

Phys Ther. 2012 March; 92(3): 388–397.

PMCID: PMC3291380

Published online 2011 November 23. doi: [10.2522/ptj.20100398](https://doi.org/10.2522/ptj.20100398)

## Sensitivity to Change and Responsiveness of Four Balance Measures for Community-Dwelling Older Adults

[Poonam K. Pardasaney](#),  [Nancy K. Latham](#), [Alan M. Jette](#), [Robert C. Wagenaar](#), [Pengsheng Ni](#), [Mary D. Slavin](#), [Jonathan F. Bean](#)

P.K. Pardasaney, PT, DPT, MS, Health and Disability Research Institute, Boston University School of Public Health, and College of Health and Rehabilitation Sciences: Sargent College, Boston University, Boston, Massachusetts.

N.K. Latham, PT, PhD, Health and Disability Research Institute, Boston University School of Public Health.


A.M. Jette, PT, PhD, FAPTA, Health and Disability Research Institute, Boston University School of Public Health.

R.C. Wagenaar, PhD, Center for Neurorehabilitation, College of Health and Rehabilitation Sciences: Sargent College, Boston University.

P. Ni, MD, MPH, Health and Disability Research Institute, Boston University School of Public Health.

M.D. Slavin, PT, PhD, Health and Disability Research Institute, Boston University School of Public Health.

J.F. Bean, MD, MPH, Spaulding Rehabilitation Hospital and Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, Massachusetts.

 Corresponding author.

Address all correspondence and reprint requests to Dr Pardasaney at Health and Disability Research Institute, Boston University School of Public Health, 715 Albany St, Talbot Building, T5W-5th Floor West, Boston, MA 02118-2526 (USA). E-mail: [poonamp@bu.edu](mailto:poonamp@bu.edu).

Received November 16, 2010; Accepted October 14, 2011.

Copyright © 2012 American Physical Therapy Association

### Abstract

#### Background

Impaired balance has a significant negative impact on mobility, functional independence, and fall risk in older adults. Although several, well-respected balance measures are currently in use, there is limited evidence regarding the most appropriate measure to assess change in community-dwelling older adults.

#### Objective

The aim of this study was to compare floor and ceiling effects, sensitivity to change, and responsiveness across the following balance measures in community-dwelling elderly people with functional limitations: Berg Balance Scale (BBS), Performance-Oriented Mobility Assessment total scale (POMA-T), POMA balance subscale (POMA-B), and Dynamic Gait Index (DGI).

#### Design

Retrospective data from a 16-week exercise trial were used. Secondary analyses were conducted on the total sample and by subgroups of baseline functional limitation or baseline balance scores.

#### Methods

Participants were 111 community-dwelling older adults 65 years of age or older, with functional limitations. Sensitivity to change was assessed using effect size, standardized response mean, and paired *t* tests. Responsiveness was assessed using minimally important difference (MID) estimates.

#### Results

No floor effects were noted. Ceiling effects were observed on all measures, including in people with

moderate to severe functional limitations. The POMA-T, POMA-B, and DGI showed significantly larger ceiling effects compared with the BBS. All measures had low sensitivity to change in total sample analyses. Subgroup analyses revealed significantly better sensitivity to change in people with lower compared with higher baseline balance scores. Although both the total sample and lower baseline balance subgroups showed statistically significant improvement from baseline to 16 weeks on all measures, only the lower balance subgroup showed change scores that consistently exceeded corresponding MID estimates.

### Limitations

This study was limited to comparing 4 measures of balance, and anchor-based methods for assessing MID could not be reported.

### Conclusions

Important limitations, including ceiling effects and relatively low sensitivity to change and responsiveness, were noted across all balance measures, highlighting their limited utility across the full spectrum of the community-dwelling elderly population. New, more challenging measures are needed for better discrimination of balance ability in community-dwelling elderly people at higher functional levels.

---

Balance, an integral component of physical function, is a fundamental area of assessment and intervention in adult rehabilitation.<sup>1</sup> Critical for normal performance of basic and advanced activities of daily living,<sup>2</sup> balance has been predictive of function, mobility, and fall risk in different patient populations and clinical settings.<sup>3–6</sup> In addition to being impaired secondary to disease and injury, a decline in balance is routinely seen with aging<sup>7,8</sup> and is a leading cause of falls in elderly people.<sup>9</sup> Given the importance of balance for function and the high morbidity and mortality associated with fall-related injuries in older adults,<sup>9</sup> considerable research has been conducted on assessment and improvement of balance in the elderly population.<sup>1,10</sup>

Clinicians and researchers routinely use standardized balance tests to diagnose balance deficits, quantify fall risk, and monitor change in balance over time. Although several balance tests are available, any given test may not necessarily address all measurement purposes in a given context.<sup>11</sup> Selection of a specific measure is ideally dependent on its appropriateness and practicality for the intended purpose and target population.<sup>11,12</sup> An appropriate measure should have essential psychometric properties, including adequate reliability, validity, measurement breadth, and minimal floor or ceiling effects, for the intended purpose and population.<sup>11,12</sup> Additionally, an appropriate measure should have adequate sensitivity to change and responsiveness when used for assessment of change<sup>12,13</sup> and adequate sensitivity and specificity when used for diagnosis.<sup>12</sup> Finally, an appropriate test should have been developed for, or tested in, the target population.<sup>12</sup>

The Berg Balance Scale (BBS),<sup>2,3,14–17</sup> Performance-Oriented Mobility Assessment (POMA),<sup>14–19</sup> and Dynamic Gait Index (DGI)<sup>14–17,20–22</sup> are among the performance-based balance measures most extensively used in older adults. Although these measures were not developed specifically for use in the community-dwelling setting, there is evidence supporting their reliability and validity in community-dwelling older adults,<sup>15,23</sup> as well as their sensitivity and specificity in predicting falls and need for clinical services.<sup>5,10,16,22,24</sup> Concern has been raised, however, regarding the potential for ceiling effects when these measures are used in community-dwelling older adults, who represent an elderly population with a higher level of functioning.<sup>6,17,18,25–27</sup> Additionally, although the BBS, POMA, and DGI have been used to assess change in balance in community-dwelling older adults,<sup>15,28</sup> their sensitivity to change and responsiveness in this population remain largely unexplored.<sup>17</sup>

*Sensitivity to change* is defined as the ability of an instrument to measure a change in state, regardless

of whether the change is relevant or meaningful to the decision-maker.<sup>13,29</sup> Although necessary, sensitivity to change has been described as insufficient for assessing change<sup>13</sup> and establishing treatment effectiveness.<sup>30</sup> *Responsiveness*, defined as the ability of an instrument to measure a meaningful or important change in a clinical state,<sup>11,13,29</sup> has been advocated as an essential property of instruments designed to measure change<sup>13,29</sup> and effectiveness of interventions.<sup>11</sup> Similar to reliability and validity, responsiveness is not considered a generalizable property and should be assessed for each population and purpose for which the measure is used.<sup>29</sup> Responsiveness is commonly reported through the minimally important difference (MID) estimate,<sup>31,32</sup> whereby a change score on a measure should equal or exceed its MID estimate to be considered important. Estimating MID of measures enhances interpretation of change scores, establishing benchmarks to help determine meaningfulness of change.<sup>30,33</sup>

Limited evidence regarding the most responsive balance measures for community-dwelling older adults<sup>5,10,14,16</sup> may at least partially explain the frequent use of a combination of tests to assess change in balance in this population.<sup>15,34,35</sup> Because the median time to complete most standardized balance tests is 15 minutes,<sup>14</sup> administration of multiple tests can be expected to place considerable burden on both patients and clinicians and result in inefficiency in clinical and research settings. A comparison of commonly used balance measures within a single sample of community-dwelling older adults is needed to allow valid conclusions regarding their relative superiority and guide test selection by clinicians and researchers.

In the present study, we sought to compare psychometric performance of 4 balance measures widely used in the community-dwelling elderly population and fill the gap in knowledge regarding their measurement breadth, sensitivity to change, and responsiveness. To our knowledge, performance of these measures has not been compared previously in a single community-dwelling elderly sample. Specifically, the aim of our study was to compare floor and ceiling effects, sensitivity to change, and responsiveness across the BBS, POMA total scale (POMA-T), POMA balance subscale (POMA-B), and DGI in a sample of community-dwelling elderly people with limitation of physical function. A sample with functional limitations was chosen to represent a population of community-dwelling older adults likely to seek physical therapy services.

## Method

### Sample and Design

The study involved secondary analysis of data from a 16-week, single-blinded randomized controlled trial comparing 2 supervised exercise programs, conducted at 2 outpatient rehabilitation facilities in the greater Boston area.<sup>36</sup> All participants gave written informed consent prior to participation in the trial. Inclusion criteria were age of 65 years or older, Short Physical Performance Battery (SPPB) score of 10 or less to rule in functional limitation, and ability to climb a flight of stairs independently with or without a device. Exclusion criteria were a Mini-Mental State Examination score of less than 23 to rule out cognitive impairment, unstable acute or chronic disease, or a positive exercise tolerance test.<sup>36</sup> Because there were no significant differences between the 2 exercise groups in age, sex, body mass index, or improvement of balance, mobility, and self-reported function, we combined the 2 exercise groups into a single sample of 111 participants for the present study.

[Table 1](#) shows the baseline descriptive characteristics of the sample. The majority of our participants were female, 28% were 80 years of age or older, 24% retrospectively reported falls in the previous year, and 4.3% reported assistive device use. Mean body mass index indicated that our sample, on average, was overweight. Number of active medical diagnoses ranged from 1 to 14, with a mean of 5.6. The SPPB scores indicated that 55% of our sample was moderately limited in baseline function, with a mean SPPB score of 8.4. Mean balance scores were in the higher ranges, implying that our sample, on

average, had mild impairment on these tests.

## Measures

Two research assistants who received standardized training administered the measures in single sessions at baseline and upon completion of the study at 16 weeks.

**SPPB.** The SPPB is a well-established, highly reliable, and valid measure of lower-extremity and physical functioning in the elderly population and consists of 3 components: standing balance, gait speed, and repeated chair rise.<sup>37,38</sup> Balance items include standing with feet side by side, semi-tandem stance, and tandem stance, scored based on time each position is held. Gait speed is scored based on time taken to walk 4 m at usual speed. Repeated chair rise is scored based on time taken to complete 5 chair-rise repetitions. Each SPPB component is scored from 0 to 4, with higher scores indicating better function. Short Physical Performance Battery scores of 10 or less have been found to be predictive of future mobility disability, with a sensitivity of 0.69 and a specificity of 0.84.<sup>39</sup>

Using baseline SPPB scores, we divided our sample into 3 subgroups to define extent of functional limitation. Participants with a baseline score of 10 were classified as having mildly limited function, those with scores of 7 to 9 were classified as having moderately limited function, and those with scores of 5 to 6 were classified as having severely limited function.<sup>38,40</sup> These subgroups were created to allow comparison of balance measures across people with varying functional limitation.

**Balance.** Balance was assessed using the BBS,<sup>2,3</sup> POMA,<sup>19</sup> and DGI,<sup>20</sup> which have established reliability and validity for use in older adults. Higher scores indicate better balance on all measures.

The BBS<sup>2,3</sup> consists of 14 items primarily assessing transfers and static standing balance, with limited dynamic activities. Examples of BBS items include sitting balance, sit-to-stand maneuver, standing with eyes open and closed, turning 360 degrees, and single-leg stance. The BBS items are rated on a 0 to 4 scale based on performance quality, performance duration, or assistance needed, with a total score range of 0 to 56 points.

The POMA version<sup>19,41</sup> used in this study consists of a balance component comprising 9 items and a gait component comprising 7 items. Like the BBS, the POMA-B primarily assesses transfers and static activities through similar items such as sitting balance, sit-to-stand maneuver, standing with eyes open and closed, and turning 360 degrees. The primary difference between the BBS and POMA-B items is in the rating scale used. Additionally, the POMA-B assesses response to an external nudge while standing. The gait component primarily consists of observational gait analysis at usual and rapid but safe speeds. The POMA items are rated on a 0 to 1 scale or a 0 to 2 scale based on performance quality or need for assistance or devices. The POMA-T score range is 0 to 28, the POMA-B score range is 0 to 16, and the POMA gait score range is 0 to 12. The POMA-T and POMA-B scores were used in this study.

The DGI<sup>20,22</sup> is an 8-item measure of dynamic balance assessing a person's ability to adapt to various gait challenges. Examples of DGI items include changes in gait speed, head turns while walking, negotiating obstacles, and negotiating stairs. The DGI items are rated on a 0 to 3 scale based on performance quality, performance speed, or need for assistance or devices, with a total score range of 0 to 24.

## Data Analysis

Statistical Analysis Software version 9.1 (SAS Institute Inc, Cary, North Carolina) was used for analysis. Descriptive analyses were performed on demographic characteristics, number of active medical diagnoses, and baseline SPPB and balance scores. Means and standard deviations were calculated for continuous variables, and frequencies and percentages were calculated for categorical and ordinal variables. Floor and ceiling effects were computed for baseline and 16-week assessments for each balance measure. Sensitivity to change and responsiveness were examined using change scores between



baseline and 16-week assessments for each balance measure. In addition to total sample analyses, subgroup analyses were conducted to examine performance of measures across participants of varying baseline physical function or varying baseline balance scores.

**Floor and ceiling effects.** Floor and ceiling effects were calculated as the percentage of participants who achieved the minimum and maximum possible scores, respectively. Floor and ceiling effects were calculated for the total sample and for the subgroups stratified by baseline SPPB functional limitation. The McNemar test for dependent data was used to formally compare floor and ceiling effects, when present, of balance measures. Because there were 6 pair-wise comparisons among the 4 balance measures, the Holm sequential Bonferroni correction was used for significance testing to maintain a family-wise error rate of .05 for comparisons within the total sample and within each functional limitation subgroup.

**Sensitivity to change.** Our choice of sensitivity to change estimates was based on our assumption that change would be homogenous across the sample.<sup>42</sup> Thus, distribution-based statistics that assume homogenous change, including Cohen effect size (ES), standardized response mean (SRM), and paired *t* tests,<sup>42</sup> were computed for the total sample and for dichotomous subgroups using mean baseline balance scores as cutpoints. We dichotomized the sample using mean baseline balance scores to examine whether sensitivity to change of balance measures varied in individuals with lower versus higher baseline scores. Effect size and SRM express change scores in terms of the underlying sampling distribution, using standard deviation estimates.<sup>43</sup> Effect size<sup>44</sup> and SRM<sup>45</sup> are standardized indicators of power of an instrument to detect true change, with larger values indicating higher sensitivity to change.<sup>13</sup>

The Cohen ES was calculated as  $(M_2 - M_1)/S_b$ ,<sup>44</sup> and the SRM was calculated as  $(M_2 - M_1)/S_\Delta$ ,<sup>45</sup> where  $M_2$  and  $M_1$  are the mean 16-week and baseline scores, respectively;  $S_b$  is the baseline standard deviation; and  $S_\Delta$  is the standard deviation of the mean change score. An ES of 0.2 reflects small change, an ES of 0.5 reflects moderate change, and an ES of 0.8 reflects large change.<sup>44</sup> Paired *t* tests were computed to examine whether a significant change in balance had occurred from baseline to 16 weeks on each measure.

The bootstrap method<sup>46</sup> was used to test for significant differences in ES and SRM estimates, respectively, of the 4 balance measures in the total sample. One thousand bootstrap samples of the difference in ES and SRM estimates, respectively, across the 4 measures were generated. The 1,000 estimates of differences in ES and SRM values were rank ordered, and the 1-sided *P* values were equal to the rank values where the difference between measures in ES and SRM estimates, respectively, was 0. The Holm sequential Bonferroni correction was used to maintain a family-wise error rate of 0.05 across the 12 ES and SRM comparisons. The bootstrap method also was used to test whether there were any significant differences between lower and higher balance subgroups in ES and SRM estimates for each balance measure, at a .05 significance level.

Finally, we calculated 95% confidence intervals for ES and SRM estimates of each balance measure by first calculating ES and SRM estimates in each of the 1,000 bootstrap samples that were generated. The bootstrap distribution of ES and SRM estimates then was compiled, with values at the 2.5th and 97.5th percentiles in the distribution representing the 95% confidence limits.<sup>46</sup>

**Responsiveness.** Distribution-based methods, which have been recommended when anchor-based methods are not feasible,<sup>31</sup> also were used to determine the MID of each balance measure for the total sample. Additionally, because responsiveness has been found to vary with baseline scores on a measure,<sup>47,48</sup> we calculated separate MIDs for dichotomous subgroups using mean baseline balance scores as cutpoints. We selected mean baseline scores as cutpoints in order to assess whether responsiveness of the balance measures varied in individuals with lower versus higher baseline scores. The MID is considered better represented by a range of values rather than a single value because any

single estimate is associated with a degree of variation or uncertainty.<sup>43</sup> Therefore, MID values were calculated using 2 commonly used ES estimates in the literature:  $0.3 \times S_b$  and  $0.5 \times S_b$ , where  $S_b$  represents standard deviation of the baseline balance score.<sup>32,49</sup> Effect size estimates are frequently used as indicators of responsiveness.<sup>11,30</sup> Change scores obtained from paired *t*-test analyses were compared with corresponding  $0.3 \times S_b$  and  $0.5 \times S_b$  MID estimates to examine whether statistically significant differences exceeded MID estimates for each measure.

### Role of the Funding Source

Dr Bean was funded by the Dennis W. Jahnigen Scholars Career Development Award, American Geriatrics Society/Hartford Foundation, a National Institutes of Health–Mentored Clinical Scientist Development Award (K23AG019663-01A2), and the Department of Physical Medicine and Rehabilitation, Harvard Medical School.

## Results

---

### Floor and Ceiling Effects

No floor effect was noted for any balance measure. [Figures 1](#) and [2](#) display baseline and 16-week ceiling effects, respectively, for the total sample and by subgroups of baseline functional limitation. Ceiling effects were higher in the 16-week assessment and in participants with higher functional levels. At baseline, the BBS showed no ceiling effect in the subgroup with severely limited function; significantly lower ceiling effects than the POMA-T, POMA-B, and DGI in the total sample and the subgroup with moderately limited function; and a significantly lower ceiling effect than the DGI in the subgroup with mildly limited function. At baseline, both the POMA-T and POMA-B showed significantly lower ceiling effects than the DGI in the total sample. No significant differences in baseline ceiling effects were noted among the POMA-T, POMA-B, and DGI in the 3 functional limitation subgroups.

In the 16-week assessment, the BBS continued to show no ceiling effect in the subgroup with severely limited function; significantly lower ceiling effects than the POMA-T, POMA-B, and DGI in the total sample and the subgroup with moderately limited function; and significantly lower ceiling effects than the POMA-B and DGI in the subgroup with mildly limited function. In addition, in the 16-week assessment, the POMA-T showed a significantly lower ceiling effect than the POMA-B in the total sample.

### Sensitivity to Change

Total sample ES values were small, ranging from 0.27 to 0.40, whereas ES values for the lower balance subgroups were moderate to large, ranging from 0.64 to 1.60. The 95% confidence intervals for ES and SRM estimates for the total sample ranged from small to moderate, whereas 95% confidence intervals for the lower balance subgroups ranged from small to large. Finally, all 95% confidence intervals for ES and SRM estimates in the higher balance subgroups contained zero, indicating a nonsignificant effect ([Tab. 2](#)).

In the total sample, there were no significant differences in ES and SRM values across the balance measures, except for a significantly larger ES for the POMA-B compared with the POMA-T ([Tabs. 2](#) and [3](#)). Subgroup analyses revealed significantly larger ES and SRM values in lower balance subgroups compared with higher balance subgroups across all balance measures ([Tabs. 2](#) and [4](#)).

Comparison of baseline and 16-week balance scores revealed significant improvement on all measures in the total sample and lower balance subgroups ([Tab. 2](#)). However, compared with the total sample, a larger mean balance improvement was consistently seen in the lower balance subgroups. The higher balance subgroups showed no significant change on any balance measure from baseline to 16 weeks.

### Responsiveness

Across all measures, MID values were largest for the total sample and smallest for the higher balance subgroups (Tab. 2). In the total sample, when change scores from paired  $t$  tests were compared with corresponding MID values, no change score exceeded the  $0.5 \times S_b$  value, whereas the POMA-B change score marginally exceeded the  $0.3 \times S_b$  value. In the lower balance subgroups, all change scores exceeded corresponding  $0.3 \times S_b$  and  $0.5 \times S_b$  MID values. In the higher balance subgroups, no change score exceeded the  $0.3 \times S_b$  or  $0.5 \times S_b$  MID values (Tab. 2).

## Discussion

The intent of our study was to provide information to guide selection of balance tests for assessment of change in community-dwelling elderly people. Our results highlight critical limitations, including ceiling effects, relatively low sensitivity to change, and limited responsiveness of 4 balance measures used extensively in the community-dwelling elderly population. Although our sample, consisting of volunteer participants, may have been biased toward older adults with higher functional levels who are motivated to exercise, their demographic and balance characteristics are similar to those in other studies of community-dwelling older adults, supporting generalizability of our results.<sup>5,15,22,25</sup>

Although similar baseline balance characteristics have been noted in other studies of community-dwelling older adults, higher<sup>7,8,21,25</sup> and lower<sup>15,23,50</sup> balance scores also have been reported, indicating that heterogeneity in balance performance exists within this population. In addition to being a result of different inclusion criteria, the range in balance performance observed across studies may be explained by the reported negative association of balance with age,<sup>7,8</sup> assistive device use,<sup>51</sup> and history of falling.<sup>21,22</sup> Because research involving community-dwelling older adults is likely to include individuals with a range of demographic and functional characteristics, any balance measure used should have adequate sensitivity to capture the broad continuum of balance associated with this population.

Although ceiling effects on the BBS were lower than on the other measures, they were nevertheless observed, particularly in the subgroup with mildly limited function. Considering the BBS is frequently used as the criterion standard to assess balance and validate mobility and balance measures,<sup>52,53</sup> ceiling effects pose an important measurement concern when the BBS is used in community-dwelling older adults. Only the subgroup with severely limited function showed no ceiling effect on the BBS, suggesting that the measure may be more appropriate for use in community-dwelling older adults with lower levels of functioning. Compared with the BBS, the larger ceiling effects on the POMA, which shares similar items, were likely due to the difference in rating scales of the 2 measures. In contrast to the 5-point BBS rating scale, the POMA has a 2- or 3-point rating scale, with a notably lower threshold to attain the maximum score on each item. Additionally, compared with the POMA and DGI, lower ceiling effects on the BBS may be the result of its larger number of items and range of scores.

When a measure is used to capture change, high baseline scores and ceiling effects limit its ability to detect improvement between 2 assessments, posing a serious concern for type II errors in clinical trials. When the more serious risk of type II errors does not occur, outcome measures with limited sensitivity to change may falsely diminish the overall magnitude of intervention effect. A potential diminished intervention effect may be reflected in our findings, as evidenced by comparison of mean balance improvement, ES, and SRM values for the total sample and balance subgroups. Across all measures, lack of significant improvement in the higher balance subgroups notably decreased mean improvement, ES, and SRM values in the total sample compared with the lower balance subgroups. Additionally, although both the total sample and lower balance subgroups showed statistically significant change scores across all balance measures, only the lower balance subgroups consistently showed change that exceeded MID estimates. Large posttest ceiling effects on measures, as noted in our study, also may diminish the magnitude of intervention effect by limiting the ability to capture continued improvement, if present, in people at the ceiling.

Our findings of high baseline scores and ceiling effects across the balance measures are in accordance with previous concerns that these measures may be susceptible to ceiling effects in elderly people with higher levels of functioning.<sup>17,25,27</sup> Ceiling effects arise due to limited higher difficulty items in a scale, reducing the ability to distinguish among higher-performing individuals who, despite a possible difference in ability, attain the maximum score. The potential limited utility of the BBS in people with minimal balance deficits was discussed by developers of the instrument.<sup>2</sup> Inability of the BBS to detect potential true differences in balance at the high end of the scale also has been demonstrated.<sup>26</sup> Furthermore, Rasch analysis of the BBS in community-dwelling elderly men has shown a tendency for a ceiling effect due to limited items to challenge people with scores above the 45/56 range.<sup>54</sup> Rasch analysis of the DGI in community-dwelling elderly men also has revealed a lack of higher difficulty items to sufficiently challenge people with the highest balance ability.<sup>55</sup>

Our finding that ceiling effects, sensitivity to change, and responsiveness of the balance measures varied based on baseline function and balance ability is supported by the literature. Responsiveness of the BBS in people with stroke has been found to progressively decrease as time since onset of stroke increases.<sup>56</sup> Although large BBS ES values have been reported in people with acute stroke, ceiling effects and low responsiveness have been reported as early as 30 to 90 days poststroke when some recovery of function is expected to have occurred. Additionally, minimal clinically important difference estimates on outcome measures have been found to vary with baseline scores on the measure.<sup>47,48</sup>

Given that our sample consisted of older adults with functional limitations, our finding of high balance scores suggests that these measures have a low threshold for screening balance impairment, with limited ability to capture hierarchical performance in the higher ranges of balance. This limitation, however, may not be a surprising finding. Due to practical constraints on the length of standardized tests, any measure designed to assess a broad range of functional status across a variety of patients and rehabilitation settings is susceptible to having inadequate precision at a given ability level, as well as ceiling and floor effects.<sup>57</sup> Designing a single traditional, fixed-form measure that maximizes precision and minimizes ceiling and floor effects across a range of patients and settings is impractical due to the large number of items required to cover the full spectrum of ability.<sup>57</sup> Computerized adaptive testing, which has the potential to overcome limitations of lengthiness, inadequate measurement breadth, and ceiling and floor effects associated with fixed-form tests,<sup>57</sup> may offer a promising new approach to balance assessment in community-dwelling older adults.<sup>58</sup>

Our study, being a secondary analysis, had some limitations. First, although the BBS, POMA, and DGI are among the most widely used balance measures in community-dwelling older adults, other measures also are used in this population. Comparative sensitivity to change and responsiveness of other relevant balance measures for community-dwelling older adults may be explored in future research. Second, because distribution-based methods to determine MID depend on sample variability, a combination of distribution-based and anchor-based methods that include participant ratings of change may have strengthened the validity of our MID estimates. Nevertheless, distribution-based methods are widely used to determine MID<sup>30,32,49,59</sup> and have been recommended when anchor-based estimates are unavailable.<sup>31</sup> Distribution-based and anchor-based estimates of MID also have been found to show convergence.<sup>31,32,60</sup> Finally, the primary application of our MID estimates was to determine whether statistically significant changes in balance scores exceeded minimally important changes at the group level, rather than to report absolute values for individual application.

## Conclusion

Based on their ceiling effects and relatively low sensitivity to change and responsiveness, the BBS, POMA, and DGI have important limitations for balance assessment across the full spectrum of the community-dwelling elderly population. There is a need for new, more challenging balance measures for community-dwelling older adults that can capture deficits in the higher ranges of balance



performance. To reduce ceiling effects on new measures, low difficulty items that may artificially inflate scores can be replaced with higher difficulty items that allow better discrimination of balance ability. Additionally, to reduce ceiling effects, the threshold to attain the highest score on rating scales can be raised. Until new measures are developed, studies using the BBS, POMA, and DGI should consider subgroup analyses to more accurately capture balance and change in balance across individuals of varying functional status. The BBS demonstrated lower ceiling effects in community-dwelling older adults compared with the POMA and DGI.

## Footnotes

---

Dr Pardasaney, Dr Latham, Dr Jette, and Dr Bean provided concept/idea/research design. Dr Pardasaney, Dr Latham, Dr Jette, Dr Wagenaar, Dr Ni, and Dr Bean provided writing. Dr Bean and Dr Slavin provided data collection. Dr Bean provided fund procurement, participants, facilities/equipment, and clerical support. Dr Pardasaney, Dr Latham, and Dr Ni provided data analysis. Dr Pardasaney and Dr Bean provided project management. Dr Latham, Dr Jette, Dr Wagenaar, Dr Slavin, and Dr Bean provided consultation (including review of manuscript before submission).

This study was approved by the institutional review boards of Boston University and Spaulding Rehabilitation Hospital, Boston, Massachusetts.

Dr Bean was funded by the Dennis W. Jahnigen Scholars Career Development Award, American Geriatrics Society/Hartford Foundation, a National Institutes of Health–Mentored Clinical Scientist Development Award (K23AG019663-01A2), and the Department of Physical Medicine and Rehabilitation, Harvard Medical School.

This study used data from the clinical trial registered as NCT00158119.

## References

---

1. Howe TE, Rochester L, Jackson A, et al. Exercise for improving balance in older people. *Cochrane Database Syst Rev*. 2009;(4):CD004963.
2. Berg KO, Wood-Dauphinée SL, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. *Physiother Can*. 1989;41:304–310.
3. Berg KO, Wood-Dauphinée SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. *Can J Public Health*. 1992;83(suppl 2):S7–S11. [PubMed: 1468055]
4. Bohannon RW, Leary KM. Standing balance and function over the course of acute rehabilitation. *Arch Phys Med Rehabil*. 1995;76:994–996. [PubMed: 7487452]
5. Muir SW, Berg K, Chesworth B, et al. Quantifying the magnitude of risk for balance impairment on falls in community-dwelling older adults: a systematic review and meta-analysis. *J Clin Epidemiol*. 2010;63:389–406. [PubMed: 19744824]
6. Lundin-Olsson L. Community-dwelling older adults with balance impairment show a moderate increase in fall risk, although further research is required to refine how balance measurement can be used in clinical practice. *Evid Based Nurs*. 2010;13:96–97. [PubMed: 20584842]
7. Steffen TM, Mollinger LA. Age- and gender-related test performance in community-dwelling adults. *J Neurol Phys Ther*. 2005;29:181–188. [PubMed: 16388685]
8. 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. [PubMed: 11856064]
9. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. *Age Ageing*. 2006;35(suppl 2):37–41. [PubMed: 16364932]
10. Gates S, Smith LA, Fisher JD, Lamb SE. Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults. *J Rehabil Res Dev*. 2008;45:1105–1116. [PubMed: 19235113]

11. Berg KO, Norman KE. Functional assessment of balance and gait. *Clin Geriatr Med*. 1996;12:705–723. [PubMed: 8890112]
12. VanSwearingen JM, Brach JS. Making geriatric assessment work: selecting useful measures. *Phys Ther*. 2001;81:1233–1252. [PubMed: 11380279]
13. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Med Care*. 2000;38(9 suppl):84–90.
14. Perell KL, Nelson A, Goldman RL, et al. Fall risk assessment measures: an analytic review. *J Gerontol A Biol Sci Med Sci*. 2001;56:M761–M766. [PubMed: 11723150]
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. [PubMed: 8996463]
16. Scott V, Votova K, Scanlan A, Close J. Multifactorial and functional mobility assessment tools for fall risk among older adults in community, home-support, long-term and acute care settings. *Age Ageing*. 2007;36:130–139. [PubMed: 17293604]
17. Hayes KW, Johnson ME. Measures of adult general performance tests: the Berg Balance Scale, Dynamic Gait Index (DGI), gait velocity, Physical Performance Test (PPT), Timed Chair Stand Test, Timed Up and Go, and Tinetti Performance-Oriented Mobility Assessment (POMA). *Arthritis Care Res*. 2003;49(suppl 5):S28–S42.
18. Faber MJ, Bosscher RJ, van Wieringen PC. Clinimetric properties of the Performance-Oriented Mobility Assessment. *Phys Ther*. 2006;86:944–954. [PubMed: 16813475]
19. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatr Soc*. 1986;34:119–126. [PubMed: 3944402]
20. Shumway-Cook A, Woollacott MH. *Motor Control: Translating Research Into Clinical Practice*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2007:390–440.
21. 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. [PMCID: PMC2709498] [PubMed: 18845439]
22. 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. [PubMed: 9256869]
23. Mecagni C, Smith JP, Roberts KE, O'Sullivan SB. Balance and ankle range of motion in community-dwelling women aged 64 to 87 years: a correlational study. *Phys Ther*. 2000;80:1004–1011. [PubMed: 11002436]
24. Harada N, Chiu V, Damron-Rodriguez J, et al. Screening for balance and mobility impairment in elderly individuals living in residential care facilities. *Phys Ther*. 1995;75:462–469. [PubMed: 7770493]
25. 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. [PubMed: 12665404]
26. Garland SJ, Stevenson TJ, Ivanova T. Postural responses to unilateral arm perturbation in young, elderly, and hemiplegic subjects. *Arch Phys Med Rehabil*. 1997;78:1072–1077. [PubMed: 9339155]
27. Rose DJ, Lucchese N, Wiersma LD. Development of a multidimensional balance scale for use with functionally independent older adults. *Arch Phys Med Rehabil*. 2006;87:1478–1485. [PubMed: 17084123]
28. Shumway-Cook A, Gruber W, Baldwin M. Reducing the likelihood for falls in the elderly: the effects

of exercise. *J Neurol Phys Ther.* 1995;19:42–44.

29. Liang MH, Lew RA, Stucki G, et al. Measuring clinically important changes with patient-oriented questionnaires. *Med Care.* 2002;40(4 suppl):45–51.

30. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care.* 1989;27(3 suppl):S178–S189. [PubMed: 2646488]

31. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61:102–109. [PubMed: 18177782]

32. Eton DT, Cella D, Yost KJ, et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *J Clin Epidemiol.* 2004;57:898–910. [PubMed: 15504633]

33. Guyatt GH, Osoba D, Wu AW, et al. ; Clinical Significance Consensus Meeting Group Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002;77:371–383. [PubMed: 11936935]

34. Beling J, Roller M. Multifactorial intervention with balance training as a core component among fall-prone older adults. *J Geriatr Phys Ther.* 2009;32:125–133. [PubMed: 20128337]

35. Banez C, Tully S, Amaral L, et al. Development, implementation, and evaluation of an interprofessional falls prevention program for older adults. *J Am Geriatr Soc.* 2008;56:1549–1555. [PubMed: 18557964]

36. Bean JF, Kiely DK, LaRose S, et al. Increased velocity exercise specific to task training versus the National Institute on Aging's strength training program: changes in limb power and mobility. *J Gerontol A Biol Sci Med Sci.* 2009;64:983–991. [PMCID: PMC2720885] [PubMed: 19414509]

37. Guralnik JM, Simonsick EM, Ferrucci L, et al. A Short Physical Performance Battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol.* 1994;49:M85–M94. [PubMed: 8126356]

38. Guralnik JM, Ferrucci L, Simonsick EM, et al. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med.* 1995;332:556–561. [PubMed: 7838189]

39. Vasunilashorn S, Coppin AK, Patel KV, et al. Use of the Short Physical Performance Battery score to predict loss of ability to walk 400 meters: analysis from the InCHIANTI Study. *J Gerontol A Biol Sci Med Sci.* 2009;64:223–229. [PMCID: PMC2655026] [PubMed: 19182232]

40. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the Short Physical Performance Battery. *J Gerontol A Biol Sci Med Sci.* 2000;55:M221–M231. [PubMed: 10811152]

41. Shumway-Cook A, Woollacott MH. *Motor Control: Translating Research Into Clinical Practice.* 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2007:257–295.

42. Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. *Health Qual Life Outcomes.* 2005;3:23. [PMCID: PMC1084357] [PubMed: 15811176]

43. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006;86:735–743. [PubMed: 16649896]

44. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

45. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care*. 1990;28:632–642. [PubMed: 2366602]
46. Spadoni GF, Stratford PW, Solomon PE, Wishart LR. The evaluation of change in pain intensity: a comparison of the P4 and single-item numeric pain rating scales. *J Orthop Sports Phys Ther*. 2004;34:187–193. [PubMed: 15128188]
47. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Phys Ther*. 1998;78:1186–1196. [PubMed: 9806623]
48. Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. *Phys Ther*. 2011;91:675–688. [PubMed: 21372203]
49. Cella D, Eton DT, Fairclough DL, et al. What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) study 5592. *J Clin Epidemiol*. 2002;55:285–295. [PubMed: 11864800]
50. Hatch J, Gill-Body KM, Portney LG. Determinants of balance confidence in community-dwelling elderly people. *Phys Ther*. 2003;83:1072–1079. [PubMed: 14640866]
51. Donoghue D, Stokes EK; Physiotherapy Research and Older People (PROP) Group How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *J Rehabil Med*. 2009;41:343–346. [PubMed: 19363567]
52. Bennie S, Bruner K, Dizon A, et al. Measurements of balance: comparison of the Timed “Up and Go” Test and Functional Reach Test with the Berg Balance Scale. *J Phys Ther Sci*. 2003;15:93–97.
53. Matjacic Z, Bohinc K, Cikajlo I. Development of an objective balance assessment method for purposes of telemonitoring and telerehabilitation in elderly population. *Disabil Rehabil*. 2010;32:259–266. [PubMed: 19479513]
54. Kornetti DL, Fritz SL, Chiu YP, et al. Rating scale analysis of the Berg Balance Scale. *Arch Phys Med Rehabil*. 2004;85:1128–1135. [PubMed: 15241763]
55. 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. [PubMed: 16737403]
56. Mao HF, Hsueh IP, Tang PF, et al. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke*. 2002;33:1022–1027. [PubMed: 11935055]
57. Jette AM, Haley SM. Longitudinal outcome monitoring across post-acute care (PAC) settings. In: *Uniform Patient Assessment for Post-Acute Care: Final Report*. Aurora, CO: Division of Health Care Policy and Research, University of Colorado at Denver and Health Sciences Center; 2006:100–120.
58. Hsueh IP, Chen JH, Wang CH, et al. Development of a computerized adaptive test for assessing balance function in patients with stroke. *Phys Ther*. 2010;90:1336–1344. [PubMed: 20592269]
59. Allen DD. Responsiveness of the Movement Ability Measure: a self-report instrument proposed for assessing the effectiveness of physical therapy intervention. *Phys Ther*. 2007;87:917–924. [PubMed: 17504827]
60. Norman GR, Sloan JA, Wywich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41:582–592. [PubMed: 12719681]

## Figures and Tables

**Table 1.**

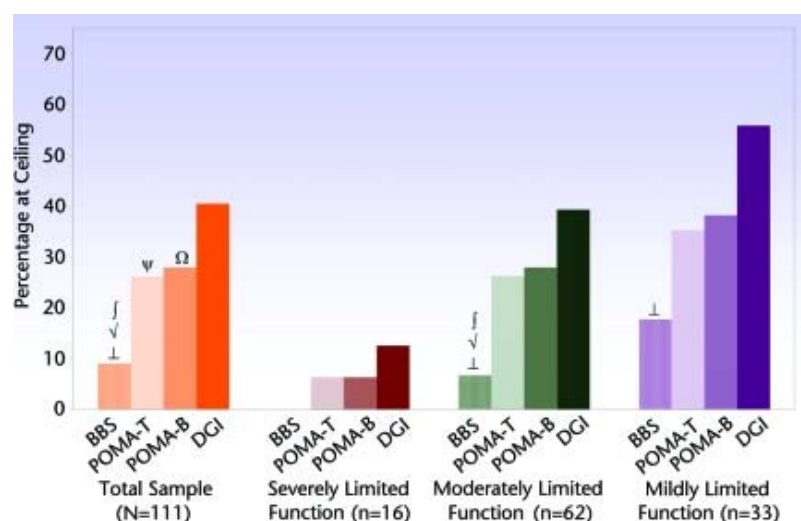


Variable	Measurement
Demographic variables	
Age (y)	75.9 (7)
Age ≥80 y	28%
Female	68.5%
Fell in previous year	24%
Assistive device user	4.3%
Body mass index (kg/m <sup>2</sup> )	27.9 (4.9)
No. of active medical diagnoses	5.6 (2.5)
SPPB	
Total score (0–12)	8.4 (1.5)
Balance subscale score (0–4)	3.1 (1)
4-m walk subscale score (0–4)	3.5 (0.7)
Chair-rise subscale score (0–4)	1.9 (0.8)
SPPB subgroups	
Mildly limited function (10)	30.6%
Moderately limited function (7–9)	55%
Severely limited function (5–6)	14.4%
Balance measures	
BBS (0–56)	50.1 (5)
POMA-T (0–28)	25.3 (3.1)
POMA-B (0–16)	14 (2)
DGI (0–24)	21 (3.7)

## Baseline Demographic Characteristics and Outcome Scores<sup>a</sup>

<sup>a</sup>Values shown are mean (SD) unless otherwise specified. N=111 for all variables except assistive device users, where n=93. SPPB=Short Physical Performance Battery, BBS=Berg Balance Scale, POMA-T=Performance-Oriented Mobility Assessment total scale, POMA-B=Performance-Oriented Mobility Assessment balance subscale, DGI=Dynamic Gait Index.

**Figure 1.**



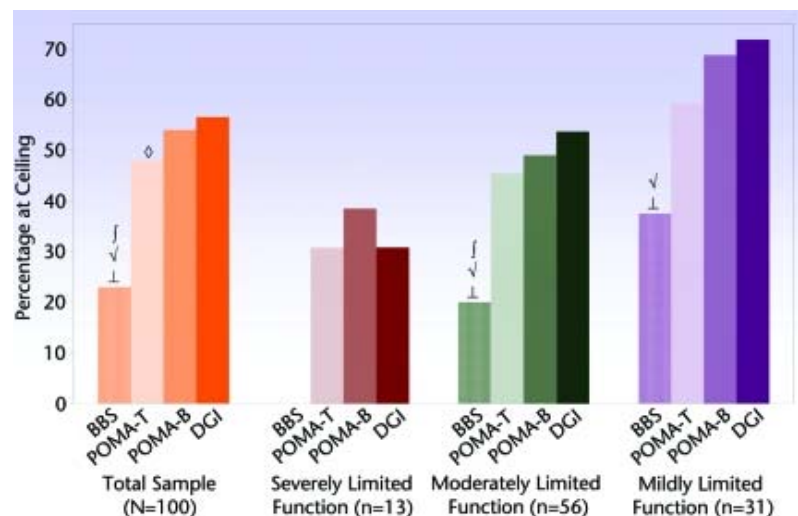
Baseline ceiling effects for the total sample and by subgroups of baseline functional limitation.

BBS=Berg Balance Scale, POMA-T=Performance-Oriented Mobility Assessment total scale, POMA-B=Performance-Oriented Mobility Assessment balance subscale, DGI=Dynamic Gait Index. Severely limited function=baseline Short Physical Performance Battery (SPPB) score of 5 to 6, moderately limited function=baseline SPPB score of 7 to 9, mildly limited function=baseline SPPB score of 10.

Significant differences: f=between BBS and POMA-T, v=between BBS and POMA-B, ⊥=between BBS

and DGI,  $\psi$ =between POMA-T and DGI,  $\Omega$ =between POMA-B and DGI.

**Figure 2.**



Sixteen-week ceiling effects for the total sample and by subgroups of baseline functional limitation. BBS=Berg Balance Scale, POMA-T=Performance-Oriented Mobility Assessment total scale, POMA-B=Performance-Oriented Mobility Assessment balance subscale, DGI=Dynamic Gait Index. Severely limited function=baseline Short Physical Performance Battery (SPPB) score of 5 to 6, moderately limited function=baseline SPPB score of 7 to 9, mildly limited function=baseline SPPB score of 10. Significant differences:  $f$ =between BBS and POMA-T,  $v$ =between BBS and POMA-B,  $l$ =between BBS and DGI,  $d$ =between POMA-T and POMA-B.

**Table 2.**

Measure	n	ES (95% CI)	SRM (95% CI)	Mean Difference <sup>b</sup> ( $S_{\Delta}$ )	MID ( $0.3 \times S_b$ )	MID ( $0.5 \times S_b$ )
BBS						
Total sample	100	0.29 (0.19, 0.44)	0.47 (0.27, 0.70)	1.4* (3.1)	1.65	2.50
Baseline BBS score <50/56	40	0.69 (0.49, 1.09)	0.85 (0.49, 1.36)	2.9* (3.5)	1.40	2.10
Baseline BBS score $\geq$ 50/56	60	0.23 (0.00, 0.52)	0.20 (-0.07, 0.55)	0.5 (2.3)	0.66	1.00
POMA-T						
Total sample	100	0.27 (0.13, 0.41)	0.40 (0.19, 0.59)	0.9* (2.1)	1.04	1.60
Baseline POMA-T score <25/28	27	0.94 (0.54, 1.37)	0.99 (0.51, 1.91)	2.6* (2.6)	0.90	1.40
Baseline POMA-T score $\geq$ 25/28	73	0.21 (-0.15, 0.51)	0.14 (-0.10, 0.39)	0.2 (1.5)	0.35	0.53
POMA-B						
Total sample	100	0.40 (0.22, 0.55)	0.44 (0.23, 0.64)	0.8* (1.8)	0.67	1.00
Baseline POMA-B score <14/16	33	1.60 (1.19, 2.42)	1.20 (0.72, 2.01)	2.2* (1.9)	0.46	0.70
Baseline POMA-B score $\geq$ 14/16	67	0.10 (-0.31, 0.49)	0.06 (-0.16, 0.37)	0.07 (1.2)	0.24	0.37
DGI						
Total sample	100	0.27 (0.16, 0.41)	0.45 (0.28, 0.61)	1.0* (2.2)	1.23	1.90
Baseline DGI score <21/24	35	0.64 (0.34, 1.22)	0.77 (0.48, 1.12)	2.3* (2.9)	1.16	1.80
Baseline DGI score $\geq$ 21/24	65	0.26 (-0.02, 0.48)	0.24 (-0.02, 0.47)	0.3 (1.2)	0.40	0.60

Sensitivity to Change and Responsiveness of Balance Measures for the Total Sample and by Subgroups of Lower and Higher Balance Scores<sup>a</sup>

<sup>a</sup>ES=effect size, CI=confidence interval, SRM=standardized response mean, MID=minimally important difference,  $S_b$ =baseline standard deviation, BBS=Berg Balance Scale, POMA-T=Performance-Oriented Mobility Assessment total scale, POMA-B=Performance-Oriented Mobility Assessment balance

subscale, DGI=Dynamic Gait Index.

<sup>b</sup> Mean difference between 16-week and baseline scores;  $S_{\Delta}$ =standard deviation of change score. Asterisk indicates significant at the .05 level using paired *t* test.

**Table 3.**

Measure	BBS	POMA-T	POMA-B	DGI
BBS		0.29	0.43	0.38
POMA-T	0.41		0.14	0.35
POMA-B	0.11	<0.001*		0.53
DGI	0.36	0.57	0.06	

Comparison of Effect Size and Standardized Response Mean Estimates of the 4 Balance Measures in the Total Sample<sup>a</sup>

<sup>a</sup>A bootstrap method was used for this analysis. Numbers above the solid line show *P* values for standardized response mean comparisons across pairs of balance measures. Numbers below the solid line show *P* values for effect size comparisons across pairs of balance measures. Asterisk indicates significant at the .004 level using the Holm sequential Bonferroni correction of .05/12. BBS=Berg Balance Scale, POMA-T=Performance-Oriented Mobility Assessment total scale, POMA-B=Performance-Oriented Mobility Assessment balance subscale, DGI=Dynamic Gait Index.

**Table 4.**

Measure	BBS	POMA-T	POMA-B	DGI
ES	.014	.002	<.001	.028
SRM	.010	<.001	<.001	.007

Comparison of Effect Size (ES) and Standardized Response Mean (SRM) Estimates Across Lower and Higher Baseline Balance Subgroups for Each Balance Measure<sup>a</sup>

<sup>a</sup>A bootstrap method was used for this analysis. Numbers shown are *P* values. All *P* values are significant at the .05 level. BBS=Berg Balance Scale, POMA-T=Performance-Oriented Mobility Assessment total scale, POMA-B=Performance-Oriented Mobility Assessment balance subscale, DGI=Dynamic Gait Index.

---

Articles from Physical Therapy are provided here courtesy of **American Physical Therapy Association**