing rating scale category effective-

ness.[37] The present rating scale analysis suggests that evaluators had no difficulty in differentiating between these 2 ratings. The psychometric stability of the ratings may have resulted from the use of the universally accepted clinical terms and explicit definitions provided in the DGI. Terms such as "severe impairment," "moderate impairment," "mild impairment," and "normal" are rooted in clinical training and are used widely across a variety of clinical instruments. Furthermore, explicit definitions, such as "normal: performs head turns smoothly with no change in gait," provide clear guidance with which to grade a person's performance.

Hierarchical Order of the DGI Tasks

The results of Rasch analysis of the DGI revealed the underlying hierarchical order of item difficulty. "Gait with horizontal head turns," "steps," and "gait with vertical head turns" were the most difficult items, whereas "gait on level surface," "change in gait speed," and "step around obstacles" were the easiest items. The degree of sensory interference, novelty, and required effort may explain the item order demonstrated. The difficulty of the items "gait with horizontal head turns" and "gait with vertical head turns" may be attributed to vestibular influences and the novelty of the tasks. In addition, tasks such as "steps" (walk up stairs, at top turn around and walk down) may have been challenging because of musculoskeletal demands. In contrast, items that have fewer sensory demands and require less effort were shown to be the least difficult items, that is, "gait on level surface" and "change in gait speed."

The hierarchical structure of the DGI may have implications for modifying the current clinical administration of the DGI. At present, several of the most difficult tasks in the DGI—that is, "gait with horizontal head turns," "gait with vertical head turns," and "pivot turn"—are presented very early in the typical administration sequence, third, fourth, and fifth, respectively. Requiring people with severe impairments to perform these relatively challenging tasks early in the assessment may lead to
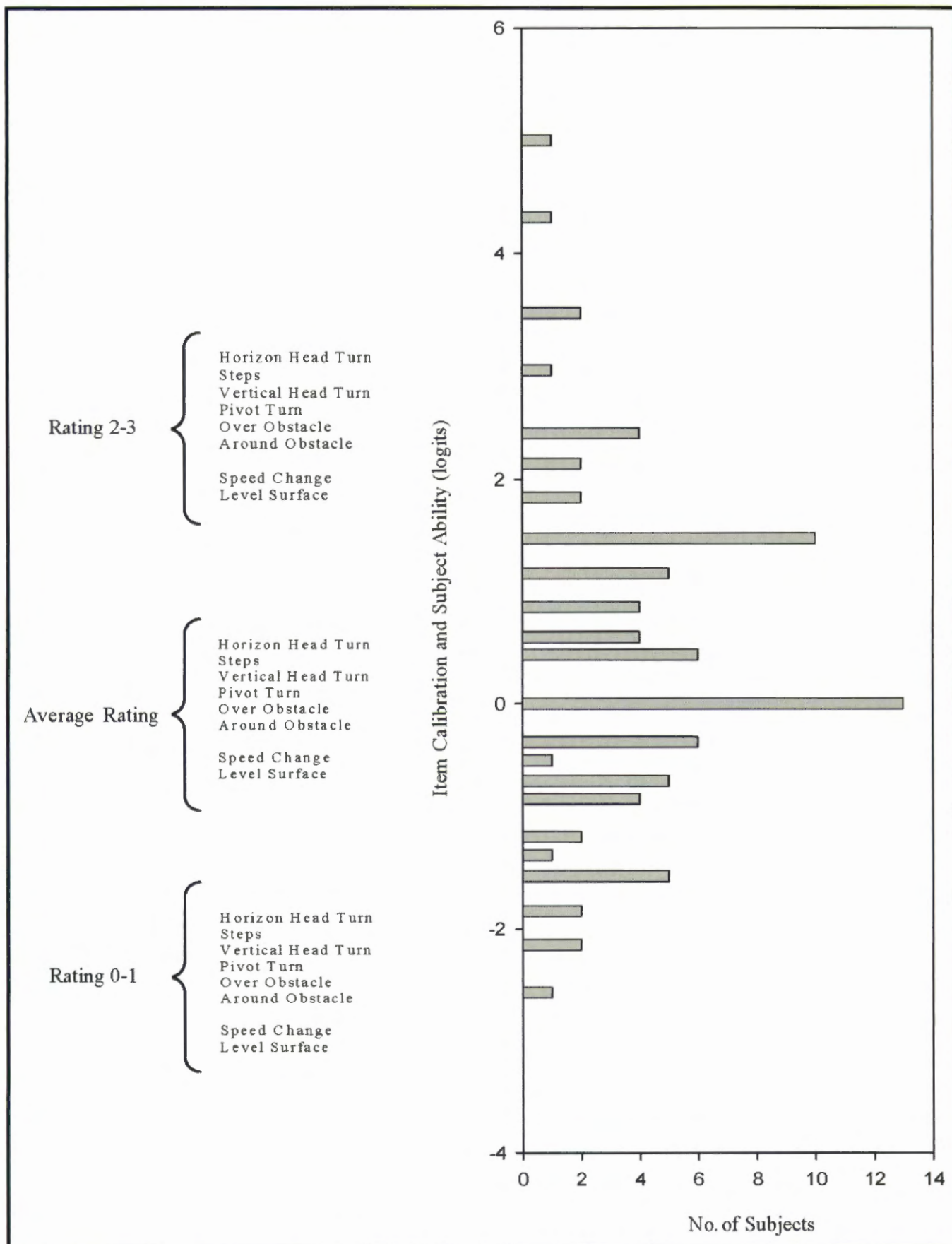
Figure.
Rasch item-subject map. Bars indicate the number of subjects (x-axis) at each ability level (y-axis). Items are presented at 3 step calibrations, corresponding to each item receiving a rating of 0 to 1, an average rating, and a rating of 2 to 3.

Table 3.
Factor Loading of 8 Dynamic Gait Index Items From the Principal Components Analysis Standardized Residuals

Item	Loading
Horizontal head turns	.68
Vertical head turns	.58
Pivot turn	.36
Speed change	−.55
Level surface	−.51
Over obstacle	−.45
Steps	−.36
Around obstacles	−.13

frustration, insecurity, and safety concerns. In addition, asking people to perform easier tasks, such as "step over obstacle" and "step around obstacles," later in the assessment (sixth and seventh items administered in the DGI) deviates from the standard administration in which tasks progress from easy tasks to challenging tasks.

Information on the item difficulty hierarchy could lead to more dramatic administration modifications. For example, on the basis of the Rasch measurement model, a person who is capable of "climbing steps" will have a high probability of being successful at "walking on a level surface." The above scenario suggests that if a person is successful at "climbing steps," a challenging item, then it would be unnecessary to test the person on "walking on a level surface," an easier item. This "modern measurement" approach of selective item administration is commonly used in developmental testing[49-51] and is the basis for computerized adaptive testing.[52] The selective administration of items on the basis of ability could dramatically reduce the burden of testing on the individual and therapist time in test administration.[53,54]

Unidimensional Construct
The unidimensionality of the DGI is supported by both the fit statistics and the PCA.[38] The infit and outfit values from overall person ability and item difficulty were both close to the ideal value of 1.0. Because of the low eigenvalue, the PCA further supports the integrity of the DGI for this sample. Often, in an effort to make an instrument all encompassing, multiple dimensions of a function or skill are combined. For example, the Functional Independence Measure combines motor and cognitive items.[55,56] This combination can lead to challenges in making clear, clinical inferences. For example, improvement in Functional Independence Measure scores may be attributable to improvement in the motor construct, the cognitive construct, or both. In contrast, the unidimensionality reflected in the present form of the DGI will support interventions that focus on a single construct representing dynamic balance.

This investigation of the dimensionality of the DGI may provide some insight into the elemental components that comprise balance. Although the PCA eigenvalue was insufficient to support multiple constructs, the factor loadings suggest that a secondary construct may be embedded in the DGI. That is, all 3 items that have significant vestibular involvement (items 3, 4, and 5) had a tendency to load in directions opposite that of the remainder of the items. Tasks with vestibular involvement represent 3 of the 4 most difficult items, suggesting that with more challenging balance tasks, the multidimensionality of dynamic balance may emerge. Furthermore, it is possible with a larger number of subjects and less variance that the vestibular factor could form a separate construct.

Several limitations in this study may have influenced the psychometric findings presented. The subjects included were community-dwelling elderly people with identified balance deficits. Furthermore, the sample consisted solely of male veterans. The homogeneity of this sample may have favored the strong psychometric outcomes in this study.[57,58] Replication of this study with a more diverse sample is warranted.

Conclusion
The results suggest that the rating scale of the DGI is used appropriately for community-dwelling elderly people with balance problems. In addition, the hierarchy of DGI item difficulty revealed may lead to a more logical administration of the instrument depending on the presenting balance skill level. Furthermore, the analysis demonstrated that the DGI fits a unidimensional construct, further supporting the use of this tool for clinical decision making. The findings support continued use of this well-constructed scale for clinical and research assessment in a community-dwelling elderly population.

References
1 Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *N Engl J Med.* 1988;319:1701–1707.

2 Rubenstein LZ, Josephson KR, Robbins AS. Falls in the nursing home. *Ann Intern Med.* 1994;121:442–451.

3 Covinsky KE, Kahana E, Kahana B, et al. History and mobility exam index to identify community-dwelling elderly persons at risk of falling. *J Gerontol A Biol Sci Med Sci.* 2001;56:M253–M259.

4 Campbell AJ, Robertson MC, Gardner MM, et al. Falls prevention over 2 years: a randomized controlled trial in women 80 years and older. *Age Ageing.* 1999;28:513–518.

5 McMurdo ME, Millar AM, Daly F. A randomized controlled trial of fall prevention strategies in old peoples' homes. *Gerontology.* 2000;46:83–87.

6 Berg KO, Maki BE, Williams JI, et al. Clinical and laboratory measures of postural balance in an elderly population. *Arch Phys Med Rehabil.* 1992;73:1073–1080.

7 Bogle Thorbahn LD, Newton RA. Use of the Berg Balance Test to predict falls in elderly persons. *Phys Ther*. 1996;76:576–583.

8 Newton RA. Balance screening of an inner city older adult population. *Arch Phys Med Rehabil*. 1997;78:587–591.

9 Shumway-Cook A, Woollacott MH. *Motor Control: Theory and Practical Applications*. Philadelphia, Pa: Lippincott Williams & Wilkins; 2001:401, 405–406.

10 Cumming RG, Kelsey JL, Nevitt MC. Methodologic issues in the study of frequent and recurrent health problems: falls in the elderly. *Ann Epidemiol*. 1990;1:49–56.

11 Cummings SR, Nevitt MC, Kidd S. Forgetting falls: the limited accuracy of recall of falls in the elderly. *J Am Geriatr Soc*. 1988;36: 613–616.

12 Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed "Up & Go" Test, and gait speeds. *Phys Ther*. 2002;82:128–137.

13 Shumway-Cook A. Retaining stability and mobility. Platform presentation at: Annual Conference of the American Physical Therapy Association; June 12–16, 1993; Cincinnati, Ohio.

14 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;83:328–339.

15 Shumway-Cook A, Gruber W, Baldwin M, Liao S. The effect of multidimensional exercises on balance, mobility, and fall risk in community-dwelling older adults. *Phys Ther*. 1997;77:46–57.

16 Shumway-Cook A, Baldwin M, Polissar NL, Gruber W. Predicting the probability for falls in community-dwelling older adults. *Phys Ther*. 1997;77:812–819.

17 Geurts AC, Mulder TW, Nienhuis B, Rijken RA. Dual-task assessment of reorganization of postural control in persons with lower limb amputation. *Arch Phys Med Rehabil*. 1991;72:1059–1064.

18 Lajoie Y, Teasdale N, Bard C, Fleury M. Attentional demands for static and dynamic equilibrium. *Exp Brain Res*. 1993;97:139–144.

19 Lundin-Olsson L, Nyberg L, Gustafson Y. "Stops walking when talking" as a predictor of falls in elderly people. *Lancet*. 1997;349:617.

20 Shumway-Cook A, Woollacott M, Kerns KA, Baldwin M. The effects of two types of cognitive tasks on postural stability in older adults with and without a history of falls. *J Gerontol A Biol Sci Med Sci*. 1997;52: M232–M240.

21 Gillespie LD, Gillespie WJ, Robertson MC, et al. Interventions for preventing falls in elderly people. *Cochrane Database Syst Rev*. 2003: CD000340.

22 Resnick B. Falls in a community of older adults: putting research into practice. *Clin Nurs Res*. 1999;8:251–266.

23 Whitney SL, Hudak MT, Marchetti GF. The dynamic gait index relates to self-reported fall history in individuals with vestibular dysfunction. *J Vestib Res*. 2000;10:99–105.

24 Whitney SL, Wrisley D, Furman J. Concurrent validity of the Berg Balance Scale and the Dynamic Gait Index in people with vestibular dysfunction. *Physiother Res Int*. 2003;8:178–186.

25 Russell DJ, Avery LM, Rosenbaum PL, et al. Improved scaling of the gross motor function measure for children with cerebral palsy: evidence of reliability and validity. *Phys Ther*. 2000;80:873–885.

26 Iyer LV, Haley SM, Watkins MP, Dumas HM. Establishing minimal clinically important differences for scores on the pediatric evaluation of disability inventory for inpatient rehabilitation. *Phys Ther*. 2003;83: 888–898.

27 Kothari DH, Haley SM, Gill-Body KM, Dumas HM. Measuring functional change in children with acquired brain injury (ABI): comparison of generic and ABI-specific scales using the Pediatric Evaluation of Disability Inventory (PEDI). *Phys Ther*. 2003;83:776–785.

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

29 Karabatsos G. A critique of Rasch residual fit statistics. *J Appl Meas*. 2000;1:152–176.

30 Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed "Up & Go" Test. *Phys Ther*. 2000;80:896–903.

31 Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. *Can J Public Health*. 1992;83(suppl 2):S7–S11.

32 Linacre JM, Wright BD. *WINSTEPS*. Chicago, Ill: MESA Press; 2000.

33 Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001:17–19, 177, 207, 208.

34 Duncan PW, Bode RK, Min Lai S, Perera S. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Arch Phys Med Rehabil*. 2003;84:950–963.

35 Campbell SK, Kolobe TH, Osten ET, et al. Construct validity of the test of infant motor performance. *Phys Ther*. 1995;75:585–596.

36 Wright BD, Stone MH. *Best Test Design*. Chicago, Ill: MESA Press; 1979:23.

37 Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas*. 2002;3:85–106.

38 Linacre JM. *Learning from Principal Components Analysis of Residuals*. Chicago, Ill: COMET; 1999.

39 Velozo CA, Peterson EW. Developing meaningful Fear of Falling Measures for community dwelling elderly. *Am J Phys Med Rehabil*. 2001;80:662–673.

40 Smith EV Jr, Johnson BD. Attention deficit hyperactivity disorder: scaling and standard setting using Rasch measurement. *J Appl Meas*. 2000;1:3–24.

41 Linacre JM. Structure in Rasch residuals: why principal components analysis? *Rasch Measurement Transactions*. 1998;12:636.

42 Linacre JM. Detecting multidimensionality: which residual datatype works best? *J Outcome Meas*. 1998;2:266–283.

43 Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas*. 2002;3:205–231.

44 Smith RM, Miao CY. Assessing unidimensionality for Rasch measurement. In: Wilson M, ed. *Objective Measurement: Theory into Practice*. Norwood, NJ: Ablex Publishing Co; 1994:316–327.

45 Moore D, McCabe G. *Introduction to the Practice of Statistics*. New York, NY: WH Freeman & Co; 1993:854.

46 Wright BD, Masters GN. *Rating Scale Analysis: Rasch Measurement*. Chicago, Ill: MESA Press; 1982:92, 106.

47 Velozo CA, Magalhaes LC, Pan AW, Leiter P. Functional scale discrimination at admission and discharge: Rasch analysis of the Level of Rehabilitation Scale-III. *Arch Phys Med Rehabil*. 1995;76:705–712.

48 Krishnan LV, O'Kane KS, Gill-Body KM. Reliability of a modified version of the Dynamic Gait Index: a pilot study. *Neurol Rep*. 2002;26: 8–14.

49 Reckase MD. Adaptive testing: the evolution of a good idea. *Educational Measurement: Issues and Practice*. 1989;8:11–15.

50 Kingsbury GG, Zara AR. Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education.* 1989;2:359–375.

51 Hambleton RK, Zaal JN. *Advances in Educational and Psychological Testing.* Boston, Mass: Kluwer Academic Publishers; 1991.

52 Wainer H, Dorans NJ. *Computerized Adaptive Testing: A Primer.* Mahwah, NJ: Lawrence Erlbaum Associates; 2000:xxiii, 335.

53 Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. *Am J Occup Ther.* 1999;53:83–90.

54 Weiss DJ. *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing.* New York, NY: Academic Press; 1983.

55 Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil.* 1987;1:6–18.

56 Hamilton BB, Granger CV. Disability outcomes following inpatient rehabilitation for stroke. *Phys Ther.* 1994;74:494–503.

57 Schultz AB, Ashton-Miller JA, Alexander NB. What leads to age and gender differences in balance maintenance and recovery? *Muscle Nerve Suppl.* 1997;5:S60–S64.

58 Wolfson L, Whipple R, Derby CA, et al. Gender differences in the balance of healthy elderly as demonstrated by dynamic posturography. *J Gerontol.* 1994;49:M160–M167.