OXFORD UNIVERSITY PRESS | Physical Therapy
Journal of the American Physical Therapy Association

Is the Brief-BESTest brief enough? Suggested modifications based on structural validity and internal consistency.

Journal:	*Physical Therapy*
Manuscript ID	PTJ-2018-0336.R2
Manuscript Category:	Original Research - Observational/Prognostic
Section:	Balance and Falls
Keywords:	Balance, Gait Disorders: Neurologic, Neurologic Manifestations, Statistics, ROC Curve, Posture, Accidental Falls, Rehabilitation, Rehabilitation Centers, Measurement: Applied

Is the Brief-BESTest brief enough? Suggested modifications based on structural validity and internal consistency

ABSTRACT

Background: The Brief Balance Evaluation Systems Test (Brief-BESTest) could be a useful tool for balance assessment. Although some psychometric characteristics have been examined, others still need to be clarified.

Objectives: To assess the structural validity, convergent validity, discriminant validity and internal consistency of the Brief-BESTest in neurological patients.

Design: Cross-sectional.

Methods: Data were from 416 patients with neurological disease and related balance disorders. Patients were assessed with the 5-levels Activities-Specific Balance Confidence Scale (ABC 5-levels), Brief-BESTest, **some simple balance tests, i.e. One-Leg Stance (OLS), Timed Up and Go (TUG) test, Functional Reach (FR),** simple balance tests and **a** fall history questionnaire. Three models of Brief-BESTest **models** were examined through confirmatory factor analysis (CFA) and the following indices calculated: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA). Convergent validity was assessed by calculating the correlation between Brief-BESTest and ABC 5-levels total scores. Receiver operating characteristics (ROC) assessed the ability of each model to differentiate between people with vs. without falls. Internal consistency was measured by Cronbach's alpha and coefficient omega.

Results: CFA showed Model 3 (CFI=0.97, TLI=0.95, RMSEA=0.05), **with item 1 removed and error covariance between items 3-4 and between items 5-6,** to have a significantly better structure than Models 1 and 2 ($p<0.001$). The correlation between Brief-BESTest and ABC 5-levels was 0.61 (Spearman's rho) for all three models. Area Under the Curve (AUC) of ROC was **showed an acceptable accuracy (0.72) in distinguishing patients with vs. without history of falls** (95% C.I.=0.66–0.78) for all models, and superior to AUCs of other simple balance tests **(OLS, TUG test,**

FR). Cronbach's alpha was good for Brief-BESTest Models 1 (0.92) and 3 (0.92), but omega was >0.80 only for Model 3.

Limitations: Heterogeneous sample size was a heterogeneous population.

Conclusions: The Brief-BESTest**, after some changes,** shows good validity and internal consistency in patients affected by different balance disorders, after applying some changes.

Contribution of the Paper:

- Although some psychometric characteristics of the Brief-BESTest have been examined in previous studies, other properties such as validity still need to be clarified.
- This study shows that the Brief-BESTest has good validity and internal consistency in patients affected by different balance disorders, after applying some changes: removal of item 1 and using an appropriate weight**ing** method for the calculation of the total score.
- This study confirms the ability of the Brief-BESTest to distinguish between people with vs. without history of falls, in contrast to other simple balance tests. Moreover, it highlights once again the superiority of a clinical scale composed of several items compared to single-item measures such as the TUG test and OLS.

Keywords: Brief-Balance Evaluation Systems Test, balance assessment, confirmatory factor analysis, structural validity, internal consistency.

INTRODUCTION

Balance disorders are a common finding in a broad spectrum of neurological disorders and are characterized by a heterogeneous set of signs and symptoms. Patients **with balance disorders experience a reduction in mobility, activities of daily living and muscle strength, leading to increased risk of falls** [1,2,3,4]. Thus, balance assessment is crucial and requires standardized measurement tools that can monitor equilibrium regardless of the pathology. Unfortunately, no gold standard exists for evaluating balance [5], and no consensus on which assessment tools to use in clinical practice [6,7].

A variety of clinical measures has been developed to evaluate different aspects of balance. While simple balance tests such as the Timed Up and Go test, One Leg Stance, and Functional Reach provide accurate evaluation of a single task, they are not able to **do not give** information on multifactorial mechanisms related to postural stability [8]. On the contrary, balance scales which include multiple tasks can provide a more complete picture of balance control in all its complexity [9,10,11].

One of the most recent balance scales is the BESTest, a 36-item scale developed to identify impairments in six balance control subsections, which it has been shown to be a valid and reliable tool [8]. However, one of its drawbacks is that it is time-consuming **to administer** [12]. For this reason, shorter versions have been proposed such as the Mini-BESTest [12] and the Brief-BESTest [13]. In particular, the Brief-BESTest, an 8-item version of the original scale, has demonstrated good to excellent psychometric properties [8,13,14,15]. It is less time-consuming [16] and more feasible than its parent scale in clinical settings [15], while it encompasses and should adequately evaluate all subsections of balance endorsed by the BESTest [13,17,18].

However, the latter hypothesis was rejected through exploratory factor analysis **dismantled the latter hypothesis**, demonstrating **that the Brief-BESTest actually has at most two subsections, or dimensions** [19]. Furthermore, Bravini et al. [14] showed by Rasch analysis that all items of the Brief-BESTest except for item 1 account for the same underlying theoretical construct and indicated that the Brief-BESTest should in fact be considered as unidimensional. Therefore, the authors suggested the adoption of a 7-item version of the test.

Although some psychometric characteristics of the Brief-BESTest, such as the internal consistency [13,14], reliability [14,17,20] and sensitivity to change [14,20], have been investigated in previous studies, other properties still need to be clarified. In particular, the Brief-BESTest structure has not yet been investigated with **undergone** confirmatory factor analysis (CFA). This statistical tool provides information on possible independent factors and can be very useful for developing shortened forms of an evaluating instrument [21,22]. Finally, the Brief-BESTest seems to have good sensitivity and accuracy in identifying retrospectively people who have had at least one fall [13,17]. However, these findings are based only on small samples of patients with multiple sclerosis [13] or Parkinson's disease [17].

The **We** aim**ed in** of the present study was to fill the existing knowledge gap by examining the structural validity, convergent validity and discriminant validity of the Brief-BESTest in a large group of patients with a variety of balance disorders. In particular, we hypothesized that:

1) among the three models of the Brief-BESTest presented in the literature [13,14,19], the 7-item version would be the one with the best structural validity;

2) in spite of its conciseness, the 7-item model [14] would have the same ability as the other two Brief-BESTest models to predict patients at risk of falls; **discriminate between people with vs. without a history of falls;**

3) in discriminating between people with vs. without a history of falls, the Brief-BESTest would be superior to other simple balance tests such as One Leg Stance, Timed Up and Go test and Functional Reach.

METHODS

Participants

This was an observational retrospective study conducted in a group of 416 patients affected by different neurological diseases: 186 females and 230 males; mean age 66.5±16.0 years (**mean ± standard deviation**) consecutively admitted for in-patient rehabilitation at the XXXXXXX between February 2014 and April 2017. **Patients' clinical and treatment data were extracted from the electronic medical record system and transferred to a specific database (Microsoft Excel).** Patients were stratified into different groups according to their diagnosis: 118 **with** Parkinson's disease, 79 **with** acute stroke, 43 **with** sensorimotor polyneuropathy, 32 **with** cerebellar ataxia, 32 **with** diffuse encephalopathy, 31 **with** chronic stroke, 21 **with** multiple sclerosis, 19 **with** traumatic brain injury, 16 **with** vestibular disorder, 13 **with** neuromuscular disorders, 12 **with** central nervous system neoplasm. Inclusion criteria were: a) ability to maintain an upright position without support for at least 5 seconds; b) ability to understand the required motor tasks; c) no hip or knee replacement surgery within the previous 6 months. Exclusion criteria were: a) musculoskeletal injury limiting the ability to walk; b) any other serious cardio-respiratory problem. The study was carried out in conformity with the Declaration of Helsinki of the World Medical Association and the guidelines for retrospective studies [23]. The local scientific and ethics committee approved the study.

Assessment tools

Patients' demographic and clinical characteristics were gathered by **A**a team of trained physical **therapists engaged in clinical practice collected patients' demographic and clinical characteristics. During the routine clinical assessment following admission to the rehabilitation department, patients underwent the following assessments:**

- Fall history

A history of falls over the past 6 months was obtained **from patients** at admission through patient interview. **A fall was defined as an unintentional event in which any part of the body came into contact with the ground [10].** Patients who reported two or more falls in the defined period were classified as 'fallers' [24]. A fall history was not recorded **taken in the case of** for patients with acute stroke at the time of admission.

- Brief-BESTest

The Brief-BESTest is an 8-item scale with each item scored on a 4-level rating scale from 0 (severe balance impairment) to 3 (no balance impairment). Its items cover the six subsections of the original BESTest (biomechanical constraints, stability limits/verticality, anticipatory postural responses, postural responses, sensory orientation, stability in gait); the maximum total score is 24 [13]. **The Brief-BESTest requires less time and equipment to administer than the BESTest and the Mini-BESTest; thus, the Brief-BESTest seems to be more feasible for clinical use [17].**

- Simple balance tests

During administration of the Brief-BESTest, we recorded the time required by patients to complete item 3 (left One Leg Stance), 4 (right One Leg Stance) and 8 (Timed Up and Go test) and the distance covered by patients during item 2 (Functional Reach). This allowed us to obtain the scores of three additional simple balance tests: One Leg Stance (OLS) [25,26], Timed Up and Go (TUG) test [27,28] and Functional Reach (FR) [29,30].

- Activities-specific Balance Confidence Scale

The short version of the Activities-specific Balance Confidence Scale (ABC) is a self-reported 16-item questionnaire that scores the perceived level of balance confidence when performing common activities of daily living [31]. We used the 5-levels **rating** version of rating (ABC 5-levels) [32] in which each item is scored from 0 (no confidence) to 4 (fully confident), giving a total score range 0-64.

Data analysis

Descriptive statistics were used to describe mean demographic and balance performance characteristics of the entire sample and of the two **smaller** subgroups, classified as fallers and non-fallers. **The analysis of discriminant validity was conducted only in these two subgroups, i.e. those patients who had a history of falls available.** These values were also determined separately both for fallers and non-fallers. For each item of the Brief-BESTest we calculated: median value, spread (25th–75th percentiles), skewness and kurtosis. Floor and ceiling effects were analyzed by calculating the percentage of individuals obtaining the lowest and the highest score for each scale item. In order to detect differences in clinical characteristics between fallers and non-fallers, the Chi-square (χ^2) test was used for two parameters, sex and use of walking aids, while the Mann-Whitney U-test was used for age, OLS, TUG test, FR, total score of Brief-

BESTest and total score of ABC 5-levels. Significance was set at $p<0.05$. All analyses were performed using STATA 13.0 software (StataCorp LLC, College Station, Texas, USA).

Structural validity

Structural validity is usually assessed through CFA [33]. CFA assesses the degree to which responses on a p x 1 vector of observable random variables (the Brief-BESTest items) can be used to assign a value to one or more unobserved variable(s) (balance subsections). For this purpose, a specific mathematical model is identified and fitted to the patients' data.

The original model (Model 1) of the Brief-BESTest [13] comprises one factor with 8 independent items that contribute with the same weights to the total score. In the recent literature, two additional models of Brief-BESTest have been presented. Model 2 was designed based on [19]. It includes two factors: one named "static balance" that comprises items 1 and 2, and another called "dynamic balance" that contains items 3 to 8. As demonstrated by the authors, items 5 and 6 showed local dependence, so Model 2 was designed allowing correlation between the errors of these items. **In** Model 3**,** was drawn up without item 1 **was dropped,** as suggested by [14]. In this 7-item model, the error of item 3 was allowed to correlate with that of item 4 and the error of item 5 with that of item 6.

For all models, the score of each item ranges from 0 to 3. Then the total score was **is** obtained by multiplying the rated score by the coefficient fitted for each model (see below formula and supplementary data). In order to allow comparison of the score models, **we adjusted** the coefficients **so** have been adjusted as to maintain a total score in the range 0 to 24. Appendix 1 summarizes the item structure of the three models and their total score.

Preliminary analysis conducted on Model 3 showed similar CFA results for patients who used a walking aid and those who did not. For this reason, we decided to consider the entire sample independently of the walking condition in creating our models.

We examined t The structural validity of these three models (from here on Model 1, 2 and 3) was examined through CFA using Structural Equation Modeling (SEM). In view of the very low occurrence of missing data (maximum 0.4%), cases with missing data were removed from the analysis. χ^2 was used to identify whether the model fitted the data well. In addition, we assessed the models' goodness of fit using the following indices: Comparative Fit Index (CFI) [34], Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) [35] and the Standardized Root Mean Square Residual (SRMR) [34]. The criteria adopted to assess goodness of fit performance were: a) CFI and TLI values ≥0.95; b) RMSEA value <0.06; and c) SRMR value ≤0.08 [36].

The goodness of fit parameters of the three models were compared by computing the χ^2 difference tests of each model pair, calculated as a $\chi^2 = \chi^2_2 - \chi^2_1$ with $df = df_2 – df_1$.

The standardized factor loadings of the models (i.e. the coefficients of the fitted model) were then transformed into weights that can be applied when scale scores for an individual are calculated. They were calculated with a non–refined method called “Weighted Sum Scores” [37]; these weights do not change the scale range [38].

Internal consistency

The internal consistency of the three Brief-BESTest models was measured by Cronbach's alpha and the coefficient omega for congeneric models [39]. Cronbach's alpha measures the extent to which the items consistently measure the same construct, with the value ≥0.80 indicating good internal consistency [40]. The coefficient omega is the ratio of the true score variance and the total variance of the scale. Interpretation of coefficient omega is similar to that of Cronbach's alpha [41].

Convergent validity

We used the correlation between the three models of Brief-BESTest and the ABC 5-levels total score to assess the convergent validity. **The** ABC scale led to rate **rates** the balance self-efficacy of patients [31]. This is associated with measures of balance [32], walking capacity [42], functional mobility [43], Activities of Daily Living performance [44], and perceived health status [45]. The choice to use the ABC as a competitor **an external criterion** was also based on the need to avoid the comparisons of**ng** the three models with a scale that had items similar to the Brief-BESTest. The correlation was assessed by means of Spearman's rho: coefficients <0.30 were interpreted as weak, those between 0.30 and 0.49 as moderate, and those ≥0.50 as strong correlations [46].

Discriminant validity

To assess the ability of the three Brief-BESTest models to distinguish between 'fallers' and 'non-fallers', receiver operating characteristic (ROC) curves were computed**,** by plotting sensitivity on the *x*-axis against 1 – specificity on the *y*-axis. In our study, sensitivity was calculated as the number of patients correctly identified as 'fallers' and specificity as the number of patients correctly identified as 'non-fallers'. The optimal cut-off value was chosen on the ROC curve at the point that jointly maximized sensitivity and specificity [47]. For each ROC, the area under the curve (AUC) and the positive and negative likelihood ratios (LR+ and LR−) were computed **to maximize the cut off scores**. Low, moderate and high accuracy of discrimination were defined respectively as AUC <0.70, 0.70< AUC <0.90, and AUC >0.90 [47]. In addition, the predictive performance of the three models was compared to that of the OLS, TUG test and FR tests by reporting the above described parameters and the percentage of correctly classified patients.

RESULTS

Fall history data**, for the analysis of discriminant validity, was available** could be collected only from **in** 295 subjects: 135 fallers (45%) and 160 non-fallers (55%). Table 1 reports mean **scores** and standard deviation for each balance measure as well as information on the use of walking aids in the overall sample and in fallers vs. non-fallers. We found a significant difference between fallers and non-fallers in the mean score of all clinical evaluations, while mean age and sex did not differ significantly. Table 2 reports descriptive statistics for each item score and total score of the original 8-item Brief-BESTest in the whole sample.

Structural validity

Analysis of the Brief-BESTest models

Figure 1A shows the standardized solution of the CFA for Model 1 that was fitted using all 8 items. We found χ^2 value of 134.0 (df = 20, p<0.001), with CFI of 0.78, TLI of 0.70, RMSEA of 0.12 (90% C.I. = 0.10–0.14) and SRMR above 0.09. Both CFI and TLI were below the cut-off value of

0.95 defined for a well-fitting model. In addition, the RMSEA value suggested that this model exhibited a poor fit of the data.

Model 2 (figure 1B) showed better values of goodness of fit with respect to the original model. In fact, Model 2 had χ^2 of 60.3 (df = 18, p<0.001), with CFI of 0.92, TLI 0.88, RMSEA 0.08 (90% C.I. = 0.06–0.10) and SRMR 0.05. However, only SRMR exhibited a value lower than the preselected well-fitting index value.

Figure 1C shows the CFA solution for Model 3. The results show a significantly better goodness of fit for this model (χ^2 = 26.2, df = 12, CFI = 0.97, TLI = 0.95, RMSEA = 0.05 (90% C.I. = 0.03–0.08), SRMR = 0.03) than for Models 1 and 2. Comparison of Model 3 (χ^2 = 26.2) to Model 1 (χ^2 = 134.0) and Model 2 (χ^2 = 60.3) yielded a difference in χ^2 value of 107.8 and 34.1 respectively and a difference of 6 degrees of freedom, suggesting that Model 3 performed better than the original Brief-BESTest (Model 1) and Model 2. Table 3 summarizes the goodness of fit indices of each model and the significance level of the comparison between each model pair.

The factor loadings of each item were significant (p<0.001) and higher than 0.6 for all three models. Item 1 (Hip/Trunk Lateral Strength), when present, and item 2 (FR Forward) had the lowest factor loading.

Internal consistency

Cronbach's alpha was good for Brief-BESTest Models 1 (0.92) and 3 (0.92), but not for Model 2 (0.56 and 0.88, respectively for static and dynamic balance factor). Coefficient omega was higher than 0.80 only for Model 3, while Model 1 (0.75) and Model 2 (0.71 and 0.61, respectively for static and dynamic balance factor) showed **a** lower coefficient, thus suggesting that only Model 3 had good internal consistency.

The total score of the Brief-BESTest needs to be weighted considering the loading coefficient of each item. Therefore, scores were adjusted to yield a uniform score range 0-24 for all three models, where 0 represents severe impairment and 24 no balance impairment. Below are the weighted total score equations used to estimate Models 2 and 3:

Brief-BESTest (Model 2) = (0.860504*item1)+(0.914286*item2)+(1.089076*item3)+(1.048739*item4)+(1.021849*item5)+(0.994958*item6)+(1.008438*item7)+(1.062185*item8)

Brief-BESTest (Model 3) = (0.963107*item2)+(1.133981*item3)+(1.118447*item4)+(1.21165*item5)+(1.165049*item6)+(1.165049*item7)+(1.242718*item8)

Convergent validity

The relationship of the total score estimated by the three different models of Brief-BESTest and the ABC 5-levels scale was rho = 0.62 (95% C.I. = 0.55-0.70) for Model 1, rho = 0.61 (95% C.I. = 0.54-0.69) for Model 2 and rho = 0.61 (95% C.I. = 0.54-0.69) for Model 3. No significant (p = 0.85) difference was found between the three Spearman correlation coefficients.

Discriminant validity

Figure 2 shows the ROC curve plotted to assess the ability of the three models of Brief-BESTest to discriminate between patients with vs. without a history of falls. Table 4 reports the discrimination parameters (**cut-off scores**) for Models 1, 2 and 3 and for the simple balance tests

(OLS, TUG test and FR). The AUC was 0.72 (95% C.I. = 0.66–0.78) for all three models. **Model 3 showed superior sensitivity, specificity and likelihood ratios compared to the other two models.** The other simple balance tests (OLS, TUG test and FR) did not reach the AUC value of 0.70, i.e. the cut-off value required to distinguish between fallers and non-fallers. In addition, their AUCs were lower than those of the Brief-BESTest models.

DISCUSSION

The main purpose of this study was to compare, in a large group of patients with balance disorders, the structural, convergent and discriminant validity of three different models of the Brief-BESTest. The Brief-BESTest, in particular Model 3, was found to be unidimensional and to have a good convergent validity with other measures of balance confidence. In addition, the Brief-BESTest confirmed its ability to distinguish subjects with a history of falling from those without a history of falling, in contrast to other simple balance tests.

Structural validity

CFA showed that measurement properties of the original Brief-BESTest scale (Model 1) [13] little fitted our data; none of CFA performance indices reached satisfactory values. As suggested in a previous study **[14]**, the original Brief-BESTest can be improved by making the following modifications: a) removing item 1; b) covarying errors between items 3 and 4 and between items 5 and 6; c) using an appropriate weighting method for the calculation of the total score. The internal structure of the scale can, in fact, be improved by removing item 1 rather than by increasing the number of factors as proposed earlier [19]. The analysis of structural validity prompted two main considerations. First, item 1 does not belong to the same construct as the other items. This finding, in line with previous studies [14,19], is also supported by the fact that “lift a leg to the side of the body” (item 1) could reflect a general reduction in strength rather than a decreased ability to maintain static balance. Second, the Brief-BESTest is unidimensional because no advantage in terms of fitting performance was found when more than one factor was taken into consideration. In other words, our study confirmed that the Brief-BESTest is not able to measure multiple dimensions of balance as claimed by [13]. The Brief-BESTest items include all subsections covered by the BESTest, but this does not mean that the two scales have the same capability to measure the different aspects underlying postural control. In accordance with other authors, one could object that a unidimensional tool is a poor representation of the balance concept, which by definition is multidimensional. We agree with this objection and believe that the Brief-BESTest can assess only certain aspects of dynamic balance. It neglects static components of balance such as those measured by the “Romberg test”. On the other hand, some aspects of balance considered independent in animal models [48], such as walking and maintenance of upright position, could be actually considered as**well** belonging to the same construct in humans. In fact, balance control and locomotion are interdependent at many different levels of the central nervous system and these functions share some common principles of organization [49,50]. Furthermore, previous studies have reported that the ability to maintain upright stance could be related to walking [51,52,53] in patients affected by balance disorders. Hence, in these patients it might be misleading for a clinical scale to investigate balance and walking as two separate factors, as also suggested by recent recommendations [54].

Internal consistency

Despite the large consensus in the psychometric literature that coefficient omega should be used when scales are not congeneric and the assumptions of Cronbach's alpha are not met [41], no previous study has reported omega values for the Brief-BESTest. This is the first study to report both the alpha and the omega coefficients. Based on the latter values, only Model 3 achieved good internal validity. On the contrary, Cronbach's alpha values were found to be good for both Model 1 and 3. Our values of alpha are similar to those reported by [13], and to other studies in both orthopaedic [55] and neurological patients [56]. It is well known that Cronbach's alpha is a function of the number of items. In this regard, it is interesting to note that Model 3, despite its lower number of items, reaches a higher value of internal consistency than the other models. This should denote a superior internal consistency of Model 3.

The discrepancy between the results for Cronbach's alpha and those for coefficient omega can be explained by the fact that the former has been frequently demonstrated to attain quite high values even when the items are measuring different latent variables [57]. On the contrary, the coefficient omega is able to highlight the presence of items that do not belong to the same latent variable. As for the Brief-BESTest, the low values of coefficient omega found for Model 1 could be due to the presence of item 1, which, as highlighted by CFA, seems not to belong to the same construct as the other items.

Convergent validity

The three models compared in this study exhibit an equivalent moderate convergent validity (Spearman's rho = 0.61) with the ABC 5-levels scale. This finding confirms previous studies in which the Brief-BESTest showed a moderate correlation with ABC in specific populations, such as chronic obstructive pulmonary disease [15], stroke [11] and cancer survivors patients [58]. This is not surprising since it is well known that balance confidence is a consequence of balance impairments [59].

Discriminant validity

The three Brief-BESTest models exhibited an equ**ally**ivalent and acceptable accuracy (AUC = 0.72) of the Brief-BESTest in distinguishing patients with vs. without history of falls. However, our AUCs are lower than those of [17], who found in patients with Parkinson's disease AUC values of 0.82, 0.86 and 0.84 respectively for the Brief-BESTest, Mini-BESTest and BESTest, thus indicating a moderate accuracy of the three scales in recognising a history of falls. Likely, the small difference is due to the heterogeneity of our patients. It **We** cannot be excluded that studying separate disease populations might improve the discriminant validity.

In accordance **line** with previous studies, **we confirm that balance scales can discriminate between fallers and non-fallers, in contrast to single balance tests which have a** we found a low level of accuracy in identifying fallers using only a single balance test [11,60,61]. Due to the heterogeneity of patients in our study, it is unlikely that a single test could accurately identify patients at risk of falls. This is in line with recommendations of [62] who suggested to use two or more clinical tests for an accurate evaluation of the various components of risk of falls in patients with Parkinson's disease. The limited association between single measures of balance and fall history could be partly explained by**due to** the multifactorial nature of causes of fall.

Study limitations

Our study was based on a heterogeneous population **sample** of patients affected by different balance disorders **that are representative of the spectrum of neurological diseases routinely**

observed in departments of rehabilitation, but may not be representative of other clinical settings. Although the Brief-BESTest assesses balance performance independently of the pathology that causes the problem, future studies should assess psychometric characteristics such as responsiveness, the minimal clinically important difference and sensitivity to change of the Brief-BESTest in specific groups of disease.

The history of falls should be considered useful only for discriminative purposes (i.e. to distinguish fallers from non-fallers) and not to predict patients who will fall in the future. In fact, falls change people's behavior and the cutoff scores for prospective prediction of falls may be very different from those reported in our study.

Implications for clinical practice and conclusions

The comparison of the three structural models of Brief-BESTest proposed in the literature clearly highlights that Model 3, i.e. the shortest version, has psychometric characteristics equal or superior to the other two. This makes Model 3 the best of the three versions to recommend for use in clinical practice, given that it is also the fastest to perform. For this reason, in the supplementary material we provide a simple calculator for the three models tested in this study. Finally, this study highlights once again the superiority of a clinical scale composed of several items compared to single-item measures.

Authors' contributions

XX and XX contributed to the concept/idea/research design. XX, XX and XX contributed to the writing and data analysis. XX and XX provided data collection. XX and XX provided project management and study participants. XX and XX provided facilities/equipment and institutional liaisons. XX and XX contributed consultation in different phases of the study (including review of manuscript before submission).

Ethical Approval

The Scientific Technical Committee of the XXXXXX approved this study with the following identification number: XXXX.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

None declared.

Bibliography

[1] Geiger RA, Allen JB, O'Keefe J, Hicks RR. Balance and mobility following stroke: effects of physical therapy interventions with and without biofeedback/forceplate training. Phys Ther 2001;81(4):995-1005. PubMed PMID: 11276182.

[2] Hyndman D, Ashburn A. People with stroke living in the community: Attention deficits, balance, ADL ability and falls. Disabil Rehabil 2003;25(15):817-22. PubMed PMID: 12851091.

[3] Horlings CG, Van Engelen BG, Allum JH, Bloem BR. A weak balance: the contribution of muscle weakness to postural instability and falls. Nat Rev Neurol. 2008;4(9):504. doi: 10.1038/ncpneuro0886.

[4] Sturnieks DL, St George R, Lord SR. Balance disorders in the elderly. Neurophysiol Clin 2008;38(6):467-478. doi: 10.1016/j.neucli.2008.09.001.

[5] Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med 2010;46:239. PubMed PMID: 20485226.

[6] 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. doi: 10.2522/ptj.20080131.

[7] 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. doi: 10.2522/ptj.20110063.

[8] Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther 2009;89:484-498. doi: 10.2522/ptj.20080071.

[9] Jacobs JV, Horak FB, Tran VK, Nutt JG. Multiple balance tests improve the assessment of postural stability in subjects with Parkinson’s disease. J Neurol Neurosurg Psychiatry 2006;1;77(3):322-6. PubMed PMID: 16484639.

[10] Jacobs JV, Earhart GM, McNeely ME. Can postural instability tests improve the prediction of future falls in people with Parkinson’s disease beyond knowing existing fall history? J Neurol 2016;263(1):133-9. doi: 10.1007/s00415-015-7950-x.

[11] 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;93:1102-15. doi: 10.2522/ptj.20120454.

[12] 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. doi: 10.2340/16501977-0537.

[13] Padgett PK, Jacobs JV and 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. doi: 10.2522/ptj.20120056.

[14] 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. PubMed PMID: 27103223.

[15] Jácome 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(11):1807-1815. doi: 10.2522/ptj.20150391.

[16] 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. doi: 10.1097/NPT.0b013e31821a620c.

[17] Duncan RP, Leddy AL, Cavanaugh JT, Dibble LE, Ellis TD, Fors MP, Foreman KB, Earhart GM. 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. doi: 10.2522/ptj.20120302.

[18] O'Hoski S, Winship B, Herridge L, Agha T, Brooks D, Beauchamp MK, Sibley KM. Increasing the clinical utility of the BESTest, Mini-BESTest, and Brief-BESTest: normative values in Canadian adults who are healthy and aged 50 years or older. Phys Ther 2014;94:334-342. doi: 10.2522/ptj.20130104.

[19] Franchignoni F, Giordano A. On "Is the BESTest at its best?...." Padgett PK, Jacobs JV, Kasser SL. Phys Ther 2012;92:1197-1207. Phys Ther 2012;92(9):1236-7. doi: 10.2522/ptj.2012.92.9.1236.

[20] Marques A, Almeida S, Carvalho J, Cruz J, Oliveira A, Jácome C. Reliability, Validity, and Ability to Identify Fall Status of the Balance Evaluation Systems Test, Mini–Balance Evaluation Systems Test, and Brief–Balance Evaluation Systems Test in Older People Living in the Community. Arch Phys Med Rehabil 2016;97:2166-2173. doi: 10.1016/j.apmr.2016.07.011.

[21] McArdle JJ. Current directions in structural factor analysis. Curr Dir Psychol Sci 1996;5:11-18.

[22] Norman GR, Streiner DL. PMPH-USA. PDQ statistics; 1st edition, 2003.

[23] Jansen AC, van Aalst-Cohen ES, Hutten BA, Büller HR, Kastelein JJ, Prins MH. Guidelines were developed for data collection from medical records for use in retrospective analyses. J Clin Epidemiol 2005;58(3):269-274. doi: 10.1016/j.jclinepi.2004.07.006.

[24] Jernigan SD, Pohl PS, Mahnken JD, Kluding PM. Diagnostic accuracy of fall risk assessment tools in people with diabetic peripheral neuropathy. Phys Ther 2012;92:1461-1470. doi: 10.2522/ptj.20120070.

484 [25] Lin MR, Hwang HF, Hu MH, Wu HDI, Wang YW, Huang FC. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. J Am Geriatr Soc 2004;52:1343-1348. doi: 10.1111/j.1532-5415.2004.52366.x.

488 [26] Vellas BJ, Wayne SJ, Romero L, Baumgartner RN, Rubenstein LZ, Garry PJ. One-leg balance is an important predictor of injurious falls in older persons. J Am Geriatr Soc 1997;45:735–738. PubMed PMID: 9180669.

491 [27] Sebastião E, Sandroff BM, Learmonth YC, Motl RW. Validity of the timed up and go test as a measure of functional mobility in persons with multiple sclerosis. Arch Phys Med Rehabil 2016;97:1072-1077. doi: 10.1016/j.apmr.2015.12.031.

494 [28] Podsiadlo D, Richardson S. The timed 'Up & Go': A test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991;39:142–148. PubMed PMID: 1991946.

496 [29] Franchignoni F, Tesio L, Martino M, Ricupero C. Reliability of four simple, quantitative tests of balance and mobility in healthy elderly females. Aging Clin Exp Res 1998;10:26-31. PubMed PMID: 9589748.

499 [30] Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. J Gerontol 1990;45(6):M192-7. PubMed PMID: 2229941.

501 [31] Powell LE, Myers AM. The activities-specific balance confidence (ABC) scale. J Gerontol A Biol Sci Med Sci 1995;50(1):M28-M34. PubMed PMID: 7814786.

503 [32] Franchignoni F, Giordano A, Ronconi G, Rabini A, Ferriero G. Rasch validation of the Activities-specific Balance Confidence Scale and its short versions in patients with Parkinson's disease. J Rehabil Med 2014;46:532-539. doi: 10.2340/16501977-1808.

[33] Atkinson TM, Rosenfeld BD, Sit L, Mendoza TR, Fruscione M, Lavene D, Shaw M, Li Y, Hay J, Cleeland CS, Scher HI, Breitbart WS, Basch E. Using confirmatory factor analysis to evaluate construct validity of the Brief Pain Inventory (BPI). J Pain Symptom Manage 2011;41(3):558-565. doi: 10.1016/j.jpainsymman.2010.05.008.

[34] Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107(2):238-46. PubMed PMID: 2320703.

[35] Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing Structural Equation Models. Newbury Park: Sage 1993;136–62.

[36] Hu LT, Bentler P. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J 1999;6:1-55. doi: 10.1080/10705519909540118.

[37] DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: Considerations for the applied researcher. PARE 2009;14(20):1-11.

[38] Schönberger M, Ponsford J, Olver J, Ponsford, M. A longitudinal study of family functioning after TBI and relatives' emotional status. Neuropsychol Rehabil 2010;20:813-829. doi: 10.1080/09602011003620077.

[39] Hancock GR, Mueller RO. Rethinking construct reliability within latent variable systems. Structural equation modeling: Present and future-A Festschrift in honor of Karl Jöreskog. Lincolnwood, IL: Scientific Software International 2001;195-216.

[40] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):297–334. PubMed PMID: 8449658.

[41] Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 2014;105:399-412. doi: 10.1111/bjop.12046.

[42] Opheim A, Jahnsen R, Olsson E, Stanghelle JK. Balance in relation to walking deterioration in adults with spastic bilateral cerebral palsy. Phys Ther 2012;92(2):279-88. doi: 10.2522/ptj.20100432.

[43] Curtze C, Nutt JG, Carlson-Kuhta P, Mancini M, Horak FB. Objective gait and balance impairments relate to balance confidence and perceived mobility in people with Parkinson disease. Phys Ther 2016;96(11):1734-1743. PubMed PMID: 27149959.

[44] Rand D. Mobility, balance and balance confidence - correlations with daily living of individuals with and without mild proprioception deficits post-stroke. NeuroRehabilitation 2018;43(2):219-226. doi: 10.3233/NRE-172398.

[45] Salbach NM, Mayo NE, Hanley JA, Richards CL, Wood-Dauphinee S. Psychometric evaluation of the original and Canadian French version of the activities-specific balance confidence scale among people with stroke. Arch Phys Med Rehabil 2006;87(12):1597-604. PubMed PMID: 17141639.

[46] Cohen, J. Statistical power analysis for the behavioral sciences. 2nd edition, 1988.

[47] Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatr 2007;96:664-647. doi: 10.1111/j.1651-2227.2006.00178.x.

[48] Hodgson JA, Roy RR, de Leon R, Dobkin B, Edgerton VR. Can the mammalian lumbar spinal cord learn a motor task? Med Sci Sports Exerc 1994;26:1491-1497. PubMed PMID: 7869884.

[49] Grasso R, Zago M, Lacquaniti F. Interactions between posture and locomotion: motor patterns in humans walking with bent posture versus erect posture. J Neurophysiol 2000;83:288-300. doi: 10.1152/jn.2000.83.1.288.

[50] Massion J. Movement, posture and equilibrium: interaction and coordination. Prog Neurobiol 1992;38(1):35-56. Review. PubMed PMID: 1736324.

[51] Giardini M, Nardone A, Godi M, Guglielmetti S, Arcolin I, Pisano F, Schieppati M. Instrumental or physical-exercise rehabilitation of balance improves both balance and gait in Parkinson's disease. Neural Plast 2018;2018:5614242. doi: 10.1155/2018/5614242.

[52] Nardone A, Godi M, Grasso M, Guglielmetti S, Schieppati M. Stabilometry is a predictor of gait performance in chronic hemiparetic stroke patients. Gait Posture 2009;30(1):5-10. doi: 10.1016/j.gaitpost.2009.02.006.

[53] Morlino P, Balbi B, Guglielmetti S, Giardini M, Grasso M, Giordano C, Schieppati M, Nardone A. Gait abnormalities of COPD are not directly related to respiratory function. Gait Posture 2017;58:352-357. doi: 10.1016/j.gaitpost.2017.08.020.

[54] Bloem BR, Marinus J, Almeida Q, Dibble L, Nieuwboer A, Post B, Ruzicka E, Goetz C, Stebbins G, Martinez-Martin P, Schrag A; Movement Disorders Society Rating Scales Committee. Measurement instruments to assess posture, gait, and balance in Parkinson's disease: Critique and recommendations. Mov Disord 2016;31(9):1342-55. doi: 10.1002/mds.26572.

[55] Chan AC, Pang MY. Assessing balance function in patients with total knee arthroplasty. Phys Ther 2015;95(10):1397-407. doi: 10.2522/ptj.20140486.

568 [56] Huang M, Pang MY. Psychometric properties of Brief-Balance Evaluation Systems Test (Brief-BESTest) in evaluating balance performance in individuals with chronic stroke. Brain Behav 2017;7(3):e00649. doi: 10.1002/brb3.649.

571 [57] Revelle W, Zinbarg RE. Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. Psychometrika 2009;74(1):145. doi:10.1007/s11336-008-9102-z.

573 [58] Huang MH, Miller K, Smith K, Fredrickson K, Shilling T. Reliability, validity, and minimal detectable change of Balance Evaluation Systems Test and its short versions in older cancer survivors: a pilot study. J Geriatr Phys Ther 2016;39(2):58-63. doi: 10.1519/JPT.0000000000000047.

577 [59] Mak MK, Pang MY. Balance confidence and functional mobility are independently associated with falls in people with Parkinson's disease. J Neurol 2009;256:742-749. doi: 10.1007/s00415-009-5007-8.

580 [60] Cameron M, Mazumder R, Murchison C, King L. Mini Balance Evaluation Systems Test in people with multiple sclerosis: reflects imbalance but may not predict falls. Gait Posture 2014;39(1):669. doi: 10.1016/j.gaitpost.2013.08.009.

583 [61] Mak MK, Auyeung MM. The mini-BESTest can predict parkinsonian recurrent fallers: a 6-month prospective study. J Rehabil Med 2013;45:565-571. doi: 10.2340/16501977-1144.

585 [62] Dibble LE, Christensen J, Ballard DJ, Foreman KB. Diagnosis of fall risk in Parkinson disease: an analysis of individual and collective clinical balance test interpretation. Phys Ther 2008;88:323-332. doi: 10.2522/ptj.20070082.

588

Table 1

Demographic data and scores on balance measures for the entire sample, and fallers vs. non-fallers.

	Entire **Total** Sample (N=416)	Fallers (N=135)	Non-fallers (N=160)	*p*
	N %	N %	N %	
Sex (M/F)	55/45	51/49	58/42	.23[a]
Walking Aid (no/yes)	54/46	47/53	68/32	<.001[a]
	Mean (SD)	Mean (SD)	Mean (SD)	
Age (y)	66.5 (14.2)	66.3 (13.4)	64.4 (14.8)	.28[b]
OLS mean (s)	4.2 (6.3)	3.3 (5.0)	6.9 (7.5)	<.001[b]
TUG test (s)	17.6 (15.3)	20.6 (20.3)	13.3 (9.5)	<.001[b]
Functional Reach (cm)	18.6 (9.7)	16.5 (9.6)	20.5 (9.8)	<.01[b]
Total score BBT	8.9 (6.4)	7.3 (5.1)	11.8 (6.9)	<.001[b]
Total score ABC 5-levels	29.6 (16.8)	26.0 (15.0)	36.2 (17.1)	<.001[b]

The total sample comprises all patients, i.e. those in whom a fall history was collected (the two subgroups fallers/non-fallers) as well as those in whom a fall history was not collected due to acute stroke event (n=79) or failure to complete the questionnaire (n=42), as this is a retrospective study. Participants were classified as fallers if they reported 1 **2** or more falls in the last 6 months.

SD, Standard Deviation; M, Male; F, Female; OLS, One Leg Stance; TUG test, Timed Up and Go test; BBT, Brief-BESTest; ABC 5-levels, Activities-specific Balance Confidence scale 5-levels of rating.

p-value was computed between fallers and non-fallers.

[a] = χ^2 test.

[b] = Mann-Whitney U-test.

Table 2

Details on each single item and total score of the original Brief-BESTest (Model 1) in the whole sample (N = 416).

	Median	25% percentile	75% percentile	**Skewness**	Kurtosis	Min score (%)	Max score (%)	Missing data (%)
Item 1	1	0	2	0.39	-1.20	42.51	7.73	0.40
Item 2	2	1	2	-0.47	-0.13	9.42	10.39	0.40
Item 3	1	0	1	0.93	-0.15	46.62	8.70	0.00
Item 4	1	0	1	0.89	-0.25	44.69	9.66	0.00
Item 5	1	0	2	0.59	-1.19	48.07	17.15	0.00
Item 6	1	0	2	0.53	-1.26	45.89	18.84	0.00
Item 7	1	0	2	0.66	-0.99	43.96	16.91	0.00
Item 8	2	0	2	0.07	-1.54	36.47	23.43	0.00
Total score	8	4	13	0.54	-0.69	6.00	8.00	

Item 1 - Hip/Trunk Lateral Strength; item 2 - Functional Reach Forward; item 3 and 4 - One Leg Stance, Left and Right; item 5 and 6 - Compensatory Stepping- Lateral, Left and Right; item 7 - stand on a foam surface; 8 - Timed Up and Go test.

Table 3

Summary of fit statistics of the specified models (N = 416)

Models	χ^2	df	RMSEA (90% C.I.)	CFI	TLI	SRMR
Model 1 *(Padgett et al., 2012)*	134.0	20	0.12 (0.10-0.14)	0.78	0.70	0.09
Model 2 *(adapted from Franchignoni & Giordano, 2012)*	60.3	18	0.08 (0.06-0.10)	0.92	0.88	0.05
Model 3 *(adapted from Bravini et al., 2016)*	26.2	12	0.05 (0.03-0.08)	0.97	0.95	0.03

Comparison of factor models	Significant difference between models[a]
Model 1 vs. Model 2	$p<0.001$
Model 2 vs. Model 3	$p<0.001$
Model 3 vs. Model 1	$p<0.001$

df, degrees of freedom; RMSEA, Root Mean-Square Error of Approximation; C.I., Confidence Interval; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; SRMSR, Standardized Root Mean Square Residual.

[a] Calculated as a $\chi^2 = \chi^2_2 - \chi^2_1$ with $df = df_2 - df_1$.

Table 4

Ability of the Brief-BESTest Models 1, 2 and 3 to discriminate fallers from non-fallers compared to other measures of balance (OLS, TUG test, Functional Reach and ABC 5-levels) **(N = 295).**

Classification	Brief-BESTest Model 1	Brief-BESTest Model 2	Brief-BESTest Model 3	OLS	TUG test	FR
Cut-off score	8	8	7	2	12	19
AUC (95% C.I.)	0.72 (0.66-0.78)	0.72 (0.66-0.78)	0.72 (0.66-0.78)	0.63 (0.55-0.70)	0.62 (0.56-0.68)	0.61 (0.52-0.69)
Sensitivity	74%	72%	74%	63%	68%	67%
Specificity	54%	58%	58%	55%	53%	49%
LR+	1.61	1.72	1.76	1.42	1.45	1.33
LR-	0.48	0.47	0.45	0.66	0.60	0.67
Correctly Classified	65%	66%	67%	59%	60%	58%

OLS, One Leg Stance; TUG test, Timed Up and Go test; FR, Functional Reach; AUC, Area Under the Curve; C.I., Confidence Interval; LR+, positive Likelihood Ratio; LR-, negative Likelihood Ratio.

Appendix 1

Items and factors of the three models of Brief-BESTest.

<table>
<tr><th>Model item
(Scoring 0-3 per item)</th><th>Model 1</th><th>Model 2</th><th>Model 3</th></tr>
<tr><td>Item 1: Hip/Trunk Lateral Strength</td><td rowspan="8">Dynamic Balance</td><td rowspan="2">Static Balance</td><td></td></tr>
<tr><td>Item 2: Functional Reach Forward</td><td rowspan="7">Dynamic Balance</td></tr>
<tr><td>Item 3: Stand on One Leg-Left</td><td rowspan="6">Dynamic Balance</td></tr>
<tr><td>Item 4: Stand on One Leg-Right</td></tr>
<tr><td>Item 5: Compensatory Stepping–Lateral, Left</td></tr>
<tr><td>Item 6: Compensatory Stepping–Lateral, Right</td></tr>
<tr><td>Item 7: Stand on foam surface with Eyes Closed</td></tr>
<tr><td>Item 8: Timed Up and Go test</td></tr>
<tr><td>Total score</td><td>0-24</td><td>0-24
(Static 0-5;
Dynamic 0-19)</td><td>0-24</td></tr>
</table>

Is the Brief-BESTest brief enough? Suggested modifications based on structural validity and internal consistency

Marco Godi [a], Marica Giardini [a,*], Ilaria Arcolin [a], Simona Ferrante [b], Antonio Nardone [a,c], Stefano Corna [a], Roberto Colombo [a]

Author affiliation:

[a] Istituti Clinici Scientifici Maugeri IRCCS, Italy

[b] Neuroengineering and Medical Robotics Laboratory, Politecnico di Milano, Italy

[c] Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Italy

[*]Corresponding author: Marica Giardini.

Istituti Clinici Scientifici Maugeri IRCCS,

Via per Revislate 13, Veruno 28010, Italy.

E-mail address: marica.giardini@icsmaugeri.it

Ph: +39 0322 884905

ABSTRACT

Background: The Brief Balance Evaluation Systems Test (Brief-BESTest) could be a useful tool for balance assessment. Although some psychometric characteristics have been examined, others still need to be clarified.

Objectives: To assess the structural validity, convergent validity, discriminant validity and internal consistency of the Brief-BESTest in neurological patients.

Design: Cross-sectional.

Methods: Data were from 416 patients with neurological disease and related balance disorders. Patients were assessed with the 5-levels Activities-Specific Balance Confidence Scale (ABC 5-levels), Brief-BESTest, **some simple balance tests, i.e. One-Leg Stance (OLS), Timed Up and Go (TUG) test, Functional Reach (FR),** simple balance tests and **a** fall history questionnaire. Three models of Brief-BESTest **models** were examined through confirmatory factor analysis (CFA) and the following indices calculated: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA). Convergent validity was assessed by calculating the correlation between Brief-BESTest and ABC 5-levels total scores. Receiver operating characteristics (ROC) assessed the ability of each model to differentiate between people with vs. without falls. Internal consistency was measured by Cronbach's alpha and coefficient omega.

Results: CFA showed Model 3 (CFI=0.97, TLI=0.95, RMSEA=0.05), **with item 1 removed and error covariance between items 3-4 and between items 5-6,** to have a significantly better structure than Models 1 and 2 ($p<0.001$). The correlation between Brief-BESTest and ABC 5-levels was 0.61 (Spearman's rho) for all three models. Area Under the Curve (AUC) of ROC was **showed an acceptable accuracy (0.72) in distinguishing patients with vs. without history of falls** (95% C.I.=0.66–0.78) for all models, and superior to AUCs of other simple balance tests **(OLS, TUG test,**

FR). Cronbach's alpha was good for Brief-BESTest Models 1 (0.92) and 3 (0.92), but omega was >0.80 only for Model 3.

Limitations: Heterogeneous sample size was a heterogeneous population.

Conclusions: The Brief-BESTest**, after some changes,** shows good validity and internal consistency in patients affected by different balance disorders, after applying some changes.

Contribution of the Paper:

- Although some psychometric characteristics of the Brief-BESTest have been examined in previous studies, other properties such as validity still need to be clarified.
- This study shows that the Brief-BESTest has good validity and internal consistency in patients affected by different balance disorders, after applying some changes: removal of item 1 and using an appropriate weight**ing** method for the calculation of the total score.
- This study confirms the ability of the Brief-BESTest to distinguish between people with vs. without history of falls, in contrast to other simple balance tests. Moreover, it highlights once again the superiority of a clinical scale composed of several items compared to single-item measures such as the TUG test and OLS.

Keywords: Brief-Balance Evaluation Systems Test, balance assessment, confirmatory factor analysis, structural validity, internal consistency.

INTRODUCTION

Balance disorders are a common finding in a broad spectrum of neurological disorders and are characterized by a heterogeneous set of signs and symptoms. Patients **with balance disorders experience a reduction in mobility, activities of daily living and muscle strength, leading to increased risk of falls** [1,2,3,4]. Thus, balance assessment is crucial and requires standardized measurement tools that can monitor equilibrium regardless of the pathology. Unfortunately, no gold standard exists for evaluating balance [5], and no consensus on which assessment tools to use in clinical practice [6,7].

A variety of clinical measures has been developed to evaluate different aspects of balance. While simple balance tests such as the Timed Up and Go test, One Leg Stance, and Functional Reach provide accurate evaluation of a single task, they are not able to **do not give** information on multifactorial mechanisms related to postural stability [8]. On the contrary, balance scales which include multiple tasks can provide a more complete picture of balance control in all its complexity [9,10,11].

One of the most recent balance scales is the BESTest, a 36-item scale developed to identify impairments in six balance control subsections, which it has been shown to be a valid and reliable tool [8]. However, one of its drawbacks is that it is time-consuming **to administer** [12]. For this reason, shorter versions have been proposed such as the Mini-BESTest [12] and the Brief-BESTest [13]. In particular, the Brief-BESTest, an 8-item version of the original scale, has demonstrated good to excellent psychometric properties [8,13,14,15]. It is less time-consuming [16] and more feasible than its parent scale in clinical settings [15], while it encompasses and should adequately evaluate all subsections of balance endorsed by the BESTest [13,17,18].

However, the latter hypothesis was rejected through exploratory factor analysis **dismantled the latter hypothesis**, demonstrating **that the Brief-BESTest actually has at most two subsections, or dimensions** [19]. Furthermore, Bravini et al. [14] showed by Rasch analysis that all items of the Brief-BESTest except for item 1 account for the same underlying theoretical construct and indicated that the Brief-BESTest should in fact be considered as unidimensional. Therefore, the authors suggested the adoption of a 7-item version of the test.

Although some psychometric characteristics of the Brief-BESTest, such as the internal consistency [13,14], reliability [14,17,20] and sensitivity to change [14,20], have been investigated in previous studies, other properties still need to be clarified. In particular, the Brief-BESTest structure has not yet been investigated with **undergone** confirmatory factor analysis (CFA). This statistical tool provides information on possible independent factors and can be very useful for developing shortened forms of an evaluating instrument [21,22]. Finally, the Brief-BESTest seems to have good sensitivity and accuracy in identifying retrospectively people who have had at least one fall [13,17]. However, these findings are based only on small samples of patients with multiple sclerosis [13] or Parkinson's disease [17].

The **We** aim**ed in** of the present study was to fill the existing knowledge gap by examining the structural validity, convergent validity and discriminant validity of the Brief-BESTest in a large group of patients with a variety of balance disorders. In particular, we hypothesized that:

1) among the three models of the Brief-BESTest presented in the literature [13,14,19], the 7-item version would be the one with the best structural validity;

2) in spite of its conciseness, the 7-item model [14] would have the same ability as the other two Brief-BESTest models to predict patients at risk of falls; **discriminate between people with vs. without a history of falls;**

3) in discriminating between people with vs. without a history of falls, the Brief-BESTest would be superior to other simple balance tests such as One Leg Stance, Timed Up and Go test and Functional Reach.

METHODS

Participants

This was an observational retrospective study conducted in a group of 416 patients affected by different neurological diseases: 186 females and 230 males; mean age 66.5±16.0 years (**mean ± standard deviation)** consecutively admitted for in-patient rehabilitation at the Istituti Clinici Scientifici Maugeri IRCCS, Institute of Veruno (Novara, Italy) between February 2014 and April 2017. **Patients' clinical and treatment data were extracted from the electronic medical record system and transferred to a specific database (Microsoft Excel).** Patients were stratified into different groups according to their diagnosis: 118 **with** Parkinson's disease, 79 **with** acute stroke, 43 **with** sensorimotor polyneuropathy, 32 **with** cerebellar ataxia, 32 **with** diffuse encephalopathy, 31 **with** chronic stroke, 21 **with** multiple sclerosis, 19 **with** traumatic brain injury, 16 **with** vestibular disorder, 13 **with** neuromuscular disorders, 12 **with** central nervous system neoplasm. Inclusion criteria were: a) ability to maintain an upright position without support for at least 5 seconds; b) ability to understand the required motor tasks; c) no hip or knee replacement surgery within the previous 6 months. Exclusion criteria were: a) musculoskeletal injury limiting the ability to walk; b) any other serious cardio-respiratory problem. The study was carried out in conformity with the Declaration of Helsinki of the World

Medical Association and the guidelines for retrospective studies [23]. The local scientific and ethics committee approved the study.

Assessment tools

Patients' demographic and clinical characteristics were gathered by **A**a team of trained physical **therapists engaged in clinical practice collected patients' demographic and clinical characteristics. During the routine clinical assessment following admission to the rehabilitation department, patients underwent the following assessments:**

- Fall history

A history of falls over the past 6 months was obtained **from patients** at admission through patient interview. **A fall was defined as an unintentional event in which any part of the body came into contact with the ground [10].** Patients who reported two or more falls in the defined period were classified as 'fallers' [24]. A fall history was not recorded **taken in the case of** for patients with acute stroke at the time of admission.

- Brief-BESTest

The Brief-BESTest is an 8-item scale with each item scored on a 4-level rating scale from 0 (severe balance impairment) to 3 (no balance impairment). Its items cover the six subsections of the original BESTest (biomechanical constraints, stability limits/verticality, anticipatory postural responses, postural responses, sensory orientation, stability in gait); the maximum total score is 24 [13]. **The Brief-BESTest requires less time and equipment to administer than the BESTest and the Mini-BESTest; thus, the Brief-BESTest seems to be more feasible for clinical use [17].**

- Simple balance tests

During administration of the Brief-BESTest, we recorded the time required by patients to complete item 3 (left One Leg Stance), 4 (right One Leg Stance) and 8 (Timed Up and Go test) and the distance covered by patients during item 2 (Functional Reach). This allowed us to obtain the scores of three additional simple balance tests: One Leg Stance (OLS) [25,26], Timed Up and Go (TUG) test [27,28] and Functional Reach (FR) [29,30].

- Activities-specific Balance Confidence Scale

The short version of the Activities-specific Balance Confidence Scale (ABC) is a self-reported 16-item questionnaire that scores the perceived level of balance confidence when performing common activities of daily living [31]. We used the 5-levels **rating** version of rating (ABC 5-levels) [32] in which each item is scored from 0 (no confidence) to 4 (fully confident), giving a total score range 0-64.

Data analysis

Descriptive statistics were used to describe mean demographic and balance performance characteristics of the entire sample and of the two **smaller** subgroups, classified as fallers and non-fallers. **The analysis of discriminant validity was conducted only in these two subgroups, i.e. those patients who had a history of falls available.** These values were also determined separately both for fallers and non-fallers. For each item of the Brief-BESTest we calculated: median value, spread (25th–75th percentiles), skewness and kurtosis. Floor and ceiling effects were analyzed by calculating the percentage of individuals obtaining the lowest and the highest score for each scale item. In order to detect differences in clinical characteristics between fallers and non-fallers, the Chi-square (χ^2) test was used for two parameters, sex and use of walking aids, while the Mann-Whitney U-test was used for age, OLS, TUG test, FR, total score of Brief-BESTest and total score of ABC 5-levels. Significance was set at $p<0.05$. All analyses were performed using STATA 13.0 software (StataCorp LLC, College Station, Texas, USA).

Structural validity

Structural validity is usually assessed through CFA [33]. CFA assesses the degree to which responses on a p x 1 vector of observable random variables (the Brief-BESTest items) can be used to assign a value to one or more unobserved variable(s) (balance subsections). For this purpose, a specific mathematical model is identified and fitted to the patients' data.

The original model (Model 1) of the Brief-BESTest [13] comprises one factor with 8 independent items that contribute with the same weights to the total score. In the recent literature, two additional models of Brief-BESTest have been presented. Model 2 was designed based on [19]. It includes two factors: one named "static balance" that comprises items 1 and 2, and another called "dynamic balance" that contains items 3 to 8. As demonstrated by the authors, items 5 and 6 showed local dependence, so Model 2 was designed allowing correlation between the errors of these items. **In** Model 3**,** was drawn up without item 1 **was dropped,** as suggested by [14]. In this 7-item model, the error of item 3 was allowed to correlate with that of item 4 and the error of item 5 with that of item 6.

For all models, the score of each item ranges from 0 to 3. Then the total score was **is** obtained by multiplying the rated score by the coefficient fitted for each model (see below formula and supplementary data). In order to allow comparison of the score models, **we adjusted** the coefficients **so** have been adjusted as to maintain a total score in the range 0 to 24. Appendix 1 summarizes the item structure of the three models and their total score.

Preliminary analysis conducted on Model 3 showed similar CFA results for patients who used a walking aid and those who did not. For this reason, we decided to consider the entire sample independently of the walking condition in creating our models.

We examined t The structural validity of these three models (from here on Model 1, 2 and 3) was examined through CFA using Structural Equation Modeling (SEM). In view of the very low occurrence of missing data (maximum 0.4%), cases with missing data were removed from the analysis. χ^2 was used to identify whether the model fitted the data well. In addition, we assessed the models' goodness of fit using the following indices: Comparative Fit Index (CFI) [34], Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) [35] and the Standardized Root Mean Square Residual (SRMR) [34]. The criteria adopted to assess goodness of fit performance were: a) CFI and TLI values ≥0.95; b) RMSEA value <0.06; and c) SRMR value ≤0.08 [36].

The goodness of fit parameters of the three models were compared by computing the χ^2 difference tests of each model pair, calculated as a $\chi^2 = \chi^2_2\text{-}\chi^2_1$ with $df = df_2–df_1$.

The standardized factor loadings of the models (i.e. the coefficients of the fitted model) were then transformed into weights that can be applied when scale scores for an individual are calculated. They were calculated with a non–refined method called "Weighted Sum Scores" [37]; these weights do not change the scale range [38].

Internal consistency

The internal consistency of the three Brief-BESTest models was measured by Cronbach's alpha and the coefficient omega for congeneric models [39]. Cronbach's alpha measures the extent to which the items consistently measure the same construct, with the value ≥0.80 indicating good internal consistency [40]. The coefficient omega is the ratio of the true score variance and the total variance of the scale. Interpretation of coefficient omega is similar to that of Cronbach's alpha [41].

Convergent validity

We used the correlation between the three models of Brief-BESTest and the ABC 5-levels total score to assess the convergent validity. **The** ABC scale led to rate **rates** the balance self-efficacy of patients [31]. This is associated with measures of balance [32], walking capacity [42], functional mobility [43], Activities of Daily Living performance [44], and perceived health status [45]. The choice to use the ABC as a competitor **an external criterion** was also based on the need to avoid the comparisons ofng the three models with a scale that had items similar to the Brief-BESTest. The correlation was assessed by means of Spearman's rho: coefficients <0.30 were interpreted as weak, those between 0.30 and 0.49 as moderate, and those ≥0.50 as strong correlations [46].

Discriminant validity

To assess the ability of the three Brief-BESTest models to distinguish between 'fallers' and 'non-fallers', receiver operating characteristic (ROC) curves were computed**,** by plotting sensitivity on the *x*-axis against 1 – specificity on the *y*-axis. In our study, sensitivity was calculated as the number of patients correctly identified as 'fallers' and specificity as the number of patients correctly identified as 'non-fallers'. The optimal cut-off value was chosen on the ROC curve at the point that jointly maximized sensitivity and specificity [47]. For each ROC, the area under the curve (AUC) and the positive and negative likelihood ratios (LR+ and LR−) were computed **to maximize the cut off scores**. Low, moderate and high accuracy of discrimination were defined respectively as AUC <0.70, 0.70< AUC <0.90, and AUC >0.90 [47]. In addition, the predictive performance of the three models was compared to that of the OLS, TUG test and FR tests by reporting the above described parameters and the percentage of correctly classified patients.

RESULTS

Fall history data**, for the analysis of discriminant validity, was available** could be collected only from **in** 295 subjects: 135 fallers (45%) and 160 non-fallers (55%). Table 1 reports mean **scores** and standard deviation for each balance measure as well as information on the use of walking aids in the overall sample and in fallers vs. non-fallers. We found a significant difference between fallers and non-fallers in the mean score of all clinical evaluations, while mean age and sex did not differ significantly. Table 2 reports descriptive statistics for each item score and total score of the original 8-item Brief-BESTest in the whole sample.

Structural validity

Analysis of the Brief-BESTest models

Figure 1A shows the standardized solution of the CFA for Model 1 that was fitted using all 8 items. We found χ^2 value of 134.0 (df = 20, p<0.001), with CFI of 0.78, TLI of 0.70, RMSEA of 0.12 (90% C.I. = 0.10–0.14) and SRMR above 0.09. Both CFI and TLI were below the cut-off value of 0.95 defined for a well-fitting model. In addition, the RMSEA value suggested that this model exhibited a poor fit of the data.

Model 2 (figure 1B) showed better values of goodness of fit with respect to the original model. In fact, Model 2 had χ^2 of 60.3 (df = 18, p<0.001), with CFI of 0.92, TLI 0.88, RMSEA 0.08 (90% C.I. = 0.06–0.10) and SRMR 0.05. However, only SRMR exhibited a value lower than the preselected well-fitting index value.

Figure 1C shows the CFA solution for Model 3. The results show a significantly better goodness of fit for this model (χ^2 = 26.2, df = 12, CFI = 0.97, TLI = 0.95, RMSEA = 0.05 (90% C.I. = 0.03–0.08), SRMR = 0.03) than for Models 1 and 2. Comparison of Model 3 (χ^2 = 26.2) to Model 1 (χ^2 = 134.0) and Model 2 (χ^2 = 60.3) yielded a difference in χ^2 value of 107.8 and 34.1 respectively and a difference of 6 degrees of freedom, suggesting that Model 3 performed better than the original Brief-BESTest (Model 1) and Model 2. Table 3 summarizes the goodness of fit indices of each model and the significance level of the comparison between each model pair.

The factor loadings of each item were significant (p<0.001) and higher than 0.6 for all three models. Item 1 (Hip/Trunk Lateral Strength), when present, and item 2 (FR Forward) had the lowest factor loading.

Internal consistency

Cronbach's alpha was good for Brief-BESTest Models 1 (0.92) and 3 (0.92), but not for Model 2 (0.56 and 0.88, respectively for static and dynamic balance factor). Coefficient omega was higher than 0.80 only for Model 3, while Model 1 (0.75) and Model 2 (0.71 and 0.61, respectively for static and dynamic balance factor) showed **a** lower coefficient, thus suggesting that only Model 3 had good internal consistency.

The total score of the Brief-BESTest needs to be weighted considering the loading coefficient of each item. Therefore, scores were adjusted to yield a uniform score range 0-24 for all three models, where 0 represents severe impairment and 24 no balance impairment. Below are the weighted total score equations used to estimate Models 2 and 3:

Brief-BESTest (Model 2) = (0.860504*item1)+(0.914286*item2)+(1.089076*item3)+(1.048739*item4)+(1.021849*item5)+(0.994958*item6)+(1.008438*item7)+(1.062185*item8)

Brief-BESTest (Model 3) = (0.963107*item2)+(1.133981*item3)+(1.118447*item4)+(1.21165*item5)+(1.165049*item6)+(1.165049*item7)+(1.242718*item8)

Convergent validity

The relationship of the total score estimated by the three different models of Brief-BESTest and the ABC 5-levels scale was rho = 0.62 (95% C.I. = 0.55-0.70) for Model 1, rho = 0.61 (95% C.I. = 0.54-0.69) for Model 2 and rho = 0.61 (95% C.I. = 0.54-0.69) for Model 3. No significant (p = 0.85) difference was found between the three Spearman correlation coefficients.

Discriminant validity

Figure 2 shows the ROC curve plotted to assess the ability of the three models of Brief-BESTest to discriminate between patients with vs. without a history of falls. Table 4 reports the discrimination parameters (**cut-off scores**) for Models 1, 2 and 3 and for the simple balance tests (OLS, TUG test and FR). The AUC was 0.72 (95% C.I. = 0.66–0.78) for all three models. **Model 3 showed superior sensitivity, specificity and likelihood ratios compared to the other two models.** The other simple balance tests (OLS, TUG test and FR) did not reach the AUC value of 0.70, i.e. the cut-off value required to distinguish between fallers and non-fallers. In addition, their AUCs were lower than those of the Brief-BESTest models.

DISCUSSION

The main purpose of this study was to compare, in a large group of patients with balance disorders, the structural, convergent and discriminant validity of three different models of the Brief-BESTest. The Brief-BESTest, in particular Model 3, was found to be unidimensional and to have a good convergent validity with other measures of balance confidence. In addition, the Brief-BESTest confirmed its ability to distinguish subjects with a history of falling from those without a history of falling, in contrast to other simple balance tests.

Structural validity

CFA showed that measurement properties of the original Brief-BESTest scale (Model 1) [13] little fitted our data; none of CFA performance indices reached satisfactory values. As suggested in a previous study **[14]**, the original Brief-BESTest can be improved by making the following modifications: a) removing item 1; b) covarying errors between items 3 and 4 and between items 5 and 6; c) using an appropriate weighting method for the calculation of the total score. The internal structure of the scale can, in fact, be improved by removing item 1 rather than by increasing the number of factors as proposed earlier [19]. The analysis of structural validity prompted two main considerations. First, item 1 does not belong to the same construct as the other items. This finding, in line with previous studies [14,19], is also supported by the fact that "lift a leg to the side of the body" (item 1) could reflect a general reduction in strength rather than a decreased ability to maintain static balance. Second, the Brief-BESTest is unidimensional because no advantage in terms of fitting performance was found when more than one factor was taken into consideration. In other words, our study confirmed that the Brief-BESTest is not able to measure multiple dimensions of balance as claimed by [13]. The Brief-BESTest items include all subsections covered by the BESTest, but this does not mean that the two scales have the same capability to measure the different aspects underlying postural control. In accordance with other authors, one could object that a unidimensional tool is a poor representation of the balance concept, which by definition is multidimensional. We agree with this objection and believe that the Brief-BESTest can assess only certain aspects of dynamic balance. It neglects static components of balance such as those measured by the "Romberg test". On the other hand, some aspects of balance considered independent in animal models [48], such as walking and maintenance of upright position, could be actually considered as**well** belonging to the same construct in humans. In fact, balance control and locomotion are interdependent at many different levels of the central nervous system and these functions share some common principles of organization [49,50]. Furthermore, previous studies have reported that the ability to maintain upright stance could be related to walking [51,52,53] in patients affected by balance disorders. Hence, in these patients it might be misleading for a clinical scale to investigate balance and walking as two separate factors, as also suggested by recent recommendations [54].

Internal consistency

Despite the large consensus in the psychometric literature that coefficient omega should be used when scales are not congeneric and the assumptions of Cronbach's alpha are not met [41], no previous study has reported omega values for the Brief-BESTest. This is the first study to report both the alpha and the omega coefficients. Based on the latter values, only Model 3 achieved good internal validity. On the contrary, Cronbach's alpha values were found to be good for both Model 1 and 3. Our values of alpha are similar to those reported by [13], and to other studies in both orthopaedic [55] and neurological patients [56]. It is well known that Cronbach's alpha is a function of the number of items. In this regard, it is interesting to note that Model 3, despite its lower number of items, reaches a higher value of internal consistency than the other models. This should denote a superior internal consistency of Model 3.

The discrepancy between the results for Cronbach's alpha and those for coefficient omega can be explained by the fact that the former has been frequently demonstrated to attain quite high values even when the items are measuring different latent variables [57]. On the contrary, the coefficient omega is able to highlight the presence of items that do not belong to the same latent variable. As for the Brief-BESTest, the low values of coefficient omega found for Model 1 could be due to the presence of item 1, which, as highlighted by CFA, seems not to belong to the same construct as the other items.

Convergent validity

The three models compared in this study exhibit an equivalent moderate convergent validity (Spearman's rho = 0.61) with the ABC 5-levels scale. This finding confirms previous studies in which the Brief-BESTest showed a moderate correlation with ABC in specific populations, such as chronic obstructive pulmonary disease [15], stroke [11] and cancer survivors patients [58]. This is not surprising since it is well known that balance confidence is a consequence of balance impairments [59].

Discriminant validity

The three Brief-BESTest models exhibited an equ**ally**ivalent and acceptable accuracy (AUC = 0.72) of the Brief-BESTest in distinguishing patients with vs. without history of falls. However, our AUCs are lower than those of [17], who found in patients with Parkinson's disease AUC values of 0.82, 0.86 and 0.84 respectively for the Brief-BESTest, Mini-BESTest and BESTest, thus indicating a moderate accuracy of the three scales in recognising a history of falls. Likely, the small difference is due to the heterogeneity of our patients. It **We** cannot be excluded that studying separate disease populations might improve the discriminant validity.

In accordance **line** with previous studies, **we confirm that balance scales can discriminate between fallers and non-fallers, in contrast to single balance tests which have a** we found a low level of accuracy in identifying fallers using only a single balance test [11,60,61]. Due to the heterogeneity of patients in our study, it is unlikely that a single test could accurately identify patients at risk of falls. This is in line with recommendations of [62] who suggested to use two or more clinical tests for an accurate evaluation of the various components of risk of falls in patients with Parkinson's disease. The limited association between single measures of balance and fall history could be partly explained by**due to** the multifactorial nature of causes of fall.

Study limitations

Our study was based on a heterogeneous population **sample** of patients affected by different balance disorders **that are representative of the spectrum of neurological diseases routinely**

observed in departments of rehabilitation, but may not be representative of other clinical settings. Although the Brief-BESTest assesses balance performance independently of the pathology that causes the problem, future studies should assess psychometric characteristics such as responsiveness, the minimal clinically important difference and sensitivity to change of the Brief-BESTest in specific groups of disease.

The history of falls should be considered useful only for discriminative purposes (i.e. to distinguish fallers from non-fallers) and not to predict patients who will fall in the future. In fact, falls change people's behavior and the cutoff scores for prospective prediction of falls may be very different from those reported in our study.

Implications for clinical practice and conclusions

The comparison of the three structural models of Brief-BESTest proposed in the literature clearly highlights that Model 3, i.e. the shortest version, has psychometric characteristics equal or superior to the other two. This makes Model 3 the best of the three versions to recommend for use in clinical practice, given that it is also the fastest to perform. For this reason, in the supplementary material we provide a simple calculator for the three models tested in this study. Finally, this study highlights once again the superiority of a clinical scale composed of several items compared to single-item measures.

Authors' contributions

MGo and AN contributed to the concept/idea/research design. MGo, MGi and IA contributed to the writing and data analysis. MGi and IA provided data collection. MGo and SC provided project management and study participants. AN and SC provided facilities/equipment and institutional liaisons. SC and RC contributed consultation in different phases of the study (including review of manuscript before submission).

Ethical Approval

The Scientific Technical Committee of the Istituti Clinici Scientifici Maugeri approved this study with the following identification number: rrf41.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

None declared.

Bibliography

[1] Geiger RA, Allen JB, O'Keefe J, Hicks RR. Balance and mobility following stroke: effects of physical therapy interventions with and without biofeedback/forceplate training. Phys Ther 2001;81(4):995-1005. PubMed PMID: 11276182.

[2] Hyndman D, Ashburn A. People with stroke living in the community: Attention deficits, balance, ADL ability and falls. Disabil Rehabil 2003;25(15):817-22. PubMed PMID: 12851091.

434 [3] Horlings CG, Van Engelen BG, Allum JH, Bloem BR. A weak balance: the contribution of muscle weakness to postural instability and falls. Nat Rev Neurol. 2008;4(9):504. doi: 10.1038/ncpneuro0886.

437 [4] Sturnieks DL, St George R, Lord SR. Balance disorders in the elderly. Neurophysiol Clin 2008;38(6):467-478. doi: 10.1016/j.neucli.2008.09.001.

439 [5] Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med 2010;46:239. PubMed PMID: 20485226.

441 [6] 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. doi: 10.2522/ptj.20080131.

444 [7] 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. doi: 10.2522/ptj.20110063.

447 [8] Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther 2009;89:484-498. doi: 10.2522/ptj.20080071.

449 [9] Jacobs JV, Horak FB, Tran VK, Nutt JG. Multiple balance tests improve the assessment of postural stability in subjects with Parkinson's disease. J Neurol Neurosurg Psychiatry 2006;1;77(3):322-6. PubMed PMID: 16484639.

452 [10] Jacobs JV, Earhart GM, McNeely ME. Can postural instability tests improve the prediction of future falls in people with Parkinson's disease beyond knowing existing fall history? J Neurol 2016;263(1):133-9. doi: 10.1007/s00415-015-7950-x.

[11] 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;93:1102-15. doi: 10.2522/ptj.20120454.

[12] 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. doi: 10.2340/16501977-0537.

[13] Padgett PK, Jacobs JV and 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. doi: 10.2522/ptj.20120056.

[14] 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. PubMed PMID: 27103223.

[15] Jácome 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(11):1807-1815. doi: 10.2522/ptj.20150391.

[16] 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. doi: 10.1097/NPT.0b013e31821a620c.

[17] Duncan RP, Leddy AL, Cavanaugh JT, Dibble LE, Ellis TD, Fors MP, Foreman KB, Earhart GM. 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. doi: 10.2522/ptj.20120302.

[18] O'Hoski S, Winship B, Herridge L, Agha T, Brooks D, Beauchamp MK, Sibley KM. Increasing the clinical utility of the BESTest, Mini-BESTest, and Brief-BESTest: normative values in Canadian adults who are healthy and aged 50 years or older. Phys Ther 2014;94:334-342. doi: 10.2522/ptj.20130104.

[19] Franchignoni F, Giordano A. On "Is the BESTest at its best?...." Padgett PK, Jacobs JV, Kasser SL. Phys Ther 2012;92:1197-1207. Phys Ther 2012;92(9):1236-7. doi: 10.2522/ptj.2012.92.9.1236.

[20] Marques A, Almeida S, Carvalho J, Cruz J, Oliveira A, Jácome C. Reliability, Validity, and Ability to Identify Fall Status of the Balance Evaluation Systems Test, Mini–Balance Evaluation Systems Test, and Brief–Balance Evaluation Systems Test in Older People Living in the Community. Arch Phys Med Rehabil 2016;97:2166-2173. doi: 10.1016/j.apmr.2016.07.011.

[21] McArdle JJ. Current directions in structural factor analysis. Curr Dir Psychol Sci 1996;5:11-18.

[22] Norman GR, Streiner DL. PMPH-USA. PDQ statistics; 1st edition, 2003.

[23] Jansen AC, van Aalst-Cohen ES, Hutten BA, Büller HR, Kastelein JJ, Prins MH. Guidelines were developed for data collection from medical records for use in retrospective analyses. J Clin Epidemiol 2005;58(3):269-274. doi: 10.1016/j.jclinepi.2004.07.006.

[24] Jernigan SD, Pohl PS, Mahnken JD, Kluding PM. Diagnostic accuracy of fall risk assessment tools in people with diabetic peripheral neuropathy. Phys Ther 2012;92:1461-1470. doi: 10.2522/ptj.20120070.

[25] Lin MR, Hwang HF, Hu MH, Wu HDI, Wang YW, Huang FC. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. J Am Geriatr Soc 2004;52:1343-1348. doi: 10.1111/j.1532-5415.2004.52366.x.

[26] Vellas BJ, Wayne SJ, Romero L, Baumgartner RN, Rubenstein LZ, Garry PJ. One-leg balance is an important predictor of injurious falls in older persons. J Am Geriatr Soc 1997;45:735–738. PubMed PMID: 9180669.

[27] Sebastião E, Sandroff BM, Learmonth YC, Motl RW. Validity of the timed up and go test as a measure of functional mobility in persons with multiple sclerosis. Arch Phys Med Rehabil 2016;97:1072-1077. doi: 10.1016/j.apmr.2015.12.031.

[28] Podsiadlo D, Richardson S. The timed 'Up & Go': A test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991;39:142–148. PubMed PMID: 1991946.

[29] Franchignoni F, Tesio L, Martino M, Ricupero C. Reliability of four simple, quantitative tests of balance and mobility in healthy elderly females. Aging Clin Exp Res 1998;10:26-31. PubMed PMID: 9589748.

[30] Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. J Gerontol 1990;45(6):M192-7. PubMed PMID: 2229941.

[31] Powell LE, Myers AM. The activities-specific balance confidence (ABC) scale. J Gerontol A Biol Sci Med Sci 1995;50(1):M28-M34. PubMed PMID: 7814786.

[32] Franchignoni F, Giordano A, Ronconi G, Rabini A, Ferriero G. Rasch validation of the Activities-specific Balance Confidence Scale and its short versions in patients with Parkinson's disease. J Rehabil Med 2014;46:532-539. doi: 10.2340/16501977-1808.

519 [33] Atkinson TM, Rosenfeld BD, Sit L, Mendoza TR, Fruscione M, Lavene D, Shaw M, Li Y, Hay J,
520 Cleeland CS, Scher HI, Breitbart WS, Basch E. Using confirmatory factor analysis to evaluate
521 construct validity of the Brief Pain Inventory (BPI). J Pain Symptom Manage 2011;41(3):558-
522 565. doi: 10.1016/j.jpainsymman.2010.05.008.

523 [34] Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107(2):238-46.
524 PubMed PMID: 2320703.

525 [35] Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS,
526 editors. Testing Structural Equation Models. Newbury Park: Sage 1993;136–62.

527 [36] Hu LT, Bentler P. Cutoff criteria for fit indices in covariance structure analysis: conventional
528 criteria versus new alternatives. Struct Equ Model Multidiscip J 1999;6:1-55. doi:
529 10.1080/10705519909540118.

530 [37] DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: Considerations for
531 the applied researcher. PARE 2009;14(20):1-11.

532 [38] Schönberger M, Ponsford J, Olver J, Ponsford, M. A longitudinal study of family functioning
533 after TBI and relatives' emotional status. Neuropsychol Rehabil 2010;20:813-829. doi:
534 10.1080/09602011003620077.

535 [39] Hancock GR, Mueller RO. Rethinking construct reliability within latent variable systems.
536 Structural equation modeling: Present and future-A Festschrift in honor of Karl Jöreskog.
537 Lincolnwood, IL: Scientific Software International 2001;195-216.

538 [40] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika
539 1951;16(3):297–334. PubMed PMID: 8449658.

540 [41] Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive
541 problem of internal consistency estimation. Br J Psychol 2014;105:399-412. doi:
542 10.1111/bjop.12046.

543 [42] Opheim A, Jahnsen R, Olsson E, Stanghelle JK. Balance in relation to walking deterioration in
544 adults with spastic bilateral cerebral palsy. Phys Ther 2012;92(2):279-88. doi:
545 10.2522/ptj.20100432.

546 [43] Curtze C, Nutt JG, Carlson-Kuhta P, Mancini M, Horak FB. Objective gait and balance
547 impairments relate to balance confidence and perceived mobility in people with Parkinson
548 disease. Phys Ther 2016;96(11):1734-1743. PubMed PMID: 27149959.

549 [44] Rand D. Mobility, balance and balance confidence - correlations with daily living of
550 individuals with and without mild proprioception deficits post-stroke. NeuroRehabilitation
551 2018;43(2):219-226. doi: 10.3233/NRE-172398.

552 [45] Salbach NM, Mayo NE, Hanley JA, Richards CL, Wood-Dauphinee S. Psychometric evaluation
553 of the original and Canadian French version of the activities-specific balance confidence scale
554 among people with stroke. Arch Phys Med Rehabil 2006;87(12):1597-604. PubMed PMID:
555 17141639.

556 [46] Cohen, J. Statistical power analysis for the behavioral sciences. 2nd edition, 1988.

557 [47] Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. Acta
558 Paediatr 2007;96:664-647. doi: 10.1111/j.1651-2227.2006.00178.x.

559 [48] Hodgson JA, Roy RR, de Leon R, Dobkin B, Edgerton VR. Can the mammalian lumbar spinal
560 cord learn a motor task? Med Sci Sports Exerc 1994;26:1491-1497. PubMed PMID: 7869884.

561 [49] Grasso R, Zago M, Lacquaniti F. Interactions between posture and locomotion: motor patterns in humans walking with bent posture versus erect posture. J Neurophysiol 2000;83:288-300. doi: 10.1152/jn.2000.83.1.288.

564 [50] Massion J. Movement, posture and equilibrium: interaction and coordination. Prog Neurobiol 1992;38(1):35-56. Review. PubMed PMID: 1736324.

566 [51] Giardini M, Nardone A, Godi M, Guglielmetti S, Arcolin I, Pisano F, Schieppati M. Instrumental or physical-exercise rehabilitation of balance improves both balance and gait in Parkinson's disease. Neural Plast 2018;2018:5614242. doi: 10.1155/2018/5614242.

569 [52] Nardone A, Godi M, Grasso M, Guglielmetti S, Schieppati M. Stabilometry is a predictor of gait performance in chronic hemiparetic stroke patients. Gait Posture 2009;30(1):5-10. doi: 10.1016/j.gaitpost.2009.02.006.

572 [53] Morlino P, Balbi B, Guglielmetti S, Giardini M, Grasso M, Giordano C, Schieppati M, Nardone A. Gait abnormalities of COPD are not directly related to respiratory function. Gait Posture 2017;58:352-357. doi: 10.1016/j.gaitpost.2017.08.020.

575 [54] Bloem BR, Marinus J, Almeida Q, Dibble L, Nieuwboer A, Post B, Ruzicka E, Goetz C, Stebbins G, Martinez-Martin P, Schrag A; Movement Disorders Society Rating Scales Committee. Measurement instruments to assess posture, gait, and balance in Parkinson's disease: Critique and recommendations. Mov Disord 2016;31(9):1342-55. doi: 10.1002/mds.26572.

579 [55] Chan AC, Pang MY. Assessing balance function in patients with total knee arthroplasty. Phys Ther 2015;95(10):1397-407. doi: 10.2522/ptj.20140486.

[56] Huang M, Pang MY. Psychometric properties of Brief-Balance Evaluation Systems Test (Brief-BESTest) in evaluating balance performance in individuals with chronic stroke. Brain Behav 2017;7(3):e00649. doi: 10.1002/brb3.649.

[57] Revelle W, Zinbarg RE. Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. Psychometrika 2009;74(1):145. doi:10.1007/s11336-008-9102-z.

[58] Huang MH, Miller K, Smith K, Fredrickson K, Shilling T. Reliability, validity, and minimal detectable change of Balance Evaluation Systems Test and its short versions in older cancer survivors: a pilot study. J Geriatr Phys Ther 2016;39(2):58-63. doi: 10.1519/JPT.0000000000000047.

[59] Mak MK, Pang MY. Balance confidence and functional mobility are independently associated with falls in people with Parkinson's disease. J Neurol 2009;256:742-749. doi: 10.1007/s00415-009-5007-8.

[60] Cameron M, Mazumder R, Murchison C, King L. Mini Balance Evaluation Systems Test in people with multiple sclerosis: reflects imbalance but may not predict falls. Gait Posture 2014;39(1):669. doi: 10.1016/j.gaitpost.2013.08.009.

[61] Mak MK, Auyeung MM. The mini-BESTest can predict parkinsonian recurrent fallers: a 6-month prospective study. J Rehabil Med 2013;45:565-571. doi: 10.2340/16501977-1144.

[62] Dibble LE, Christensen J, Ballard DJ, Foreman KB. Diagnosis of fall risk in Parkinson disease: an analysis of individual and collective clinical balance test interpretation. Phys Ther 2008;88:323-332. doi: 10.2522/ptj.20070082.

Table 1

Demographic data and scores on balance measures for the entire sample, and fallers vs. non-fallers.

	Entire **Total** Sample (N=416)	Fallers (N=135)	Non-fallers (N=160)	*p*
	N %	N %	N %	
Sex (M/F)	55/45	51/49	58/42	.23[a]
Walking Aid (no/yes)	54/46	47/53	68/32	<.001[a]
	Mean (SD)	Mean (SD)	Mean (SD)	
Age (y)	66.5 (14.2)	66.3 (13.4)	64.4 (14.8)	.28[b]
OLS mean (s)	4.2 (6.3)	3.3 (5.0)	6.9 (7.5)	<.001[b]
TUG test (s)	17.6 (15.3)	20.6 (20.3)	13.3 (9.5)	<.001[b]
Functional Reach (cm)	18.6 (9.7)	16.5 (9.6)	20.5 (9.8)	<.01[b]
Total score BBT	8.9 (6.4)	7.3 (5.1)	11.8 (6.9)	<.001[b]
Total score ABC 5-levels	29.6 (16.8)	26.0 (15.0)	36.2 (17.1)	<.001[b]

The total sample comprises all patients, i.e. those in whom a fall history was collected (the two subgroups fallers/non-fallers) as well as those in whom a fall history was not collected due to acute stroke event (n=79) or failure to complete the questionnaire (n=42), as this is a retrospective study. Participants were classified as fallers if they reported 1 **2** or more falls in the last 6 months.

SD, Standard Deviation; M, Male; F, Female; OLS, One Leg Stance; TUG test, Timed Up and Go test; BBT, Brief-BESTest; ABC 5-levels, Activities-specific Balance Confidence scale 5-levels of rating.

p-value was computed between fallers and non-fallers.

[a] = χ^2 test.

[b] = Mann-Whitney U-test.

Table 2

Details on each single item and total score of the original Brief-BESTest (Model 1) in the whole sample (N = 416).

	Median	25% percentile	75% percentile	**Skewness**	Kurtosis	Min score (%)	Max score (%)	Missing data (%)
Item 1	1	0	2	0.39	-1.20	42.51	7.73	0.40
Item 2	2	1	2	-0.47	-0.13	9.42	10.39	0.40
Item 3	1	0	1	0.93	-0.15	46.62	8.70	0.00
Item 4	1	0	1	0.89	-0.25	44.69	9.66	0.00
Item 5	1	0	2	0.59	-1.19	48.07	17.15	0.00
Item 6	1	0	2	0.53	-1.26	45.89	18.84	0.00
Item 7	1	0	2	0.66	-0.99	43.96	16.91	0.00
Item 8	2	0	2	0.07	-1.54	36.47	23.43	0.00
Total score	8	4	13	0.54	-0.69	6.00	8.00	

Item 1 - Hip/Trunk Lateral Strength; item 2 - Functional Reach Forward; item 3 and 4 - One Leg Stance, Left and Right; item 5 and 6 - Compensatory Stepping- Lateral, Left and Right; item 7 - stand on a foam surface; 8 - Timed Up and Go test.

Table 3

Summary of fit statistics of the specified models (N = 416)

Models	χ^2	df	RMSEA (90% C.I.)	CFI	TLI	SRMR
Model 1 *(Padgett et al., 2012)*	134.0	20	0.12 (0.10-0.14)	0.78	0.70	0.09
Model 2 *(adapted from Franchignoni & Giordano, 2012)*	60.3	18	0.08 (0.06-0.10)	0.92	0.88	0.05
Model 3 *(adapted from Bravini et al., 2016)*	26.2	12	0.05 (0.03-0.08)	0.97	0.95	0.03

Comparison of factor models	Significant difference between models[a]
Model 1 vs. Model 2	p<0.001
Model 2 vs. Model 3	p<0.001
Model 3 vs. Model 1	p<0.001

df, degrees of freedom; RMSEA, Root Mean-Square Error of Approximation; C.I., Confidence Interval; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; SRMSR, Standardized Root Mean Square Residual.

[a] Calculated as a $\chi^2 = \chi^2_2 – \chi^2_1$ with df = df_2–df_1.

Table 4

Ability of the Brief-BESTest Models 1, 2 and 3 to discriminate fallers from non-fallers compared to other measures of balance (OLS, TUG test, Functional Reach and ABC 5-levels) **(N = 295).**

Classification	Brief-BESTest Model 1	Brief-BESTest Model 2	Brief-BESTest Model 3	OLS	TUG test	FR
Cut-off score	8	8	7	2	12	19
AUC (95% C.I.)	0.72 (0.66-0.78)	0.72 (0.66-0.78)	0.72 (0.66-0.78)	0.63 (0.55-0.70)	0.62 (0.56-0.68)	0.61 (0.52-0.69)
Sensitivity	74%	72%	74%	63%	68%	67%
Specificity	54%	58%	58%	55%	53%	49%
LR+	1.61	1.72	1.76	1.42	1.45	1.33
LR-	0.48	0.47	0.45	0.66	0.60	0.67
Correctly Classified	65%	66%	67%	59%	60%	58%

OLS, One Leg Stance; TUG test, Timed Up and Go test; FR, Functional Reach; AUC, Area Under the Curve; C.I., Confidence Interval; LR+, positive Likelihood Ratio; LR-, negative Likelihood Ratio.

Appendix 1

Items and factors of the three models of Brief-BESTest.

<table>
<tr><th>Model item
(Scoring 0-3 per item)</th><th>Model 1</th><th>Model 2</th><th>Model 3</th></tr>
<tr><td>Item 1: Hip/Trunk Lateral Strength</td><td rowspan="8">Dynamic Balance</td><td rowspan="2">Static Balance</td><td></td></tr>
<tr><td>Item 2: Functional Reach Forward</td><td rowspan="7">Dynamic Balance</td></tr>
<tr><td>Item 3: Stand on One Leg-Left</td><td rowspan="6">Dynamic Balance</td></tr>
<tr><td>Item 4: Stand on One Leg-Right</td></tr>
<tr><td>Item 5: Compensatory Stepping–Lateral, Left</td></tr>
<tr><td>Item 6: Compensatory Stepping–Lateral, Right</td></tr>
<tr><td>Item 7: Stand on foam surface with Eyes Closed</td></tr>
<tr><td>Item 8: Timed Up and Go test</td></tr>
<tr><td>Total score</td><td>0-24</td><td>0-24
(Static 0-5;
Dynamic 0-19)</td><td>0-24</td></tr>
</table>

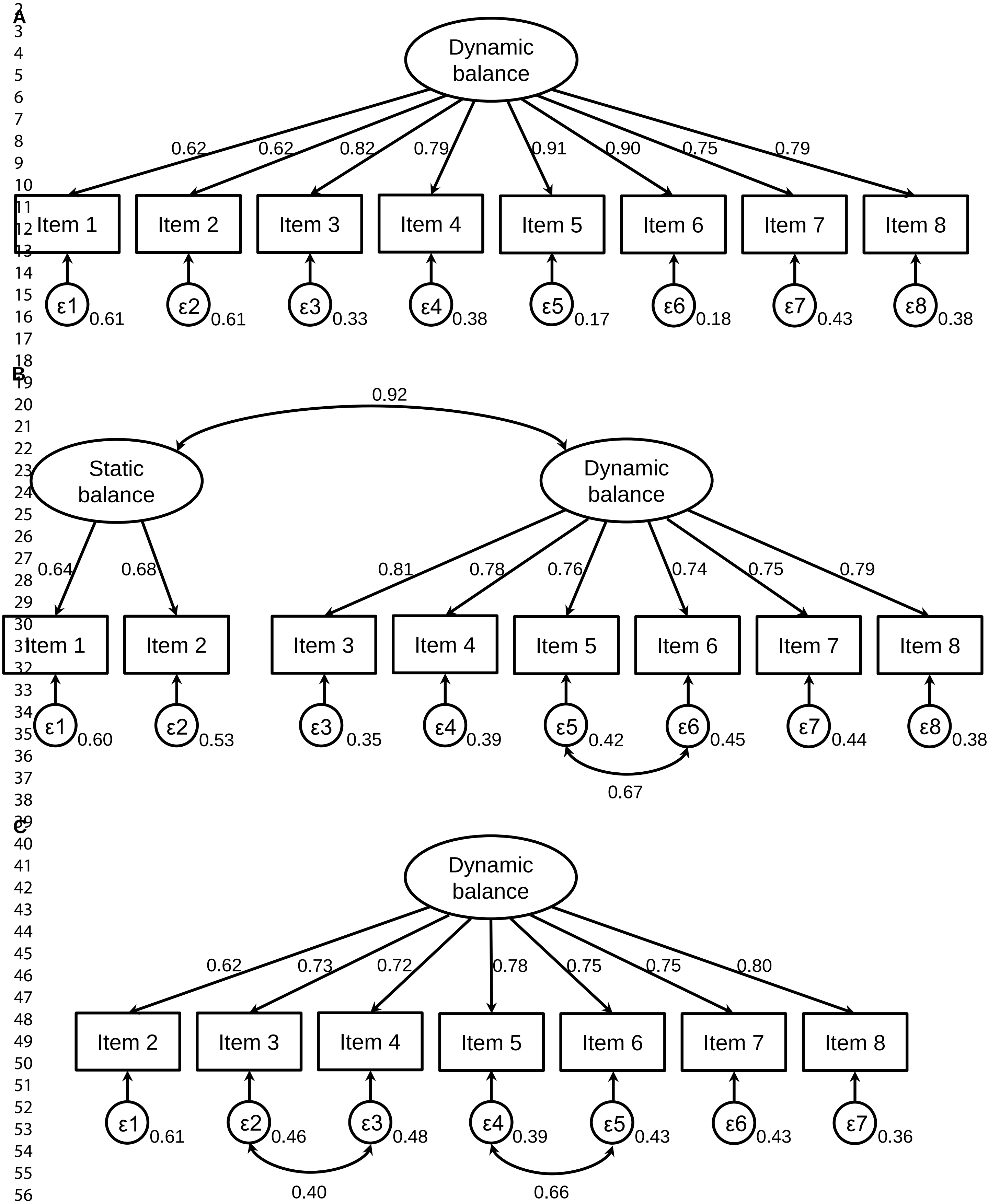
Items
Item 1: Hip/Trunk Lateral Strength
Item 2: Functional Reach Forward
Item 3: Stand on One Leg-Left
Item 4: Stand on One Leg-Right
Item 5: Compensatory Stepping–Lateral, Left
Item 6: Compensatory Stepping–Lateral, Right
Item 7: Stand on foam surface with Eyes Closed
Item 8: Timed Up and Go test

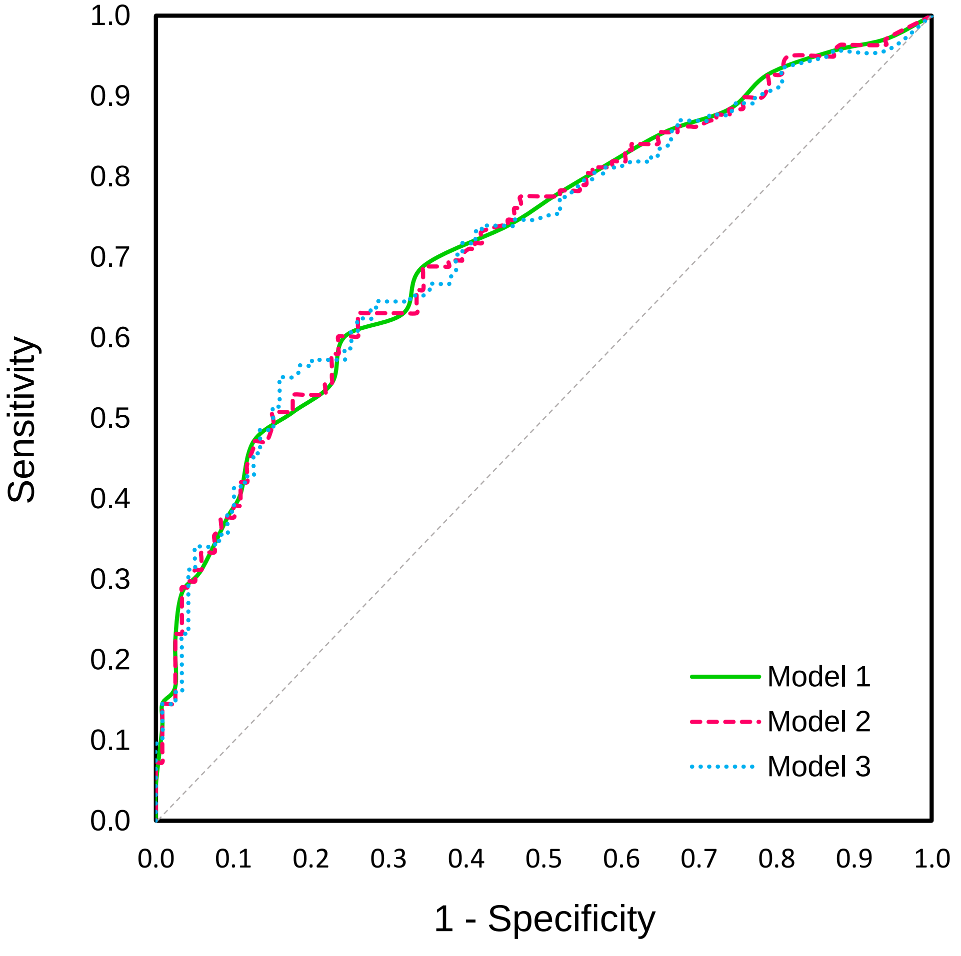
Total score (range 0-24)
Static Balance (range 0-5)
Dynamic Balance (range 0-19)

* to obtain the total Score of Model 3 is not necessary the score of item 1

Insert the score of each items in the cells below
3
0
0
2
3
1
3
2

Model 1	Model 2	Model 3*
	Static Balance	
Dynamic Balance	*Dynamic Balance*	*Dynamic Balance*
14.00	**13.95**	**13.01**
	2.46 11.52	


A
Dynamic balance
0.62
0.62
0.82
0.79
0.91
0.90
0.75
0.79
Item 1
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7
Item 8
ε1 0.61
ε2 0.61
ε3 0.33
ε4 0.38
ε5 0.17
ε6 0.18
ε7 0.43
ε8 0.38
B
0.92
Static balance
Dynamic balance
0.64
0.68
0.81
0.78
0.76
0.74
0.75
0.79
Item 1
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7
Item 8
ε1 0.60
ε2 0.53
ε3 0.35
ε4 0.39
ε5 0.42
ε6 0.45
ε7 0.44
ε8 0.38
0.67
C
Dynamic balance
0.62
0.73
0.72
0.78
0.75
0.75
0.80
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7
Item 8
ε1 0.61
ε2 0.46
ε3 0.48
ε4 0.39
ε5 0.43
ε6 0.43
ε7 0.36
0.40
0.66


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Sensitivity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1 - Specificity
Model 1
Model 2
Model 3