Reliability and Validity of the Balance Evaluation Systems Test (BESTest) in People With Subacute Stroke

Butsara Chinsongkram, Nithinun Chaikeeree, Vitoon Saengsirisuwan, Nitaya Viriyatharakij, Fay B. Horak, Rumpa Boonsinsukh

Background. The Balance Evaluation Systems Test (BESTest) is a new clinical balance assessment tool, but it has never been validated in patients with subacute stroke.

Objective. The purpose of this study was to examine the reliability and validity of the BESTest in patients with subacute stroke.

Design. This was an observational reliability and validity study.

Methods. Twelve patients participated in the interrater and intrarater reliability study. Convergent validity was investigated in 70 patients using the Berg Balance Scale (BBS), Postural Assessment Scale for Stroke (PASS), Community Balance and Mobility Scale (CB&M), and Mini-BESTest. The receiver operating characteristic curve was used to calculate the sensitivity, specificity, and accuracy of the BESTest, Mini-BESTest, and BBS in classifying participants into low functional ability (LFA) and high functional ability (HFA) groups based on Fugl-Meyer Assessment motor subscale scores.

Results. The BESTest showed excellent intrarater reliability and interrater reliability (intraclass correlation coefficient = .99) and was highly correlated with the BBS (Spearman \( r = .96 \)), PASS (\( r = .96 \)), CB&M (\( r = .91 \)), and Mini-BESTest (\( r = .96 \)), indicating excellent convergent validity. No floor or ceiling effects were observed with the BESTest. In contrast, the Mini-BESTest and CB&M had a floor effect in the LFA group, and the BBS and PASS demonstrated responsive ceiling effects in the HFA group. In addition, the BESTest showed high accuracy as the BBS and Mini-BESTest in separating participants into HFA and LFA groups.

Limitation. Whether the results are generalizable to patients with chronic stroke is unknown.

Conclusion. The BESTest is reliable, valid, sensitive, and specific in assessing balance in people with subacute stroke across all levels of functional disability.
Balance control is a complex integration of multiple systems, including sensory and motor systems, sensorimotor integration, and high-level premotor processing—all of which can be affected following stroke.1 Laboratory studies have shown that muscle paresis and spasticity can impair postural alignment, base of foot support, and forces necessary to correct postural equilibrium.2–4 Most patients with stroke also demonstrate decreased ankle proprioception and somatosensation at the feet, which are critical for balance control.2,4 In addition, patients with stroke have difficulties in central processing for sensory integration and sensory reweighting.5 They demonstrate more postural sway in the conditions of altered somatosensory information and altered vision,6 such as walking on uneven surfaces and walking in the dark. Anticipatory postural adjustments and automatic postural responses to restore the body’s equilibrium during perturbations are reduced on the paretic side.7–10 Cognitive impairments can lead to inadequate allocation of attention and to greater fall risk in this population.2,11 These studies suggest that postural control impairments in patients with stroke arise from diverse mechanisms involved in balance control. Therefore, a balance assessment that could capture all underlying balance impairments is necessary.

At present, balance evaluation for assessing patients with stroke in clinical settings involves ordinal scales to assess a set of functional tasks that require balance while maintaining posture or changing position.12 Examples of the most commonly used functional scales in patients with stroke include the Berg Balance Scale (BBS) and the Postural Assessment Scale for Stroke Patients (PASS).13 These functional balance scales are useful in reporting the presence of balance impairments, but they cannot identify the type of impairments underlying balance problems.

Recently, the Balance Evaluation Systems Test (BESTest) has been developed based on a conceptual model of balance control in which 6 different systems or domains contribute to balance control. The BESTest evaluates the following systems: biomechanical constraints, stability limits and verticality, anticipatory postural adjustments, automatic postural responses, sensory organization, and stability in gait.14 The 36 items of the BESTest are scored from 0 (severe impairment) to 3 (no impairment). The total score is calculated as a percentage of the points scored out of 108 total points.14 The advantage of the BESTest is that it includes tasks involving many different mechanisms underlying the control of balance so that clinicians can determine the type of balance retraining intervention specific to the underlying causes of balance problems. The BESTest has shown excellent interrater reliability and validity and the potential to differentiate types of balance deficits in patients with Parkinson disease (PD), vestibular loss, and peripheral neuropathy.14 However, the BESTest has not yet been validated to assess balance performance in patients with stroke.

The Mini-BESTest, the shorter version of the BESTest, was developed to reduce the assessment time.15 Item redundancy and the first 2 sections (biomechanical constraints and stability limits with verticality) from the BESTest were removed based on Rasch analysis to yield the assessment of a common construct, “dynamic balance,” in the Mini-BESTest.15 The 14-item Mini-BESTest is scored on a 3-level ordinal scale from 0 (poor balance performance) to 2 (no balance impairment). Several studies have shown that the Mini-BESTest also was reliable and valid16–18 and useful for fall prediction in patients with PD.19,20 The use of the Mini-BESTest in community-dwelling people with chronic stroke has recently been reported, with excellent interrater and intrarater reliability and high correlations with other balance measures such as the BBS and the Timed “Up & Go” Test but without the floor and ceiling effects seen in these traditional balance tests.21 However, the feasibility of the Mini-BESTest for assessing balance across lower levels of subacute stroke impairment severity has not been reported.

The purposes of this study were: (1) to assess the reliability and convergent validity of the BESTest in patients with subacute stroke and (2) to determine whether the BESTest could be used to identify patients with low and high functional ability, as classified with the Fugl-Meyer Stroke Assessment motor subscale (FM-motor). We hypothesized that the BESTest would be reliable and well correlated with other balance scales such as the BBS, PASS, and Mini-BESTest. Furthermore, we predicted that the BESTest would have fewer floor and ceiling effects than the other clinical balance tests, so that the BESTest will be more feasible for people with stroke across a wider range of motor disability. Finally, we hypothesized that the lack of musculoskeletal and limits of stability/verticality domains in the Mini-BESTest may lead to a floor effect in patients...
Validation of BESTest in Stroke

who have low functional ability because these domains are commonly affected in patients with subacute stroke.

Method

Participants

Participants were patients with subacute stroke (defined as stroke of less than 4 months since onset\(^2\)) referred for physical therapist services at Prasart Neurological Institute in Thailand from November 1, 2012, to August 31, 2013. Individuals were included in this study if they passed the screening tests by a physician using the following criteria: diagnosis of cerebral hemorrhage or cerebral infarction with stable medical conditions, age between 25 and 90 years, first unilateral hemispheric stroke, and able to follow instructions to complete the assessment. Individuals were excluded from the study if they had cognitive impairment (defined as having a Mini-Mental State Examination [MMSE] score of less than 24),\(^{25,24}\) cerebral aneurysm, a lesion at the brain stem involving the sleep-wake and respiratory control center or cerebellum, aphasia as diagnosed by a physician using a Thai adaptation of the Western Aphasia Battery,\(^{25}\) a neurological disorder other than stroke, or the presence of a major peripheral neuropathy or musculoskeletal problem that was sufficient to disturb balance. All participants signed a consent form approved by the Human Research Protection Committee at Prasart Neurological Institute Research Center, Thailand.

Outcome Measures

To determine whether the balance tools could be used to separate participants into high- and low-functioning groups, the FM-motor was selected to measure functional recovery from stroke.\(^{26}\) The FM-motor is scored from 0 to 100 points, with each item scored on a 3-point ordinal scale ranging from 0 (cannot perform) to 2 (performs fully). The FM-motor score is used to classify patients with stroke into 4 groups according a level of disability: a score of 0 to 35 represents severe disability, a score of 36 to 55 represents moderately severe disability, a score of 56 to 79 represents moderate disability, and a score of 80 or higher represents mild disability.\(^{27}\) In this study, we referred to patients with stroke who had FM-motor scores between 0 and 55 as the low functional ability group (LFA group) and to those who had FM-motor scores greater than 55 as the high functional ability group (HFA group).

Five balance evaluation scales—BESTest, Mini-BESTest, BBS, PASS, and Community Balance and Mobility Scale (CB&M)—were administered to each participant. The BBS is a widely used clinical scale for functional balance in the clinic and is currently considered as a reference standard for assessing balance in patients with stroke. The BBS consists of 14 functional balance items that focus on the ability to maintain a position and postural adjustments to voluntary movements.\(^{28}\) The total possible score on the BBS is 56 points, with ratings ranging from 0 (unable to perform) to 4 (able to complete the task).\(^{13}\) The PASS is a 12-item test designed to measure balance function in patients with stroke, especially those with very poor motor function. The 4-point items, creating the total score of 36, assess functional balance performance in maintaining or changing lying, sitting, and standing postures.\(^{29}\) The CB&M consists of 19 tasks, including advanced functional balance and mobility activities aimed at ambulatory or community-dwelling people with stroke.\(^{30}\) Items are scored on ordinal scale ranging from 0 (unable to perform) to 5 (able to perform independently).

Procedure

Reliability. Interrater and intrarater reliability were determined by 5 physical therapists. Raters were a convenient sample of 3 physical therapists from the stroke rehabilitation center with stroke rehabilitation experience of 1, 5, and 10 years, and 2 PhD physical therapist students from Srinakharinwirot University and Mahidol University. Prior to the study, all raters participated in 2 training workshops on using the BESTest. The training workshops were carried out by a therapist who received training from the BESTest developer. Each item of the BESTest was demonstrated on individuals who were healthy, and the rating criteria were discussed. All raters also scored each participant’s performance from a demonstration video, and rating criteria were discussed again if there was a discrepancy. When there were differences in the rating scores, the raters were asked to watch the videotape with the trainer, who encouraged the discussion until all raters reached agreement.

After the training workshops, a BESTest reliability trial was performed in 12 patients with stroke. The sample size calculation for this study was based on a power of 0.80 and an alpha level of .05. A null intraclass correlation coefficient (ICC) of .60\(^{31}\) and expected correlation coefficients in our study were based on the correlation coefficient of \(r=0.93\) found in a previous balance measure reliability study.\(^{32}\) The BESTest was administered by one rater (B.C.), and the procedures were video recorded. The evaluation was performed in a laboratory setting at Prasart Neurological Institute, and videotape recording was performed in the same view for all participants. The rater used the same testing instructions to standardize performance. Before testing, the vital signs of the participants were evaluated to

Downloaded from https://academic.oup.com/ptj/article-abstract/94/11/1632/2735450 on 27 July 2018
Validation of BESTest in Stroke

ensure stable medical status. All participants received the same verbal instructions. Participants were permitted to rest as needed during the test.

All raters scored each participant’s performance from videotape on 2 separate occasions. The second scoring was performed 7 days after the first scoring. Intrarater reliability of total scores and section scores was assessed by comparing the score of occasion 1 and score of occasion 2 for each rater. Interrater reliability was determined by comparing the scores from occasion 1 for all 5 raters. Each rater scored the videotape on separate scoring sheets for each occasion and did not discuss scoring among participants and occasions.

Validity. Prior to the validity study, the raters (B.C., N.C.) received additional training for using the Mini-BESTest, BBS, PASS, CB&M, and FM-motor. The convergent validity sample size was calculated from a null correlation coefficient of .50 and acceptable correlation of .80 from a previous study, yielding at least 29 participants per group. After signing the approved consent forms, each participant’s demographic and clinical information, including Barthel Index (BI) score and FM-motor score, was recorded by one rater (N.C.). Seventy patients, 35 in each group, participated in this study. A balance assessment of each participant was then performed by the second rater (B.C.), who was blinded to the basic characteristics and the FM-motor scores of the participants. The sequence of the balance tests was randomly performed following the order of the starting position: from lying to sitting to standing and walking. Any test item that was duplicated between the tests was performed only once and scored using the criteria for each test. The evaluation was performed in the same laboratory setting, and all participants received the same verbal instructions. Rest was allowed during performance of each item to avoid fatigue. Total assessment time was approximately 1.5 hours, but if the testing could not be completed in 1 day, it was continued on the next day. To ensure the accuracy of the scoring, the rater (B.C.) reviewed and scored each participant’s test performance from the videotape again. In case of disagreement between concurrent rating and video rating, the second rater (N.C.) was asked to score those items from the videotape to decide the final score.

Data Analysis
A descriptive statistical analysis of demographic and baseline clinical characteristics of the participants was conducted. An independent-sample t test was used to compare age, time since stroke, BI score, and BBS score, whereas the Mann-Whitney U test was used to compare MMSE score, FM-motor score, and other balance measurement scores between the LFA and HFA groups. Interrater and intrarater reliability values were calculated based on the participants’ ICCs. The ICC model 2,k was used for interrater reliability, and the ICC model 3,k was used for intrarater reliability. An ICC value of .80 or higher indicates excellent correlation (good reliability), values of .60 to .80 indicate adequate correlation (moderate reliability), and values of .40 to .60 indicate poor correlation (weak reliability).

To determine the convergent validity of the BESTest, the strength of the relationship among the BESTest, Mini-BESTest, PASS, BBS, and CB&M was examined using Spearman rank order correlations. Correlation coefficients of .00 to .49 were interpreted as poor, those of .50 to .79 as moderate, and those of 0.8 or higher as excellent. The floor and ceiling effects were calculated as the percentage of the sample scoring the minimum or maximum possible scores, respectively. Ceiling and floor effects of 20% or greater are considered significant. In this study, we also recorded the “responsive ceiling effect,” which represents the 20% or greater percentage of participants who scored within the top 10% of the test and allows a sufficient number of scores to measure change over time. The percentage of minimum and maximum scores for each item of the BESTest was calculated as the percentage of participants who received a minimum score (0) or a maximum score (3) for each item. These percentages represented the difficulty of the item, such that a high percentage (>80%) of minimum score reflected a difficult item and a high percentage of maximum score indicated an easy item. Receiver operating characteristic (ROC) curves were used to determine the relative performance of the BESTest score, the Mini-BESTest score, and the BBS score for classifying participants into 2 groups based on their FM-motor score to discriminate between low functional ability (FM-motor score ≤55) and high functional ability (FM-motor score >55). The accuracy for each balance test was assessed using the area under the curve (AUC), which can be interpreted as the probability of correctly identifying participants with low or high function. An AUC value of 0.9 and above indicates high accuracy, 0.7 to 0.9 indicates moderate accuracy, 0.5 to 0.7 indicates low accuracy, and 0.5 and below indicates test due to chance. The cutoff point value was chosen by selecting the score that provides the best balance between high sensitivity and high specificity. Furthermore, likelihood ratios and accuracy were used to determine posttest probabilities that helped to improve certainty for confirming the diagnosis or abandoning it. A positive likelihood ratio...
Validation of BESTest in Stroke

Table 1.
Demographic and Clinical Characteristics of Participants in the Validity Study, Classified by Functional Ability

| Characteristic               | All Participants (N=70) | Functional Ability Group |
|-----------------------------|-------------------------|-------------------------|
|                             | Low (n=35)              | High (n=35)             |
| Age (y)                     | 57.01 (12.23)           | 54.85 (12.09)           | 59.17 (12.16) |
| Sex (male/female), n        | 35/35                   | 21/14                   | 17/18          |
| Time since stroke (mo)      | 1.11 (2.00)             | 1.74 (2.60)*            | 0.48 (0.75)*   |
| Type of stroke (ischemic/hemorrhagic), n | 54/16 | 23/12 | 31/4 |
| Affected side (right/left), n | 39/31                  | 22/13                   | 17/18          |
| MMSE (/30)                  | 28.00 (2.39)            | 27.06 (2.38)            | 28.77 (2.10)   |
| BI (/100)                   | 61.93 (24.76)           | 48.86 (18.87)*          | 75.00 (23.17)* |
| FM-motor (/100)             | 55.53 (34.00)           | 24.37 (12.59)*          | 86.69 (13.72)* |
| BESTest (/100%)             | 41.70 (28.19)           | 23.89 (18.44)*          | 59.52 (24.82)* |
| Mini-BESTest (/28)          | 9.07 (8.53)             | 4.06 (5.44)*            | 14.09 (8.14)*  |
| BBS (/56)                   | 29.46 (19.17)           | 18.37 (15.92)*          | 40.57 (15.49)* |
| PASS (/36)                  | 26.70 (7.79)            | 22.03 (7.7)*            | 31.37 (4.33)*  |
| CB&B&M (/96)                | 11.64 (18.98)           | 1.37 (3.95)*            | 21.91 (22.32)* |

*All values are presented as mean (SD), unless otherwise noted. Significant difference between low and high functional ability groups (*P<.01, †P<.001).

Functional ability was classified with the Fugl-Meyer Stroke Assessment–motor subscale (FM-motor) as low (0–55) or high (>55). MMSE=Mental State Examination, BI=Bartel Index, BESTest=Balance Evaluation System Test, Mini-BESTest=Mini-Balance Evaluation System Test, BBS=Berg Balance Scale, CB&B&M=Community Balance and Mobility Scale.

(LR+) was the probability of a participant having a score above the cutoff point, whereas a negative likelihood ratio (LR−) was the probability of a participant having a score below the cutoff point. An LR+ over 5 and an LR− less than 0.2 indicate that the test is useful due to its high probability of correctly classifying patients into functional ability groups, but likelihood ratios close to 1.00 indicate that the tests are useless, as the probability of correctly or incorrectly classifying patients into groups is the same.55 The accuracy of the selected cutoff point in classifying patients into functional ability groups was calculated as the proportion of the participants in the correctly identified group when using that cutoff point for all participants.55

Role of the Funding Source
This study was supported by the Thailand Research Fund, the Office of Higher Education Commission, and Srinakharinwirot University (grant no. RSA5580002).

Results

Reliability
Twelve patients with subacute hemorrhagic or ischemic stroke (4 female, 8 male; average age=58.24 years) participated in the reliability testing. The intrarater reliability of the BESTest total scores was excellent, with an ICC of .99 (95% confidence interval [95% CI]=.99, 1.00) and subsection ICCs ranging from .95 (95% CI=.90, .98) to .99 (95% CI=.98, .99). The interrater reliability of the BESTest total scores was also excellent, with an ICC of .99 (95% CI=.98, .99) and subsection ICCs ranging from .87 (95% CI=.72, .96) to .98 (95% CI=.95, .99).

Convergent Validity
The eFigure (available at ptjournal.apta.org) depicts the number of eligible participants, included participants, and excluded participants with the reasons for exclusion. One hundred thirteen patients with subacute stroke met the eligibility criteria, and 70 patients consented to participate in the validity testing. Of the 43 patients who were not included in this study, 30 did not require continued rehabilitation, 5 had cognitive impairment or severe motor aphasia, 2 had severe internal carotid artery occlusion that required urgent surgery, and 6 had major musculoskeletal problems, including osteoarthritis and gouty arthritis. The majority of the participants (n=67) completed the validity testing within 1 day. Demographic and clinical characteristics of participants in the validity study are shown in Table 1. Participants in the LFA and HFA groups were similar in age, sex, and cognitive impairment as measured with the MMSE, but those in the LFA group had a longer time since stroke and significantly lower scores on all of the clinical balance tests.

The convergent validity of the BESTest in patients with subacute stroke was assessed with clinical reference scales that have been validated in populations with stroke (ie, BBS, PASS, CB&B&M, and Mini-BESTest). Figure 1 shows high correlations of the
total scores from the BESTest with the BBS ($r = 0.96$), PASS ($r = 0.96$), CB&M ($r = 0.91$), and Mini-BESTest ($r = 0.96$), indicating excellent convergent validity.

**Score Distribution Across Levels of Functional Ability**

The distribution of participants’ scores for the balance measures across levels of functional ability is shown in Figure 2. The distribution of scores from all observed scales except the CB&M covered the whole range of functional ability. However, for the Mini-BESTest and PASS, scores were clustered at one end of the graph, indicating the potential floor or ceiling effects. The analysis of the floor effects demonstrated that 34.3% ($n = 12$) of the participants in the LFA group (17% of all participants with subacute stroke) received the lowest possible score of 0 on the Mini-BESTest, suggesting the floor effect of the Mini-BESTest specifically in the LFA group. Likewise, 77.1% ($n = 27$) of the participants in the LFA group and 11.4% ($n = 4$) of the participants in the HFA group (44.3% of all participants with subacute stroke) had a score of 0 on the CB&M. None of the participants had a score of 0 on the BESTest, BBS, or PASS.

For the ceiling effect, 4.3% ($n = 3$) of the participants received the maximum score on the BBS and PASS. However, when calculating by using the scores within the top 10%, 25.7% of all participants (5.7% of the participants in the LFA group and 45.7% of the participants in the HFA group) scored within the top 10% on the BBS, indicating a responsive ceiling effect especially in the HFA group. Similarly, 37.1% of all participants (14.3% of the participants in the LFA group and 60.0% of the participants in the HFA group) received a score within the top 10% on the PASS. To the contrary, no responsive ceiling effect was observed in the rest of the balance scales, as very few participants received a score within the top 10% (BESTest: 4.3%, Mini-BESTest: 4.3%, and CB&M: 1.4%).

**Item Scores of the BESTest**

The scores of each item and section of the BESTest compared between the LFA and HFA groups are shown in Table 2. In general, the section scores were significantly lower in the LFA group, suggesting that the BESTest was able to differentiate between the 2 groups. Several items in the BESTest were too hard to perform for the LFA group, as shown by the percentage of minimum score of more than 80%. These items included sit on floor, rise to toe, stand on one leg, alternate stair touching, compensatory postural response in all directions, stand on foam with eyes closed, and all items in the “stability in gait” section, with the exception of change in gait speed. In contrast, there were only 2
items (base of support and sitting verticality) that were shown to be easy (>80% of maximum score) in both LFA and HFA groups.

Comparison of the percentage of minimum and maximum scores between the LFA and HFA groups also suggested the feasibility of each item to be used in all levels of functional ability. For example, 31.4% of the participants in the LFA group compared with 71.4% of the participants in the HFA group received the maximum score on item 6 (lateral lean). A similar trend was evident with the minimum score, where there was a decrease in percentage of minimum score in the HFA group compared with the LFA group. Such a pattern was found for all items of the BESTest, except for 3 items (base of support and sitting verticality, left and right).

Identification of Functional Ability Level

The ROC plots for the BESTest, Mini-BESTest, and BBS are shown in Figure 3. All 3 scales showed high AUCs in identifying patients with stroke in the LFA and HFA groups: 0.88 for the BESTest, 0.85 for the Mini-BESTest, and 0.85 for the BBS (Tab. 3). The suggested cutoff scores for low functional ability were 49% for the BESTest, 9/28 for the Mini-BESTest, and 19/56 for the BBS (Tab. 3). Based on these cutoff scores, comparable sensitivity and specificity was found in all 3 measures, with specificity highest for the BESTest, whereas the BBS had the highest sensitivity but the lowest specificity. All 3 scales also showed high posttest accuracy (0.77–0.81) when applying the suggested cutoff point in classifying participants into the LFA and HFA groups. Although all 3 tests showed similar AUC, sensitivity, specificity, and posttest accuracy, the BESTest and Mini-BESTest had LR+ ratios over 5 and LR− ratios close to 0.2, suggesting the BESTest and Mini-
### Table 2.
Item Scores on the Balance Evaluation Systems Test (BESTest)*

| Section Item | BESTest | LFA Group | HFA Group |
|--------------|---------|-----------|-----------|
|              | Median  | Minimum  | Maximum   | Median  | Minimum  | Maximum   |
| Biomechanical constraints | 31.43 (17.08)* | 0.0 | 82.9 | 3 | 0.0 | 91.4 |
| 1. Base of support | 3 | 37.1 | 5.7 | 2 | 2.9 | 48.6 |
| 2. CoM alignment | 1 | 71.4 | 5.7 | 3 | 1.5 | 28.6 |
| 3. Ankle strength/ROM | 0 | 80.0 | 0.0 | 0 | 60.0 | 8.6 |
| 4. Hip/trunk strength | 0 | 85.7 | 5.7 | 2 | 42.9 | 25.7 |
| 5. Sit on floor/stand up | 0 | 85.7 | 5.7 | 2 | 42.9 | 25.7 |
| Limits of stability | 50.95 (19.57)* | 5.7 | 82.9 | 3 | 0.0 | 100.0 |
| 6A. Sitting verticality, right† | 3 | 5.7 | 82.9 | 3 | 0.0 | 100.0 |
| 6B. Lateral lean, right† | 2 | 5.7 | 31.4 | 3 | 0.0 | 71.4 |
| 7. FR forward | 0 | 65.7 | 0.0 | 2 | 17.1 | 5.7 |
| 8. FR right† | 0 | 68.6 | 2.9 | 2 | 22.9 | 8.6 |
| Anticipatory adjustments | 18.41 (20.73)* | 28.6 | 40.0 | 3 | 0.0 | 85.7 |
| 9. Sit to stand | 2 | 85.7 | 0.0 | 2 | 25.7 | 22.9 |
| 10. Rise to toe | 0 | 88.6 | 0.0 | 1 | 40.0 | 11.4 |
| 11. Stand on one leg, right† | 0 | 91.4 | 0.0 | 2 | 31.4 | 31.4 |
| 12. Alternate stair touching | 0 | 91.4 | 0.0 | 2 | 31.4 | 31.4 |
| 13. Standing arm raise | 0 | 91.4 | 0.0 | 2 | 31.4 | 31.4 |
| Postural responses | 12.86 (26.40)* | 71.4 | 20.0 | 3 | 14.3 | 71.4 |
| 14. In-place, front | 0 | 74.3 | 14.3 | 3 | 17.1 | 62.9 |
| 15. In-place, back | 0 | 91.4 | 5.7 | 2 | 31.4 | 42.9 |
| 16. Compensatory, front | 0 | 91.4 | 5.7 | 2 | 31.4 | 42.9 |
| 17. Compensatory, back | 0 | 91.4 | 5.7 | 2 | 31.4 | 31.4 |
| 18. Compensatory, right† | 0 | 91.4 | 5.7 | 2 | 31.4 | 31.4 |
| Sensory orientation | 23.24 (28.58)* | 45.7 | 34.3 | 3 | 2.9 | 80.0 |
| 19A. Firm surface, EO | 1 | 54.3 | 14.3 | 3 | 2.9 | 60.0 |
| 19B. Firm surface, EC | 0 | 77.1 | 2.9 | 2 | 17.1 | 34.3 |
| 19C. Foam surface, EO | 0 | 91.4 | 0.0 | 0.5 | 51.4 | 8.6 |
| 20. Incline, EC | 0 | 62.9 | 17.1 | 3 | 20.0 | 60.0 |
| Stability in gait | 6.26 (15.61)* | 82.9 | 0.0 | 2 | 25.7 | 5.7 |
| 21. Gait, level surface | 0 | 80.0 | 0.0 | 2 | 25.7 | 22.9 |
| 22. Change in gait speed | 0 | 82.9 | 0.0 | 1 | 42.9 | 14.3 |
| 23. Walk head turn, lateral | 0 | 91.4 | 2.9 | 2 | 28.6 | 22.9 |
| 24. Walk with pivot turns | 0 | 91.4 | 0.0 | 2 | 34.3 | 25.7 |
| 25. Step over obstacles | 0 | 91.4 | 0.0 | 2 | 34.3 | 25.7 |
| 26. TUG | 0 | 91.4 | 0.0 | 2 | 34.3 | 25.7 |
| 27. TUG with dual task | 0 | 82.9 | 0.0 | 1 | 25.7 | 0.0 |

*All section scores are presented as mean (SD). LFA=low functional ability, HFA=high functional ability, FR=functional reach, EO=eyes open, EC=eyes closed, TUG=Timed "Get Up and Go" test, CoM=center of mass, ROM=range of motion. †Significant difference between LFA and HFA groups at P<.01. †Item that evaluated both right and left sides, but with similar results on each side, so data for only the right side are presented.
BESTest had higher probability of correctly classifying patients according to level of functional ability compared with the BBS.

**Discussion**

This is the first study to examine whether the BESTest is reliable and valid to assess balance impairments in patients with subacute stroke. Our results confirmed that the BESTest has excellent interrater and intrarater reliability, as well as excellent convergent validity, compared with commonly used clinical balance scales in patients with stroke. The high intrarater and interrater reliability of the BESTest in patients with subacute stroke agreed with previous studies in patients with neurological conditions or healthy elderly people. For example, the ICC for total score interrater reliability was .91, with section score ICCs ranging from .79 to .96 in patients with neurological diseases and healthy people. Similarly, the ICC for total score interrater reliability was .96, with section score ICCs ranging from .79 to .96 in patients with PD. The high correlations (r > .90) between the BESTest and other clinical balance measures indicated that, to some extent, the BESTest measures similar balance constructs as the BBS, PASS, CB&M, and Mini-BESTest. Our results are in line with previous findings of a high correlation among the BESTest and BBS (r = .87) and the Mini-BESTest (r = .96) in patients with PD.

Table 3. Cutoff Points With Associated Sensitivity, Specificity, Likelihood Ratios, and Area Under the Curve (AUC) of the Receiver Operating Characteristic Curve Plot for the Balance Evaluation Systems Test (BESTest), Mini-Balance Evaluation Systems Test (Mini-BESTest), and Berg Balance Scale (BBS)

| Distinguishing HFA From LFA | BESTest | Mini-BESTest | BBS |
|-----------------------------|---------|--------------|-----|
| Cutoff point                | >49%    | >9           | >19 |
| AUC (95% CI)                | 0.88 (0.80, 0.95) | 0.85 (0.77, 0.93) | 0.85 (0.78, 0.93) |
| Sensitivity (95% CI)        | 71.4 (54.7, 83.7) | 71.4 (54.7, 83.7) | 85.7 (70.0, 94.1) |
| Specificity (95% CI)        | 91.4 (76.7, 97.7) | 85.7 (70.0, 94.1) | 68.6 (51.9, 81.5) |
| LR+ (95% CI)                | 8.33 (2.77, 25.09) | 5.00 (2.16, 11.56) | 2.73 (1.64, 4.53) |
| LR− (95% CI)                | 0.31 (0.18, 0.53) | 0.21 (0.09, 0.48) | 0.33 (0.19, 0.57) |
| Accuracy                    | 0.81    | 0.79         | 0.77 |

*HFA = high functional ability, LFA = low functional ability, 95% CI = 95% confidence interval, LR+ = positive likelihood ratio, LR− = negative likelihood ratio.*

Although all of the clinical balance measures observed in this study can be used to assess balance impairments in patients with subacute stroke, this study demonstrated that the BESTest may be more desirable than other balance scales due to its lack of floor and ceiling effects. The BESTest was shown to be suitable for patients with subacute stroke across many levels of functional ability, as evidenced by the distribution of BESTest scores that covered the
whole range of functional ability. In contrast, a significant floor effect was found for the Mini-BESTest, particularly in patients with low functional ability, and a significant floor effect was present in all patients with subacute stroke using the CB&M. The Mini-BESTest included dynamic balance in standing and walking, but most patients in the LFA group had not yet regained their ability to stand or walk independently; thus, they received scores of 0 on these items of the Mini-BESTest, resulting in a floor effect (34.3%). However, no floor effect has been reported when using the Mini-BESTest in community-dwelling people with chronic stroke, indicating that this assessment scale is more appropriate for community-dwelling people poststroke.21 Similarly, the CB&M has been recommended for ambulatory patients with stroke who have moderate to mild poststroke deficits,40 and our results supported that the assessment tasks in the CB&M were too difficult for the patients with low functional ability.

In this study, the calculation of the ceiling effect was performed 2 ways: by use of maximum score and by use of the score within top 10% of test (responsive ceiling effect). Our results are in agreement with Leddy et al,34 who found that the use of the score within the top 10% would give more information about the ability of the scale to measure balance progress over time rather than using only the maximum score. Our results demonstrated the responsive ceiling effect of the BBS (45.7%) and PASS (60%) in the HFA group, suggesting that the BBS and PASS may have limited ability to detect balance progress over time, which is very important for evaluating the effectiveness of balance rehabilitation in patients with high functional ability.40 The BBS and PASS had scores within the top 10% of 51 to 56 and 34 to 36, respectively, and the scores must change at least 5.8 points41 on the BBS and by 2.2 points29 on the PASS to indicate a true change that is not due to measurement error. Therefore, the BBS and PASS would not be proper to administer in patients with high functional stroke, as they have insufficient room to address true balance progression.

The BESTest was developed to provide information on which particular balance systems were underlying balance impairments. Results from our study confirmed that balance deficits in patients with subacute stroke covered impairments of all postural control subsystems represented in the BESTest (Tab. 2), but the amount of subsystem deficit varied depending on the level of motor impairment. For example, patients with LFA had deficits in all postural control systems (mean score of the BESTest was less than 50%), except the stability limit and verticality subsystems. In contrast, patients in the HFA group showed more deficits in stability in gait. Therefore, the use of the BESTest in patients with subacute stroke can guide clinicians to tailor their intervention to impairments of specific postural control subsystems. Although we did not study the responsiveness of the BESTest, the differences in the item scores between the LFA and HFA groups indicated the potential of the BESTest to detect change in performance. All items in the BESTest, with the exception of base of support and sitting verticality (left and right), showed high median scores and high percentage of maximum scores in the HFA group compared with the LFA group. Our participants received a daily rehabilitation program that included interventions to maintain joint range of motion and prevent deformity. Therefore, deformity or pain in the feet was rare in our participants, resulting in an initially high score on the base of support item. Inaccurate representation of verticality is often found in patients with stroke who have neglect or pusher syndrome.42,43 However, none of our participants had been diagnosed with either neglect or pusher syndrome, and they all had perfect scores on the verticality items.

In this study, we also questioned whether the scores from the BESTest, Mini-BESTest, and BBS could classify patients with a high versus low level of functional ability as measured with the FM-motor. This study demonstrated that the scores from all 3 balance scales could be used to classify patients with stroke into 2 functional ability groups. However, the slightly larger LR+ values suggest the BESTest may be preferable to the BBS and Mini-BESTest for functional classification. Although the time for administering the BESTest (30–35 minutes) is twice as long as for the Mini-BESTest or the BBS (10–15 minutes), the information gathered with the BESTest is more useful compared with the BBS and Mini-BESTest in customizing balance intervention and more sensitive than the Mini-BESTest to detect changes in balance performance. However, we question whether some items in the BESTest that had maximum scores in the first assessment may have a floor effect and could limit item responsiveness in patients with subacute stroke. Therefore, further analysis, such as Rasch analysis, may be required on BESTest data from patients with subacute stroke to eliminate the least sensitive and redundant items, which will shorten the assessment time. Future studies about sensitivity to detect change and minimal detectable changes of the BESTest also are essential for evaluating the effectiveness of treatment programs. Furthermore, our study was carried out only in patients with early stroke; thus, the results cannot be generalized to patients with chronic stroke.
Validation of BESTest in Stroke

In conclusion, the BESTest is a reliable, valid, and comprehensive measure of balance that can be used in people with subacute stroke throughout different levels of functional ability without floor and ceiling effects.

This study was supported by the Thailand Research Fund, the Office of Higher Education Commission, and Srinakharinwirot University (grant no. RSA5580002).

DOI: 10.2522/ptj.20130558

References
1 Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med. 2010;46:239–248.
2 deHaart M, Geurts AC, Huidekoper SC, de Geus EO, de la Rie MA, van Wijck FL. The Berg Balance Scale: a reliable tool for measuring balance in stroke patients: a rehabilitation cohort study. Arch Phys Med Rehabil. 2004;85:886–895.
3 Mizrahi J, Solzi P, Ring H, Nisell R. Postural stability in stroke patients: vectorial expression of asymmetry, sway activity and relative sequence of reactive forces. Med Biol Eng Comput. 1989;27:181–190.
4 Tyson SF, Hanley M, Chilla J, et al. Balance disability after stroke. Phys Ther. 2006;86:30–38.
5 Bonan IV, Colle FM, Guichard JP, et al. Reliance on visual information after stroke, part I: balance on dynamic posturography. Arch Phys Med Rehabil. 2004;85:268–273.
6 Di Fabio RP, Badke MB. Stance duration under sensory conflict conditions in patients with hemiplegia. Arch Phys Med Rehabil. 1991;72:292–295.
7 Shumway-Cook A, Woollacott MH. Motor Control: Translating Research to Practice. 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2012.
8 Rogers MW, Hedman LD, Pai YC. Kinetic analysis of dynamic transitions in stance support accompanying voluntary leg flexion movements in hemiparetic adults. Arch Phys Med Rehabil. 1995;76:19–25.
9 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.
10 Marigold DS, Eng JJ. Altered timing of postural reflexes contributes to falling in persons with chronic stroke. Exp Brain Res. 2006;171:459–468.
11 Brown LA, Sleik RJ, Winder TR. Attentional demands for static postural control after stroke. Arch Phys Med Rehabil. 2002;83:1752–1755.
12 Horak FB. Clinical assessment of balance disorders. Gait Posture. 1997;5:76–84.
13 Blum L, Kornier-Bitseny N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. Phys Ther. 2008;88:559–566.
14 Horak FB, Wirsel DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther. 2009;89:484–498.
15 Franchignoni F, Horak F, Godi M, et al. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med. 2010;42:525–531.
16 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. 2011 Oct 4 [Epub ahead of print]. doi: 10.1155/2012/375419.
17 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.
18 Godi M, Franchignoni F, Caligari M, et al. Comparison of reliability, validity, and responsiveness of the mini-BESTest and Berg Balance Scale in patients with balance disorders. Phys Ther. 2013;93:158–167.
19 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. 2011 Nov 30 [Epub ahead of print]. doi: 10.1155/2012/237673.
20 Duncan RP, Leddy AL, Cavanaugh JT, et al. Comparative utility of the BESTest, Mini-BESTest, and Brief-BESTest for predicting falls in individuals with Parkinson disease: a cohort study. Phys Ther. 2013;93:542–550.
21 Tsang CS, Liao LR, Chung RC, Pang MY. Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. Phys Ther. 2015;95:1102–1115.
22 Sullivan KJ. On “Modified constraint-induced therapy…” Phys Ther. 2007;87:1560.
23 Dick JP, Guillof RJ, Stewart A, et al. Mini-Mental State Examination in neurological patients. J Neurol Neurosurg Psychiatry. 1984;47:496–499.
24 Tombah TN, McIntyre NJ. The Mini-Mental State Examination: a comprehensive review. J Am Geriatr Soc. 1992;40:922–935.
25 Shewan CM, Kertesz A. Reliability and validity characteristics of the Western Aphasia Battery (WAB). J Speech Hear Disord. 1980;45:508–524.
26 Gladstone DJ, Danells CJ, Black SE. The Fifteen minute assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair. 2002;16:232–240.
27 Duncan PW, Goldstein LB, Matchar D, et al. Measurement of motor recovery after 3rd ed. outcome assessment and sample size requirements. Stroke. 1992:23:1084–1089.
28 Berg KO, Maki B, Williams JL, et al. Clinical and laboratory measures of postural balance in an elderly population. Arch Phys Med Rehabil. 1992;75:1073–1080.
29 Chien CW, Hu MH, Tang PF, et al. A comparison of psychometric properties of the smart balance master system and the postural assessment scale for stroke in people who have had mild stroke. Arch Phys Med Rehabil. 2007;88:374–380.
30 Howe JA, Inness EL, Venturini A, et al. The Community Balance and Mobility Scale: a balance measure for individuals with traumatic brain injury. Clin Rehabil. 2006;20:885–895.
31 Jorstad EC, Hauer K, Becker C, et al. Measuring the psychological outcomes of falling: a systematic review. J Am Geriatr Soc. 2005;53:501–510.
32 de Oliveira CB, de Medeiros IR, Frota NA, et al. Balance control in hemiparetic stroke patients: main tools for evaluation. J Rehabil Res Dev. 2008;45:1215–1226.
33 Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
34 Leddy AL, Crownover BE, Earhart GM. Functional Gait Assessment and Balance Evaluation Systems Test: reliability, validity, sensitivity, and specificity for identifying individuals with Parkinson disease who fall. Phys Ther. 2011;91:102–113.
35 Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 3rd ed. Norwalk, CT: Appleton & Lange; 1993.
36 Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil. 2000;81:515–520.
37 McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). III: tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care. 1994;32:40–66.
38 Akobeng AK. Understanding diagnosis tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr.* 2007;96:487–491.

39 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.

40 Knorr S, Brouwer B, Garland SJ. Validity of the Community Balance and Mobility Scale in community-dwelling persons after stroke. *Arch Phys Med Rehabil.* 2010;91:890–896.

41 Stevenson TJ. Detecting change in patients with stroke using the Berg Balance Scale. *Aust J Physiother.* 2001;47:29–38.

42 Ting DSJ, Pollock A, Dutton GN, et al. Visual neglect following stroke: current concepts and future focus. *Surv Ophthalmol.* 2011;56:114–134.

43 Abe H, Kondo T, Oouchida Y, et al. Prevalence and length of recovery of Pusher syndrome based on cerebral hemispheric lesion side in patients with acute stroke. *Stroke.* 2012;43:1654–1656.