J Rehabil Med 2020; 52: jrm00079

ORIGINAL REPORT

COMPARISON OF THE STRUCTURAL VALIDITY OF THREE BALANCE EVALUATION SYSTEMS TESTS IN OLDER ADULTS WITH FEMORAL OR VERTEBRAL FRACTURES

Kazuhiro MIYATA, PhD, PT[1], Satoshi HASEGAWA, MS, PT[2,3], Hiroki IWAMOTO, MS, PT[4], Tomohiro OTANI, MS, PT[5], Yoichi KAIZU, MS, PT[5], Tomoyuki SHINOHARA, PhD, PT[6] and Shigeru USUDA, PhD, PT[2]

From the [1]Department of Physical Therapy, Ibaraki Prefectural University of Health Science, Ami-Machi, [2]Department of Rehabilitation Sciences, Gunma University Graduate School of Health Sciences, Maebashi, [3]Department of Rehabilitation, Public Nanokaichi Hospital, Tomioka, [4]Department of Rehabilitation Center, Hidaka Rehabilitation Hospital, [5]Department of Rehabilitation Center, Hidaka Hospital and [6]Department of Physical Therapy, Faculty of Health Care, Takasaki University of Health and Welfare, Takasaki, Japan

Objective: **To clarify and compare the structural validity of 3 Balance Evaluation Systems Tests (BESTest, Mini-BESTest, and Brief-BESTest) in older adults with femoral or vertebral fractures.**

Design: **Cross-sectional study.**

Subjects: Ninety-four older adults (age ≥65 years) with femoral or vertebral fractures, who could walk without physical assistance.

Methods: **Four BESTest models (BESTest, one-factor Mini-BESTest, four-factor Mini-BESTest, and Brief-BESTest) were examined using confirmatory factor analysis, and the models' goodness-of-fit was assessed. Unidimensionality of the best-fitting model was confirmed by Rasch principal component analysis on the residuals.**

Results: **Confirmatory factor analysis showed that the four-factor Mini-BESTest model (comparative fit index = 0.952; Tucker-Lewis index = 0.937; root-mean square error of approximation = 0.060; standardized root-mean-square residual = 0.062) has a better structure than other models. The principal component analysis of standardized residuals showed that the variance attributable to Rasch factor was good, with eigenvalues < 2, confirming the factor's unidimensionality.**

Conclusion: **The four-factor Mini-BESTest model shows good structural validity in older adults with femoral or vertebral fracture. Evaluating dynamic balance by focusing on 4 components (anticipatory postural adjustments, postural responses, sensory orientation, and stability in gait) may help therapists in making clinical decisions.**

Key words: validity; femoral fracture; vertebral fracture; BESTest.

Accepted Jun 8, 2020; Epub ahead of print Jun 16, 2020

J Rehabil Med 2020; 52: jrm00079

Correspondence address: Kazuhiro Miyata, Ami-Machi, Inashiki-gun, Ibaraki, Japan. E-mail: miyatak@ipu.ac.jp

LAY ABSTRACT

The Balance Evaluation Systems Test (BESTest), a clinical balance measure, categorizes balance into 6 factors. The aim of this study was to determine which short versions of the BESTest, Mini-BESTest, and Brief-BESTest, were most appropriate for assessing balance impairments in older adults with femoral or vertebral fracture. Subjects were older adults (age ≥65 years), who could walk without physical assistance. The models' fitting was evaluated for four BESTest models (BESTest, one-factor Mini-BESTest, four-factor Mini-BESTest, and Brief-BESTest). The four-factor Mini-BESTest model was the only model that had a good fit and reflected the balance ability in older adults with femoral or vertebral fracture. We suggest that the using the four-factor Mini-BESTest model and seeing dynamic balance as composed of four factors may help therapists in making clinical decisions.

Loss of balance control is a major cause of falls and fractures in older adults. In Japan, the incidence rate of hip fractures is 249 per 100,000 person-years and that of vertebral fractures 352 per 100,000 person-years, and most of these are sustained due to falls (1). Thus, in older adults who return to community living after a fracture, fall prevention is important. Many clinical balance assessment tools are typically used to identify balance limitations and the risk of falls. Some of the balance assessment tools commonly used in older adult rehabilitation are the Berg Balance Scale, Timed "Up & Go" test, and Functional Reach Test. However, no gold standard exists for evaluating balance (2), and there is no consensus regarding which assessment tools to use in clinical practice (3, 4).

One of the most recent balance assessment tools is the Balance Evaluation Systems Test (BESTest), the scale developed from a theoretical understanding of 6 postural control systems (5). In a 2015 review, the BESTest was identified as the only standardized balance measurement tool that evaluates all components of balance consistent with established conceptual models (6). However, its administration is time-consuming (7). Thus, shorter versions have been proposed, such as the Mini-BESTest and the Brief-BESTest (7, 8). The Mini-BESTest was developed to reflect the unidimensional construct "dynamic balance" through psychometric methods. In addition, the Mini-BESTest has recently been recommended as a method of choice for evaluating standing balance in adults (9). The Brief-BESTest includes items to adequately evaluate

doi: 10.2340/16501977-2709

all postural control systems of balance endorsed by the BESTest (8).

Although some psychometric properties of the 3 BESTests, such as reliability, internal consistency, criterion validity, and construct validity, have been investigated in previous studies (5, 8, 10–12), other properties still need to be clarified for older adults. In particular, in the development of the Mini-BESTest, all items representing the 2 postural control systems were omitted; the remaining items, instead of representing different subsystems, were now collectively thought to reflect the unidimensional construct. However, the test items are arranged in 4 subsystems, and scores may be calculated separately for each subscale (13), which is somewhat contradictory to the concept of a unidimensional construct. Consequently, studies in which the scores for each subsystem are analysed and reported separately are emerging (14–17).

There is currently no unified view on the structure of the 3 BESTests, and it is not clear which BESTest is suitable for evaluating older adults with fractures. In rehabilitation, evaluation and intervention of balance are facilitated by a clear definition of the structure of the balance component associated with various diseases. Therefore, understanding the structure of the 3 BESTests in older adults with fractures may provide better guidance for a more directed clinical decision-making. The objective of this study was to compare the structural validity of the 3 BESTests in older adults with femoral or vertebral fractures and to determine a suitable balance assessment tool.

METHODS

Subjects

This cross-sectional study analysed 94 older adults with fractures who participated in a rehabilitation programme at 3 hospitals with convalescent rehabilitation wards. The inclusion criteria were: (*i*) age ≥65 years; (*ii*) history of femoral or vertebral fracture due to fall; (*iii*) able to walk without physical assistance from another person (Functional Ambulation Categories; FAC ≥ 3); (*iv*) independence in basic daily activities; and (*v*) discharged to home. The exclusion criteria were: (*i*) vestibular disorder; (*ii*) cognitive impairment (Hasegawa Dementia Scale-Revised, <21/30) (18); (*iii*) visual problem in daily living; and (*iv*) history of neurological disease, such as stroke or Parkinson's disease. All subjects received a conventional orthopaedic rehabilitation programme, prescribed by a doctor, with physical therapists, and occupational therapists, as required. The therapies were customized and included muscle strength, balance, gait, activities of daily living, and cognitive training. Therapy was carried out 7 days a week, for 1–1.5 h per day on weekdays and 40–60 min on Saturdays, Sundays, and national holidays.

Gunma University Ethical Review Board for Medical Research Involving Human Subjects (No. 15–73) and the ethics committees of Public Nanokaichi Hospital (20160208), Hidaka Rehabilitation Hospital (No. 151101), and Hidaka Hospital (No. 112) approved this study. The reporting of this study conforms to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.

Data collection procedures

Data were collected on age, type of fracture, time since admission, use of walking aid to transfer in the hospital from the subject's medical records and the rehabilitation centre's database at discharge. The BESTest was performed at discharge as indices. Based on the subject's performance on the BESTest, the therapist provided a rating according to specific scoring criteria of the Mini-BESTest and Brief-BESTest.

Instrumentations

Table I shows the outline and items of the 3 BESTests.

BESTest. This test consists of 36 items; each item is scored on a 4-level ordinal scale from 0 to 3. The score for the total, as well as that for each section, is expressed as a percentage of the total points. The BESTest items are categorized into 6 postural control systems: biomechanical constraints, stability limits and verticality, anticipatory postural adjustments, postural response, sensory orientation, and stability in gait (5).

Mini-BESTest. This test consists of 14 items, each scored on a 3-level ordinal scale from 0 to 2, and the maximum score was 28 points. The Mini-BESTest items derived from the 4 of the 6 subsystems of the BESTest were: anticipatory postural adjustments, postural responses, sensory orientation, and stability in gait (7).

Brief-BESTest. This test consists of 8 items. The scoring method for each item was the same as that in the BESTest, and the maximum score was 24 points. The Brief-BESTest items also corresponded to the 6 subsystems of the BESTest (8).

Data analysis

Structural validity is defined as "the degree to which the scores of a measurement instrument are an adequate reflection of the dimensionality of the construct to be measured" and is regarded as 1 aspect of construct validity (19). This is usually assessed through confirmatory factor analysis (CFA) (20). CFA assesses the degree to which responses on a $p \times 1$ vector of observable random variables (3 BESTest items) can be used to assign a value to 1 or more unobserved variables (balance subsections). For this purpose, a specific mathematical model is identified and fitted to the subject's data. Features of CFA include the ability to compare several models, relatively low dependence on data, and the need for a clear hypothesis about factor structure (21).

The structures of BESTest, Mini-BESTest, and Brief-BESTest were modelled from each original and related literature. Model 1 was designed based on Horak et al.'s (5) BESTest, which includes 6 factors with 36 items. Model 2 was the original Mini-BESTest shown by Franchignoni et al. (7), consisting of one factor with 14 items. As demonstrated by the authors, this model has been confirmed to have one-factor (unidimensional) structure in neurological diseases causing balance impairment. Model 3, which is another model of Mini-BESTest, consisting of 4 factors with 14 items, and the properties of these 4 factors are clearly different in several studies (14–17). The final model (model 4) is the Brief-BESTest presented by Padgett et al. (8), consisting of one factor with 8 items.

This study examined the structural validity of these 4 models (termed below as: models 1, 2, 3, and 4) through CFA with robust maximum likelihood estimation methods using structural

Table I. Items of the Balance Evaluation Systems Test and short versions

Section	Item		BESTest	Mini-BESTest	Brief-BESTest
Biomechanical constraints	1	Base of support	/		
	2	CoM alignment	/		
	3	Ankle strength and ROM	/		
	4	Hip/trunk lateral strength	/		/
	5	Sit to floor and stand up	/		
Stability limits	6a	Sitting verticality, a/s	/		
	6b	Sitting verticality, s/s	/		
	6c	Sitting lateral lean, a/s	/		
	6d	Sitting lateral lean, s/s	/		
	7	Functional reach forward	/		/
	8a	Functional reach lateral, a/s	/		
	8b	Functional reach lateral, s/s	/		
Anticipatory postural adjustments	9	Sit to stand	/	/	
	10	Rise to toes	/	/	
	11a	Stand on one leg, a/s	/	/	/
	11b	Stand on one leg, s/s	/	/	/
	12	Alternate stair touching	/		
	13	Standing arm raise	/		
Postural response	14	In-place response, forward	/		
	15	In-place response, backward	/		
	16	Compensatory stepping correction, forward	/	/	
	17	Compensatory stepping correction, backward	/	/	
	18a	Compensatory stepping correction, lateral, a/s	/	/	/
	18b	Compensatory stepping correction, lateral, s/s	/	/	/
Sensory orientation	19a	Stance on firm surface, EO	/	/	
	19b	Stance on firm surface, EC	/		
	19c	Stance on foam surface, EO	/		
	19d	Stance on foam surface, EC	/	/	/
	20	Incline, EC	/	/	
Stability in gait	21	Gait, level surface	/		
	22	Change in gait speed	/	/	
	23	Walk with head turns, horizontal	/	/	
	24	Walk with pivot turns	/	/	
	25	Step over obstacles	/	/	
	26	Timed "Get Up & Go" test	/		/
	27	Timed "Get Up & Go" test with dual task	/	/	

CoM: centre of mass; ROM: range of motion; EO: eyes open; EC: eyes closed.

equation modelling. In addition, the models' goodness-of-fit was assessed using the following indexes: comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root-mean-square residual (SRMR). Goodness-of-fit indices were examined in light of the following standards used in past literature: CFI and TLI values should be >0.9 for acceptable fit and >0.95 for good fit (22, 23), RMSEA value should be <0.10 for acceptable fit and <0.06 for good fit (23), and SRMR values should be <0.10 for acceptable fit and <0.08 for good fit (21, 23). The CFA was conducted using Mplus 8.3 statistical software (Muthen & Muthen, Los Angeles, CA, USA).

Rasch analysis (partial credit model) was used to examine unidimensionality, internal construct validity, and item-difficulty hierarchy of the best-fitting model. If the best-fitting model was multidimensional, Rasch analysis was performed separately. The unidimensionality of the best-fitting model is evaluated using principal component analysis (PCA) of the standardized residuals. The subsequent criteria were variance explained by the measured construct (called "Rasch factor") greater than 40% and the eigenvalue of the first residual factor ≤2 (24). Internal construct validity was assessed by determining how well the empirical data fit the Rasch model. Fit statistics were calculated as infit, which is the most sensitive for rating items located close to the ability of a person, and outfit, which is more influenced by the rating of off-target items. Fit statistics are routinely reported as means of squared residuals (MNSQ). According to Linacre (24), a MNSQ of 0.5–1.5 could be considered a reasonable range for infit and outfit measures in clinical observations. Rasch modelling provides estimates of the level of difficulty achieved for each item (item difficulty measure), and this parameter is expressed on a common interval scale in logit units. The sequencing of difficulty thresholds and ceiling/floor effects were evaluated by the person-item maps, using a standardized logit unit to place item difficulty and person ability in the same continuum. Rasch analysis was performed using WINSTEPS software 4.2.0 (Winsteps Rasch measurement computer program, Beaverton, OR, USA).

RESULTS

Complete sets of data were obtained from all 94 subjects, and Table II shows their assessment outcomes. The item "stance on firm surface, eyes open" was excluded from the subsequent analysis because the overall score of all subjects was perfect.

Structural validity

Table III summarizes the goodness-of-fit of each model. Fig. 1 shows the CFA solution for model 3 (four-factor model) of the Mini-BESTest. Model 3 shows significantly better goodness-of-fit (CFI = 0.952, TLI = 0.937, RMSEA = 0.060, and SRMR = 0.062) than other models. The CFI and SRMR values reached the good-fitting index values, and the TLI and RMSEA

Table II. Characteristics of study subjects

Characteristics	All ($n=94$)	Hospital A ($n=36$)	Hospital B ($n=35$)	Hospital C ($n=23$)
Age, years, mean (SD) [range]	80.0 (6.4) [65–92]	82.4 (6.2) [70–92]	80.1 (5.6) [65–88]	76.2 (6.4) [66–87]
Fracture site (femoral/vertebral), *n*	66/28	29/7	24/11	13/10
Time since onset, days, mean (SD) [range]	75.1 (26.2) [31–133]	79.7 (23.9) [37–131]	77.4 (29.0) [31–133]	64.5 (23.8) [36–117]
Walking aid (use/none), *n*	76/18	28/8	29/6	19/4
FAC (3/4/5), *n*	6/81/7	1/35/0	5/25/5	0/21/2
BESTest (/100%), mean (SD) [range]	72.2 (13.8) [30.6–95.4]	67.5 (14.7) [30.6–90.7]	72.0 (12.8) [29.6–90.7]	80.0 (10.5) [46.3–95.4]
Brief-BESTest (/24), mean (SD) [range]	12.8 (5.0) [3–23]	11.4 (5.3) [3–23]	12.4 (4.2) [3–21]	15.6 (4.9) [3–23]
Mini-BESTest (/28), mean (SD) [range]	18.4 (5.0) [4–26]	16.1 (5.4) [4–24]	18.8 (4.2) [5–25]	21.4 (3.8) [9–26]
CWS, m/s, mean (SD) [range]	0.83 (0.32) [0.19–1.39]	0.73 (0.29) [0.19–1.30]	0.83 (0.28) [0.26–1.35]	1.01 (0.33) [0.33–1.39]

BESTest: Balance Evaluation Systems Test; Brief-BESTest: Brief-Balance Evaluation Systems Test; CWS: comfortable walking speed; FAC: Functional Ambulation Categories; Mini-BESTest: Mini-Balance Evaluation Systems Test.

Table III. Summary of fit statistics for models of BESTest and BESTest short versions

	Model	CFI	TLI	RMSEA (90% CI)	SRMR
1	BESTest	0.686	0.657	0.102 (0.093–0.110)	0.126
2	Mini-BESTest (1 factor)	0.841	0.809	0.104 (0.078–0.130)	0.080
3	Mini-BESTest (4 factor)	0.952	0.937	0.060 (0.011–0.092)	0.062
4	Brief-BESTest	0.710	0.594	0.214 (0.175–0.255)	0.170

CFI: comparative fit index; TLI: Tucker-Lewis index; RMSEA: root-mean square error of approximation; 90% CI: 90% confidence interval; SRMR: standardized root-mean-square residual; BESTest: Balance Evaluation Systems Test; Mini-BESTest: Mini-Balance Evaluation Systems Test; Brief-BESTest: Brief-Balance Evaluation Systems Test.

values have also reached the accept-fitting index values. The factor loadings of each item were significant ($p<0.001$) and higher than 0.4 for model 3. The item "timed get up & go test with dual task" had the lowest factor loading.

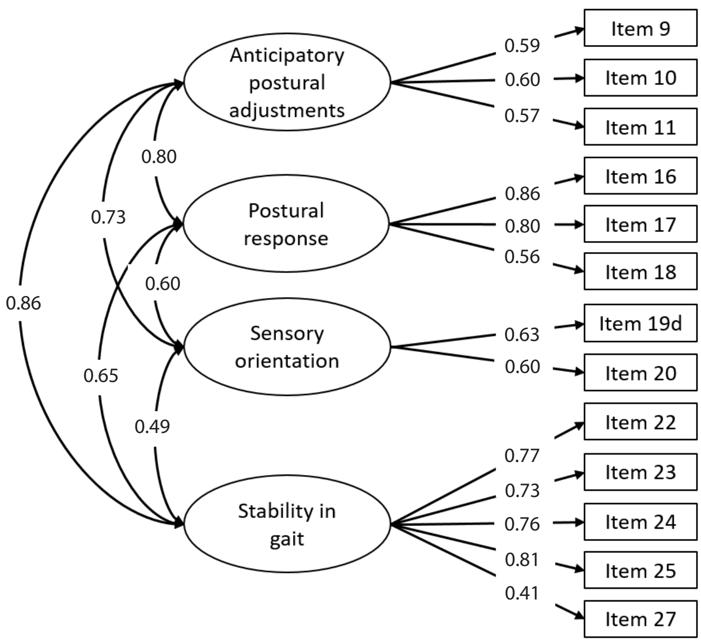

Fig. 1. Confirmatory factor analysis of the four-factor model of Mini-BESTest. The comparative fit index was 0.952, the Tucker-Lewis index was 0.937, the root-mean square error of approximation was 0.060, and the standardized root-mean-square residual was 0.062.

Rasch analysis

In 4 separate Rasch analyses for the proposed factors of model 3, the PCA of standardized residuals showed that the variance attributable to Rasch factor was good (56.9–72.9%), and the eigenvalues of the unexplained variance in the first construct were 2.0, 1.8, 0.1, and 1.5 for anticipatory postural adjustments, postural response, sensory orientation, and stability in gait, respectively, confirming the factor's unidimensionality.

Table IV shows item difficulty measures and fit information of each factor. One item of model 3 had low outfit values, including the item "sit to stand" (outfit MNSQ=0.12). The difficulty level of the item range from –3.91 to 4.04, –2.33 to 2.39, –1.21 to 1.21, and –3.74 to 4.71 logit for anticipatory postural adjustments, postural response, sensory orientation, and stability in gait, respectively. Items "stand on one leg" (4.04±0.29 logits), "compensatory stepping correction, lateral" (2.39±0.26), "stance on foam surface, eyes closed" (1.21±0.43), and "timed get up & go test with dual task" (4.75±0.44) were the most difficult items. Conversely, items "sit to stand" (–3.91±0.59), "compensatory stepping correction, forward" (–2.33±0.30), "incline, eyes closed" (–1.21±0.43), and "change in gait speed" (–3.74±0.31) were the easiest items.

Person ability ranged from –5.19 to 8.46, –6.51 to 5.67, –3.45 to 4.38, and –8.79 to 6.26 logit for anticipatory postural adjustments, postural response, sensory orientation, and stability in gait, respectively (Fig. 2). The sensory orientation factor was the only factor for which a majority of the subjects (69.1%) had a maximum raw score, indicating a ceiling effect

DISCUSSION

This study compared the structural validity of the 2 different short versions of the BESTest

Table IV. Summary of Rasch analyses of the four-factor Mini-BESTest model containing item difficulty measure and fit information

Factor	Item		Measure	SE	Infit MNSQ	Outfit MNSQ
Anticipatory postural adjustments	9	Sit to stand	−3.91	0.59	0.75	0.12
	10	Rise to toes	−0.13	0.27	1.02	0.93
	11	Stand on one leg	4.04	0.29	1.01	0.83
Postural response	16	Compensatory stepping correction, forward	−2.33	0.30	0.68	0.73
	17	Compensatory stepping correction, backward	−0.06	0.28	1.08	1.08
	18	Compensatory stepping correction, lateral	2.39	0.26	1.18	1.16
Sensory orientation	19d	Stance on foam surface, EC	1.21	0.43	0.95	0.91
	20	Incline, EC	−1.21	0.43	0.97	1.13
Stability in gait	22	Change in gait speed	−3.74	0.31	0.86	0.69
	23	Walk with head turns, horizontal	−1.66	0.28	1.20	1.12
	24	Walk with pivot turns	1.65	0.27	0.87	0.79
	25	Step over obstacles	−1.00	0.27	1.18	0.83
	27	Timed "Get Up & Go" test with dual task	4.75	0.44	0.96	0.80

EC: eyes closed; SE: standard error; MNSQ: means of squared residuals.

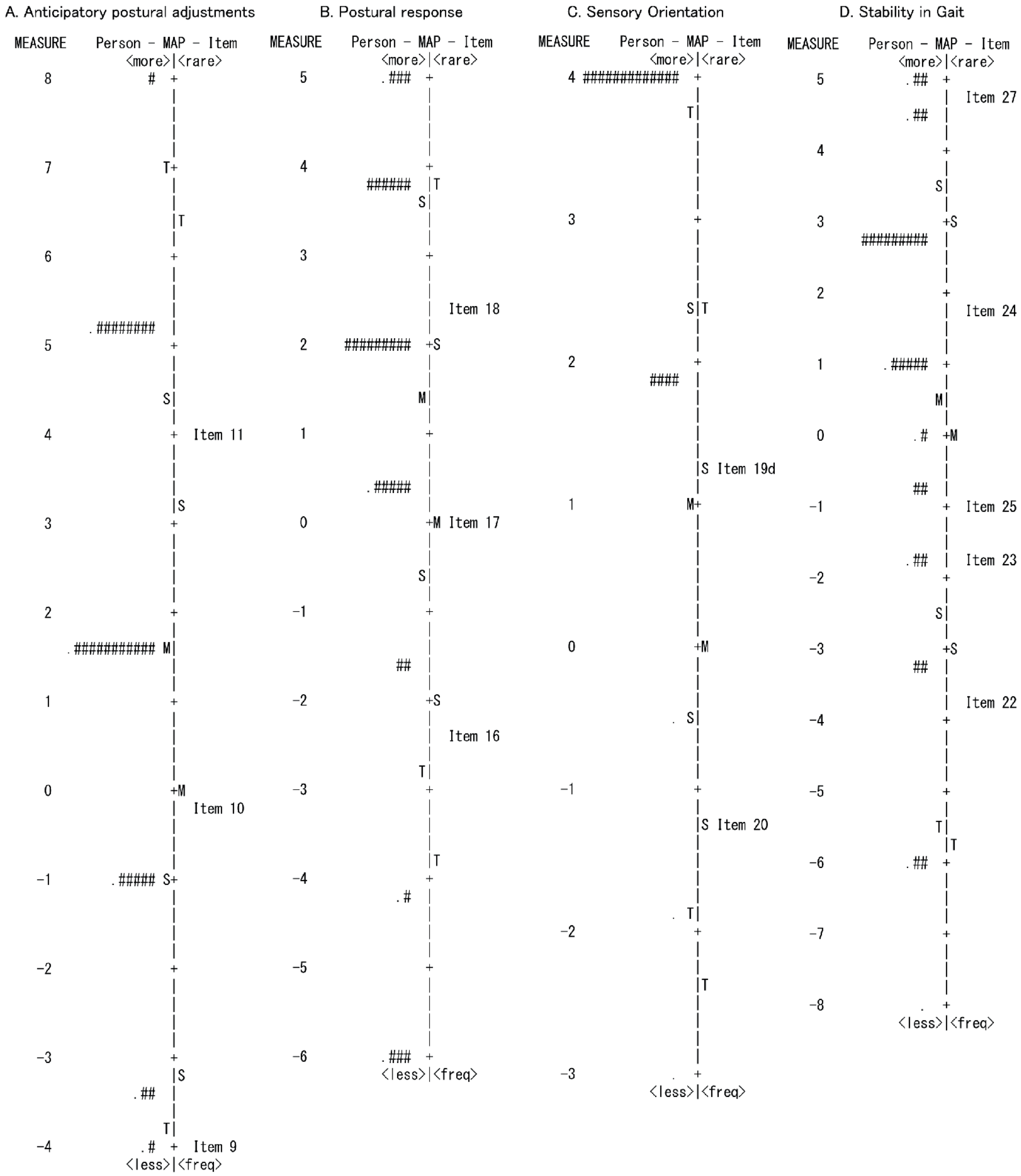

Fig. 2. Person-item map of each factor of the Mini-BESTest from the Rasch analysis. (A) Anticipatory postural adjustments. (B) Postural response. (C) Sensory orientation. (D) Stability in gait.

(Mini-BESTest and Brief-BESTest) with the original BESTest in older adults with femoral or vertebral fractures. The four-factor Mini-BESTest model was the only model that had a good fit and reflected the balance ability in older adults with fractures. To the best of our knowledge, this study is the first to investigate the BESTest and its short versions to identify the best assessment tool that can be recommended for assessing the balance of older adults with femoral or vertebral fractures.

Results from the CFA and Rasch PCA showed the structural validity of the four-factor Mini-BESTest model in older adults with femoral or vertebral fractures. Therefore, this study clarified that the four-factor Mini-BESTest model was structurally more stable than the one-factor model. Although the Mini-BESTest is widely used to assess balance in older adults (11, 12, 16), the structural validity of the assessment tool has not been investigated in this population. The Mini-BESTest was developed by Franchignoni et al. (7) as a one-factor balance assessment tool to evaluate "dynamic balance." Previous studies have also confirmed that the Mini-BESTest is a unidimensional scale for neurological disorders such as stroke (25, 26). However, the unidimensional structure was reported not suitable for people with mild to moderate Parkinson's disease (27). Therefore, there is a possibility that a different factor structure is exhibited depending on the populations.

In general, balance is a composite ability that involves rapid, automatic, anticipatory, reactive integration and sensory strategies based on information derived from several systems (28). In older adults, several aspects are related to balance that decrease with ageing (29, 30), which is especially true for those who had a history of falls (16). Because the concept of balance includes many components, a balance assessment tool should be multidimensional. These results suggest that the four-factor Mini-BESTest model can evaluate multiple aspects of dynamic balance in older adults with femoral or vertebral fractures. This study suggested that using the four-factor Mini-BESTest model and considering dynamic balance as composed of 4 factors may help therapists in making clinical decisions, specifically intervening in factors that indicated decline in function.

In 4 separate Rasch analyses for the 4 proposed factors of the Mini-BESTest, the fit was good except for the item "sit to stand," and each factor was unidimensional. Rasch analysis also revealed the difficulty of each factor task. This was in agreement with a previous finding that items "stand on one leg," "compensatory stepping correction, lateral," and "timed get up & go test with dual task" were difficult tasks in people with subacute stroke (26). Therefore, the difficulty of the Mini-BESTest may not be dependent on the disease. In this study, the difficulty level of the problem was clarified within the four-factor Mini-BESTest model. This identifies the next challenge to achieve among factors and will help therapists to set goals for intervention. Our results showed only that the sensory orientation factor had a ceiling effect. The sensory orientation factor is composed of static standing tasks, such as items "stance on foam surface, eyes closed" and "incline, eyes closed", and evaluates any increase in body sway during stance associated with altering visual or surface somatosensory information for control of standing balance (5, 7). Neurological disorders, such as stroke, may reduce sensory orientation ability; that is, standing postural control depends on ones' vision, which is affected by somatosensory impairment, muscle weakness, and asymmetric muscle tone (31). Ageing also decreases somatosensory and increases postural sway in older adults (32, 33). However, the sensory orientation factor was unable to distinguish the decline in somatosensory in older adults, suggesting that it was very easy to perform. In addition, since the study subjects were those who could walk without assistance, sensory orientation ability may be essential for walking.

This study confirmed that the structural validity of the original model of the BESTest and Brief-BESTest in older adults with femoral or vertebral fractures was not sufficient. In recent years, studies have reported the structural validity of the BESTest and Brief-BESTest in neurological diseases and proposed different structures from the original model (34, 35). Godi et al. compared 3 Brief-BESTest models, and found that the model with the item "hip/trunk lateral strength" removed showed better fit than the original model (35). Since it is important to understand the structure of the balance component of the subject, further research on various diseases is needed. In this study, the one-factor model of Mini-BESTest was more stable in structure than the BESTest and Brief-BESTest. This may involve the number of scoring levels on the assessment tool. In fact, the number of scoring levels is different between Mini-BESTest, BESTest, and Brief-BESTest, with Mini-BESTest having 3 levels and BESTest and Brief-BESTest having 4 levels. Previous studies have indicated that the BESTest and Brief-BESTest scoring stages are redundant (7, 36). This redundancy also affects the structure, suggesting that the Mini-BESTest showed better fit than the BESTest and Brief-BESTest.

The findings of this study have the following implications for clinical rehabilitation practice. The structure of the balance component in older adults suggests that it is effective to evaluate the four-factor structure of Mini-BESTest, including anticipatory postural adjustments, postural response, sensory orientation, and stability in gait. This clarifies the problematic balance component and facilitates selection of intervention. Furthermore, since the difficulty level of each factor is shown, the difficulty level of practice can easily be adjusted. Thus, providing effective interventions in these balance abilities may lead to the prevention of secondary fractures.

Study limitations

This study has a number of limitations. First, the sample size was small. It is desirable to exceed 100 samples in structural equation modelling (37); therefore, large-scale investigations are warranted. Secondly, subjects with femoral or vertebral fractures were analysed together. Femoral and vertebral fractures are osteoporosis-related fractures, and no previous studies have shown differences in the balance ability of patients. Therefore, they were analysed together, but separate analysis is more appropriate if a sufficient sample size can be obtained. Thirdly, our subjects were inpatients at a rehabilitation hospital, had no cognitive function decline, and were able to walk independently. It is unclear whether the results of the structural validity of the Mini-BESTest obtained herein can be applied to community-based older adults with fractures and cognitive function decline and need physical assistance to walk. Fourthly, the item "stance on firm surface, eyes open" was excluded from the analysis because all subjects had a perfect score. Thus, the BESTest and Mini-BESTest containing the item "stance on firm surface, eyes open" were analysed, with the exception of the item "stance on firm surface, eyes open." Therefore, generalization of the results may be limited. Further studies involving large samples, participants with different walking abilities, and community-based older adults with femoral or vertebral fractures are needed.

Conclusion

The structural validities of the 3 BESTest models were compared in older adults with femoral or vertebral fractures. This study showed that the four-factor Mini-BESTest model adequately reflects balance ability and has a good fit. Based on these results, we suggest that using the four-factor Mini-BESTest model and seeing dynamic balance as composed of 4 factors, namely, anticipatory postural adjustments, postural responses, sensory orientation, and stability in gait, may help therapists in making clinical decisions.

ACKNOWLEDGEMENTS

This work was supported by the Sasakawa Scientific Research Grant from The Japan Science Society and JSPS KAKENHI Grant Number 18K17724.

The authors have no conflicts of interest to declare.

REFERENCES

1. Imai N, Endo N, Shobugawa Y, Oinuma T, Takahashi Y, Suzuki K, et al. Incidence of four major types of osteoporotic fragility fractures among elderly individuals in Sado, Japan, in 2015. J Bone Miner Metab 2019; 37: 484–490.
2. Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med 2010; 46: 239–248.
3. McGinnis PQ, Hack LM, Nixon-Cave K, Michlovitz SL. Factors that influence the clinical decision making of physical therapists in choosing a balance assessment approach. Phys Ther 2009; 89: 233–247.
4. Sibley KM, Straus SE, Inness EL, Salbach NM, Jaglal SB. Balance assessment practices and use of standardized balance measures among Ontario physical therapists. Phys Ther 2011; 91: 1583–1591.
5. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther 2009; 89: 484–498.
6. Sibley KM, Beauchamp MK, Van Ooteghem K, Straus SE, Jaglal SB. Using the systems framework for postural control to analyze the components of balance evaluated in standardized balance measures: a scoping review. Arch Phys Med Rehabil 2015; 96: 122–132.
7. 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.
8. 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.
9. Sibley KM, Howe T, Lamb SE, Lord SR, Maki BE, Rose DJ, et al. Recommendations for a core outcome set for measuring standing balance in adult populations: a consensus-based approach. PLoS One 2015; 10: e0120568.
10. Di Carlo S, Bravini E, Vercelli S, Massazza G, Ferriero G. The Mini-BESTest: a review of psychometric properties. Int J Rehabil Res 2016; 39: 97–105.
11. Viveiro LAP, Gomes GCV, Bacha JMR, Carvas Junior N, Kallas ME, Reis M, et al. Reliability, validity, and ability to identity fall status of the Berg Balance Scale, Balance Evaluation Systems Test (BESTest), Mini-BESTest, and Brief-BESTest in older adults who live in nursing homes. J Geriatr Phys Ther 2019; 42: E45–E54.
12. O'Hoski S, Sibley KM, Brooks D, Beauchamp MK. Construct validity of the BESTest, mini-BESTest and brief-BESTest in adults aged 50 years and older. Gait Posture 2015; 42: 301–305.
13. King L, Horak FB. On the mini-BESTest: scoring and the reporting of total scores. Phys Ther 2013; 93: 571–575.
14. Mak MK, Auyeung MM. The mini-BESTest can predict parkinsonian recurrent fallers: a 6-month prospective study. J Rehabil Med 2013; 45: 565–571.
15. Löfgren N, Lenholm E, Conradsson D, Ståhle A, Franzén E. The Mini-BESTest--a clinically reproducible tool for balance evaluations in mild to moderate Parkinson's disease? BMC Neurol 2014; 14: 235.
16. Yingyongyudha A, Saengsirisuwan V, Panichaporn W,

Boonsinsukh R. The Mini-Balance Evaluation Systems Test (Mini-BESTest) demonstrates higher accuracy in identifying older adult participants with history of falls than do the BESTest, Berg Balance Scale, or timed up and go test. J Geriatr Phys Ther 2016; 39: 64–70.
17. Löfgren N, Benka Wallén M, Sorjonen K, Conradsson D, Franzén E. Investigating the Mini-BESTest's construct validity in elderly with Parkinson's disease. Acta Neurol Scand 2017; 135: 614–621.
18. Imai Y, Hasegawa K. The Revised Hasegawa's Dementia Scale (HDS-R) – evaluation of its usefulness as a screening test for dementia. J Hong Kong Psychiatr 1994; 4: 20–24.
19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63: 737–745.
20. Atkinson TM, Rosenfeld BD, Sit L, Mendoza TR, Fruscione M, Levene D, et al. Using confirmatory factor analysis to evaluate construct validity of the Brief Pain Inventory (BPI). J Pain Symptom Manage 2011; 41: 558–565.
21. Kline RB. Principles and practice of structural equation modeling. 4th edn. New York: Guilford Press; 2015.
22. Bentler PM, Bonnet DC. Significance test and goodness of fit in the analysis of covariance structures. Psychol Bull 1980; 88: 588–606.
23. Hu LT, Bentler PM. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. Psychol Meth 1998; 3: 424–453.
24. Linacre JM. A user's guide to Winsteps Ministep 4.4.5: Rasch model computer programs. Chicago: Winsteps; 2019.
25. Franchignoni F, Godi M, Guglielmetti S, Nardone A, Giordano A. Enhancing the usefulness of the Mini-BESTest for measuring dynamic balance: a Rasch validation study. Eur J Phys Rehabil Med 2015; 51: 429–437.
26. Goljar N, Giordano A, Schnurrer Luke Vrbanić T, Rudolf M, Banicek-Sosa I, Albensi C, et al. Rasch validation and comparison of Slovenian, Croatian, and Italian versions of the Mini-BESTest in patients with subacute stroke. Int J Rehabil Res 2017; 40: 232–239.
27. Benka Wallén M, Sorjonen K, Löfgren N, Franzén E. Structural validity of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in people with mild to moderate Parkinson disease. Phys Ther 2016; 96: 1799–1806.
28. Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? Age Ageing 2006; 35: ii7–ii11.
29. Larsson L. Histochemical characteristics of human skeletal muscle during aging. Acta Physiol Scand 1983; 117: 469–471.
30. Horak FB, Shupert CL, Mirka A. Components of postural dyscontrol in the elderly: a review. Neurobiol Aging 1989; 10: 727–738.
31. Bonan IV, Colle FM, Guichard JP, Vicaut E, Eisenfisz M, Tran Ba Huy P, et al. Reliance on visual information after stroke. Part I: Balance on dynamic posturography. Arch Phys Med Rehabil 2004; 85: 268–273.
32. Perret E, Regli F. Age and the perceptual threshold for vibratory stimli. Eur Neurol 1970; 4: 65–76.
33. McChesney JW, Woollacott MH. The effect of age-related declines in proprioception and total knee replacement on postural control. J Gerontol A Biol Sci Med Sci 2000; 55: M658–M666.
34. Miyata K, Hasegawa S, Iwamoto H, Otani T, Kaizu Y, Shinohara T, et al. Structural validity of Balance Evaluation Systems Test assessed using factor and Rasch analyses in patients with stroke. J Phys Ther Sci 2018; 30: 1446–1454.
35. Godi M, Giardini M, Arcolin I, Ferrante S, Nardone A, Corna S, et al. Is the Brief-BESTest brief enough? Suggested modifications based on structural validity and internal consistency. Phys Ther 2019; 99: 1562–1573.
36. Bravini E, Nardone A, Godi M, Guglielmetti S, Franchignoni F, Giordano A. Does the Brief-BESTest meet classical test theory and Rasch analysis requirements for balance assessment in people with neurological disorders? Phys Ther 2016; 96: 1610–1619.
37. Reinartz W, Haenlein M, Henseler J. An empirical comparison of the efficacy of covariance-based and variance-based SEM. Int J Res Mark 2009; 26: 332–344.