Expanding the Scoring System for the Dynamic Gait Index

Anne Shumway-Cook, Catherine S. Taylor, Patricia Noritake Matsuda, Michael T. Studer, Brady K. Whetten

Background. The Dynamic Gait Index (DGI) measures the capacity to adapt gait to complex tasks. The current scoring system combining gait pattern (GP) and level of assistance (LOA) lacks clarity, and the test has a limited range of measurement.

Objective. This study developed a new scoring system based on 3 facets of performance (LOA, GP, and time) and examined the psychometric properties of the modified DGI (mDGI).

Design. A cross-sectional, descriptive study was conducted.

Methods. Nine hundred ninety-five participants (855 patients with neurologic pathology and mobility impairments [MI group] and 140 patients without neurological impairment [control group]) were tested. Interrater reliability was calculated using kappa coefficients. Internal consistency was computed using the Cronbach alpha coefficient. Factor analysis and Rasch analysis investigated unidimensionality and range of difficulty. Internal validity was determined by comparing groups using multiple *t* tests. Minimal detectable change (MDC) was calculated for total score and 3 facet scores using the reliability estimate for the alpha coefficients.

Results. Interrater agreement was strong, with kappa coefficients ranging from 90% to 98% for time scores, 59% to 88% for GP scores, and 84% to 100% for LOA scores. Test-retest correlations (*r*) for time, GP, and LOA were .91, .91, and .87, respectively. Three factors (time, LOA, GP) had eigenvalues greater than 1.3 and explained 79% of the variance in scores. All group differences were significant, with moderate to large effect sizes. The 95% minimal detectable change (MDC_{95}) was 4 for the mDGI total score, 2 for the time and GP total scores, and 1 for the LOA total score.

Limitations. The limitations included uneven sample sizes in the 2 groups. The MI group were patients receiving physical therapy; therefore, they may not be representative of this population.

Conclusions. The mDGI, with its expanded scoring system, improves the range, discrimination, and facets of measurement related to walking function. The strength of the psychometric properties of the mDGI warrants its adoption for both clinical and research purposes.

A. Shumway-Cook, PT, PhD, FAPTA, Department of Rehabilitation Medicine, Division of Physical Therapy, University of Washington, 1959 NE Pacific St, Box 356490, Seattle, WA 98195 (USA). Address all correspondence to Dr Shumway-Cook at: ashumway@u.washington.edu.

C.S. Taylor, PhD, College of Education, University of Washington.

P.N. Matsuda, PT, DPT, PhD, Department of Rehabilitation Medicine, Division of Physical Therapy, University of Washington.

M.T. Studer, PT, MHS, NCS, Northwest Rehabilitation Associates, Salem, Oregon.

B.K. Whetten, PT, DPT, GCS, Northwest Rehabilitation Associates.

[Shumway-Cook A, Taylor CS, Matsuda PN, et al. Expanding the scoring system for the Dynamic Gait Index. *Phys Ther.* 2013;93:1493–1506.]

Published Ahead of Print: June 27, 2013
Accepted: June 24, 2013
Submitted: January 26, 2013

The Dynamic Gait Index (DGI) is a commonly used clinical measure that evaluates the capacity to adapt gait to complex walking tasks commonly encountered in daily life.[1] The DGI is based on a person-environment model of mobility disability (defined as reduced participation in the mobility domain) in which environmental demands are categorized into 8 dimensions—distance, temporal, ambient, terrain, physical load, attention, postural transitions, and density—representing the external demands that have to be met for an individual to be mobile within a particular environment.[2] Item 1 of the DGI is considered the reference or baseline task and examines the ability to walk under low-challenge conditions (self-paced, level surface, gait). The remaining 7 items examine a person's ability to adapt gait to task demands in 4 environmental dimensions: temporal (changing speed), postural transition (change in direction using a pivot turn, change in head orientation including horizontal and vertical head turns), terrain (climbing stairs), and density dimension (stepping over and around obstacles for collision avoidance). Performance on each item is evaluated using an ordinal scale with criteria based on a combination of gait pattern (GP), speed, and level of assistance (LOA). Ordinal scores for each item range from 0 (severe GP impairment, unable to perform without the physical assistance of another person) to 3 (normal GP, in a timely manner, no assistance required), with a total score range from 0 (worst/poor) to 24 (best function).

Available With This Article at ptjournal.apta.org

- eAppendix: Modified Dynamic Gait Index

The psychometric properties of the original DGI have been studied in a wide variety of patient populations, including older adults[3–7] and those with stroke,[8,9] Parkinson disease,[10] multiple sclerosis,[11–13] and vestibular deficits.[14,15] Interrater reliability is reported to be high.[8,11,14] Test-retest reliability values also were high, with intraclass correlation coefficients ranging from .84 to .96.[8–10] Concurrent validity coefficients were found to be moderate to high with the Timed "Up & Go" Test (TUG) and Timed Walking Test in people with stroke[8] and the 10-Meter Walk Test and Postural Assessment Scale for Stroke Patients.[9] Performance on the DGI has been used as a measure of fall risk; however, the cutpoint differentiating fallers from nonfallers varies by patient population, ranging from 19 to 23.[16,17]

Two studies have used Rasch analyses to examine DGI function in adult male veterans[6] and individuals with vestibular deficits.[15] Results suggest the DGI is unidimensional, with a range of item difficulty from low (walking on level surface, changing speed) to high (horizontal head turns, stairs). Multiple items (eg, walking at usual pace, walking around obstacles) have a comparable level of difficulty, leading to the suggestion that easier and psychometrically redundant items be dropped to shorten the test.[6,15]

Despite its psychometric strengths, the original DGI has a number of limitations. First, performance is rated using an ordinal score that incorporates 3 facets of performance (assessment of GP relative to normal, LOA, and, in some items, time) without a clear explanation of the relative contribution of each of these facets of performance to the task score. Putting all 3 facets of performance into a single score prevents therapists from monitoring improvement or decline in mobility function relative to these different facets of performance. Second, a ceiling effect has been reported when using the DGI to evaluate walking in community-dwelling older adults with relatively high mobility function, thus limiting responsiveness and sensitivity in this population.[7] Both of these limitations reduce the utility of the DGI as an outcome measure.

The purposes of this study were: (1) to develop a new scoring system for the DGI based on 3 facets of performance (ie, LOA, GP, and time) and (2) to examine the psychometric properties of a modified version of the original DGI (mDGI). One goal of the new scoring system was to provide a greater range of measurement so that scores could be used to reliably monitor improvement or deterioration of mobility in patients with a broader range of function than is currently possible using the original scoring rubric. Additionally, the new scoring system would enable monitoring of patients' improvement or deterioration in any 1 of the 3 identified facets of performance (ie, LOA, GP, and time) in each walking task. A second goal of the study was to investigate the psychometric properties of the mDGI scoring system, including dimensionality and potential redundancy of measurement.

Method

In order to conduct the planned Rasch analyses, sufficient numbers of participants were needed to obtain stable item parameters. We estimated that 10 participants were needed for each score parameter; because the mDGI included 64 score parameters, a minimum of 640 participants was needed.

Recruitment

Identification of patient participants began with recruitment of potential clinical testing sites. This process involved posting an e-mail on the

Section of Neurology listserve for the American Physical Therapy Association (APTA) requesting assistance from clinical sites currently using the DGI as a part of current clinical evaluation procedures. A letter of agreement was signed by each participating center. Within each site, study participants were recruited from among patients with neurological impairments, currently receiving physical therapist services for balance and mobility deficits, who would be evaluated using the DGI as part of usual care. Additional inclusion criteria included ability to walk 6.1 m (20 m) without the physical assistance of another person, although use of an assistive device was permitted.

A convenience sample of adults with no neurologic diagnosis was recruited from volunteers responding to flyers posted at the Department of Rehabilitation Medicine, University of Washington. In addition, letters were sent to retirement communities in the greater Seattle and Bellevue area. Criteria for inclusion in the control cohort were: age between 15 and 99 years, no neurologic diagnosis, ability to walk without the physical assistance of another person for a distance of 6.1 m, and ability to give informed consent. Potential participants went through an initial telephone screen to determine eligibility prior to being tested at 1 of 5 community sites. All participants gave informed consent prior to testing.

Modifying the DGI

Task modifications. The original 8 tasks were retained in the mDGI; however, minor modifications were made to facilitate timing and to clarify procedures for several of the tasks. To enable timing, a 6.1-m distance was imposed for all tasks. The change pace task included an acceleration phase ("walk as quickly as you safely can") but no deceleration phase, as it was difficult to impose both of these task modifications in a distance of 6.1 m. The over obstacle task changed the dimensions of the obstacles and specified the dimensions and placement of the 2 obstacles. Obstacle dimensions were 76 cm long, 12 cm wide, and 5 cm thick. The shoebox was eliminated due to anecdotal reports that many patients completed the task by sliding a foot around the obstacle rather that stepping over the obstacle (Anne Shumway-Cook, unpublished reports). The first obstacle was placed 2.4 m (8 ft) from the starting point, with the 12-cm side flat on the floor, requiring patients to increase step length to clear the obstacle. The second obstacle was placed with the 12-cm side up, requiring participants to increase their step height to clear the obstacle. This obstacle was placed 2.4 m past the first obstacle (about 4.9 m [16 ft] from the start). These modifications were based on research on obstacle crossing in patients with hemiplegia, indicating that some patients have difficulty increasing step length and that others have difficulty with step height[18]; thus, both conditions were included in the obstacle task. The pivot turn task included a turn halfway through the course with a return to the starting position, rather than a stop, in order to complete the 6.1-m distance. The stairs task measured performance while going up but not down stairs.

Scoring modifications. Scoring for original DGI was limited to a single ordinal score for each item ranging from 0 to 3. For the mDGI, 3 separate scores were applied to a participant's performance. Therapists measured the time it took for the individual to complete each task performed over a 6.1-m course. Level of assistance was scored using a 3-level scale (2=no assistance; 1=uses an assistive device [not including orthosis or braces]; and 0=requires the physical assistance of another person, including contact guard). Gait pattern was scored on a 4-level scale that differed slightly for each task. For example, the 4-level DGI scale for the usual pace walk on level surface task was:

(3) Normal: walks 6.1 m, normal GP, no evidence for imbalance

(2) Mild impairment: walks 6.1 m, mild gait deviations

(1) Moderate impairment: walks 6.1 m, moderate gait deviations, evidence of imbalance but recovers independently

(0) Severe impairment: cannot walk 6.1 m or walks with severe gait deviations or imbalance

A copy of the mDGI is shown in the eAppendix (available at ptjournal.apta.org). Data for each participant were collected for both the original DGI and the mDGI. Prior to data collection, therapists participating in this study received training on the new scoring system for the DGI. Three patient training videos were created for this purpose, with detailed verbal and written instructions on the new scoring system. All therapists participating in this study completed these training videos and had the opportunity to ask questions regarding the new scoring system. In addition, they scored 1 training video and submitted their data to study coordinators, who provided feedback to them regarding their scores. Finally, all sites submitted data on the first 5 patients for review by study investigators (P.N.M. and A.S-C.).

Converting time to an ordinal score. In order to investigate dimensionality using factor analysis and Rasch analysis, all item scores had to be on a dichotomous or ordinal scale. Therefore, time measure-

Table 1.
Proposed Scoring for the Modified Dynamic Gait Index

Task	Facets of Performance			
	Time	Gait Pattern	Level of Assistance	Total Task Score
Temporal				
Usual pace	0–3	0–3	0–2	0–8
Change pace	0–3	0–3	0–2	0–8
Postural				
Horizontal head turns	0–3	0–3	0–2	0–8
Vertical head turns	0–3	0–3	0–2	0–8
Pivot turn	0–3	0–3	0–2	0–8
Terrain				
Stairs	0–3	0–3	0–2	0–8
Density				
Stepping over obstacles	0–3	0–3	0–2	0–8
Stepping around obstacles	0–3	0–3	0–2	0–8
Total facet score	0–24	0–24	0–16	
Total mDGI score				0–64

ments were converted to a 4-level ordinal scale. All time measurements were converted to meters per second. The time for the usual pace task was converted to a 4-level scale based on walking speeds documented in the literature[19–23]:

(3) Normal walking speed: >1.1 m/s

(2) Community ambulatory: 0.8 to ≤1.1 m/s

(1) Limited ambulatory: 0.4 to ≤0.79 m/s

(0) Household ambulatory: ≤0.39 m/s

Time-level scores for tasks 2 through 8 were determined through an examination of the time distributions on the remaining 7 tasks for individuals at each time level on the usual pace task. Time distributions were examined graphically through histograms and numerically through descriptive statistics. Initial cut scores for time levels on the remaining 7 tasks were based on the points of intersection for these distributions.

The points of intersection were modeled as follows:

1. For participants at each of the 4 levels on the usual pace task, means and standard deviations of time and speed (in meters per second) were computed for tasks 2 through 8.

2. A time band, based on mean (±1 standard deviation), was computed for participants at each usual pace time level on each of the other 7 tasks.

3. To set the cutpoint between time level 1 and time level 2 on tasks 2 through 8, the midpoint was computed between the highest point in the time band for participants at level 1 and the lowest time in the time band for participants at level 2. The midpoint became the time level cutpoint between level 1 and level 2 on the given task.

4. This process was repeated to set the cutpoints between subsequent levels for the remaining tasks.

Proposed Scoring for the mDGI

Table 1 summarizes the proposed scoring for the mDGI. Scores are calculated at the task, facets of performance, and total score levels. To calculate a performance score at the task level, scores for time, GP, and LOA within a task are summed, for a score range of 0 to 8. An individual who is unable or refuses to complete a task is scored 0 for all 3 performance indicators. In addition to the 8 individual task scores, a total score for each of the 3 facets of performance is calculated, enabling walking performance to be characterized with respect to time (range=0–24), GP (range=0–24), and LOA (range=0–16). Finally, an mDGI total score is calculated by combining the 3 performance scores for a total score range of 0 to 64. Also included in Table 1 are the environmental dimensions reflected in the mDGI.

Reliability

Score reliability was investigated in 3 ways: an interrater reliability (IRR) study, a test-retest reliability study, and an internal consistency reliability analysis. To investigate IRR, 9 volunteer participants (7 in the mobility-impaired group and 2 in the control group) were videotaped performing the DGI. Videos were posted to the research website, where they were viewed and scored by 8 physical therapists who participated in collecting data for the study. Time measures were converted to the time-level ordinal scale for each of the tasks. Interrater reliability was evaluated for both individual task scores and total scores for each of the 3 scoring facets. The percentage of times physical therapists gave exactly the same score (percent exact agreement) as the criterion score (scores assigned by the test developer [A.S-C.]) for time, LOA, and GP for each task were calculated. Kappa coefficients were computed to evaluate the degree to which agreement with criterion scores was over and above what would be expected by chance alone.[24] Interrater reliability at the performance facet level also was assessed by computing total facet scores for time, GP, and LOA and correlating each score with the criterion facet scores.

To investigate test-retest reliability, 257 participants (253 in the mobility-impaired group and 4 in the control group) were retested using both the original DGI and the mDGI. For patients with rapidly changing function (eg, patients in acute and subacute settings), the optimal timeframe for retesting was within 24 hours. For patients whose function was changing at a slower rate (eg, outpatients with chronic conditions), the optimal timeframe for retesting was between 3 and 7 days following the initial examination. Internal consistency was computed for task scores, 3 facets of performance scores, and the total scale score from all 995 participants using Cronbach alpha coefficients.

Investigating Performance Facets

Performance facets were investigated using both factor analysis and Rasch analysis. Exploratory factor analysis was conducted for all task scores using direct Oblimin rotation to allow for correlated factors. Oblique rotation was applied in recognition of the inevitable intercorrelations among scores. We expected that both GP and LOA were likely to affect the time taken for an individual to complete a task.

The Rasch model was used for the item response theory (IRT) analysis. The Rasch model focuses on only 1 of several possible task score parameters—that of the task difficulty. For tasks with rubric scores, the Rasch model is expanded to address the difficulty of moving from a lower score to a higher score on a rubric, referred to as "step difficulties."

Infit and outfit statistics were used to evaluate the fit of a model to the data. Infit statistics were used to indicate how well the model fit the data for participants whose abilities were similar to the level of difficulty for that task. Outfit statistics indicated how well the model fit the data for all examinees, even those with abilities very different from the difficulty of task scores. Fit statistics that ranged between 0.70 and 1.3 were considered adequate fit to the data, that is, indicated the degree to which scores "fit together" in a single measure.[25] A less conservative range of 0.5 to 1.5 also has been recommended.[26]

The first stage of IRT analyses was intended to ascertain whether a partial credit model (PCM)[27] or a rating scale model (RSM) was a better fit to the data.[28,29] The PCM allows the parameters from all task scores to freely vary. In contrast, the RSM assumes that, when rating scales have the same meaning across items (eg, all time level scores), the distances between score levels are the same across items. For example, the difficulty of moving from a time-level score of 0 to a time-level score of 1 (or from time level 1 to time level 2) is the same across all tasks.

To assess the performance facets, all item scores (time, GP, LOA) for each task were entered into the same PCM or RSM analyses. Fit statistics were evaluated to select the model that best fit the data. Once the best-fitting model was selected, performance facet scores were analyzed in separate analyses to evaluate whether a single facet or a multifaceted model of performance better fit the data. Finally, correlations were used to examine the relationship between the original DGI and the mDGI.

Comparisons of Mobility-Impaired and Control Groups

Internal evidence for the validity of mDGI scores was determined by comparing performance scores in the adults with mobility impairments and the control cohort using multiple *t* tests. We expected that the adults with mobility impairments would have consistently lower scores for all score categories (task, performance facet, and total scale score). Finally, the minimal detectable change (MDC) was calculated for mDGI total score and for each of the 3 performance facet scores for the control cohort and the mobility-impaired group using the reliability estimate for the alpha coefficients for each group.

Results

Participants

Forty-two physical therapists from 17 sites (5 university-affiliated departments, 5 medical centers, 3 rehabilitation hospitals, and 4 outpatient centers) agreed to collect data. The mean age of the therapists was 37 years (range=24–64), 85% were female, and 24% were board-certified clinical specialists (Neurology Certified Specialists=8, Geriatric Certified Specialists=3). A total of 995 participants (140 controls and 855 patients) participated. As shown in Table 2, the mean age of the entire sample was 65.3 years (range=15–99), 54.5% were female, and 84.9% were white. Half of the sample used assistive devices when walking, and almost half reported 1 or more falls in the previous 6 months. The sample included individuals with a wide range of diagnoses, with the largest group being those with stroke.

It is important to note that sample sizes were comparable across 7 of the 8 mDGI tasks. However, 100 cases were lost in the analyses of the stairs task because not all clinics had 10 steps available for the standardized conditions of the task. Analyses

Table 2.
Sociodemographics of Sample (N=995)[a]

Measure	Data
Age (y)	
$\overline{X}$	65.3
SD	18.0
Range	15–99
Sex, n (%)	
Female	542 (54.5)
Male	453 (45.5)
Ethnicity, n (%)	
White	845 (84.9)
African American	80 (8.0)
Latino	40 (4.0)
Asian	26 (2.6)
Other	2 (0.2)
Diagnosis, n (%)	
Stroke	239 (24.0)
Vestibular dysfunction	140 (14.1)
Closed head injury, traumatic brain injury, brain tumor	100 (10.1)
Gait abnormality, imbalance, falls, geriatric weakness	91 (9.1)
Parkinson disease	84 (8.4)
Neuropathy	41 (4.1)
Multiple sclerosis	31 (3.1)
Brain tumor	24 (2.4)
Spinal cord injury	22 (2.2)
Cerebellar ataxia	10 (1.0)
Other	97 (9.7)
Control cohort	140 (14.1)
Gait device, n (%)	
None	446 (47.4)
Cane	213 (22.7)
Walker	281 (29.9)
Falls, n (%)	
None	446 (47.4)
1	213 (22.7)
2 or more	281 (29.9)

[a] Subtotals of less than 995 are due to missing data.

for the stairs task and total scores reflect this decrease in number of cases.

Conversion of Time to Ordinal Scores

Prior to conducting the performance facet analysis, the interval time scale was converted to an ordinal scale that could be incorporated into factor analyses and IRT analyses. Table 3 summarizes the distribution of time and speed for the ordinal scores related to time across the eight DGI tasks. Time ranges associated with the ordinal score of 3 are fairly consistent across all tasks, with the exception of changing pace. Task-specific differences in time ranges (and speed) were present for the lower ordinal scores (0–2).

Reliability

Interrater agreement at the task level. Before the analysis related to performance facets could be conducted, it was important to obtain evidence to support the reliability of task scores through interrater agreement analysis. For all tasks, with the exception of stairs, the percent exact agreement between physical therapists and the criterion scores was high for both time and LOA, ranging from 94% to 100%. Percent exact agreement for the stairs task was 75% for time and 97% for LOA. Percent exact agreement was lower, although still acceptable, across all tasks for GP scores, ranging from 71% to 91%. Raters consistently rated GP slightly higher than the criterion scores. Kappa coefficients also were high for all 3 facets of scoring. Kappa coefficients ranged from .90 to .98 for time, from .59 to .88 for GP, and from .84 to 1.00 for LOA. These data suggest that task scores were sufficiently reliable to be used to investigate facets of performance of mDGI.

Interrater agreement of performance facet scores. Interrater agreement also was strong for all 3 performance facet scores. For time total scores, the mean criterion score was 14.0 (SD=8.61) compared with the raters' mean score of 14.1 (SD=8.3); the correlation between these scores was r=1.0. For the GP total score, the criterion mean was 15.8 (SD=7.2) compared with the raters' mean score of 16.96 (SD=6.2); the correlation between these scores was r=.94. The average GP total score for the physical therapists was 1.2 points higher than the mean criterion GP total score, and the stan-

Table 3.
Cutpoints (Time and Speed) for Time Levels for the 8 Dynamic Gait Index Tasks

Measure	Usual Pace	Change Pace	Horizontal Head Turns	Vertical Head Turns	Pivot Turn	Stepping Over Obstacles	Stepping Around Obstacles	Stairs
Time (s)								
3	<6.0	<4.9	<6.2	<6.0	<6.9	<6.0	<6.0	<6.1
2	6.0–7.6	4.9–6.8	6.2–8.5	6.0–8.2	6.9–9.4	6.0–8.5	6.0–8.2	6.1–9.0
1	7.7–15.2	6.9–11.7	8.6–14.5	8.3–13.9	9.5–16.9	8.6–17.4	8.3–14.5	9.1–19.7
0	>15.2	>11.7	>14.5	>13.9	>16.9	>17.4	>14.5	>19.7
Speed (m/s)								
3	>1.01	>1.24	>0.98	>1.01	>0.88	>1.02	>1.03	>1.0
2	0.8–1.01	0.90–1.24	0.71–0.98	0.74–1.01	0.65–0.88	0.72–1.02	0.74–1.03	0.68–1.0
1	0.4–0.79	0.52–0.89	0.42–0.70	0.44–0.73	0.36–0.64	0.35–0.71	0.42–0.73	0.31–0.67
0	<0.4	<0.52	<0.42	<0.44	<0.36	<0.35	<0.42	<0.31

dard deviation of the physical therapists' GP scores was 1 point smaller than that of the criterion GP scores. It appears from these data that the physical therapists judged participants to have better GPs with narrower distribution of scores compared with the criterion scores. Finally, the criterion mean score for LOA was 12.56 (SD=4.94), and the raters' mean score was 12.91 (SD=4.62) with a correlation of r=.90.

Test-retest reliability. To investigate test-retest reliability, 257 participants were retested an average of 3.65 days (SD=5.20) after initial testing. The Pearson correlation for test-retest scores for time, GP, and LOA was .91, .91, and .87, respectively. The test-retest correlation for the DGI total score was .92. Test-retest correlations for task total scores ranged from .86 to .90. These data suggest that scores for the mDGI were quite stable over time.

Performance Facet Analyses

Factor analysis. Table 4 presents the factor pattern matrix from the factor analysis. Three factors with eigenvalues greater than 1.3 explained 79% of the variability in scores. Factor 1 included all of the time-level scores (shaded scores), fac-

Table 4.
Factor Pattern Matrix for Modified Dynamic Gait Index Task Scores

Measure	Time	Assistance	Gait Pattern
Usual pace, time level	.951	.023	−.022
Usual pace, gait pattern	.126	−.058	.714
Usual pace, level of assistance	.021	−.974	−.075
Change pace, time level	.933	−.003	−.014
Change pace, gait pattern	.201	−.045	.656
Change pace, level of assistance	.011	−.993	−.075
Horizontal head turns, time level	.887	.003	.028
Horizontal head turns, gait pattern	−.096	.044	.912
Horizontal head turns, level of assistance	−.066	−.943	.061
Vertical head turns, time level	.905	.011	.035
Vertical head turns, gait pattern	−.096	.014	.926
Vertical head turns, level of assistance	−.063	−.930	.071
Pivot turn, time level	.890	.013	.033
Pivot turn, gait pattern	.107	−.099	.666
Pivot turn, level of assistance	−.051	−.969	.021
Stepping over obstacles, time level	.913	−.023	.007
Stepping over obstacles, gait pattern	.341	−.142	.453
Stepping over obstacles, level of assistance	.114	−.806	.012
Stepping around obstacles, time level	.933	−.010	−.019
Stepping around obstacles, gait pattern	.136	−.039	.684
Stepping around obstacles, level of assistance	−.024	−.980	−.034
Stairs, time level	.800	−.068	.026
Stair, gait pattern	.392	−.156	.295
Stairs, level of assistance	.117	−.698	.080

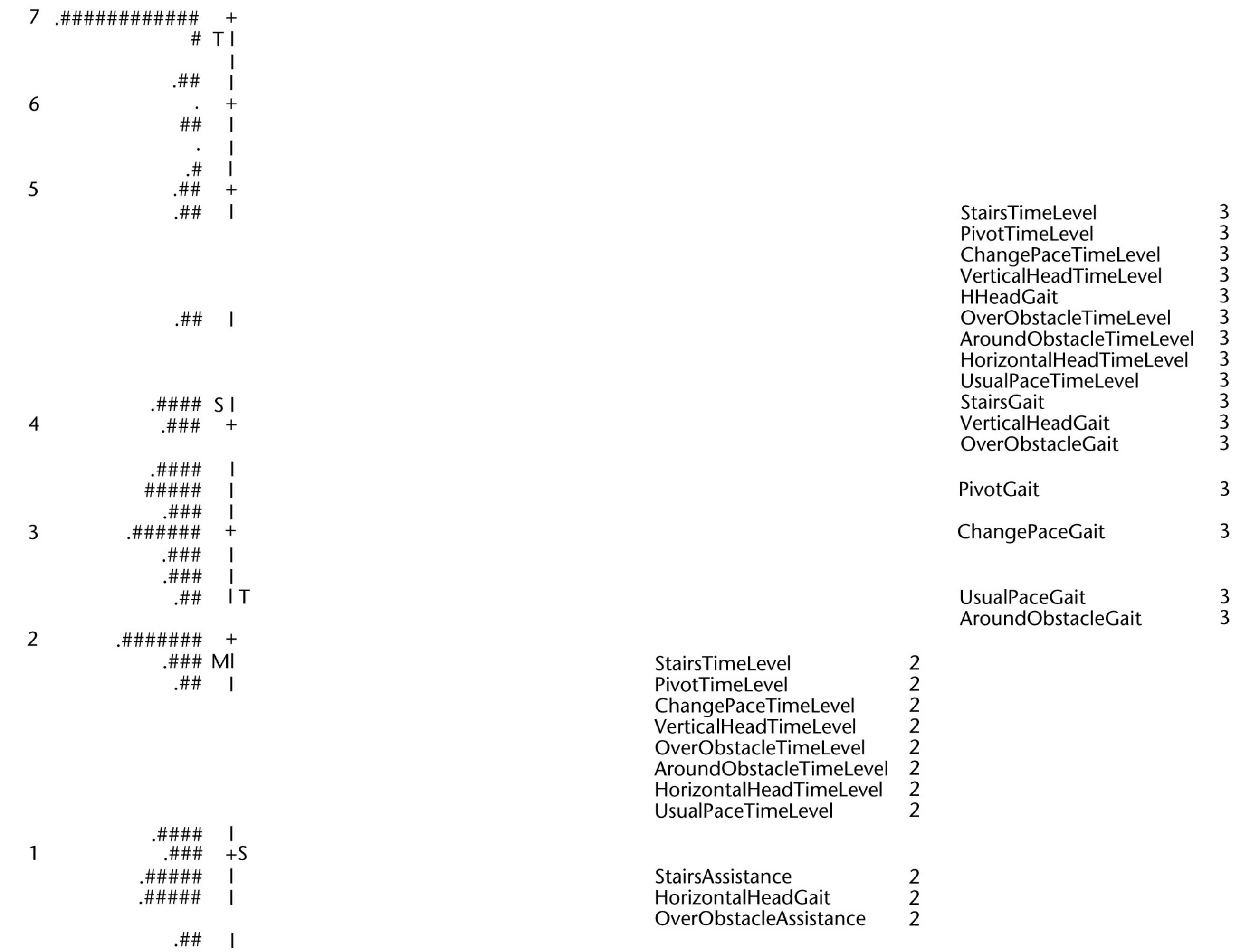

Figure.
Rasch person-item map. The underlying scale is represented as a vertical column of plus signs and vertical bars. Plus signs represent integer locations on the underlying scale. Vertical bars represent locations 0.25, 0.50, and 0.75 above whole positive numbers or below whole negative numbers. The scale ranges from negative 6.0 to 7.0. Item names are on the right-hand side of the scale. Persons are on the left-most column, with 7 people designated by a pound sign. Each location on the scale represents the point on the scale in which examinees have an equal probability of adjacent scores. For example, location −2.75 is the point on the scale where examinees have an equal probability of scoring 0 or 1 on the horizontal head turns gait pattern.

tor 2 included all of the LOA scores (shaded scores), and factor 3 included all the GP scores (shaded scores) except for scores from the stairs and the over obstacles tasks. Correlations among the factors were moderate. The Pearson correlation between time-level scores and LOA scores was .61; the correlation between time-level scores and GP scores was .70, and the correlation between GP and LOA scores was .57. After rotation, each factor explained approximately 11% to 13% of unique variance. These results suggest that the mDGI measured 3 different but correlated facets of performance. The unique and relatively equal variance explained by each facet suggested support for an additive model for generating the mDGI total score.

Rasch analysis. To evaluate relative difficulty of task scores and score dimensionality, all scores were analyzed together in a single analysis. Two strategies were used during the analysis. First, item scores for GP, time, and LOA were grouped separately prior to analysis so that an RSM could be established for each performance facet. For the second analysis, step parameters were allowed to vary across all scores using a PCM. During this stage of analysis, it was determined that an RSM was a better fit to the data; therefore, all subsequent analyses reflect an RSM.

The Figure presents the relative difficulty of each score when all facets of performance are combined into a

```
 0       ####  +M                                        StairsGait                 2
                                                         HorizontalHeadAssistance   2
                                                         VerticalHeadGait           2
                                                         VerticalHeadAssistance     2
                                                         OverObstacleGait           2
       .#####   |                                        PivotAssistance            2
                                                         ChangePaceAssistance       2
                                                         AroundObstacleAssist       2
       .#####   |                                        UsualPaceAssist            2
        ##### S |                                        PivotGait                  2
-1        .##  +S  StairsAssist                   1
         .###   |                                        ChangePaceGait             2
           .#   |  OverObstacleAssistance         1
         .###   |  HorizontalHeadAssistance       1
                   VerticalHeadAssistance         1
                   StairsTimeLevel                1
                   PivotAssistance                1      UsualPaceGait              2
-2         .#   +  PivotTimeLevel                 1      AroundObstacleGait         2
                   ChangePaceTimeLevel            1
                   VerticalHeadTimeLevel          1
                   OverObstacleTimeLevel          1
                   AroundObstacleTimeLevel        1
                   ChangePaceAssistance           1
                   HorizontalHeadTimeLevel        1
                   AroundObstacleAssistance       1
                   UsualPaceTimeLevel             1
                   UsualPaceAssistance            1
          .##   | T
           .#   |  HorizontalHeadGait             1
            .   |  StairsGait                     1
-3         .#   +  VerticalHeadGait               1
            .      OverObstacleGait               1
            .  T |
            .   |  PivotGait                      1
                |
-4          .   +
            .   |  ChangePaceGait                 1
            .   |
                |  UsualPaceGait                  1
                   AroundObstacleGait             1
-5              +
                |
                |
                |
-6              +
```

Figure.
Continued

single scale. In the Figure, the underlying scale is represented as a vertical column of plus signs and vertical bars. Plus signs represent integer locations on the underlying scale. Vertical bars represent locations 0.25, 0.50, and 0.75 above whole positive numbers or below whole negative numbers. The scale ranges from negative 6.0 to 7.0. Item names are on the right-hand side of the scale. Persons are on the left column, with 7 people designated by a pound sign, and a dot representing 1 to 6 people. Each location on the scale represents the point on the scale in which examinees have an equal probability of adjacent scores. For example, location −2.75 is the point on the scale where examinees have an equal probability of scoring 0 or 1 on the horizontal head turns GP score. This person and item map shows that the range of scores for the mDGI is appropriate for individuals with a wide range of mobility function and that there are very few gaps in measurement along the Rasch scale.

As shown in the Figure, when a group of scores are located within the same scale band, they are considered to have comparable level of difficulty and thus are listed together. For example, earning a score of 3 for time level on the pivot task is located in the same scale band as earning a score of 3 for time level on the change of pace, vertical head turn, and stepping over obstacles tasks, suggesting a comparable level of difficulty. In contrast, earning a score of 3 for time level on pivot task is located on a lower scale band than a score for time level on stairs, suggesting a lower level of difficulty.

The Figure also shows the added benefits of measuring 3 facets of performance. As shown in the Figure, it is more difficult to earn a score of 3 for time level on most tasks than it is to complete the task with a normal GP (score of 3). Thus, the addition of time extends the range of measurement of the original DGI (represented by the range of scores of 0–3 for GP for the 8 tasks) and, therefore, capacity to discriminate levels of mobility function. For example, among individuals who earn a score of 3 for GP on the usual pace task (item 1), the time-level score for this task could conceivably range from 0 to 3, providing further discrimination among these individuals who would have scored 3 on the original DGI usual pace item.

When all task scores were combined into a single analysis, 3 scores had infit and outfit statistics greater than 1.3 (GP scores for horizontal head turns, vertical head turns, and stepping over obstacles). Two scores had infit statistics less than 0.70 (LOA scores for usual pace and change of pace). One score (LOA on stairs) had an adequate infit statistic, but the outfit statistic was greater than 1.3. These incongruities in infit and outfit statistics suggest that, although a unidimensional scale (eg, DGI total score) is psychometrically acceptable and conceptually meaningful, further information about mobility function is gained by examining scores related to facets of performance (time, GP, and LOA), as well as individual task scores, which carry information about mobility function with respect to demands in the 4 environmental dimensions.

For the performance facet analyses, when scores for time level were analyzed separately, all task scores fit 1 RSM model except for the time-level scores on the stairs task, suggesting that individuals with impaired mobility required more time to complete this task. For GP, all task scores fit 1 RSM model. Finally, when LOA scores were analyzed together, 2 scores did not fit the RSM model: LOA stepping over obstacles and LOA on stairs, suggesting that individuals with impaired mobility require more assistance for these tasks. These results highlight the fact that environmental dimensions influence the pattern of scores within each facet of performance.

Internal Consistency Reliability for Task, Performance Facet, and Total Scores

The internal consistency of the mDGI total score was quite strong, with an alpha coefficient of .97. For the 3 performance facet scores, the alpha coefficients also were high (.97 for time and LOA and .92 for GP), suggesting that the performance facet scores were internally consistent. The alpha coefficients for the 8 DGI tasks ranged from .75 (horizontal head turns) to .85 (stepping over obstacles), suggesting moderately strong internal consistency within the tasks themselves. The fact that the alpha coefficients were not stronger supports the suggestion that the 3 facets of performance provide unique variance within each task.

Modified DGI Score Comparisons for Participants With Mobility Impairments and the Control Cohort

Table 5 compares performance on the mDGI (8 tasks, 3 performance facets, and the total score) between participants with mobility impairments and those in the control cohort. Differences in means for all scores were significantly lower for the participants with impaired mobility than for the control participants ($P<.001$ for all comparisons). Effect sizes (d) suggest that all differences were moderate ($0.50\leq d<0.80$) to large ($d\geq0.80$) except for the LOA performance score ($d=0.45$). Given the fact that an inclusion criterion for both groups was the ability to walk without the physical assistance of another on the usual pace task, the lower effect size for LOA was not surprising.

Minimal detectable change at the 95% level of confidence for each group was as follows: for the control cohort, the mDGI total score was 4, the time total score was 2, the GP total score was 2, and the LOA total score was 1. For the mobility-impaired cohort, the MDC for the mDGI total score was 5, the time total score was 2, the GP total score was 3, and the LOA total score was 2.

Relationship Between Original DGI and mDGI

Correlations (r) between the original DGI scores and the mDGI GP scores ranged from .90 to .99, with a median correlation of .97. Correlations (r) between the original DGI scores and mDGI time-level item scores ranged from .56 to .79 with a median correlation of .63. Correlations (r) between the original DGI item scores and mDGI LOA scores ranged from .49 to .68, with a median correlation of .53. The correlation (r) between original DGI total scores and mDGI GP score was .99, suggesting that the original DGI scores captured primarily GP information. Correlations (r) between original DGI total scores and mDGI time-level and LOA performance scores were .81 and .68, respectively. The Pearson correlation between DGI total scores and mDGI total scores was .92.

These analyses, including factor analysis, Rasch analysis, internal consis-

Table 5.
Description of Participants' Performance on Modified Dynamic Gait Index (mDGI)

Measure	Mobility Impaired Group					Control Group					d
	n	Mininum	Maximum	$\overline{X}$	SD	n	Minimum	Maximum	$\overline{X}$	SD	
mDGI total score	698	0	64	40.74	14.53	117	13	64	53.16	14.01	.87
Facets of performance scores											
Time score	701	0	24	12.30	6.49	117	2	24	18.44	6.34	.96
Gait pattern score	702	0	24	15.94	4.78	118	5	24	20.10	5.16	.84
Level of assistance score	705	0	16	12.52	5.08	118	3	16	14.48	3.38	.45
Task scores											
Usual pace	857	0	8	5.55	1.73	138	2	8	6.84	1.59	.78
Change pace	856	0	8	5.38	1.88	138	1	8	6.83	1.61	.83
Horizontal head turns	856	0	8	4.95	1.91	138	1	8	6.46	1.93	.79
Vertical head turns	857	0	8	5.08	1.92	138	1	8	6.46	2.08	.69
Pivot turn	857	0	8	5.22	1.89	138	1	8	6.53	1.89	.69
Stepping over obstacles	854	0	8	5.03	2.23	136	0	8	6.48	2.33	.64
Stepping around obstacles	856	0	8	5.53	1.80	138	2	8	6.92	1.50	.84
Stairs	702	0	8	4.75	2.19	118	0	8	6.31	2.36	.69

tency, and correlation between original DGI and mDGI scores, suggest that, although the original DGI item scoring rubrics referred to time and LOA, the original DGI scores may not have captured these 2 aspects of performance. Instead, the original DGI scores appear to have captured primarily information on GP.

Discussion

The purpose of this study was to develop and examine the psychometric properties of a new scoring system for the DGI. We expected that expanding the parameters for evaluating performance would enable clinicians to reliably monitor change in mobility in patients with a wider range of function and to quantify improvement or deterioration in any 1 of the 3 facets of performance (ie, GP, LOA, and time taken to complete the 8 walking tasks). The factor analysis identified 3 separate factors—time, LOA and GP—that measured unique but correlated aspects of walking performance. Rasch analyses verified that the mDGI allowed measurement of performance over a wider range of mobility function than the original DGI and supported the hierarchical structure of the item tasks. The Rasch analysis shows good measurement at all locations on the underlying scale. Previous studies suggest that the original DGI scores resulted in a ceiling effect for individuals with mobility impairments who were relatively high functioning, as well as fairly sizeable breaks between scores, suggesting gaps in measurement.[6–8,15] The original purpose for developing a revised scoring system was to improve measurement capability of the DGI by eliminating both the ceiling effects and the gaps in measurement. The analyses presented suggest that these goals have been achieved in the mDGI. The mDGI showed sound psychometric properties, including strong interrater reliability, strong test-retest reliability, good internal consistency, and evidence for discriminative validity at the individual task, performance, and total score levels.

The major strength of this study was the large and diverse sample of participants, enabling the reliable use of Rasch analysis to validate a new scoring system for the DGI. A total of 995 adults participated (855 patients with neurologic deficits currently receiving physical therapy for mobility deficits and 140 individuals without neurologic impairment). Only 2 studies have used Rasch and factor analyses to investigate properties of the original DGI. Chiu and colleagues[6] used a Rasch analysis to validate the 4 rating scale categories and to verify the single construct and hierarchical order of the scale. A limitation of that study was the relatively small and homogeneous sample (140 male veterans with impaired balance); thus, there were not sufficient cases for consideration of step parameters in their Rasch difficulty estimations. This may have been the rationale for collapsing the DGI task scoring categories from 4 to 2 levels. The study by Marchetti and Whitney[15] included 123 participants with balance and vestibular disorders and 103 control participants to construct and validate a 4-item DGI using Rasch and factor analyses. That study also was limited by the relative homogeneity and size of the samples, which may have been the reason that only location parameters (average of the step difficulties) for the tasks were used. The current study has addressed both of these limitations through the heterogeneity of the samples (participants representing a variety of mobility disorders as well as participants without neurologic impairment) and the size of the samples, making the item and step difficulty estimates quite stable.

Rating Scale

The current study suggests that the rating scale developed for the original DGI captures the performance facet of GP alone. In contrast, the modified scoring system, which includes time, GP, and LOA, measures distinct but correlated facets of performance that operate together as an individual completes each task. These 3 facets of performance, although distinct, are nonetheless correlated and thus can be incorporated into a single scale. Individually, each performance facet contributes about 11% of unique variance, and together they explain about 80% of the total variance in scores. The new scoring system will enable clinicians to measure changes in 3 aspects of walking function: the GP used for a walking task, the LOA needed, and the time required to complete each of the tasks. Measuring time and level of assistance in addition to GP is important in light of research indicating that for many patients, improvement in walking function is not always associated with changes in underlying GP.[30,31] Thus, locomotor rehabilitation may improve some aspects of walking performance, such as time or LOA, but not others, such as the pattern used to walk. The new scoring system for the mDGI also will allow clinicians and researchers to examine the association between changes in these 3 facets of performance and changes in walking function. Finally, expanding the facets of measurement expands the range of ability measured by the scale, thus reducing the likelihood of a ceiling effect.

Unidimensional and Hierarchical Order of Tasks

The studies by Chiu et al[6] and Marchetti and Whitney[15] determined that the original DGI was unidimensional and measured a single factor referred to as "dynamic balance." These studies reported a hierarchical order of task difficulty, using the performance facet of GP. The specific order of tasks based on difficulty varied slightly, probably due to differences in samples tested. Both studies confirmed that the 3 most difficult tasks were walking with horizontal head turns, walking up and down stairs, and walking with vertical head turns; the order of remaining tasks varied by study. The 3 least difficult tasks were walking on level surfaces,[6] walking around obstacles,[15] and walking and changing speeds.[6,15] Both studies recommended administrative changes in the DGI, including shortening the test to eliminate psychometric redundancy and the need for equipment or reordering and possibly eliminating some tasks based on both psychometric redundancy and hierarchical ordering.

Although the DGI was originally designed to capture the construct of dynamic balance, it was based on a conceptual framework that recognized the multidimensional nature of dynamic balance, including the ability to adapt the sensory, motor, and cognitive systems controlling balance and gait to specific demands within distinct environmental dimensions.[1,2] The 8 tasks in the DGI were designed to assess the capacity to adapt gait to demands within 4 environmental dimensions (temporal, postural transitions, terrain, and density) and thus to specify the constellation of environmental dimensions contributing to mobility disability in an individual patient. This information was considered critical to developing targeted interventions designed to reduce or reverse mobility disability. Item response theory identifies score patterns based on level of difficulty; however, redundancy in measurement is not the same as redundancy in construct. Just as a mathematics test must assess all dimensions of the construct of mathematics (eg, algebra, geometry, probability, and statistics) to provide diagnostic information about students, each of the DGI tasks was designed to assess walking in relationship to different environmental dimensions, thus providing insight into the determinants of mobility disability within an individual patient. Although there may be redundancy in the level of Rasch difficulty of the scores, it does not mean that tasks or performance scores should be dropped. If 2 tasks demonstrate the same level of difficulty but capture different dimensions of gait adaptation, they should be retained because unique patterns of scores can help therapists identify an individual patient's response in the different environmental dimensions that affect walking performance.

It is important to note that the Rasch step parameters represent theoretical rather than empirical results and the item difficulties found in this study and previous studies represent expected rather than absolute differences in difficulty for task scores. An individual patient could have a score pattern that deviates from the theoretical model, which is an important distinction when selecting the appropriate therapy and measuring the success of therapy, particularly given the environmental dimensions of mobility function measured by mDGI.

In addition, score patterns may be unique for different populations, based on the specific constellation of sensory, motor, and cognitive impairments that limit adaptations to specific task and environmental demands. Previous research has shown individuals with impaired mobility report avoiding specific features in the environment perceived as challenging to safety when walking in the community.[32,33] However, level of avoidance varies by environmental dimension; thus, not all aspects of the environment are perceived as equally challenging to safe walking. Among people with stroke, avoidance in the temporal and postural dimensions was significantly associated with frequency of community walking.[32] In contrast, avoidance in the distance, temporal, terrain, and postural dimensions was associated with frequency of participation in community walking in older adults with mobility disability.[33]

Even the fit statistics reflect the uniqueness of different tasks in time, GP, and LOA needed. The pattern of fit statistics for the unidimensional Rasch analysis can be explained, in part, by the different environmental dimensions represented within the DGI. The LOA scores for 2 tasks (usual pace and change pace) had fit statistics less than the criterion of .70. These tasks reflect the temporal dimension of the DGI. Because of the nature of these tasks, they require less assistance than any of the other tasks in the other environmental dimensions. The GP scores for vertical head turns, horizontal head turns, stepping over obstacles, and stairs tasks had fit statistics greater than the criterion of 1.30. These tasks reflect the postural, density, and terrain dimensions of the DGI. The GP for each of these tasks would be expected to differ from the GPs used in the other 4 tasks. The pattern of fit statistics for the Rasch analyses of the 3 facets of performance scores also pick up the unique contributions of each task. Within the time facet, time-level scores on the stairs task had infit and outfit statistics that suggest the stairs task affects time-level scores differently than other tasks. For the Rasch analysis of the LOA scores, fit statistics for the stepping over obstacles task and for the stairs LOA differ for tasks measuring the terrain environmental dimension. These results provide support for retaining the unique contributions of the different tasks from the original DGI in the mDGI.

Limitations

The main limitations in this study were the uneven sample sizes between the 2 groups and the sampling process. Individuals in the control cohort were drawn from among volunteers between the ages of 15 and 99 years, with no known neurologic diagnosis. The sample size of 140 was relatively small to be truly representative of this population. Participants with mobility impairments were recruited from patients with neurologic diagnoses actively receiving physical therapy from clinical sites already using the DGI as part of their clinical assessment protocols. Therefore, these patients may not represent the broader population of individuals with mobility impairments. Another limitation was that a number of clinical sites did not have 10 steps for the stair task; therefore, 100 cases were lost in the analyses of the stairs task and the total score analyses.

Future Research

The next phase of analysis in this study will focus on differential analysis of the mDGI on the different sample populations to determine whether order of score difficulty varies in patients with different pathologies. Future research is needed to determine the responsiveness of the new measure and the relationship between the mDGI and fall risk.

Conclusions

The mDGI retains the 8 items from the original test but modifies the scoring system, enabling measurement over a broader range of mobility function and allowing clinicians to characterize changes in 3 facets of performance: time, LOA, and GP. The expanded range and specificity of measurement should improve the utility of the DGI as an outcome measure. We believe the strength of the psychometric properties of the mDGI warrant its adoption for both clinical and research purposes.

All authors provided concept/idea/research design. Dr Shumway-Cook, Dr Taylor, and Dr Matsuda provided writing and data analysis. Dr Shumway-Cook, Dr Matsuda, Mr Studer, and Dr Whetten provided data collection. Dr Shumway-Cook provided fund procurement. Mr Studer and Dr Whetten provided study participants. Mr Studer provided facilities/equipment. Dr Matsuda provided institutional liaisons. Dr Taylor provided consultation (including review of manuscript before submission). The authors thank the testing sites, physical therapists, and all participants for their contributions to this study.

All procedures were approved by the University of Washington Human Subjects Division.

A summary of this research was presented in poster format at the Combined Sections Meeting of the American Physical Therapy Association; January 21–24, 2013; San Diego, California.

This study was supported by a grant from the Walter C. and Anita C. Stolov Research Fund, University of Washington.

DOI: 10.2522/ptj.20130035

References

1 Shumway-Cook A, Woollacott M. Motor Control: *Translating Research Into Clinical Practice*. 4th ed. Baltimore MD: Lippincott Williams & Wilkins; 2011.

2 Patla A, Shumway-Cook A. Dimensions of mobility: defining the complexity and difficulty associated with community mobility. *J Aging Phys Act*. 1999;7:7–19.

3 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.

4 Badke MB, Sherman J, Boyne P, et al. Tongue-based biofeedback for balance in stroke: results of an 8-week pilot study. *Arch Phys Med Rehabil*. 2011;92:1364–1370.

5 Herman T, Inbar-Borovsky N, Brozgol M, et al. The Dynamic Gait Index in healthy older adults: the role of stair climbing, fear of falling, and gender. *Gait Posture*.2009; 29:237–241.

6 Chiu Y, Fritz SL, Light KE, Velozo CA. Use of item response analysis to investigate measurement properties and clinical validity and data for the Dynamic Gait Index. *Phys Ther*. 2006;66:779–787.

7 Pardasaney PK, Latham NK, Jette AM, et al. Sensitivity to change and responsiveness of four balance measures for community dwelling older adults. *Phys Ther*. 2012;92: 388–397.

8 Jonsdottir J, Cattaneo D. Reliability and validity of the Dynamic Gait Index in persons with chronic stroke. *Arch Phys Med Rehabil*. 2007;88:1410–1415.

9 Lin JH, Hsu MJ, Hsu HW, et al. Psychometric comparisons of three functional ambulation measures for patients with stroke. *Stroke*. 2010;41:2021–2025.

10 Huang S-L, Hsieh C-L, Wu R-M, et al. Minimal detectable change of the Timed "Up & Go" Test and the Dynamic Gait Index in people with Parkinson disease. *Phys Ther*. 2011;91:114–121.

11 McConvey J, Bennett SE. Reliability of the Dynamic Gait Index in individuals with multiple sclerosis. *Arch Phys Med Rehabil*. 2005;86:130–133.

12 Cattaneo D, Jonsdottir J, Repetti S. Reliability of four scales on balance disorders in persons with multiple sclerosis. *Disabil Rehabil*. 2007;29:1920–1925.

13 Cattaneo D, Regola A, Meotti M. Validity of six balance disorders scales in persons with multiple sclerosis. *Disabil Rehabil*. 2006;28:789–795.

14 Wrisley DM, Walker ML, Echternach JL, Strasnick B. Reliability of the Dynamic Gait Index in people with vestibular disorders. *Arch Phys Med Rehabil*. 2003;84:1528–1533.

15 Marchetti GF, Whitney SL. Construction and validation of the 4-item Dynamic Gait Index. *Phys Ther*. 2006;86:1651–1660.

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

17 Dibble LE, Lange M. Predicting falls in individuals with Parkinson disease: a reconsideration of clinical balance measures. *J Neurol Phys Ther*. 2006;30:60–67.

18 Said CM, Goldie PA, Culham E, et al. Control of lead and trail limbs during obstacle crossing following stroke. *Phys Ther*. 2005;85:413–427.

19 Perry J, Garrett M, Gronley JK, Mulroy SJ. Classification of walking handicap in the stroke population. *Stroke*. 1995;26:982–989.

20 Bohannon RW. Comfortable and walking speed of adults aged 20–79 years: reference values and determinants. *Age Aging*. 1997;26:15–19.

21 Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. *J Am Geriatr Soc*. 2006;54:743–748.

22 Schmid A, Duncan PW, Studenski S, et al. Improvements in speed-based gait classifications are meaningful. *Stroke*. 2007;38: 2096–2100.

23 Studenski S, Perera S, Patel K, et al, Gait speed and survival in older adults. *JAMA*. 2011;305:50–58.

24 Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960;20:37–46.

25 Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Measurement Transactions*. 1994;8:370–376.

26 Linacre JM. Winsteps help for Rasch analysis: misfit diagnosis. Winsteps Manual. 2005. Available at: http://www.winsteps.com/winman/index.htm?diagnosingmisfit.htm. Accessed January 3, 2013.

27 Masters GN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47: 149–174.

28 Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43:561–573.

29 Wright BD, Masters GN. *Rating Scale Analysis*. Chicago, IL: MESA Press; 1982.

30 Den Otter AR, Geurts AC, Mulder T, Duysens J. Gait recovery is not associated with changes in the temporal patterning of muscle activity during treadmill walking in patients with post-stroke hemiparesis. *Clin Neurophysiol*. 2006;117:4–15.

31 Burke JH, Nene AV, Kwakkel G, et al. Recovery of gait after stroke: what changes? *Neurorehabil Neural Repair*. 2008;22:676–683.

32 Robinson C, Matsuda PN, Ciol M, Shumway-Cook A. Participation in community walking following stroke: the influence of self-perceived environmental barriers. *Phys Ther*. 2013;93:620–627.

33 Shumway-Cook A, Patla A, Stewart A, et al. Environmental components of mobility disability in community-living older persons. *J Am Ger Soc*. 2003;51:393–398.