



COMPARISON OF MEASUREMENT PROPERTIES OF THREE SHORTENED VERSIONS OF THE BALANCE EVALUATION SYSTEM TEST (BESTest) IN PEOPLE WITH SUBACUTE STROKE

Thitimard WINAIRUK, BS¹, Marco Y. C. PANG, PhD², Vitoon SAENGSRISUWAN, PhD³, Fay B. HORAK, PhD⁴ and Rumpa BOONSINSUKH, PhD¹

From the ¹Division of Physical Therapy, Faculty of Physical Therapy, Srinakharinwirot University, Nakhon Nayok, Thailand, ²Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, ³Department of Physiology, Faculty of Science, Mahidol University, Bangkok, Thailand, ⁴Department of Neurology, Oregon, Oregon Health and Science University, Beaverton, OR, USA

Objectives: To examine the reliability, validity and responsiveness of 3 different short versions of the Balance Evaluation Systems Test (BESTest: S-BESTest, Brief-BESTest and Mini-BESTest) in patients with subacute stroke.

Design: A prospective cohort study.

Participants: Patients with subacute stroke.

Methods: Patients were assessed using the full BESTest. Scores of 3 short-form BESTests were later extracted. The intra-rater and inter-rater reliability ($n=12$) were gathered from 5 raters. Concurrent validity was assessed with the Berg Balance Scale (BBS). Floor/ceiling effect, internal responsiveness and external responsiveness with the BBS ($n=70$) were assessed at baseline, 2 weeks and 4 weeks post-rehabilitation.

Results and conclusion: All short-form BESTests demonstrated excellent intra-rater and inter-rater reliability (intraclass correlation coefficient (ICC) = 0.95–0.99) and excellent concurrent validity ($r=0.93$ – 0.96). Unlike the Brief-BESTest and Mini-BESTest, the S-BESTest and BESTest had no significant floor/ceiling effects (<20%). The standardized response mean of all 4 BESTest versions were large, ranging between 1.19 and 1.57, indicating sufficient internal responsiveness. The area under the curve of the S-BESTest and BESTest were significantly higher than the Brief-BESTest and Mini-BESTest, reflecting better accuracy of the S-BESTest and BESTest in identifying patients with subacute stroke who had balance improvement using the minimal clinically important difference of 6 and 16 points, respectively. These findings suggest that the S-BESTest is a short-form BESTest that is appropriate for assessing balance impairments in patients with subacute stroke.

Key words: psychometric; physical therapist; postural balance; minimal clinically important difference; patient-reported outcome measures; cerebrovascular disease.

Accepted Aug 9, 2019; Epub ahead of print Aug 23, 2019

J Rehabil Med 2019; 51: 683–691

Correspondence address: Rumpa Boonsinsukh, Division of Physical Therapy, Faculty of Physical Therapy, Srinakharinwirot University, 63 moo 7 Rungsit-Nakhonnayok, Ongkharuk, Nakhon Nayok, Thailand 26120. E-mail: rumpa@g.swu.ac.th

LAY ABSTRACT

The aim of this study is to determine which short versions of the Balance Evaluation System Test (BESTest), S-BESTest, Brief-BESTest and Mini-BESTest, are most appropriate for assessing balance impairments in patients with subacute stroke. Participants were patients with subacute stroke: 12 in the reliability assessment and 70 in the validity testing. Patients were assessed using 3 short-form BESTests. All short-form BESTests demonstrated excellent reliability and excellent validity, but the S-BESTest demonstrated better accuracy in identifying patients with subacute stroke who had balance improvement using the cutoff score of 6 points. These findings suggest that the S-BESTest is a short-form BESTest that is appropriate for assessing balance impairments in patients with subacute stroke.

Postural control involves complex co-operation between several systems to maintain the centre of body mass above its base of support (postural equilibrium) and to control body alignment with reference to itself and the environment (postural orientation). These systems include musculoskeletal components and neuromuscular synergies, individual sensory systems and sensory strategies, internal representations, adaptive and anticipatory mechanisms (1). One problem commonly found in individuals post-stroke is postural control or balance impairment. This balance problem can be related to impairment in each postural control system; for example, prolonged anticipatory reaction time during affected side stepping (2), diminished and delayed adaptive responses on the affected side (3–5), and abnormal sensory integration (6). To assess balance impairments in patients with stroke, the Berg Balance Scale (BBS), a 14-item functional balance test, is commonly used as a gold standard (7). Although the BBS is useful in reporting the presence of balance impairments, it cannot specify which system of postural control is impaired.

The Balance Evaluation System Test (BESTest) is a clinical scale developed to assess the systems of postural control through 6 domains: biomechanical constraints, stability limits/verticality, anticipatory postural adjustments, postural responses, sensory

orientation, and gait stability. The BESTest has been validated to assess postural control impairments in various populations (8–11). In individuals with subacute stroke, the BESTest showed excellent intra-rater and inter-rater reliability as well as significant correlation with the BBS, Postural Assessment Scale for Stroke (PASS), and Community Balance and Mobility scale (CB&M) suggesting concurrent validity (12). With high sensitivity (80.8%), specificity (87.5%) and post-test accuracy (84%), the BESTest demonstrated precision in specifying patients with stroke who have balance improvements, using a 10% increase in score as an indicator (13). The BESTest showed an advantage over other balance assessment tools in patients with stroke when it did not have floor or ceiling effects, but the long administration time (35 min) can limit its practicality in the clinic.

Two short versions of the BESTest, which can reduce the assessment time are currently available. The brief-BESTest, which contains only 6 items, one for each domain of the BESTest, was validated in patients with chronic stroke, but has not been validated in patients with subacute stroke (14, 15). The Mini-BESTest is another short version, which deletes the first and second domains of the BESTest to evaluate the dynamic component of postural control (16). The Mini-BESTest showed excellent internal consistency for community-dwelling patients with chronic stroke as well as excellent reliability and concurrent validity in patients with subacute stroke (12, 17). However, the Mini-BESTest had a floor effect in patients with subacute stroke who had low functional ability during day 27 through day 94 (12), limiting its use in this group of patients.

The S-BESTest is our newly developed short version of the BESTest for patients with subacute stroke aiming to reduce the assessment time and floor effect while retaining all domains of the BESTest. Using Rasch analysis partial credit model to reduce the items (18), the S-BESTest contains 13 items (total 39 points) using a similar scoring system as the original BESTest (see Table I for comparison of the original and 3 shortened BESTest). The construct validity of the S-BESTest was confirmed by performing hypothesis testing on the known group (19), but other psychometric properties of the S-BESTest, such as reliability, validity, floor and ceiling effect, and responsiveness have not been assessed. Therefore, it is unclear which short version of the BESTest is most appropriate, in terms of having highest responsiveness and lowest floor/ceiling effect, for assessing patients with subacute stroke. This study, therefore, aimed to compare the reliability, validity, floor and ceiling effect and responsiveness of 3 shortened versions of the BESTest (S-BESTest, the Brief-BESTest, the

Mini-BESTest) and the original BESTest in patients with subacute stroke. To reduce the learning effect and recall bias of the assessor due to repeated scoring, the original BESTest was administered to each patient and scores of 3 shortened versions of the BESTest were extrapolated from the BESTest scores.

METHODS

Participants

Twelve patients with subacute stroke were recruited from departments of physical therapy at Lerdsin Hospital in Bangkok, Thailand for assessing the reliability of the scales. The sample size calculation was estimated from a power of 0.80 and alpha level of 0.05. A null intraclass correlation coefficient (ICC) of 0.60 and expected correlation coefficients of 0.93 were determined by a previous study (20, 21). The inclusion criteria were: diagnosis of a first unilateral hemispheric stroke, onset within 4 months, stable vital signs, and ability to follow instructions. Participants were excluded if they had any neurological disorder other than stroke, unstable epilepsy, lesion at the brainstem involving sleep-wake and respiratory control centres or cerebellum, cerebral aneurysm, visual problems that have not been resolved with glasses, and cognitive impairment as measured by the Mini-Mental State Examination (MMSE score ≤ 23) (22, 23).

Another 70 patients were recruited at the same hospital for the assessment of validity and responsiveness using similar inclusion and exclusion criteria. Sample size calculation was based on a power of 0.80, alpha level of 0.05, correlation coefficient (r) of 0.78 and an expected correlation coefficient of 0.8 (21). Since the level of functional ability influences the recovery process, to ensure that this study represents sufficient low and high level of functional ability, the lower extremity motor function domain of the Fugl-Meyer Assessment (FM-LE) was used to classify the subjects into 2 functional level groups (35 patients in each group) for recruitment purposes. A FM-LE score of 0–14 was classified as low functional ability and a score of higher than 14 was classified as high functional ability (24). The Institutional Review Board of Lerdsin Hospital (number 0306/13/127) approved the study protocol and all patients gave written consent prior to participation.

Data collection

Prior to the tests, all raters were first trained to score healthy subjects, and then patients with stroke. The reliability was assessed through videotape rating to ensure consistency of performance and reduce the error from movement variability. Validation for using the videotapes was first determined by one physical therapist who had 10 years of experience in stroke rehabilitation. This rater scored the patient's performance both at the time of the test and also at 7 days later from videotape to confirm that the results from concurrent scoring and videotape scoring were not different. Intra-rater and inter-rater reliability were later assessed using 5 physical therapists: 3 from Lerdsin Hospital, with stroke rehabilitation experience of 1, 5, and 10 years, respectively and 2 PhD physical therapy students. 5 raters scored each patient's performance from videotape on 2 separate occasions within 7 days. Each rater did not discuss scoring among themselves and scored the patients' performance on separate scoring worksheets on each occasion. Intra-rater reliability of total scores and domain scores were determined

Table I. Summary of the Balance Evaluation Systems Test (BESTest: S-BESTest, Brief-BESTest and Mini-BESTest)

	BESTest	Mini-BESTest	Brief-BESTest	S-BESTest
Number of tasks	27	13	6	11
Number of items	36	16	8	13
Domains and Items of the scale				
Domain I: Biomechanical Constraints				
1. Base of support	/			
2. CoM alignment	/			
3. Ankle strength and range	/			
4. Hip/trunk lateral strength	/		/	/
5. Sit on floor and standup	/			
Domain II: Stability limits				
6. Lateral Lean Paretic Side	/			
7. Lateral lean non-paretic side	/			
8. Verticality paretic side	/			
9. Verticality non-paretic side	/			
10. Functional reach-forward	/		/	
11. Functional reach-lateral paretic side	/			
12. Functional reach-lateral non-paretic side	/			/
Domain III: Anticipatory postural adjustment				
13. Sit to stand	/	/		
14. Rise to toes	/	/		/
15. Stand on paretic leg	/	/	/	/
16. Stand on non-paretic leg	/	/	/	/
17. Alternate stair touching	/			
18. Standing arm raise	/			/
Domain IV: Reactive postural response				
19. In place response-forward	/			
20. In place response-backward	/			
21. Compensatory stepping correction-forward	/	/		
22. Compensatory stepping correction-backward	/	/		
23. Compensatory stepping correction-lateral paretic side	/	/	/	/
24. Compensatory stepping correction-lateral non-paretic side	/	/	/	
Domain V: Sensory orientation				
25. Eyes open, firm surface	/	/		
26. Eyes closed, firm surface	/			/
27. Eyes open, foam surface	/			/
28. Eyes closed, foam surface	/	/	/	
29. Incline-eyes closed	/	/		/
Domain VI: Stability in gait				
30. Gait-level surface	/			
31. Change in gait speed	/	/		/
32. Walk with head turns	/	/		/
33. Walk with pivot turns	/	/		
34. Step over obstacle	/	/		
35. Timed up and go (TUG)	/		/	
36. TUG with dual task	/	/		/
Scoring scale	0–3	0–2	Same as BESTest	Same as BESTest
Total score	108	28	24	39
Estimated time to complete, min	30–45	10–15	5–7	7–10

by comparing the score of time 1 and time 2 for each rater. The S-BESTest and Brief-BESTest scores were extracted from the relevant subset of the BESTest items. As scoring scale of the Mini-BESTest was different from other BESTest versions (Table I), the reliability of the Mini-BESTest was assessed separately 1 month later to prevent recall bias by using the same videotape and repeated in 7 days.

For the assessment of validity and responsiveness, 70 patients with stroke who were referred to physical therapy at Lerdsin Hospital were assessed using the BESTest and BBS 3 times: at baseline, at 2 weeks and at 4 weeks after rehabilitation. This allowed for comparison of psychometric properties across time. Baseline demographic and clinical information was gathered from the patients and charts. Before testing, vital signs were monitored to ensure stable medical status. A single therapist administered the BESTest before the BBS on all 70 patients at 3 time-points. The scoring methods of the S-BESTest, Brief-BESTest and Mini-BESTest were similar to those employed in the reliability section. A 15-point Global Rating of Change

(GRC) where -7 indicates a very great deterioration, $+7$ indicates a very great improvement and 0 indicates no change, was also administered to the patient. Patients independently rated the overall change (using GRC criteria) in their balance performance at 2 weeks and at 4 weeks after rehabilitation (25). All patients received the same verbal instruction and were allowed to rest as long as they required. Patient's performance was videotaped for further review and analysis. Total assessment time was approximately 1.5 h. If the test could not be completed in 1 day, it was continued on the next day.

Data analysis

The correlations between the scores from the test and videotape were examined using the Spearman rank-order correlations. Correlation coefficients of 0.80 or higher indicated excellent correlation, 0.50 – 0.79 indicated moderate correlation, and 0.00 – 0.49 indicated poor correlation (26). Intra-rater and inter-rater reliability were calculated using ICC model 3, k

Table II. Intra-rater reliability of the S-BESTest, Brief-BESTest, and Mini-BESTest in people with subacute stroke ($n=12$)

	Intra-rater reliability					
	S-BESTest		Brief-BESTest		Mini-BESTest	
	ICC (3,5)	95% CI	ICC (3,5)	95% CI	ICC (3,5)	95% CI
Total	0.98	0.97–0.99	0.98	0.97–0.99	0.98	0.97–0.99
Domain I: Biomechanical constraints	0.97	0.93–0.99	0.97	0.93–0.99	–	–
Domain II: Stability limits	0.98	0.96–0.99	0.99	0.98–0.99	–	–
Domain III: Anticipatory postural adjustment	0.96	0.92–0.99	0.96	0.91–0.99	0.95	0.90–0.99
Domain IV: Reactive postural response	0.96	0.92–0.99	0.98	0.95–0.99	0.97	0.94–0.99
Domain V: Sensory orientation	0.97	0.94–0.99	0.99	0.98–0.99	0.95	0.90–0.98
Domain VI: Stability in gait	0.96	0.92–0.99	0.97	0.94–0.99	0.98	0.96–0.99

All intraclass correlation coefficient (ICCs) were significant, with p -value of <0.001 . 95% CI: 95% confidence interval.

and 2, k , respectively, for the S-BESTest, Brief-BESTest and Mini-BESTest (27). The ICC values were interpreted using the criteria: 0.8 indicates good reliability, 0.8–0.6 indicates moderate reliability and 0.6–0.4 indicates weak reliability (20, 27).

The concurrent validity of the S-BESTest, Brief-BESTest and Mini-BESTest was assessed with the BBS using the Spearman rank-order correlations. Floor and ceiling effect of S-BESTest, Brief-BESTest, Mini-BESTest, and BESTest were calculated as the percentage for minimum or maximum possible scores of the sample scoring, respectively. Floor and ceiling effects greater than or equal to 20% were interpreted as significant (28). Comparisons of balance scores between baseline and 2 weeks post-rehabilitation and between 2 and 4 weeks post-rehabilitation were analysed using paired t -test with significance level $p < 0.05$.

Internal and external responsiveness of the S-BESTest, Brief-BESTest, Mini-BESTest, and BESTest were assessed. Internal responsiveness refers to the possibility of detecting any change before and after a known treatment. Internal responsiveness was examined using the standardized response mean (SRM) and minimal detectable change (MDC) (29, 30). SRM of 0.8 or greater represented a large change, values from 0.5 to 0.8 represented moderate change, and values of 0.2–0.5 represented small change. MDC was calculated as the standard error of the mean (SEM) multiplied by $1.96(\sqrt{2})$ (29, 30). SEM was calculated as standard deviation (SD) multiplied by $\sqrt{(1-\text{reliability})}$. The limitation of internal responsiveness is that it lacks information on the quality of changes, such as worsening or improvement (31).

In contrast, external responsiveness is associated with the concept of clinical relevance, which depends on the choice of external standard (32). In this study, 2 external standard scales, the BBS and the GRC were selected to compare between the change in performance and patient's own perception. External responsiveness was assessed by using receiver operator curve (ROC) analysis to establish which version of the BESTest could best identify patients whose balance had improved using a change in BBS score of 7 points as the milestone value for deciding if change had occurred (33, 34). The area under the curve (AUC) value was used to reflect this. The AUC values were compared across test versions using a t -test and significance level of 0.05. The ROC analysis was repeated using the change in GRC (5 points) (25). The AUC was used to interpret the probability of correctly discriminating between patients with and without balance improvement (29). An AUC of 0.8 or greater indicated excellent discrimination (27). Paired t -test was used to compare the AUC between 2 testing scales with significance level at $p < 0.01$. A likelihood ratio demonstrates accuracy of post-test probabilities; values of LR+ above 5 and values of LR– below 0.2 were considered meaningful (27). The optimal cut-off score was also chosen from the sensitivity and specificity (27).

RESULTS

Reliability

A total of 12 patients with stroke (8 males and 4 females) were included in the reliability assessment. The mean age of patients was 58.42 years (SD 13.41 years) with a mean time since stroke onset of 40.60 days (SD 45.39 days). Correlation between concurrent test with videotape scoring of the S-BESTest total scores ($r=0.97$) and domain scores (r from 0.90 to 1.0) were excellent. The mean and SD of the S-BESTest scores at time 1 (day 1) and time 2 (day 7) were 20.45 (SD 0.66) and 20.53 (SD 1.13), respectively. Those scores of the Mini-BESTest were 12.62 (SD 1.11) (day 1) and 12.52 (SD 1.46) (day 7) and the scores of the Brief-BESTest were 8.32 (SD 0.53) (day 1) and 8.23 (SD 0.80) (day 7). The intra-rater and inter-rater reliability of the total score and domain score of the 3 short-form BESTests were excellent (ICC=0.86–0.99) (Tables II and III).

Concurrent validity and floor-ceiling effect

Demographic and clinical characteristics of 70 patients with subacute stroke selected for the validity and responsiveness study are presented in Table IV. All patients had lower extremity and balance impairment based on the FM-LE and BESTest scores. High correlations between the BBS and the BESTest ($r=0.96$), the S-BESTest ($r=0.95$), the Brief-BESTest ($r=0.93$) and the Mini-BESTest ($r=0.95$) were observed, indicating excellent concurrent validity of all versions of the BESTest. Table V shows floor and ceiling effect using all BESTest scales across 3 intervals. It can be seen that the number of patients with minimum scores decreased between baseline and 4 weeks post-rehabilitation, while the number of patients with maximum scores increased over time. The Brief-BESTest and the Mini-BESTest demonstrated a significant floor effect at baseline, but the S-BESTest and the BESTest showed no significant floor effect ($<20\%$). Although all 4 balance scales showed no ceiling effect at 4 weeks,

Table III. Inter-rater reliability of the S-BESTest, Brief-BESTest, and Mini-BESTest in people with subacute stroke (*n* = 12)

	Inter-rater reliability					
	S-BESTest		Brief-BESTest		Mini-BESTest	
	ICC (2,5)	95% CI	ICC (2,5)	95% CI	ICC (2,5)	95% CI
Total	0.95	0.91–0.98	0.98	0.97–0.99	0.98	0.96–0.99
Domain I: Biomechanical constraints	0.91	0.78–0.97	0.97	0.93–0.99	–	–
Domain II: Stability limits	0.96	0.90–0.99	0.99	0.98–0.99	–	–
Domain III: Anticipatory postural adjustment	0.88	0.75–0.96	0.96	0.91–0.99	0.92	0.84–0.97
Domain IV: Reactive postural Response	0.90	0.78–0.97	0.98	0.95–0.99	0.97	0.94–0.99
Domain V: Sensory orientation	0.91	0.82–0.97	0.99	0.98–0.99	0.86	0.72–0.95
Domain VI: Stability in gait	0.91	0.83–0.97	0.97	0.94–0.99	0.97	0.95–0.99

All intraclass correlation coefficient (ICCs) were significant, with *p*-value of <0.001. 95% CI: 95% confidence interval.

the Brief-BESTest had the highest number of patients who had the maximum score (15.7%).

Internal responsiveness

After 2 and 4 weeks of rehabilitation, the majority of patients showed improvement of balance performance as demonstrated by significant increases in total scores of the S-BESTest, Brief-BESTest, Mini-BESTest, and BESTest (Table VI). Those score changes were larger

during 0–2 weeks than 2–4 weeks in all BESTest versions, indicating larger recovery at 0–2 weeks. The number of patients who had no change in score post-rehabilitation was seen only when assessing with the BESTest. The highest number of those who had no change in score was found when using the Brief-BESTest at 2–4 weeks post-rehabilitation. The SRM of the S-BESTest, Brief-BESTest, Mini-BESTest, and BESTest were large, ranging from 1.19 to 1.57, indicating sufficient internal responsiveness. Minimal

Table IV. Demographic and clinical characteristics of patients with subacute stroke (*n* = 70) in validity and responsiveness study

Characteristics	Mean (SD)	Range
Age, years	55.24 (12.11)	30–77
Sex, M/F, <i>n</i>	44/26	
Time since stroke, days	15.81 (15.6)	7–103
Type of stroke ischaemic/haemorrhagic, <i>n</i>	64/6	
Affected side (right/left), <i>n</i>	37/33	
MMSE (/30)	27.33 (1.87)	24–30
FM-LE (/34)	19.39 (10.06)	2–34
BBS (/56)	31.24 (19.96)	0–56
S-BESTest (/39)	17.41 (12.73)	0–39
Brief-BESTest (/24)	8.80 (7.46)	0–23
Mini-BESTest (/28)	10.93 (8.58)	0–26
BESTest (/108)	55.26 (34.15)	0–104

MMSE: Mini-Mental State Examination; FM-LE: Fugl-Meyer Stroke Assessment-lower extremity motor subscale; BBS: Berg Balance Scale; S-BESTest: Stroke-Balance Evaluation Systems Test; Brief-BESTest: Brief-Balance Evaluation Systems Test; Mini-BESTest: Mini-Balance Evaluation Systems Test; BESTest: Balance Evaluation Systems Test; SD: standard deviation.

Table V. Floor and ceiling effect of the S-BESTest, Brief-BESTest, Mini-BESTest, and BESTest

Patients with subacute stroke (<i>n</i> = 70)	Baseline <i>n</i> (%)	2 weeks <i>n</i> (%)	4 weeks <i>n</i> (%)
S-BESTest			
Floor effect	13 (18.6)	3 (4.3)	0 (0.0)
Ceiling effect	0 (0.0)	3 (4.3)	8 (11.4)
Brief-BESTest			
Floor effect	14 (20)	2 (2.9)	0 (0.0)
Ceiling effect	0 (0.0)	3 (4.3)	11 (15.7)
Mini-BESTest			
Floor effect	15 (21.4)	4 (5.7)	0 (0.0)
Ceiling effect	0 (0.0)	2 (2.9)	5 (7.1)
BESTest			
Floor effect	4 (5.7)	0 (0.0)	0 (0.0)
Ceiling effect	0 (0.0)	0 (0.0)	3 (4.3)

S-BESTest: Stroke-Balance Evaluation Systems Test; Brief-BESTest: Brief-Balance Evaluation Systems Test; Mini-BESTest: Mini-Balance Evaluation Systems Test; BESTest: Balance Evaluation Systems Test

Table VI. Internal responsiveness of the S-BESTest total score, the Brief-BESTest total score, the Mini-BESTest total score, and the BESTest total score measure at 0–2 weeks and 2–4 weeks after physical therapy rehabilitation

Balance assessment	Before Mean (SD)	After Mean (SD)	Change Mean (SD)	SRM	No change <i>n</i> (%)	MDC95
S-BESTest (/39)						
0–2 weeks	17.41 (12.73)	25.27 (10.93)	7.86 (6.14)*	1.28	2 (2.86)	4.99
2–4 weeks	25.27 (10.93)	30.31 (7.93)	5.04 (3.91)*	1.29	5 (7.14)	4.28
Brief-BESTest (/24)						
0–2 weeks	8.79 (7.46)	14.24 (7.00)	5.46 (3.47)*	1.57	3 (4.28)	2.92
2–4 weeks	14.24 (7.00)	17.99 (5.19)	3.74 (2.82)*	1.33	9 (12.86)	2.74
Mini-BESTest (/28)						
0–2 weeks	10.93 (8.58)	16.14 (8.58)	5.21 (4.06)*	1.28	4 (5.71)	3.35
2–4 weeks	16.14 (8.58)	19.94 (6.62)	3.80 (3.18)*	1.19	4 (5.71)	3.35
BESTest (/108)						
0–2 weeks	55.23 (34.15)	77.29 (26.32)	22.06 (17.90)*	1.23	1 (1.43)	9.47
2–4 weeks	77.29 (26.32)	90.66 (16.98)	13.37 (10.85)*	1.23	0 (0)	7.29

*Significant difference between before and after rehabilitation (*p* < 0.001). SD: standard deviation; SRM: standardized response mean; *n* (%) no change: number of patients showed no change in score; MDC95: minimal detectable change at 95% confidence interval.

Table VII. Minimal clinically important difference (MCID) and related parameters of the S-BESTest, the Brief-BESTest, the Mini-BESTest, and the BESTest to identify balance improvement using the Berg Balance Scale (BBS) as reference

External responsiveness	S-BESTest (/39)	Brief-BESTest (/24)	Mini-BESTest (/28)	BESTest (/108)
0–2 weeks				
MCID	6.5	5.5	7.5	18.5
AUC	0.84 (0.75–0.94)	0.77*** (0.66–0.88)	0.68*** (0.56–0.81)	0.89 (0.82–0.98)
Sensitivity	0.78 (0.60–0.91)	0.63 (0.46–0.78)	0.91 (0.75–0.98)	0.79 (0.63–0.90)
Specificity	0.82 (0.66–0.92)	0.84 (0.67–0.95)	0.39 (0.24–0.57)	0.94 (0.79–0.99)
LR+	4.24	4.04	1.50	12.63
LR–	0.27	0.44	0.24	0.22
Post-test accuracy, %	80	70	63	83
2–4 weeks				
MCID	5.5	4.5	4.5	13.5
AUC	0.89 (0.82–0.97)	0.79*** (0.68–0.91)	0.79*** (0.67–0.91)	0.89 (0.81–0.98)
Sensitivity	0.78 (0.65–0.89)	0.74 (0.49–0.91)	0.78 (0.65–0.89)	0.89 (0.67–0.99)
Specificity	0.84 (0.60–0.97)	0.72 (0.58–0.84)	0.74 (0.49–0.91)	0.86 (0.74–0.94)
LR+	4.97	2.68	2.98	6.52
LR–	0.26	0.36	0.29	0.12
Post-test accuracy, %	80	66	77	87

*Significant difference of AUC with the BESTest ($p < 0.01$).

**Significant difference of AUC with the S-BESTest ($p < 0.01$).

CI: confidence interval; AUC: area under the curve; LR+: positive likelihood ratio; LR–: negative likelihood ratio.

detectable change at 95% confidence interval (95% CI) (MDC_{95}) of the S-BESTest, Brief-BESTest, and Mini-BESTest measured at 0–2 weeks were comparable to those measured at 2–4 weeks, except the MDC_{95} of the BESTest at 0–2 weeks was higher than that at 2–4 weeks.

External responsiveness

The external responsiveness using the BBS as reference is shown in Table VII. MCID of all BESTest versions at 0–2 weeks were higher than at 2–4 weeks, indicating larger recovery at 0–2 weeks. The AUC of the S-BESTest and BESTest were significantly higher than the Brief-BESTest and Mini-BESTest, reflecting better accuracy of the S-BESTest and BESTest in identifying patients with subacute stroke who had balance improvement using the proposed MCID (Fig. 1). This

was also supported by higher post-test accuracy (80% or more) and meaningful LR+/LR– of the S-BESTest and BESTest.

Table VIII shows the external responsiveness, using the GRC as reference. It can be seen that the AUC values using the GRC as reference were lower than AUC values calculated by using the BBS as reference. None of the AUC using the GRC as reference showed excellent discrimination to identify patients with subacute stroke who had balance improvement (< 0.8). The proposed MCID was varied among all BESTest versions with low post-test accuracy, especially at 2–4 weeks, and the LR+/LR– fell outside of the acceptable range. These results indicated that the MCID using the GRC as reference had low accuracy in correctly identifying patients with subacute stroke who have balance improvement.

Table VIII. Minimal clinically important difference (MCID) and related parameters of the S-BESTest, the Brief-BESTest, the Mini-BESTest, and the BESTest to identify balance improvement using the global rating of change (GRC) as reference

External responsiveness	S-BESTest (/39)	Brief-BESTest (/24)	Mini-BESTest (/28)	BESTest (/108)
0–2 weeks				
MCID	2.5	4.5	1.5	8.5
AUC	0.73 (0.55–0.91)	0.73 (0.56–0.90)	0.71 (0.54–0.88)	0.71 (0.53–0.88)
Sensitivity	0.50 (0.23–0.77)	0.66 (0.52–0.78)	0.93 (0.83–0.98)	0.86 (0.74–0.94)
Specificity	0.91 (0.80–0.97)	0.71 (0.42–0.92)	0.43 (0.18–0.71)	0.57 (0.29–0.82)
LR+	5.60	2.31	1.62	2.00
LR–	0.55	0.47	0.17	0.25
Post-test accuracy, %	83	74	83	77
2–4 weeks				
MCID	4.5	3.5	3.5	2.5
AUC	0.69 (0.38–1)	0.68 (0.42–0.93)	0.69 (0.40–1)	0.50 (–0.35–1)
Sensitivity	1 (0.16–1)	1 (0.16–1)	0.47 (0.35–0.60)	0.50 (0.01–0.99)
Specificity	0.44 (0.32–0.57)	0.51 (0.39–0.64)	1 (0.16–1)	0.91 (0.82–0.97)
LR+	1.79	2.06	0	5.67
LR–	0	0	0.53	0.55
Post-test accuracy, %	56	41	53	5

CI: confidence interval; AUC: area under the curve; LR+: positive likelihood ratio; LR–: negative likelihood ratio.

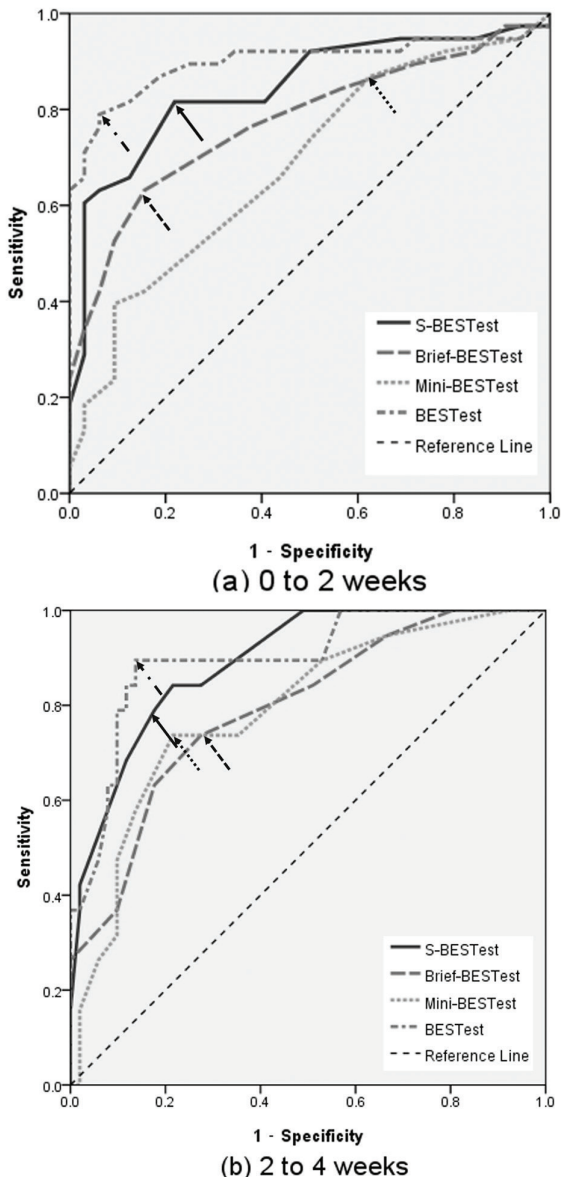


Fig. 1. Receiver operating characteristic (ROC) plot of the S-BESTest, the Brief-BESTest, the Mini-BESTest, and the BESTest scores using the Berg Balance Scale (BBS) as reference, measured at (a) 0–2 weeks (b) 2–4 weeks, in people with subacute stroke. Arrow depicts cut-off score (minimal clinically important difference; MCID).

DISCUSSION

This study compared the reliability, validity and responsiveness of 3 different short versions of the BESTest (S-BESTest, the Brief-BESTest and the Mini-BESTest) with the original BESTest in patients with subacute stroke. Our findings showed that all short forms of the BESTest were reliable and valid in this group of the population, but only the S-BESTest demonstrated no significant floor-ceiling effect with high external responsiveness in accurately identification of balance improvement similar to the original BESTest,

suggesting that the S-BESTest was more appropriate as a short version of the BESTest for assessing postural control in patients with subacute stroke.

Excellent intra-rater and inter-rater reliability of the S-BESTest, the Brief-BESTest, and the Mini-BESTest in patients with subacute stroke were consistent with a previous study in people with balance disorders, including patients with subacute stroke (12). In patients with subacute stroke, excellent concurrent validity of the S-BESTest, Brief-BESTest, Mini-BESTest and original BESTest with the BBS, the clinical gold standard of balance tests, suggest that all versions of the BESTest assess the same balance constructs as the BBS. Our finding of a strong correlation between the BESTest and the BBS ($r=0.96$) was also in agreement with previous findings (12).

This study confirmed that the original BESTest had no floor and ceiling effect similar to those reported in the previous studies (12). Although the S-BESTest did not reach significant floor effect ($<20\%$), the magnitude of floor effect at baseline of the S-BESTest was still high (18.6%) compared with the original BESTest (5.7%), suggesting that the original BESTest outperformed other shortened versions, including S-BESTest, Brief-BESTest and Mini-BESTest. We recommended the use of original BESTest in assessing balance impairment in patients with subacute stroke when there is no time constraint. However, when there is the need for reducing assessment time or the need to identify patients with balance improvement, the S-BESTest is the most appropriate choice for balance assessment in patients with subacute stroke. In contrast, the Brief-BESTest and the Mini-BESTest may be more suitable for assessing postural control in patients with chronic stroke, as previous studies found no floor effect in patients with chronic stroke using the Brief-BESTest and the Mini-BESTest (15, 17). The floor effect of the Brief-BESTest and Mini-BESTest in subacute stage may be due to the fact that these 2 scales contain items that are more difficult for patients with subacute stroke to perform. For example, the items of compensatory stepping correction in a backward direction, standing with eyes closed on a foam surface and walking with pivot turn showed no change in score when measured at 0–2 weeks. A previous study showed that protective steps in a backward direction require more supraspinal control than protective steps in a forward direction; thus, this backward protective mechanism could be more impaired in patients with stroke (35). In the same way, patients with stroke showed larger postural sway velocity while standing on foam with eyes closed, compared with other conditions of sensory orientation testing (36).

The S-BESTest, the Mini-BESTest, the Brief-BESTest, and the BESTest demonstrated good internal

responsiveness (large SRM), suggesting all 4 versions of the BESTest were sensitive in detecting the effectiveness of rehabilitation in subacute stroke. Results regarding acceptable SRM were in accordance with previous studies on psychometric properties of the BESTest in patients with subacute and community-dwelling stroke (13, 15, 17). However, when considering the external responsiveness using the BBS as the reference, our study showed that only the S-BESTest and the original BESTest demonstrated higher accuracy in identifying the patients with subacute stroke who had balance improvement with the mean MCID of 6 and 16 points, respectively. Better external responsiveness of the S-BESTest in the patients with subacute stroke, compared with the Brief-BESTest and the Mini-BESTest, may be due to the fact that the S-BESTest was developed specifically for patients with stroke. With 13 items that preserve all 6 domains of the original BESTest, the S-BESTest is likely to be better at representing impairments and activity limitations of patients with stroke than the other short forms of the BESTest. For example, patients with stroke demonstrated larger mediolateral postural swaying than healthy subjects, while antero-posterior swaying was similar between the groups (37). The item “functional reach lateral on non-paretic side” was therefore included in the S-BESTest to represent the impairment of the paretic trunk muscles to maintain posture when reaching toward the non-paretic side, whereas the brief-BESTest contains the item “functional reach forward”. Another example was the item “rise to toes”, which was included in the S-BESTest. This item can capture another common problem in stroke, where impairment of tibialis anterior would limit its anticipatory action, leading to inability to perform rise to toes in patients with stroke (38).

Although previous studies reported a high recovery rate in the lower extremity at 1–4 weeks post-stroke and a plateau phase of recovery after 6 months with smaller score changes (39, 40), this study found that internal responsiveness and external responsiveness (using BBS as reference) were lower at 2–4 weeks compared with the first 2 weeks post-rehabilitation. The larger recovery occurring in the first 2 weeks post-rehabilitation could be due to the spontaneous recovery of body functions from cerebrovascular injury, such as reduction in inflammation and swelling in the brain, which occurs intensively during the first few weeks post-stroke, together with the body adaptation from rehabilitation (39, 40). We also demonstrated that the minimal clinically important difference (MCID) of the S-BESTest, the Brief-BESTest, the Mini-BESTest, and the BESTest calculated using the BBS score change was more accurate than those calculated using the GRC score change. The GRC score was obtained from patient’s perception that may result in underestimation or overestimation from recall bias,

personal experiences and the ability to understand the context of improvement (33).

This study has several limitations concerning generalization to different cluster groups and location. The S-BESTest was developed from patients with stroke who were able to stand independently for at least 3 s. This scale should be investigated further in patients who are unable to stand or dependence in standing. Total assessment time for each patient was approximately 1.5 h. Such lengthy assessment may lead to fatigue of individual, which may interfere with test results, but the administration of the S-BESTest in a clinical setting will be free of fatigue due to shorter assessment time. Moreover, the clinician should be aware that the method of scoring the shortened version of the BESTest in this study may not be the same as that in real practice, as we extracted the scores of shortened versions of the BESTest from the scores of the full BESTest. Lastly, the responsiveness of our study was carried out for a duration of 4 weeks post-rehabilitation, when the recovery process is still occurring although slowly. It will be useful to determine the MCID for the later phase of the recovery period.

In conclusion, all 4 versions of the BESTest, the S-BESTest, Brief-BESTest, Mini-BESTest and BESTest, had excellent reliability, high concurrent validity with the BBS in patients with subacute stroke and high internal responsiveness. However, only the S-BESTest demonstrated no significant floor-ceiling effect and high external responsiveness to identify balance improvement similar to the original BESTest. Therefore, the S-BESTest is a short version of the BESTest that is appropriate for use in assessing postural control impairments in patients with subacute stroke.

ACKNOWLEDGEMENTS

The authors thank the physical therapists at Lerdsin Hospital in Bangkok, Thailand for offering their facilities and a location for patient recruitment.

Funding: This study was supported by the Thailand Research Fund (grant number RSA5580002) and the National Research Council of Thailand (grant number 073/2560).

The authors have no conflicts of interest to declare.

REFERENCES

1. Horak FB, Henry SM, Shumway-Cook A. Postural perturbations: new insights for treatment of balance disorders. *Phys Ther* 1997; 77: 517–533.
2. Chang WH, Tang PF, Wang YH, Lin KH, Chiu MJ, Chen SH. Role of the premotor cortex in leg selection and anticipatory postural adjustments associated with a rapid stepping task in patients with stroke. *Gait Posture* 2010; 32: 487–493.
3. de Kam D, Roelofs JMB, Bruijnes A, Geurts ACH, Weerdesteijn V. The next step in understanding impaired reactive balance control in people with stroke: the role of defective

- early automatic postural responses. *Neurorehabil Neural Repair* 2017; 31: 708–716.
4. Mansfield A, Inness EL, Wong JS, Fraser JE, McIlroy WE. Is impaired control of reactive stepping related to falls during inpatient stroke rehabilitation? *Neurorehabil Neural Repair* 2013; 27: 526–533.
 5. Salot P, Patel P, Bhatt T. Reactive balance in individuals with chronic stroke: biomechanical factors related to perturbation-induced backward falling. *Phys Ther* 2016; 96: 338–347.
 6. Boonsinsukh R, Panichareon L, Phansuwan-Pujito P. Light touch cue through a cane improves pelvic stability during walking in stroke. *Arch Phys Med Rehabil* 2009; 90: 919–926.
 7. Berg KO, Wood-Dauphinee SL, Williams JJ. The Balance Scale: reliability assessment with elderly residents and patients with an acute stroke. *Scand J Rehabil Med* 1995; 27: 27–36.
 8. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Phys Ther* 2009; 89: 484–498.
 9. Dewar R, Claus AP, Tucker K, Ware R, Johnston LM. Reproducibility of the Balance Evaluation Systems Test (BESTest) and the Mini-BESTest in school-aged children. *Gait Posture* 2017; 55: 68–74.
 10. Jacome C, Cruz J, Oliveira A, Marques A. Validity, reliability, and ability to identify fall status of the Berg Balance Scale, BESTest, Mini-BESTest, and Brief-BESTest in patients with COPD. *Phys Ther* 2016; 96: 1807–1815.
 11. Tamura T, Otaka Y, Konno S, Sadashima K, Tomatsu T, Machida S. The impaired balance systems identified by the BESTest in older patients with knee osteoarthritis. *PM R* 2016; 8: 869–875.
 12. Chinsongkram B, Chaikeeree N, Saengsirisuwan V, Viriyatharakij N, Horak FB, Boonsinsukh R. Reliability and validity of the Balance Evaluation Systems Test (BESTest) in people with subacute stroke. *Phys Ther* 2014; 94: 1632–1643.
 13. Chinsongkram B, Chaikeeree N, Saengsirisuwan V, Horak FB, Boonsinsukh R. Responsiveness of the Balance Evaluation Systems Test (BESTest) in people with subacute stroke. *Phys Ther* 2016; 96: 1638–1647.
 14. 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.
 15. Huang M, Pang MYC. Psychometric properties of Brief-Balance Evaluation Systems Test (Brief-BESTest) in evaluating balance performance in individuals with chronic stroke. *Brain Behav* 2017; 7: e00649.
 16. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. *J Rehabil Med* 2010; 42: 323–331.
 17. 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* 2013; 93: 1102–1115.
 18. Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982; 47: 149–174.
 19. Winairuk T, Chinsongkram B, Lekskulchai R, Pang M, Boonsinsukh R. Rasch analysis of the Balance Evaluation System Test (BESTest) in persons with stroke. *Proceedings of WCPT-AWP& PTAT Congress 2017: moving towards health, longevity, and sustainability*; 2017, June 28–30. Bangkok: Bangkok International Trade & Exhibition Centre (BITEC); 2017.
 20. Jorstad EC, Hauer K, Becker C, Lamb SE. Measuring the psychological outcomes of falling: a systematic review. *J Am Geriatr Soc* 2005; 53: 501–510.
 21. de Oliveira CB, de Medeiros IR, Frota NA, Greters ME, Conforto AB. Balance control in hemiparetic stroke patients: main tools for evaluation. *J Rehabil Res Dev* 2008; 45: 1215–1226.
 22. Dick JP, Guiloff RJ, Stewart A, Blackstock J, Bielawska C, Paul EA, et al. Mini-Mental State Examination in neurological patients. *J Neurol Neurosurg Psychiatry* 1984; 47: 496–499.
 23. Tombaugh TN, McIntyre NJ. The Mini-Mental State Examination: a comprehensive review. *J Am Geriatr Soc* 1992; 40: 922–935.
 24. Boonsinsukh R, Panichareon L, Saengsirisuwan V, Phansuwan-Pujito P. Clinical identification for the use of light touch cues with a cane in gait rehabilitation poststroke. *Top Stroke Rehabil* 2011; 18 Suppl 1: 633–642.
 25. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther* 2009; 17: 163–170.
 26. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000; 81 Suppl 2: S15–S20.
 27. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. 3rd ed. Upper Saddle River, N.J.: Pearson/Prentice Hall; 2009.
 28. 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.
 29. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010; 63: 524–534.
 30. Copay AG, Subach BR, Glassman SD, Polly Jr DW, Schuller TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007; 7: 541–546.
 31. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; 56: 395–407.
 32. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000; 53: 459–468.
 33. Godi M, Franchignoni F, Caligari M, Giordano A, Turcato AM, Nardone A. 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.
 34. Stevenson TJ. Detecting change in patients with stroke using the Berg Balance Scale. *Aust J Physiother* 2001; 47: 29–38.
 35. Marigold DS, Eng JJ. Altered timing of postural reflexes contributes to falling in persons with chronic stroke. *Exp Brain Res* 2006; 171: 459–468.
 36. Lee HH, Jung SH. Prediction of post-stroke falls by quantitative assessment of balance. *Ann Phys Rehabil Med* 2017; 41: 339–346.
 37. Dickstein R, Abulaffio N. Postural sway of the affected and nonaffected pelvis and leg in stance of hemiparetic patients. *Arch Phys Med Rehabil* 2000; 81: 364–367.
 38. Nardone A, Schieppati M. Postural adjustments associated with voluntary contraction of leg muscles in standing man. *Exp Brain Res* 1988; 69: 469–480.
 39. Kollen B, van de Port I, Lindeman E, Twisk J, Kwakkel G. Predicting improvement in gait after stroke: a longitudinal prospective study. *Stroke* 2005; 36: 2676–2680.
 40. Kwakkel G, Kollen B, Twisk J. Impact of time on improvement of outcome after stroke. *Stroke* 2006; 37: 2348–2353.