

# Physical Therapy

Journal of the American Physical Therapy Association and



de Fysiotherapeut

Royal Dutch Society for Physical Therapy



## Is the BESTest at Its Best? A Suggested Brief Version Based on Interrater Reliability, Validity, Internal Consistency, and Theoretical Construct

Parminder K. Padgett, Jesse V. Jacobs and Susan L. Kasser

*PHYS THER.* 2012; 92:1197-1207.

Originally published online June 7, 2012

doi: 10.2522/ptj.20120056

The online version of this article, along with updated information and services, can be found online at: <http://ptjournal.apta.org/content/92/9/1197>

### Collections

This article, along with others on similar topics, appears in the following collection(s):

[Balance](#)

[Falls and Falls Prevention](#)

[Tests and Measurements](#)

### e-Letters

3 e-letter(s) have been posted to this article, which can be accessed for free at:

<http://ptjournal.apta.org/cgi/eletters/92/9/1197>

To submit an e-Letter on this article, click [here](#) or click on "Submit a response" in the right-hand menu under "Responses" in the online version of this article.

### E-mail alerts

Sign up [here](#) to receive free e-mail alerts

## Is the BESTest at Its Best? A Suggested Brief Version Based on Interrater Reliability, Validity, Internal Consistency, and Theoretical Construct

Parminder K. Padgett, Jesse V. Jacobs, Susan L. Kasser

**Background.** The Balance Evaluation Systems Test (BESTest) and Mini-BESTest are clinical examinations of balance impairment, but the tests are lengthy and the Mini-BESTest is theoretically inconsistent with the BESTest.

**Objective.** The purpose of this study was to generate an alternative version of the BESTest that is valid, reliable, time efficient, and founded upon the same theoretical underpinnings as the original test.

**Design.** This was a cross-sectional study.

**Methods.** Three raters evaluated 20 people with and without a neurological diagnosis. Test items with the highest item-section correlations defined the new Brief-BESTest. The validity of the BESTest, the Mini-BESTest, and the new Brief-BESTest to identify people with or without a neurological diagnosis was compared. Interrater reliability of the test versions was evaluated by intraclass correlation coefficients. Validity was further investigated by determining the ability of each version of the examination to identify the fall status of a second cohort of 26 people with and without multiple sclerosis.

**Results.** Items of hip abductor strength, functional reach, one-leg stance, lateral push-and-release, standing on foam with eyes closed, and the Timed "Up & Go" Test defined the Brief-BESTest. Intraclass correlation coefficients for all examination versions were greater than .98. The accuracy of identifying people from the first cohort with or without a neurological diagnosis was 78% for the BESTest versus 72% for the Mini-BESTest or Brief-BESTest. The sensitivity to fallers from the second cohort was 100% for the Brief-BESTest, 71% for the Mini-BESTest, and 86% for the BESTest, and all versions exhibited specificity of 95% to 100% to identify nonfallers.

**Limitations.** Further testing is needed to improve the generalizability of findings.

**Conclusions.** Although preliminary, the Brief-BESTest demonstrated reliability comparable to that of the Mini-BESTest and potentially superior sensitivity while requiring half the items of the Mini-BESTest and representing all theoretically based sections of the original BESTest.

P.K. Padgett, DPT, Department of Rehabilitation and Movement Science, University of Vermont, Burlington, Vermont.

J.V. Jacobs, PhD, Department of Rehabilitation and Movement Science, University of Vermont, 305 Rowell Bldg, 106 Carrigan Dr, Burlington, VT 05405 (USA). Address all correspondence to Dr Jacobs at: [jjacobs@uvm.edu](mailto:jjacobs@uvm.edu).

S.L. Kasser, PhD, Department of Rehabilitation and Movement Science, University of Vermont.

[Padgett PK, Jacobs JV, Kasser SL. Is the BESTest at its best? A suggested brief version based on interrater reliability, validity, internal consistency, and theoretical construct. *Phys Ther.* 2012;92:1197-1207.]

© 2012 American Physical Therapy Association

Published Ahead of Print:  
June 7, 2012

Accepted: May 30, 2012

Submitted: February 12, 2012



Post a Rapid Response to  
this article at:  
[ptjournal.apta.org](http://ptjournal.apta.org)

Injuries resulting from falls contribute to decreased health status and increased mortality, particularly for individuals of advanced age or with chronic disease.<sup>1</sup> In addition, falls are linked to a reduction in overall functioning and to early admission to long-term care facilities.<sup>2-5</sup> Balance impairments, which often lead to injurious falls, can be quantified in a clinical setting in order to direct therapeutic rehabilitation aimed at mitigating an individual's specific impairments and minimizing the risk of falls.

To aid in balance assessment and therapeutic prescription, several reliable clinical tools have been developed.<sup>3,6-8</sup> Although widely used, most of these assessments provide a measure of stability based on a single context of balance impairment.<sup>7-9</sup> Research, however, has demonstrated that postural impairments may be evident across several contexts of behavior.<sup>7,10</sup> A recently developed balance assessment tool, the Balance Evaluation Systems Test (BESTest), examines balance performance in 6 specific contexts (or systems, as termed by the test's developers) of postural control: mechanical constraints, limits of stability, anticipatory postural adjustments, postural responses to an induced loss of balance, sensory orientation, and gait.<sup>9</sup> In some cases, balance control may be compromised by a single balance system or subset of systems. The BESTest allows for the identification of specific balance systems responsible for poor balance performance and, therefore, can help direct clinical interventions.<sup>9</sup>

The BESTest has been found to be reliable across raters evaluating a cohort of individuals with and without various neurological diagnoses, and its validity was initially confirmed on the basis that BESTest scores correlate with reported scores of balance confidence.<sup>9,11</sup>

BESTest scores also have been validated to differentiate people with and without fibromyalgia, chronic obstructive pulmonary disease, and multiple sclerosis (MS).<sup>12-14</sup> In addition, the BESTest exhibited high test-retest and interrater reliability when used to evaluate participants with Parkinson disease, and BESTest scores were sensitive to these participants' prospective or retrospective fall reports.<sup>15-17</sup> This initial literature shows promise for the BESTest, but its clinical feasibility is extremely limited due to the time required to complete all 36 items.

To address potential limitations of the BESTest's redundancy and lengthy test duration, Franchignoni et al<sup>18</sup> identified a subset of the original BESTest items (ie, the Mini-BESTest), which consists of 16 items that can be administered in approximately 15 to 20 minutes. The Mini-BESTest has been reported to be just as reliable and capable of identifying fall status as the BESTest for individuals with Parkinson disease.<sup>16</sup> In addition, the Mini-BESTest exhibited superior psychometric properties compared with the Berg Balance Scale for identifying motor impairments in people with Parkinson disease.<sup>19</sup>

Although the reduced time to administer the Mini-BESTest renders the examination more efficient than the original BESTest, anecdotal reports suggest this assessment remains too lengthy, given increasing constraints on patient contact time in the clinic. In addition, although the Rasch analysis used to define the test offers a powerful technique to generate an examination consisting of nonredundant items that measure a correlated construct and represent a range of difficulty to prevent ceiling or floor effects, the result was contrary to the theoretical basis of the BESTest. Specifically, the items defining the Mini-BESTest represent a singular construct (termed "dynamic

balance" by the authors) identified by the Rasch analysis, but excluded items related to mechanical constraints and to limits of stability.

The construction of the Mini-BESTest thus implies that postural control represents a single construct and that a clinical assessment need only to evaluate this construct, as opposed to the original BESTest, which sought a global assessment of multiple constructs that influence balance impairment. Multiple constructs influence postural control and may be important for a broader clinical assessment of balance impairment intended for use across diverse clinical populations. Specifically, the Mini-BESTest's lack of items assessing mechanical constraints or limits of stability could inhibit its sensitivity when applied to people with musculoskeletal impairment or impaired limits of stability. Furthermore, without assessing these contexts of postural control, clinicians would be uninformed to direct interventions on the basis of these impairments. Indeed, the existing literature on the BESTest suggests that mechanical constraints or limits of stability differentiate groups with and without clinical health conditions or groups with and without a fall history.<sup>9,12-14,16</sup> Alternative methods, such as classical test theory, therefore, may offer another approach to shortening the BESTest based on its original theoretical underpinnings, thereby advancing the clinical goal of generating a time-efficient balance assessment across several influential constructs of postural control for use across multiple clinical populations.

This preliminary study, therefore, evaluated the internal consistency of items in each section of the BESTest and used item-total correlations to identify each section's most representative item. Each section's most representative item then was

included in a new Brief-BESTest examination. We evaluated each test version for interrater reliability and its validity to identify individuals with and without a neurological diagnosis. We further tested each examination version's validity by evaluating its ability to identify the reported fall history of people with or without MS. This second cohort was chosen because lower-limb strength and postural performance at the limits of stability have been found to represent significant predictors of future falls in people with MS and because the BESTest subsections on mechanical constraints and limits of stability are significantly affected in people with MS.<sup>14,20</sup> Thus, evaluating the different examination versions' ability to identify fall status in people with and without MS would provide insight into whether retaining an evaluation of mechanical constraints and limits of stability is important to the utility of the Brief-BESTest versus the Mini-BESTest. We hypothesized that the Brief-BESTest would exhibit psychometric properties comparable or superior to those of the Mini-BESTest, but with fewer items and with representation from every section of the original BESTest.

## Method

### Participants

Twenty participants were included in the first cohort (Tab. 1). Recruitment occurred with the intent to include participants with a wide range of balance abilities. Participants were included if they were able to stand independently and ambulate 6.1 m (20 ft) with or without an assistive device, and were willing to complete the BESTest (45–60 minutes). No other criteria were applied. Five of the 20 participants in the first cohort reported at least 1 fall in the previous 2 months, for a total of 7 falls (range=0–2). As part of a larger study,<sup>14,21</sup> thus representing a secondary analysis within this study, the second cohort

**Table 1.**

Participant Characteristics of the First Cohort<sup>a</sup>

| Participant No. | Sex | Age (y) | Neurological Diagnosis                      |
|-----------------|-----|---------|---|
| 1               | F   | 60      | N/A   |
| 2               | F   | 51      | N/A   |
| 3               | F   | 56      | N/A   |
| 4               | M   | 75      | N/A   |
| 5               | F   | 70      | Cerebrovascular accident (stroke)           |
| 6               | M   | 63      | Parkinson disease                           |
| 7               | M   | 83      | Essential tremor                            |
| 8               | M   | 60      | N/A   |
| 9               | M   | 68      | Parkinson disease                           |
| 10              | M   | 63      | Multiple sclerosis                          |
| 11              | F   | 57      | Multiple sclerosis                          |
| 12              | M   | 64      | Parkinson disease                           |
| 13              | M   | 75      | N/A   |
| 14              | M   | 60      | Peripheral neuropathy secondary to diabetes |
| 15              | F   | 57      | Multiple sclerosis                          |
| 16              | F   | 64      | N/A   |
| 17              | M   | 65      | Multiple sclerosis                          |
| 18              | M   | 69      | Parkinson disease                           |
| 19              | M   | 51      | N/A   |
| 20              | F   | 61      | N/A   |

<sup>a</sup> N/A=not applicable, F=female, M=male.

included 13 people with MS (8 women and 5 men; mean age=50 years, range=31–64) and 13 people without MS (8 women and 5 men; mean age=50 years, range=31–66). People with MS were recruited by advertisement in the local chapter of the National MS Society and were included if they: (1) had neurologist-diagnosed MS, (2) had an Expanded Disability Status Scale score of less than 6, and (3) had no uncorrected hearing or visual impairments. People without MS were recruited by advertising within the local community and were included if they: (1) had no self-reported neurological, musculoskeletal, or psychiatric disorders; (2) had no uncorrected hearing or visual impairments; and (3) were matched to the individuals with MS according to sex, similar height and weight, and within 2 years of age. The disease severity of

the participants with MS ranged from 0 to 4.5 on the Expanded Disability Status Scale. Seven of the participants with MS in cohort 2 reported at least one fall in the previous 3 months, for a total of 18 reported falls (range=0–6 falls; 5 participants reported multiple falls). The participants without MS reported no falls. All individuals gave written informed consent to participate in the study.

### Instrument

The BESTest consists of 36 items grouped into 6 specific postural control systems: biomechanical constraints, stability limits and verticality, anticipatory postural adjustments, postural responses to external perturbations, sensory orientation during stance, and stability in gait. Each item is scored based on a 4-level ordinal scale from 0 to 3. A score of 0

---

## The Brief-BESTest: A Suggested Brief Version of the BESTest

---

indicates failure or inability to complete the task, and a score of 3 indicates successful completion of the task according to all scoring criteria. As such, total scores range from 0 to 108, with subsection totals ranging from 0 to 15–21 (depending on the number of items in the respective subsection).

The Mini-BESTest is a subset of 14 tasks (16 items due to bilateral assessment) from sections of the BESTest related to anticipatory postural adjustments, reactive postural responses, sensory orientation, and stability in gait. The Mini-BESTest's items are scored on a 3-level ordinal scale from 0 to 2.

### Training

The raters were a student in a Doctor of Physical Therapy program and 2 doctorate researchers with expertise in the postural control of individuals with balance disorders. Raters prepared to administer the BESTest by reviewing the written version of the test and viewing the accompanying DVD provided by the test developer. In addition, one of the BESTest's original developers provided a 2-hour training session to the raters. The raters practiced administering and scoring the BESTest on student and community volunteers. These practice sessions allowed all raters to become familiar and comfortable with the implementation and scoring of the BESTest prior to the studies.

### Data Collection

Following the training sessions, the full 36-item BESTest was administered to all participants, regardless of cohort, although all 3 raters concurrently rated only the first cohort. The space was organized to facilitate transitions from one item to the next in order to minimize fatigue and mobility requirements. Five rest periods were offered at regular intervals, and participants were instructed to request additional rest if needed.

For the first cohort, 1 of the 3 raters administered the test while that rater or another rater served as a spotter for the participant during task performance in order to minimize the risk of falling. All raters independently scored the test for each participant. One rater evaluated the participants' BESTest performance for the second cohort.

### Data Analysis

The data from the first cohort were analyzed for internal consistency, validity, and interrater reliability using PASW version 18 (SPSS Inc, Chicago, Illinois). First, Cronbach alpha and item-total correlations were generated for each BESTest section and rater. The item with the highest average item-total correlation for each section was selected for inclusion in a new Brief-BESTest. If the item assessed a lateralized behavior with a companion item that assesses behavior on the other side (ie, items that are performed on the left and right sides), both items were included in the Brief-BESTest.

After establishing the Brief-BESTest, total scores were calculated for each version of the BESTest. For the Mini-BESTest, total scores were generated by transforming scores from the BESTest's 4-point ordinal scale to the Mini-BESTest's 3-point scale. Item-total correlations and Cronbach alpha also were reported for the Brief-BESTest and Mini-BESTest to confirm internal consistency and each item's contribution to the respective examination's total score. Interrater reliability of each test version was analyzed using 2-way, mixed-model intraclass correlation coefficients (ICC) testing for absolute agreement. Validity was initially assessed from the data of the first cohort by single-variable logistic regression models to determine the sensitivity, specificity, overall accuracy, and positive and negative likelihood ratios of each rater's total

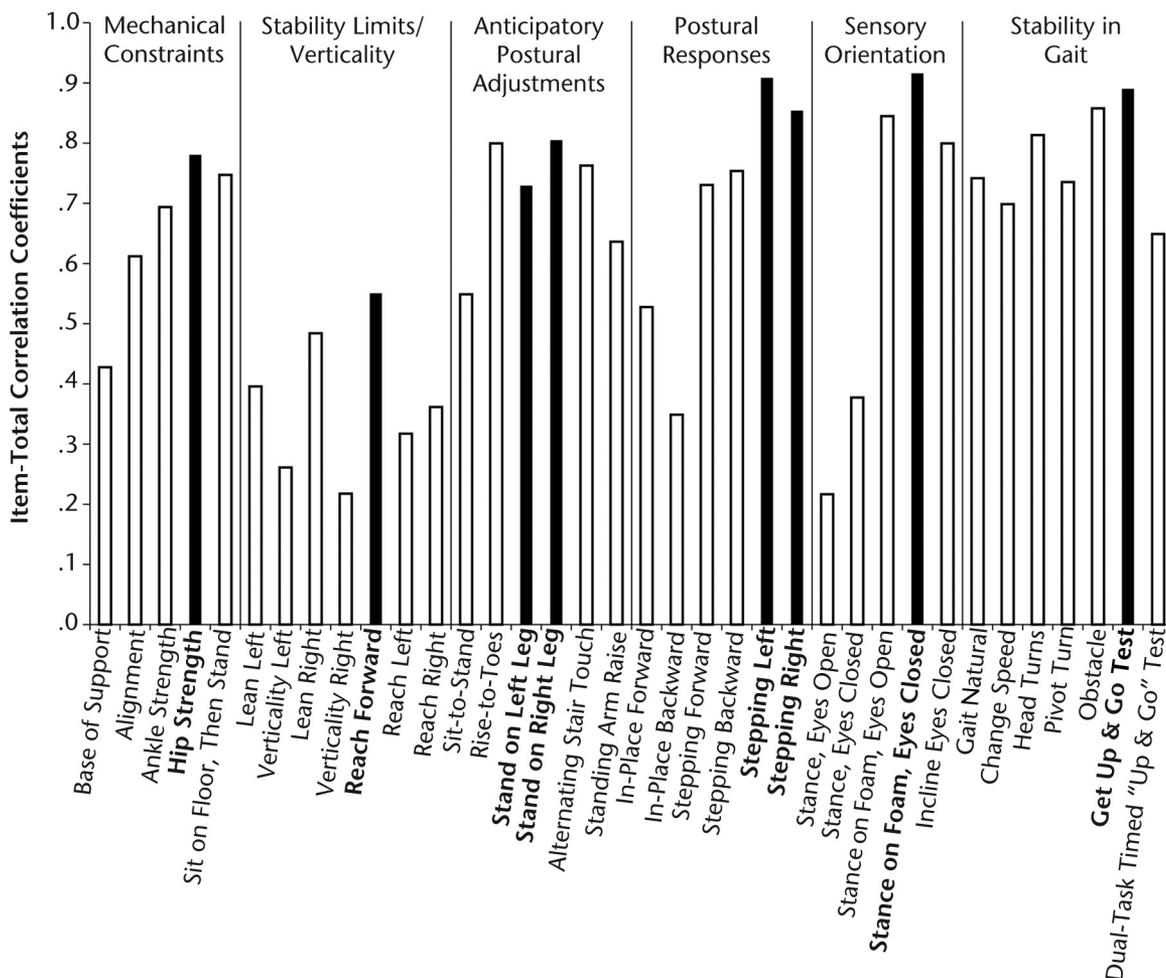
BESTest, Mini-BESTest, and Brief-BESTest scores to identify participants with or without a neurological diagnosis. Similar logistic regression models were used to determine the sensitivity, specificity, overall accuracy, and positive and negative likelihood ratios of the second cohort's scores for each examination version to identify participants with or without a reported fall history (ie, whether they reported experiencing at least 1 fall in the previous 3 months).

## Results

### Internal Consistency and Item-Total Correlations

The average Cronbach alpha coefficients for each section of the BESTest were .839, .621, .874, .863, .813, and .920 for mechanical constraints, limits of stability and verticality, anticipatory postural adjustments, postural responses, sensory orientation, and gait, respectively. The items with the highest item-total correlation coefficients to their respective section totals were hip abduction, forward reaching, single-leg stance, lateral compensatory stepping, standing with eyes closed on foam, and the Timed "Up & Go" Test (Fig. 1). These items, therefore, defined the Brief-BESTest. It should be noted that single-leg stance was selected as the representative item for the section on anticipatory postural adjustments based on an item-total correlation of .805, which was just slightly higher than the value of .800 elicited by the rise-to-toes item.

Cronbach alpha and item-total correlations for the Mini-BESTest and Brief-BESTest are identified in Table 2. Cronbach alpha was higher for the Mini-BESTest than for the Brief-BESTest in both the first and second cohorts, but both versions exhibited values above .85. On average, the item-total correlations were .732 for the Mini-BESTest and .737 for the Brief-BESTest in the first cohort and



**Figure 1.**

Item-total correlation coefficients for the items of the Balance Evaluation Systems Test (BESTest) to their respective section totals. The black bars represent the 8 items that were included in the Brief-BESTest on the basis of having the highest item-total correlations.

.617 for both the Mini-BESTest and Brief-BESTest in the second cohort.

**Interrater Reliability and Validity**

All 3 versions of the examination (score distributions identified in Fig. 2) exhibited very strong levels of interrater reliability: ICC (95% confidence interval [CI])=.985 (.959-.994) for the BESTest, .995 (.988-.998) for the Mini-BESTest, and .994 (.986-.997) for the Brief-BESTest. Total BESTest, Mini-BESTest, and Brief-BESTest scores significantly differentiated people from the first cohort with and without diagnosed neurological disorders or injuries (Tab. 3). BESTest scores were more

sensitive than the mini or brief versions to identify people with neurological disorders, whereas levels of specificity were similar among all versions of the examination. The relative sensitivity and specificity of the Mini-BESTest versus the Brief-BESTest depended on the rater, but the Brief-BESTest’s average sensitivity and specificity were 3% higher and 4% lower, respectively, than the Mini-BESTest’s sensitivity and specificity.

For the second cohort of 26 individuals with and without MS, the Brief-BESTest was 100% accurate in identifying whether the participants

reported no falls or at least 1 fall in the previous 3 months. The Mini-BESTest and original BESTest also provided high levels of specificity for people without a fall history, but exhibited lower sensitivities for people with a fall history than the Brief-BESTest (Tab. 4).

**Discussion**

The results support our hypothesis that the Brief-BESTest, defined from items with the highest item-total correlations, exhibits psychometric properties comparable or superior to those of the Mini-BESTest. Although the original BESTest better identified people with a neurological diagnosis

## The Brief-BESTest: A Suggested Brief Version of the BESTest

**Table 2.**

Internal Consistency of the Mini-BESTest and Brief-BESTest Based on Cronbach Alpha and Item-Total Correlations

| Item   | Item-Total Correlation Coefficients |          |
|--|-------------------------------------|----------|
|  | Cohort 1                            | Cohort 2 |
| Mini-BESTest (Cronbach alpha=.949 for cohort 1 and .905 for cohort 2)  |                                     |          |
| Sit-to-stand   | .626                                | .434     |
| Rise-to-toes   | .753                                | .684     |
| Left single-leg stance   | .653                                | .725     |
| Right single-leg stance  | .742                                | .699     |
| Forward compensatory stepping  | .807                                | .858     |
| Backward compensatory stepping   | .631                                | .591     |
| Leftward compensatory stepping   | .875                                | .699     |
| Rightward compensatory stepping  | .834                                | .647     |
| Standing with eyes open on firm surface                                | Excluded due to lack of variability |          |
| Standing with eyes closed on foam surface                              | .740                                | .534     |
| Incline stance   | .765                                | .627     |
| Change gait speed  | .611                                | .630     |
| Gait with head turns   | .813                                | .569     |
| Gait with pivot turn   | .746                                | .837     |
| Gait over obstacle   | .888                                | .591     |
| Dual-task Timed "Up & Go" Test   | .495                                | .124     |
| Brief-BESTest (Cronbach alpha=.917 for cohort 1 and .856 for cohort 2) |                                     |          |
| Hip abduction  | .755                                | .579     |
| Forward functional reach   | .330                                | .495     |
| Left single-leg stance   | .708                                | .866     |
| Right single-leg stance  | .760                                | .810     |
| Leftward compensatory stepping   | .904                                | .605     |
| Rightward compensatory stepping  | .820                                | .619     |
| Standing with eyes closed on foam surface                              | .820                                | .401     |
| Timed "Up & Go" Test   | .799                                | .562     |

compared with either abbreviated version of the examination, the Brief-BESTest and Mini-BESTest exhibited very similar levels of overall accuracy. The Brief-BESTest offered the highest sensitivity and overall accuracy to identify people with and without MS who reported at least 1 fall in the previous 3 months. In addition, all 3 versions of the examination exhibited very high levels of interrater reliability. Thus, the Brief-BESTest offers an even more abbreviated alternative to the Mini-BESTest with similar or superior psychometric properties.

The primary objective of this study was to shorten the BESTest in a theoretically consistent manner, with secondary objectives to provide preliminary comparisons of each version's psychometric properties. The preliminary validity analysis to identify people with and without diagnosed neurological disorders is not intended to suggest that any version of the BESTest would be used to diagnose the existence of neurological disorders. The analysis was instead chosen to demonstrate each examination version's ability to differentiate people with and without a

neurological diagnosis regardless of pathology, thereby supporting their use as a generalized balance assessment across the sampled clinical diagnoses.

The 8 selected items of the Brief-BESTest also exhibit face validity for representing each context of balance impairment beyond providing the highest item-total correlations for their respective contexts of impairment. In previously published research, each item has been shown to discriminate groups with and without neurological diagnoses, to associate with or predict falls or fractures, or to associate with ecological tasks or limited participation in activities of daily living.<sup>20,22-40</sup> In addition, the Brief-BESTest includes many assessments reported to be most frequently executed by physical therapists (eg, one-leg stance, functional reach, and Timed "Up & Go" Test).<sup>41</sup> Thus, the Brief-BESTest items appear to provide valid representative assessments for each context of balance impairment assessed by the original BESTest.

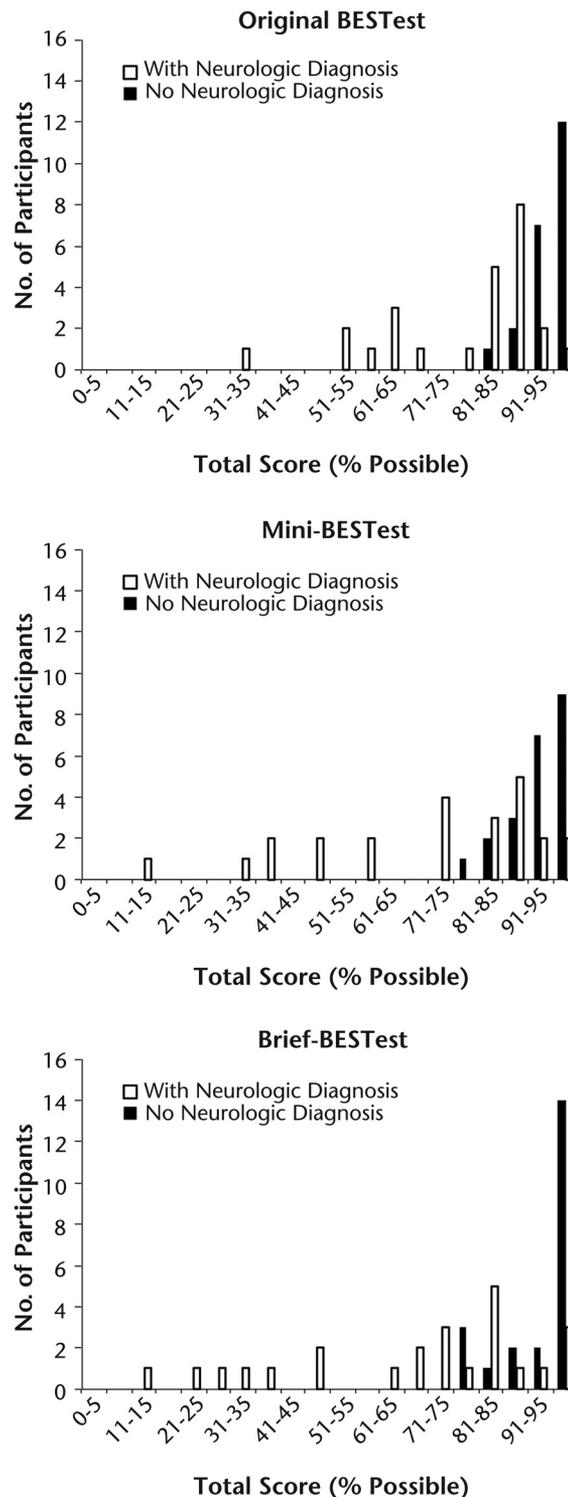
Both the Brief-BESTest and Mini-BESTest elicited a Cronbach alpha of greater than .85, regardless of cohort, although this measure of internal consistency was higher for the Mini-BESTest. This finding is not surprising, as the Mini-BESTest was derived based on a Rasch analysis designed to identify items representing a single construct, and our analysis confirms its internal consistency. In contrast, the Brief-BESTest was defined from the items most strongly representative of each of the BESTest's section scores that represent different contexts of postural control. The Brief-BESTest's items appear similarly associated to the total score as those of the Mini-BESTest are to its total score. Thus, although including items from all 6 contexts of postural control examined by the original BESTest, the

selected examination items appear appropriate for assessing balance impairment.

The analysis of fall history on people with and without MS demonstrates either the importance of the Brief-BESTest's retention of all 6 contexts of postural control or that the Mini-BESTest includes additional items that diminish its sensitivity to falls. Although further testing is necessary to identify which is true, measures of lower-limb strength and the ability to maintain balance at the limits of stability are associated with falling in people with MS, and the 2 BESTest subsections are significantly different between people with and without MS.<sup>14,20</sup> Thus, although the improved sensitivity of the Brief-BESTest may have been due to the removal of insensitive items rather than the retention of items related to mechanical constraints or limits of stability, including relevant contexts of balance impairment in people with MS likely contributed to its combined sensitivity and specificity to fall history.

The Mini-BESTest and BESTest have previously been reported to elicit 86% and 84% accuracy, respectively, in identifying fallers and nonfallers with Parkinson disease.<sup>16</sup> Thus, our results suggest all versions of the BESTest could provide similar or higher levels of accuracy to identify the fall status of people with MS. A larger study is needed to confirm this finding, and further testing across multiple patient populations remains necessary to determine each test version's relative capability to serve as a falls screening tool.

When evaluating the internal consistency of the BESTest, items with particularly low correlations to their section totals were generally those with little variability across participants and that exhibited ceiling effects (eg, base of support, standing



**Figure 2.** Frequency histograms of scores (percentage of possible maximum) from the first and second cohorts on the original Balance Evaluation Systems Test (BESTest) (top), Mini-BESTest (middle), and the proposed Brief-BESTest (bottom). With the combined cohorts, scores represent 24 individuals with a neurological diagnosis (white bars) and 22 without a neurological diagnosis (black bars).

## The Brief-BESTest: A Suggested Brief Version of the BESTest

**Table 3.**

Ability of the BESTest, Mini-BESTest, and Brief-BESTest to Differentiate People With and Without a Neurological Diagnosis for Each Rater

| Rater No.                 | Mean (95% Confidence Interval)<br>Percent Total Score |   | Regression Model<br>$\chi^2$ Value<br>(P Value) | % Sensitivity,<br>% Specificity,<br>% Accuracy | Positive, Negative<br>Likelihood Ratios |
|---------------------------|---|---|---|--|---|
|                           | Group Without<br>Neurological<br>Diagnosis            | Group With<br>Neurological<br>Diagnosis |   |  |   |
| BESTest total score       |   |   |   |  |   |
| 1                         | 95 (91–98)  | 74 (61–88)                              | 13.60 (.00023)                                  | 82, 67, 75                                     | 2.48, 0.27                              |
| 2                         | 89 (87–91)  | 72 (59–84)                              | 11.61 (.00066)                                  | 82, 78, 80                                     | 3.73, 0.23                              |
| 3                         | 92 (88–97)  | 73 (61–86)                              | 10.37 (.0013)                                   | 82, 78, 80                                     | 3.73, 0.23                              |
| Mini-BESTest total score  |   |   |   |  |   |
| 1                         | 90 (86–95)  | 64 (46–81)                              | 11.54 (.00068)                                  | 64, 78, 70                                     | 2.91, 0.46                              |
| 2                         | 90 (85–94)  | 61 (43–78)                              | 10.07 (.0015)                                   | 64, 78, 70                                     | 2.91, 0.46                              |
| 3                         | 90 (84–96)  | 59 (41–77)                              | 13.09 (.00030)                                  | 64, 89, 75                                     | 5.82, 0.40                              |
| Brief-BESTest total score |   |   |   |  |   |
| 1                         | 91 (83–100)   | 59 (39–79)                              | 10.04 (.0015)                                   | 64, 78, 70                                     | 2.91, 0.46                              |
| 2                         | 84 (80–87)  | 57 (38–76)                              | 9.72 (.0018)                                    | 64, 89, 75                                     | 5.82, 0.40                              |
| 3                         | 90 (82–97)  | 61 (43–80)                              | 9.50 (.0021)                                    | 74, 67, 70                                     | 2.24, 0.39                              |

**Table 4.**

Ability of the BESTest, Mini-BESTest, and Brief-BESTest to Differentiate People With and Without a Self-Reported Recent Fall History From the Second Cohort

| Test Version  | Mean (95% Confidence Interval)<br>Percent Total Score |                               | Regression Model<br>$\chi^2$ Value<br>(P Value) | % Sensitivity,<br>% Specificity,<br>% Accuracy | Positive, Negative<br>Likelihood Ratios |
|---------------|---|-------------------------------|---|--|---|
|               | Group With<br>Fall History                            | Group Without<br>Fall History |   |  |   |
| Brief-BESTest | 59 (45–73)  | 95 (92–98)                    | 30.29 (<.00001)                                 | 100, 100, 100                                  | Infinity, 0.00                          |
| Mini-BESTest  | 68 (52–83)  | 94 (91–96)                    | 20.61 (<.00001)                                 | 71, 100, 92                                    | Infinity, 0.29                          |
| BESTest       | 77 (66–87)  | 96 (94–98)                    | 24.47 (<.00001)                                 | 86, 95, 92                                     | 17.20, 0.46                             |

on a firm surface, feet-in-place responses). In addition, although the assessments of verticality and stability limits are included in the same section, those on verticality did not correlate well with their section scores, nor did they correlate well in an exploratory analysis with the sensory orientation scores (not shown). The interrater reliability of the verticality items also was reportedly low in previous research,<sup>9</sup> suggesting these items do not provide very meaningful contributions to the examination.

Unexpectedly, the dual-task Timed “Up & Go” Test provided the lowest item-total correlations for the gait

section despite the Timed “Up & Go” Test providing the highest correlations. One potential reason may be that the examination was challenging for most participants, with dual-task costs on either walking speed or counting being evident for most participants. Instructional standardization for attentional focus also may be an important factor,<sup>42</sup> as the participants could have differentially prioritized either task. Given that the scoring of this item diminishes based on impaired performance of either or both tasks and our itemwise interrater reliability for the item was adequate (Kendall  $W=.76$ ), this was not a likely cause of low item-total correlations. Alternatively, the addition

of a dual task may represent another system of impairment that represents cognitive-motor interaction. It would be of interest to evaluate dual-task analogs to multiple tasks (reaching, one-leg stance, stance on firm and foam surfaces, and the Timed “Up & Go” Test) in order to determine whether dual-task impairment represents a unique context of impairment.

## Conclusions

Given the economics of clinical evaluation allowing an extremely limited amount of patient-clinician contact time, it becomes even more imperative to develop an efficient examination. At the same time, the presenta-

tion of balance impairment is multifactorial, and the consequences of impaired balance are deserving of adequate assessment. This study confirms the validity and reliability of both the BESTest and Mini-BESTest for raters in research and raters in training as physical therapists, but also provides initial support for the Brief-BESTest. The Brief-BESTest elicited equally high levels of interrater reliability as the existing versions of the examination, its ability to differentiate individuals with and without diagnosed neurologic disorders or injuries was similar to that of the Mini-BESTest, and its ability to differentiate people with and without MS based on fall history was superior to that of either existing version. These psychometric properties are balanced by the most clinically feasible combination of only 8 scored items (compared with the 16 or 36 items of the other versions) while remaining theoretically intact with representative items from all 6 systems of postural control assessed by the original BESTest. The findings, however, are preliminary, representing a modest number of participants in both cohorts. Further use and examination of the Brief-BESTest are needed to confirm its validity and generalizability, as well as to identify clinically useful cutoff scores.

All authors provided concept/idea/research design, writing, and data collection and analysis. Dr Jacobs provided project management. Dr Padgett and Dr Kasser provided study participants. Dr Jacobs and Dr Kasser provided facilities/equipment. The authors acknowledge Dr Fay Horak for providing training.

These data were presented as a poster and abstract at the Joint World Congress of the International Society for Posture & Gait Research (ISPGR) and Gait & Mental Function; June 24–28, 2012; Trondheim, Norway.

The study was funded by the University of Vermont's Department of Rehabilitation and Movement Science.

DOI: 10.2522/ptj.20120056

## References

- Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System (WISQARS). 2005. Available at: <http://www.cdc.gov/ncipc/wisqars>. Accessed March 24, 2010.
- Brown AP. Reducing falls in elderly people: a review of exercise interventions. *Physiother Theory Pract*. 1999;15:59–68.
- Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatrics Soc*. 1986;34:119–126.
- Stevens JA, Corso PS, Finkelstein EA, Miller TR. The costs of fatal and nonfatal falls among older adults. *Inj Prev*. 2006;12:290–295.
- Tinetti ME, Williams CS. Falls, injuries due to falls, and the risk of admission to a nursing home. *N Engl J Med*. 1997;337:1279–1284.
- Tyson SF. How to measure balance in clinical practice: a systematic review of the psychometrics and clinical utility of measures of balance activity for neurological conditions. *Clin Rehabil*. 2009;23:824–840.
- Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. *Eur J Phys Rehabil Med*. 2010;46:239–248.
- Yelnik A, Bonan I. Clinical tools for assessing balance disorders. *Clin Neurophysiol*. 2008;38:439–445.
- Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Phys Ther*. 2009;89:484–498.
- Shumway-Cook A, Woollcott MH. *Motor Control: Theory and Practical Applications*. Baltimore, MD: Williams & Wilkins; 1995:119–140.
- Opheim A, Jahnsen R, Olsson E, Stanghelle JK. Balance in relation to walking deterioration in adults with spastic bilateral cerebral palsy. *Phys Ther*. 2012;92:279–288.
- Jones KD, Horak FB, Winters-Stone K, et al. Fibromyalgia is associated with impaired balance and falls. *J Clin Rheumatol*. 2009;15:16–21.
- Beauchamp MK, Sibley KM, Lakhani B, et al. Impairments in systems underlying control of balance in COPD. *Chest*. 2011 Nov 23 [Epub ahead of print].
- Jacobs JV, Kasser SL. Balance Impairment in people with multiple sclerosis: preliminary evidence for the Balance Evaluation Systems Test. *Gait Posture*. 2012 Apr 21 [Epub ahead of print].
- Leddy AL, Crowner BE, Earhart GM. Functional Gait Assessment and Balance Evaluation System Test: reliability, validity, sensitivity, and specificity for identifying individuals with Parkinson disease who fall. *Phys Ther*. 2011;91:102–113.
- Leddy AL, Crowner BE, Earhart GM. Utility of the Mini-BESTest, BESTest, and BESTest sections for balance assessments in individuals with Parkinson disease. *J Neurol Phys Ther*. 2011;35:90–97.
- Duncan RP, Leddy AL, Cavanaugh JT, et al. Accuracy of fall prediction in Parkinson disease: six-month and 12-month prospective analyses. *Parkinsons Dis*. 2012;2012:237673.
- Franchignoni FB, Horak FB, Godi M, et al. Using psychometric techniques to improve the Balance Evaluations Systems Test: the Mini-BESTest. *J Rehabil Med*. 2010;42:323–331.
- King LA, Priest KC, Salarian A, et al. Comparing the Mini-BESTest with the Berg Balance Scale to evaluate balance disorders in Parkinson's disease. *Parkinsons Dis*. 2012;2012:375419.
- Kasser SL, Jacobs JV, Foley JT, et al. A prospective evaluation of balance, gait, and strength to predict falling in women with multiple sclerosis. *Arch Phys Med Rehabil*. 2011;92:1840–1846.
- Jacobs JV, Kasser SL. Effects of dual tasking on the postural performance of people with and without multiple sclerosis: a pilot study. *J Neurol*. 2012;259:1166–1176.
- Hilliard MJ, Martinez KM, Janssen I, et al. Lateral balance factors predict future falls in community-living older adults. *Arch Phys Med Rehabil*. 2008;89:1708–1713.
- Nadollek H, Brauer S, Isles R. Outcomes after trans-tibial amputation: the relationship between quiet stance ability, strength of hip abductor muscles and gait. *Physiother Res Int*. 2002;7:203–214.
- Jenkins ME, Johnson AM, Holmes JD, et al. Predictive validity of the UPDRS postural stability score and the Functional Reach Test, when compared with ecologically valid reaching tasks. *Parkinsonism Relat Disord*. 2010;16:409–411.
- Weiner DK, Duncan PW, Chandler J, Studenski SA. Functional reach: a marker of physical frailty. *J Am Geriatr Soc*. 1992;40:203–207.
- Frzovic D, Morris ME, Vowels L. Clinical tests of standing balance: performance of persons with multiple sclerosis. *Arch Phys Med Rehabil*. 2000;81:215–221.
- Mak MK, Pang MY. Parkinsonian single fallers versus recurrent fallers: different fall characteristics and clinical features. *J Neurol*. 2010;257:1543–1551.
- Muir SW, Berg K, Chesworth B, et al. Balance impairment as a risk factor for falls in community-dwelling older adults who are high functioning: a prospective study. *Phys Ther*. 2010;90:338–347.
- Komatsu T, Jung Kim K, Kaminai T, et al. Clinical factors as predictors of the risk of falls and subsequent bone fractures due to osteoporosis in postmenopausal women. *J Bone Miner Metab*. 2006;24:419–424.
- Jacobs JV, Horak FB, Tran VK, Nutt JG. Multiple balance tests improve the assessment of postural stability in subjects with Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2006;77:322–326.
- Maki BE, Edmondstone MA, McIlroy WE. Age-related differences in laterally directed compensatory stepping behavior. *J Gerontol A Biol Sci Med Sci*. 2000;55:M270–M277.

---

## The Brief-BESTest: A Suggested Brief Version of the BESTest

---

- 32 Mansfield A, Peters AL, Liu BA, Maki BE. Effect of a perturbation-based balance training program on compensatory stepping and grasping reactions in older adults: a randomized controlled trial. *Phys Ther*. 2010;90:476–491.
- 33 King LA, Horak FB. Lateral stepping for postural correction in Parkinson's disease. *Arch Phys Med Rehabil*. 2008;89:492–499.
- 34 Winters-Stone KM, Torgrimson B, Horak FB, et al. Identifying factors associated with falls in postmenopausal breast cancer survivors: a multi-disciplinary approach. *Arch Phys Med Rehabil*. 2011;92:646–652.
- 35 Frenklach A, Louie S, Koop MM, Bronte-Stewart H. Excessive postural sway and the risk of falls at different stages of Parkinson's disease. *Mov Disord*. 2009;24:377–385.
- 36 Buatois S, Gueguen R, Gauchard GC, et al. Posturography and risk of recurrent falls in healthy non-institutionalized persons aged over 65. *Gerontology*. 2006;52:345–352.
- 37 Cohen HS, Kimball KT. Usefulness of some current balance tests for identifying individuals with disequilibrium due to vestibular impairments. *J Vestib Res*. 2008;18:295–303.
- 38 Herman T, Giladi N, Hausdorff JM. Properties of the “timed up and go” test: more than meets the eye. *Gerontology*. 2011;57:203–210.
- 39 Montes J, Cheng B, Diamond B, et al. The Timed Up and Go test: predicting falls in ALS. *Amyotroph Lateral Scler*. 2007;8:292–295.
- 40 Dite W, Connor HJ, Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial amputation. *Arch Phys Med Rehabil*. 2007;88:109–114.
- 41 Sibley KM, Straus SE, Inness EL, et al. Balance assessment practices and use of standardized balance measures among Ontario physical therapists. *Phys Ther*. 2011;91:1583–1591.
- 42 Abernethy B. Dual-task methodology and motor skills research: some methodological constraints. *Journal of Human Movement Studies*. 1988;14:101–132.

## The Brief-BESTest: A Suggested Brief Version of the BESTest

### Appendix.

Scoring Form for the Brief Balance Evaluation Systems Test (Brief-BESTest).<sup>a</sup>

#### Brief Balance Evaluation Systems Test

Patient/Subject: \_\_\_\_\_

Examiner: \_\_\_\_\_

Date: \_\_\_\_\_

General Note: "instability" is defined as using more than an ankle strategy to maintain balance (eg, a hip strategy is used).

| Section I. Biomechanical Constraints  |  |  |  |
|---|--|--|--|
| Item 1: Hip/Trunk Lateral Strength<br>"Rest fingertips in my hands while you lift your leg to the side and hold, keep trunk vertical. You will hold for 10 s."<br>Count 10 s, watch for straight knee; if they use moderate force on your hands, score as "without keeping trunk vertical."   | (3) Normal (10 s with trunk vertical)<br>(2) Mild (10 s without trunk vertical)<br>(1) Moderate (1 hip abducts with trunk vertical)<br>(0) Severe (neither hip, 10 s and vertical or not vertical) |  |  |
| Section II. Stability Limits  |  |  |  |
| Item 2: Functional Reach Forward<br>"Stand normally; lift both arms straight in front of you; reach as far forward as you can with arms parallel to the ruler without lifting your heels."<br>2 attempts<br>Observe that patient does not lift heels, rotate trunk, or protract scapula.<br>Watch for vertical initial alignment. Record best reach.  | (3) >32 cm (12.5 in)<br>(2) 16.5–32 cm (6.5–12.5 in)<br>(1) <16.5 cm (6.5 in)<br>(0) No measurable lean (or must be caught)  | Trial 1 (cm or in)<br>_____<br>Trial 2 (cm or in)<br>_____ |  |
| Section III. Transitions–Anticipatory Postural Adjustment   |  |  |  |
| Items 3 and 4: Stand on One Leg–Left and Right<br>"Look ahead; hands must stay on hips; bend one leg behind you; stand on 1 leg as long as you can for up to 30 s. Do not let your lifted leg touch the other leg."<br>Allow 2 attempts, record best attempt; record time up to 30 s (stop time if hands off hips or leg on floor or leg touches supporting leg).   | (3) Normal (stable >20 s)<br>(2) Trunk motion OR 10–20 s<br>(1) Stand 2–10 s<br>(0) Unable   | Left<br>Seconds<br>_____<br>Right<br>Seconds<br>_____      |  |
| Section IV. Reactive Postural Response  |  |  |  |
| Items 5 and 6: Compensatory Stepping–Lateral, Left and Right<br>"Stand with feet nearly together; lean into my hands; I will remove my hands; do whatever necessary to keep balance, trying to take 1 step."<br>Note: Stand next to and behind participant. Place hand on greater trochanter and brace yourself to hold the person's weight shifted to supported leg.   | (3) Recovers with 1 side/crossover step<br>(2) Several steps to recover independently<br>(1) Steps but needs assist to prevent fall<br>(0) No step OR falls  | Left<br>Right<br>_____<br>_____                            |  |
| Section V. Sensory Orientation  |  |  |  |
| Item 7: Stance With Eyes Closed, on Foam Surface<br>"Stand on foam with your eyes closed, your hands on your hips, and your feet close but not touching. Start by looking straight ahead, and I will start timing when you close your eyes. Stay as stable as possible and try to keep your eyes closed for the entire time. The goal is 30 s."<br>Two trials, if necessary. Patient must step off foam between trials. | (3) 30 s stable<br>(2) 30 s unstable<br>(1) <30 s<br>(0) Unable  | Trial 1 (s)<br>_____<br>Trial 2 (s)<br>_____               |  |
| Section VI. Stability in Gait   |  |  |  |
| Item 8: Timed "Up & Go" Test<br>"When I say 'go,' stand up and walk quickly but safely to the tape, turn, and walk back and sit in chair."<br>Start with back against chair, stop timing when buttocks hit the chair; chair should have arms to push from, if necessary. Imbalance might include trips or lateral/backward stumbles or crossovers.  | (3) Fast, <11 s, good balance<br>(2) Slow, >11 s, good balance<br>(1) Fast, <11 s, imbalance<br>(0) Slow, >11 s, imbalance   | Time (s)<br>_____  |  |
| <b>TOTAL:</b>   |  |  |  |

<sup>a</sup> The scoring form for the Brief-BESTest examination may not be used or reproduced without written permission of the authors.

# Physical Therapy

Journal of the American Physical Therapy Association and



**Is the BESTest at Its Best? A Suggested Brief Version Based on Interrater Reliability, Validity, Internal Consistency, and Theoretical Construct**  
Parminder K. Padgett, Jesse V. Jacobs and Susan L. Kasser  
*PHYS THER.* 2012; 92:1197-1207.  
Originally published online June 7, 2012  
doi: 10.2522/ptj.20120056

---

## References

This article cites 38 articles, 10 of which you can access for free at:  
<http://ptjournal.apta.org/content/92/9/1197#BIBL>

## Cited by

This article has been cited by 6 HighWire-hosted articles:  
<http://ptjournal.apta.org/content/92/9/1197#otherarticles>

## Subscription Information

<http://ptjournal.apta.org/subscriptions/>

## Permissions and Reprints

<http://ptjournal.apta.org/site/misc/terms.xhtml>

## Information for Authors

<http://ptjournal.apta.org/site/misc/ifora.xhtml>

---