

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

Journal:	<i>Physical Therapy</i>
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

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

For Peer Review Only

1  
2  
3 9 **ABSTRACT**  
4  
5  
6 10  
7

8 11 **Background:** The Brief Balance Evaluation Systems Test (Brief-BESTest) could be a useful tool for  
9  
10 12 balance assessment. Although some psychometric characteristics have been examined, others  
11 13 still need to be clarified.  
12 14

15 14 **Objectives:** To assess the structural validity, convergent validity, discriminant validity and  
16 15 internal consistency of the Brief-BESTest in neurological patients.  
17 16

18 16 **Design:** Cross-sectional.  
19 17

20 17 **Methods:** Data were from 416 patients with neurological disease and related balance disorders.  
21 18

22 18 Patients were assessed with the 5-levels Activities-Specific Balance Confidence Scale (ABC 5-  
23 19

24 19 levels), Brief-BESTest, **some simple balance tests, i.e. One-Leg Stance (OLS), Timed Up and Go**  
25 20

26 20 **(TUG) test, Functional Reach (FR), simple balance tests** and a fall history questionnaire. Three  
27 21

28 21 **models of Brief-BESTest models** were examined through confirmatory factor analysis (CFA) and  
29 22

30 22 the following indices calculated: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean  
31 23

32 23 Square Error of Approximation (RMSEA). Convergent validity was assessed by calculating the  
33 24

34 24 correlation between Brief-BESTest and ABC 5-levels total scores. Receiver operating  
35 25

36 25 characteristics (ROC) assessed the ability of each model to differentiate between people with vs.  
37 26

38 26 without falls. Internal consistency was measured by Cronbach's alpha and coefficient omega.  
39 27

40 27 **Results:** CFA showed Model 3 (CFI=0.97, TLI=0.95, RMSEA=0.05), **with item 1 removed and error**  
41 28

42 28 **covariance between items 3-4 and between items 5-6**, to have a significantly better structure  
43 29

44 29 than Models 1 and 2 ( $p < 0.001$ ). The correlation between Brief-BESTest and ABC 5-levels was 0.61  
45 30

46 30 (Spearman's rho) for all three models. Area Under the Curve (AUC) of ROC **was showed an**  
47 31

48 31 **acceptable accuracy (0.72) in distinguishing patients with vs. without history of falls** (95%  
49 32

50 32 C.I.=0.66–0.78) for all models, and superior to AUCs of other simple balance tests (**OLS, TUG test,**  
51 33  
52 34  
53 35  
54 36  
55 37  
56 38  
57 39  
58 40  
59 41  
60 42

1  
2  
3 33 **FR).** Cronbach's alpha was good for Brief-BESTest Models 1 (0.92) and 3 (0.92), but omega was  
4  
5  
6 34 >0.80 only for Model 3.

7  
8 35 **Limitations:** Heterogeneous sample size was a heterogeneous population.

9  
10 36 **Conclusions:** The Brief-BESTest, after some changes, shows good validity and internal  
11  
12  
13 37 consistency in patients affected by different balance disorders, after applying some changes.  
14

15 38

16  
17  
18 39 **Contribution of the Paper:**

- 19  
20 40 • Although some psychometric characteristics of the Brief-BESTest have been examined in  
21  
22  
23 41 previous studies, other properties such as validity still need to be clarified.  
24  
25 42 • This study shows that the Brief-BESTest has good validity and internal consistency in  
26  
27  
28 43 patients affected by different balance disorders, after applying some changes: removal of item 1  
29  
30 44 and using an appropriate weighting method for the calculation of the total score.  
31  
32 45 • This study confirms the ability of the Brief-BESTest to distinguish between people with vs.  
33  
34  
35 46 without history of falls, in contrast to other simple balance tests. Moreover, it highlights once  
36  
37 47 again the superiority of a clinical scale composed of several items compared to single-item  
38  
39  
40 48 measures such as the TUG test and OLS.

41  
42 49

43  
44  
45 50

46  
47 51 **Keywords:** Brief-Balance Evaluation Systems Test, balance assessment, confirmatory factor  
48  
49 52 analysis, structural validity, internal consistency.  
50

51  
52 53

53  
54  
55 54

56  
57  
58  
59  
60

## 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].

1  
2  
3 78 However, the latter hypothesis was rejected through exploratory factor analysis **dismantled the**  
4  
5 79 **latter hypothesis**, demonstrating **that the Brief-BESTest actually has at most two subsections,**  
6  
7  
8 80 **or dimensions** [19]. Furthermore, Bravini et al. [14] showed by Rasch analysis that all items of  
9  
10 81 the Brief-BESTest except for item 1 account for the same underlying theoretical construct and  
11  
12  
13 82 indicated that the Brief-BESTest should in fact be considered as unidimensional. Therefore, the  
14  
15 83 authors suggested the adoption of a 7-item version of the test.

16  
17  
18 84 Although some psychometric characteristics of the Brief-BESTest, such as the internal  
19  
20  
21 85 consistency [13,14], reliability [14,17,20] and sensitivity to change [14,20], have been  
22  
23  
24 86 investigated in previous studies, other properties still need to be clarified. In particular, the Brief-  
25  
26 87 BESTest structure has not yet **been investigated with undergone** confirmatory factor analysis  
27  
28 88 (CFA). This statistical tool provides information on possible independent factors and can be very  
29  
30  
31 89 useful for developing shortened forms of an **evaluating** instrument [21,22]. Finally, the Brief-  
32  
33  
34 90 BESTest seems to have good sensitivity and accuracy in identifying retrospectively people who  
35  
36 91 have had at least one fall [13,17]. However, these findings are based only on small samples of  
37  
38 92 patients with multiple sclerosis [13] or Parkinson's disease [17].

39  
40  
41 93 **The We aimed in of** the present study **was** to fill the existing knowledge gap by examining the  
42  
43  
44 94 structural validity, convergent validity and discriminant validity of the Brief-BESTest in a large  
45  
46  
47 95 group of patients with a variety of balance disorders. In particular, we hypothesized that:

48  
49  
50 96 1) among the three models of the Brief-BESTest presented in the literature [13,14,19], the 7-item  
51  
52  
53 97 version would be the one with the best structural validity;  
54  
55  
56 98 2) in spite of its conciseness, the 7-item model [14] would have the same ability as the other two  
57  
58 99 Brief-BESTest models to **predict patients at risk of falls; discriminate between people with vs.**  
59  
60  
100 **without a history of falls;**

1  
2  
3 101 3) in discriminating between people with vs. without a history of falls, the Brief-BESTest would be  
4  
5  
6 102 superior to other simple balance tests such as One Leg Stance, Timed Up and Go test and  
7  
8 103 Functional Reach.  
9

## 10 11 104 12 13 14 15 105 **METHODS**

### 16 17 106 18 19 20 21 107 **Participants**

22  
23  
24 108 This was an observational retrospective study conducted in a group of 416 patients affected by  
25  
26  
27 109 different neurological diseases: 186 females and 230 males; mean age  $66.5 \pm 16.0$  years (**mean  $\pm$**   
28  
29 110 standard deviation) consecutively admitted for in-patient rehabilitation at the XXXXXXX between  
30  
31  
32 111 February 2014 and April 2017. **Patients' clinical and treatment data were extracted from the**  
33  
34 112 **electronic medical record system and transferred to a specific database (Microsoft Excel).**  
35  
36 113 Patients were stratified into different groups according to their diagnosis: 118 **with** Parkinson's  
37  
38  
39 114 disease, 79 **with** acute stroke, 43 **with** sensorimotor polyneuropathy, 32 **with** cerebellar ataxia,  
40  
41 115 32 **with** diffuse encephalopathy, 31 **with** chronic stroke, 21 **with** multiple sclerosis, 19 **with**  
42  
43  
44 116 traumatic brain injury, 16 **with** vestibular disorder, 13 **with** neuromuscular disorders, 12 **with**  
45  
46 117 central nervous system neoplasm. Inclusion criteria were: a) ability to maintain an upright  
47  
48  
49 118 position without support for at least 5 seconds; b) ability to understand the required motor  
50  
51 119 tasks; c) no hip or knee replacement surgery within the previous 6 months. Exclusion criteria  
52  
53  
54 120 were: a) musculoskeletal injury limiting the ability to walk; b) any other serious cardio-respiratory  
55  
56 121 problem. The study was carried out in conformity with the Declaration of Helsinki of the World  
57  
58  
59 122 Medical Association and the guidelines for retrospective studies [23]. The local scientific and  
60  
123 ethics committee approved the study.

1  
2  
3 124  
4  
5  
6  
7 125 **Assessment tools**

8  
9  
10 126 Patients' demographic and clinical characteristics were gathered by a team of trained physical  
11  
12 127 therapists engaged in clinical practice collected patients' demographic and clinical  
13  
14  
15 128 characteristics. During the routine clinical assessment following admission to the rehabilitation  
16  
17 129 department, patients underwent the following assessments:

18  
19  
20  
21 130 **- Fall history**

22  
23  
24 131 A history of falls over the past 6 months was obtained from patients at admission through  
25  
26 132 patient interview. A fall was defined as an unintentional event in which any part of the body  
27  
28  
29 133 came into contact with the ground [10]. Patients who reported two or more falls in the defined  
30  
31  
32 134 period were classified as 'fallers' [24]. A fall history was not recorded taken in the case of  
33  
34 135 patients with acute stroke at the time of admission.

35  
36  
37 136 **- Brief-BESTest**

38  
39  
40  
41 137 The Brief-BESTest is an 8-item scale with each item scored on a 4-level rating scale from 0 (severe  
42  
43 138 balance impairment) to 3 (no balance impairment). Its items cover the six subsections of the  
44  
45  
46 139 original BESTest (biomechanical constraints, stability limits/verticality, anticipatory postural  
47  
48 140 responses, postural responses, sensory orientation, stability in gait); the maximum total score is  
49  
50  
51 141 24 [13]. The Brief-BESTest requires less time and equipment to administer than the BESTest  
52  
53 142 and the Mini-BESTest; thus, the Brief-BESTest seems to be more feasible for clinical use [17].

54  
55  
56 143 **- Simple balance tests**

57  
58  
59  
60



1  
2  
3 144 During administration of the Brief-BESTest, we recorded the time required by patients to  
4  
5  
6 145 complete item 3 (left One Leg Stance), 4 (right One Leg Stance) and 8 (Timed Up and Go test) and  
7  
8 146 the distance covered by patients during item 2 (Functional Reach). This allowed us to obtain the  
9  
10  
11 147 scores of three additional simple balance tests: One Leg Stance (OLS) [25,26], Timed Up and Go  
12  
13 148 (TUG) test [27,28] and Functional Reach (FR) [29,30].

### 16 149 - *Activities-specific Balance Confidence Scale*

18  
19 150 The short version of the Activities-specific Balance Confidence Scale (ABC) is a self-reported 16-  
20  
21 151 item questionnaire that scores the perceived level of balance confidence when performing  
22  
23  
24 152 common activities of daily living [31]. We used the 5-levels **rating** version of rating (ABC 5-levels)  
25  
26 153 [32] in which each item is scored from 0 (no confidence) to 4 (fully confident), giving a total score  
27  
28  
29 154 range 0-64.

### 35 156 **Data analysis**

37  
38  
39 157 Descriptive statistics were used to describe mean demographic and balance performance  
40  
41 158 characteristics of the entire sample and of the two **smaller** subgroups, classified as fallers and  
42  
43  
44 159 non-fallers. **The analysis of discriminant validity was conducted only in these two subgroups,**  
45  
46 160 **i.e. those patients who had a history of falls available.** These values were also determined  
47  
48  
49 161 **separately both for fallers and non-fallers.** For each item of the Brief-BESTest we calculated:  
50  
51 162 median value, spread (25<sup>th</sup>–75<sup>th</sup> percentiles), skewness and kurtosis. Floor and ceiling effects  
52  
53  
54 163 were analyzed by calculating the percentage of individuals obtaining the lowest and the highest  
55  
56 164 score for each scale item. In order to detect differences in clinical characteristics between fallers  
57  
58  
59 165 and non-fallers, the Chi-square ( $\chi^2$ ) test was used for two parameters, sex and use of walking  
60  
166 aids, while the Mann-Whitney U-test was used for age, OLS, TUG test, FR, total score of Brief-

1  
2  
3 167 BESTest and total score of ABC 5-levels. Significance was set at  $p < 0.05$ . All analyses were  
4  
5  
6 168 performed using STATA 13.0 software (StataCorp LLC, College Station, Texas, USA).  
7  
8

9 169

10

11

12 170

13

14

15

16 171

17

18 172

19

20

21 173

22

23 174

24

25

26

27 175

28

29 176

30

31

32 177

33

34 178

35

36 179

37

38

39 180

40

41 181

42

43

44 182

45

46 183

47

48

49 184

50

51

52 185

53

54 186

55

56

57 187

58

59 188

60

### Structural validity

Structural validity is usually assessed through CFA [33]. CFA assesses the degree to which responses on a  $p \times 1$  vector of observable random variables (the Brief-BESTest items) can be used to assign a value to one or more unobserved variable(s) (latent 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.

1  
2  
3 189 **Preliminary analysis conducted on Model 3 showed similar CFA results for patients who used a**  
4  
5  
6 190 **walking aid and those who did not. For this reason, we decided to consider the entire sample**  
7  
8 191 **independently of the walking condition in creating our models.**

9  
10  
11 192 **We examined t**he structural validity of these three models (from here on Model 1, 2 and 3) **was**  
12 **examined** through CFA using Structural Equation Modeling (SEM). In view of the very low  
13  
14 193 occurrence of missing data (maximum 0.4%), cases with missing data were removed from the  
15  
16 194 analysis.  $\chi^2$  was used to identify whether the model fitted the data well. In addition, we assessed  
17  
18 195 the models' goodness of fit using the following indices: Comparative Fit Index (CFI) [34], Tucker-  
19  
20 196 Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) [35] and the Standardized  
21  
22 197 Root Mean Square Residual (SRMR) [34]. The criteria adopted to assess goodness of fit  
23  
24 198 performance were: a) CFI and TLI values  $\geq 0.95$ ; b) RMSEA value  $< 0.06$ ; and c) SRMR value  $\leq 0.08$   
25  
26 199 [36].  
27  
28  
29  
30  
31 200

32  
33  
34 201 The goodness of fit parameters of the three models were compared by computing the  $\chi^2$   
35  
36 202 difference tests of each model pair, calculated as a  $\chi^2 = \chi^2_2 - \chi^2_1$  with  $df = df_2 - df_1$ .

37  
38  
39  
40 203 The standardized factor loadings of the models (i.e. the coefficients of the fitted model) were  
41  
42 204 then transformed into weights that can be applied when scale scores for an individual are  
43  
44 205 calculated. They were calculated with a non-refined method called "Weighted Sum Scores" [37];  
45  
46 206 these weights do not change the scale range [38].  
47  
48  
49

### 51 207 **Internal consistency**

52  
53  
54 208 The internal consistency of the three Brief-BESTest models was measured by Cronbach's alpha  
55  
56 209 and the coefficient omega for congeneric models [39]. Cronbach's alpha measures the extent to  
57  
58 210 which the items consistently measure the same construct, with the value  $\geq 0.80$  indicating good  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14

211 internal consistency [40]. The coefficient omega is the ratio of the true score variance and the  
212 total variance of the scale. Interpretation of coefficient omega is similar to that of Cronbach's  
213 alpha [41].

15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

### 215 **Convergent validity**

216 We used the correlation between the three models of Brief-BESTest and the ABC 5-levels total  
217 score to assess the convergent validity. The ABC scale led to rate rates the balance self-efficacy  
218 of patients [31]. This is associated with measures of balance [32], walking capacity [42],  
219 functional mobility [43], Activities of Daily Living performance [44], and perceived health status  
220 [45]. The choice to use the ABC as a competitor an external criterion was also based on the need  
221 to avoid the comparisons of the three models with a scale that had items similar to the Brief-  
222 BESTest. The correlation was assessed by means of Spearman's rho: coefficients <0.30 were  
223 interpreted as weak, those between 0.30 and 0.49 as moderate, and those  $\geq 0.50$  as strong  
224 correlations [46].

33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

### 226 **Discriminant validity**

227 To assess the ability of the three Brief-BESTest models to distinguish between 'fallers' and 'non-  
228 fallers', receiver operating characteristic (ROC) curves were computed, by plotting sensitivity on  
229 the x-axis against 1 – specificity on the y-axis. In our study, sensitivity was calculated as the  
230 number of patients correctly identified as 'fallers' and specificity as the number of patients  
231 correctly identified as 'non-fallers'. The optimal cut-off value was chosen on the ROC curve at the  
232 point that jointly maximized sensitivity and specificity [47]. For each ROC, the area under the

1  
2  
3 233 curve (AUC) and the positive and negative likelihood ratios (LR+ and LR-) were computed to  
4  
5  
6 234 **maximize the cut off scores.** Low, moderate and high accuracy of discrimination were defined  
7  
8 235 respectively as AUC <0.70, 0.70 < AUC <0.90, and AUC >0.90 [47]. In addition, the predictive  
9  
10 236 performance of the three models was compared to that of the OLS, TUG test and FR tests by  
11  
12  
13 237 reporting the above described parameters and the percentage of correctly classified patients.  
14  
15

## 16 238

## 17 239 **RESULTS**

18  
19  
20 239  
21  
22  
23 240 Fall history data, **for the analysis of discriminant validity, was available** could be collected only  
24  
25  
26 241 **from** in 295 subjects: 135 fallers (45%) and 160 non-fallers (55%). Table 1 reports mean **scores**  
27  
28 242 and standard deviation for each balance measure as well as information on the use of walking  
29  
30  
31 243 aids in the overall sample and in fallers vs. non-fallers. We found a significant difference between  
32  
33 244 fallers and non-fallers in the mean score of all clinical evaluations, while mean age and sex did  
34  
35  
36 245 not differ significantly. Table 2 reports descriptive statistics for each item score and total score of  
37  
38 246 the original 8-item Brief-BESTest in the whole sample.  
39  
40

### 41 247

### 42 248 **Structural validity**

#### 43 44 45 249

#### 46 250 ***Analysis of the Brief-BESTest models***

47  
48 251 Figure 1A shows the standardized solution of the CFA for Model 1 that was fitted using all 8  
49  
50  
51  
52 252 items. We found  $\chi^2$  value of 134.0 (df = 20, p<0.001), with CFI of 0.78, TLI of 0.70, RMSEA of 0.12  
53  
54  
55 253 (90% C.I. = 0.10–0.14) and SRMR above 0.09. Both CFI and TLI were below the cut-off value of

1  
2  
3 254 0.95 defined for a well-fitting model. In addition, the RMSEA value suggested that this model  
4  
5  
6 255 exhibited a poor fit of the data.  
7  
8  
9 256 Model 2 (figure 1B) showed better values of goodness of fit with respect to the original model. In  
10  
11  
12 257 fact, Model 2 had  $\chi^2$  of 60.3 (df = 18,  $p < 0.001$ ), with CFI of 0.92, TLI 0.88, RMSEA 0.08 (90% C.I. =  
13  
14 258 0.06–0.10) and SRMR 0.05. However, only SRMR exhibited a value lower than the preselected  
15  
16 259 well-fitting index value.  
17  
18  
19  
20 260 Figure 1C shows the CFA solution for Model 3. The results show a significantly better goodness of  
21  
22 261 fit for this model ( $\chi^2 = 26.2$ , df = 12, CFI = 0.97, TLI = 0.95, RMSEA = 0.05 (90% C.I. = 0.03–0.08),  
23  
24  
25 262 SRMR = 0.03) than for Models 1 and 2. Comparison of Model 3 ( $\chi^2 = 26.2$ ) to Model 1 ( $\chi^2 = 134.0$ )  
26  
27 263 and Model 2 ( $\chi^2 = 60.3$ ) yielded a difference in  $\chi^2$  value of 107.8 and 34.1 respectively and a  
28  
29  
30 264 difference of 6 degrees of freedom, suggesting that Model 3 performed better than the original  
31  
32 265 Brief-BESTest (Model 1) and Model 2. Table 3 summarizes the goodness of fit indices of each  
33  
34  
35 266 model and the significance level of the comparison between each model pair.  
36  
37  
38 267 The factor loadings of each item were significant ( $p < 0.001$ ) and higher than 0.6 for all three  
39  
40 268 models. Item 1 (Hip/Trunk Lateral Strength), when present, and item 2 (FR Forward) had the  
41  
42  
43 269 lowest factor loading.  
44  
45  
46 270  
47  
48  
49

### 50 271 **Internal consistency**

51  
52  
53 272 Cronbach's alpha was good for Brief-BESTest Models 1 (0.92) and 3 (0.92), but not for Model 2  
54  
55  
56 273 (0.56 and 0.88, respectively for static and dynamic balance factor). Coefficient omega was higher  
57  
58 274 than 0.80 only for Model 3, while Model 1 (0.75) and Model 2 (0.71 and 0.61, respectively for  
59  
60

1  
2  
3 275 static and dynamic balance factor) showed a lower coefficient, thus suggesting that only Model 3  
4  
5  
6 276 had good internal consistency.

7  
8  
9 277 The total score of the Brief-BESTest needs to be weighted considering the loading coefficient of  
10  
11 278 each item. Therefore, scores were adjusted to yield a uniform score range 0-24 for all three  
12  
13  
14 279 models, where 0 represents severe impairment and 24 no balance impairment. Below are the  
15  
16 280 weighted total score equations used to estimate Models 2 and 3:

17  
18  
19  
20 Brief-BESTest (Model 2) =  
21  $(0.860504 \cdot \text{item1}) + (0.914286 \cdot \text{item2}) + (1.089076 \cdot \text{item3}) + (1.048739 \cdot \text{item4}) + (1.021849 \cdot \text{item5}) +$   
22  $(0.994958 \cdot \text{item6}) + (1.008438 \cdot \text{item7}) + (1.062185 \cdot \text{item8})$   
23 281

24  
25  
26  
27 Brief-BESTest (Model 3) =  
28  $(0.963107 \cdot \text{item2}) + (1.133981 \cdot \text{item3}) + (1.118447 \cdot \text{item4}) + (1.21165 \cdot \text{item5}) + (1.165049 \cdot \text{item6}) +$   
29  $(1.165049 \cdot \text{item7}) + (1.242718 \cdot \text{item8})$   
30 282

### 31 32 33 283 **Convergent validity**

34  
35  
36  
37 284 The relationship of the total score estimated by the three different models of Brief-BESTest and  
38  
39 285 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. =  
40  
41 286 0.54-0.69) for Model 2 and  $\rho = 0.61$  (95% C.I. = 0.54-0.69) for Model 3. No significant ( $p = 0.85$ )  
42  
43  
44 287 difference was found between the three Spearman correlation coefficients.

### 45 46 47 288 48 49 50 51 289 **Discriminant validity**

52  
53  
54 290 Figure 2 shows the ROC curve plotted to assess the ability of the three models of Brief-BESTest to  
55  
56  
57 291 discriminate between patients with vs. without a history of falls. Table 4 reports the  
58  
59 292 discrimination parameters (**cut-off scores**) for Models 1, 2 and 3 and for the simple balance tests  
60

1  
2  
3 293 (OLS, TUG test and FR). The AUC was 0.72 (95% C.I. = 0.66–0.78) for all three models. **Model 3**  
4  
5  
6 294 **showed superior sensitivity, specificity and likelihood ratios compared to the other two**  
7  
8 295 **models.** The other simple balance tests (OLS, TUG test and FR) did not reach the AUC value of  
9  
10 296 0.70, i.e. the cut-off value required to distinguish between fallers and non-fallers. In addition,  
11  
12  
13 297 their AUCs were lower than those of the Brief-BESTest models.

16 298

## 20 299 **DISCUSSION**

23 300

27 301 The main purpose of this study was to compare, in a large group of patients with balance  
28  
29 302 disorders, the structural, convergent and discriminant validity of three different models of the  
30  
31 303 Brief-BESTest. The Brief-BESTest, in particular Model 3, was found to be unidimensional and to  
32  
33  
34 304 have a good convergent validity with other measures of balance confidence. In addition, the  
35  
36 305 Brief-BESTest confirmed its ability to distinguish subjects with a history of falling from those  
37  
38  
39 306 without a history of falling, in contrast to other simple balance tests.

### 42 307 ***Structural validity***

46 308 CFA showed that measurement properties of the original Brief-BESTest scale (Model 1) [13] little  
47  
48 309 fitted our data; none of CFA performance indices reached satisfactory values. As suggested in a  
49  
50 310 previous study [14], the original Brief-BESTest can be improved by making the following  
51  
52  
53 311 modifications: a) removing item 1; b) covarying errors between items 3 and 4 and between items  
54  
55  
56 312 5 and 6; c) using an appropriate weighting method for the calculation of the total score. The  
57  
58 313 internal structure of the scale can, in fact, be improved by removing item 1 rather than by  
59  
60 314 increasing the number of factors as proposed earlier [19]. The analysis of structural validity



1  
2  
3 315 prompted two main considerations. First, item 1 does not belong to the same construct as the  
4  
5  
6 316 other items. This finding, in line with previous studies [14,19], is also supported by the fact that  
7  
8 317 “lift a leg to the side of the body” (item 1) could reflect a general reduction in strength rather  
9  
10 318 than a decreased ability to maintain static balance. Second, the Brief-BESTest is unidimensional  
11  
12  
13 319 because no advantage in terms of fitting performance was found when more than one factor was  
14  
15 320 taken into consideration. In other words, our study confirmed that the Brief-BESTest is not able  
16  
17  
18 321 to measure multiple dimensions of balance as claimed by [13]. The Brief-BESTest items include all  
19  
20 322 subsections covered by the BESTest, but this does not mean that the two scales have the same  
21  
22  
23 323 capability to measure the different aspects underlying postural control. In accordance with other  
24  
25 324 authors, one could object that a unidimensional tool is a poor representation of the balance  
26  
27  
28 325 concept, which by definition is multidimensional. We agree with this objection and believe that  
29  
30 326 the Brief-BESTest can assess only certain aspects of dynamic balance. It neglects static  
31  
32  
33 327 components of balance such as those measured by the “Romberg test”. On the other hand, some  
34  
35 328 aspects of balance considered independent in animal models [48], such as walking and  
36  
37 329 maintenance of upright position, could be actually considered as well belonging to the same  
38  
39  
40 330 construct in humans. In fact, balance control and locomotion are interdependent at many  
41  
42 331 different levels of the central nervous system and these functions share some common principles  
43  
44  
45 332 of organization [49,50]. Furthermore, previous studies have reported that the ability to maintain  
46  
47 333 upright stance could be related to walking [51,52,53] in patients affected by balance disorders.  
48  
49  
50 334 Hence, in these patients it might be misleading for a clinical scale to investigate balance and  
51  
52 335 walking as two separate factors, as also suggested by recent recommendations [54].  
53  
54

### 55 336 ***Internal consistency***

56  
57  
58  
59  
60

1  
2  
3 337 Despite the large consensus in the psychometric literature that coefficient omega should be used  
4  
5  
6 338 when scales are not congeneric and the assumptions of Cronbach's alpha are not met [41], no  
7  
8 339 previous study has reported omega values for the Brief-BESTest. This is the first study to report  
9  
10 340 both the alpha and the omega coefficients. Based on the latter values, only Model 3 achieved  
11  
12  
13 341 good internal validity. On the contrary, Cronbach's alpha values were found to be good for both  
14  
15 342 Model 1 and 3. Our values of alpha are similar to those reported by [13], and to other studies in  
16  
17  
18 343 both orthopaedic [55] and neurological patients [56]. It is well known that Cronbach's alpha is a  
19  
20 344 function of the number of items. In this regard, it is interesting to note that Model 3, despite its  
21  
22  
23 345 lower number of items, reaches a higher value of internal consistency than the other models.  
24  
25 346 This should denote a superior internal consistency of Model 3.  
26  
27  
28  
29 347 The discrepancy between the results for Cronbach's alpha and those for coefficient omega can  
30  
31 348 be explained by the fact that the former has been frequently demonstrated to attain quite high  
32  
33  
34 349 values even when the items are measuring different latent variables [57]. On the contrary, the  
35  
36 350 coefficient omega is able to highlight the presence of items that do not belong to the same latent  
37  
38 351 variable. As for the Brief-BESTest, the low values of coefficient omega found for Model 1 could  
39  
40  
41 352 be due to the presence of item 1, which, as highlighted by CFA, seems not to belong to the same  
42  
43 353 construct as the other items.

#### 46 47 354 ***Convergent validity***

48  
49  
50 355 The three models compared in this study exhibit an equivalent moderate convergent validity  
51  
52 356 (Spearman's rho = 0.61) with the ABC 5-levels scale. This finding confirms previous studies in  
53  
54  
55 357 which the Brief-BESTest showed a moderate correlation with ABC in specific populations, such as  
56  
57 358 chronic obstructive pulmonary disease [15], stroke [11] and cancer survivors patients [58]. This is  
58  
59  
60

1  
2  
3 359 not surprising since it is well known that balance confidence is a consequence of balance  
4  
5  
6 360 impairments [59].  
7  
8

### 9 361 ***Discriminant validity***

10  
11  
12 362 The three Brief-BESTest models exhibited an equally **ivalent and** acceptable accuracy (AUC = 0.72)  
13  
14  
15 363 **of the Brief-BESTest** in distinguishing patients with vs. without history of falls. However, our AUCs  
16  
17 364 are lower than those of [17], who found in patients with Parkinson's disease AUC values of 0.82,  
18  
19  
20 365 0.86 and 0.84 respectively for the Brief-BESTest, Mini-BESTest and BESTest, **thus** indicating a  
21  
22 366 moderate accuracy of the three scales in recognising a history of falls. Likely, the small difference  
23  
24  
25 367 is due to the heterogeneity of our patients. **It We cannot be** excluded that studying separate  
26  
27 368 disease populations might improve the discriminant validity.  
28  
29

30  
31 369 In **accordance line** with previous studies, **we confirm that balance scales can discriminate**  
32  
33 370 **between fallers and non-fallers, in contrast to single balance tests which have a** **we found a** low  
34  
35 371 level of accuracy in identifying fallers **using only a single balance test** [11,60,61]. Due to the  
36  
37  
38 372 heterogeneity of patients in our study, it is unlikely that a single test could accurately identify  
39  
40 373 patients at risk of falls. This is in line with recommendations of [62] **who suggested** to use two or  
41  
42  
43 374 more clinical tests for an accurate evaluation of the various components of risk of falls in patients  
44  
45 375 with Parkinson's disease. The limited association between single measures of balance and fall  
46  
47  
48 376 history could be partly **explained by** **due to** the multifactorial nature of causes of fall.  
49  
50

51 377

### 52 53 54 378 **Study limitations**

55  
56  
57  
58 379 Our study was based on a heterogeneous **population sample** of patients affected by different  
59  
60 380 balance disorders **that are representative of the spectrum of neurological diseases routinely**

1  
2  
3 381 **observed in departments of rehabilitation, but may not be representative of other clinical**  
4  
5  
6 382 **settings. Although the Brief-BESTest assesses balance performance independently of the**  
7  
8 383 **pathology that causes the problem, future studies should assess psychometric characteristics**  
9  
10 384 **such as responsiveness, the minimal clinically important difference and sensitivity to change of**  
11  
12  
13 385 **the Brief-BESTest in specific groups of disease.**

14  
15  
16 386 The history of falls should be considered useful only for discriminative purposes (i.e. to  
17  
18 387 distinguish fallers from non-fallers) and not to predict patients who will fall in the future. In fact,  
19  
20  
21 388 falls change people's behavior and the cutoff scores for prospective prediction of falls may be  
22  
23  
24 389 very different from those reported in our study.

### 27 390 **Implications for clinical practice and conclusions**

28  
29  
30 391 The comparison of the three structural models of Brief-BESTest proposed in the literature clearly  
31  
32 392 highlights that Model 3, i.e. the shortest version, has psychometric characteristics equal or  
33  
34  
35 393 superior to the other two. This makes Model 3 the best of the three versions to recommend for  
36  
37  
38 394 use in clinical practice, given that it is also the fastest to perform. For this reason, in the  
39  
40 395 supplementary material we provide a simple calculator for the three models tested in this study.  
41  
42  
43 396 Finally, this study highlights once again the superiority of a clinical scale composed of several  
44  
45 397 items compared to single-item measures.

### 52 399 **Authors' contributions**

53  
54  
55 400 XX and XX contributed to the concept/idea/research design. XX, XX and XX contributed to the  
56  
57  
58 401 writing and data analysis. XX and XX provided data collection. XX and XX provided project  
59  
60 402 management and study participants. XX and XX provided facilities/equipment and institutional

1  
2  
3 403 liaisons. XX and XX contributed consultation in different phases of the study (including review of  
4  
5  
6 404 manuscript before submission).

7  
8  
9 405 **Ethical Approval**

10  
11  
12 406 The Scientific Technical Committee of the XXXXXX approved this study with the following  
13  
14  
15 407 identification number: XXXX.

16  
17  
18 408 **Funding**

19  
20  
21  
22 409 This research did not receive any specific grant from funding agencies in the public, commercial,  
23  
24 410 or not-for-profit sectors.

25  
26  
27  
28 411 **Conflict of Interest**

29  
30  
31 412 None declared.

32  
33  
34  
35 413  
36  
37  
38 414  
39  
40  
41 415 **Bibliography**

- 42  
43 416 [1] Geiger RA, Allen JB, O'Keefe J, Hicks RR. Balance and mobility following stroke: effects of  
44  
45  
46 417 physical therapy interventions with and without biofeedback/forceplate training. Phys Ther  
47  
48 418 2001;81(4):995-1005. PubMed PMID: 11276182.  
49  
50  
51 419 [2] Hyndman D, Ashburn A. People with stroke living in the community: Attention deficits,  
52  
53  
54 420 balance, ADL ability and falls. Disabil Rehabil 2003;25(15):817-22. PubMed PMID: 12851091.  
55  
56  
57  
58  
59  
60

- 1  
2  
3 421 [3] Horlings CG, Van Engelen BG, Allum JH, Bloem BR. A weak balance: the contribution of muscle  
4  
5  
6 422 weakness to postural instability and falls. *Nat Rev Neurol*. 2008;4(9):504. doi:  
7  
8 423 10.1038/ncpneuro0886.  
9  
10  
11 424 [4] Sturnieks DL, St George R, Lord SR. Balance disorders in the elderly. *Neurophysiol Clin*  
12  
13  
14 425 2008;38(6):467-478. doi: 10.1016/j.neucli.2008.09.001.  
15  
16  
17 426 [5] Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate  
18  
19  
20 427 balance deficits. *Eur J Phys Rehabil Med* 2010;46:239. PubMed PMID: 20485226.  
21  
22  
23 428 [6] McGinnis PQ, Hack LM, Nixon-Cave K, Michlovitz SL. Factors that influence the clinical  
24  
25 429 decision making of physical therapists in choosing a balance assessment approach. *Phys Ther*  
26  
27  
28 430 2009;89:233. doi: 10.2522/ptj.20080131.  
29  
30  
31 431 [7] Sibley KM, Straus SE, Inness EL, Salbach NM, Jaglal SB. Balance assessment practices and use  
32  
33 432 of standardized balance measures among Ontario physical therapists. *Phys Ther*  
34  
35  
36 433 2011;91:1583. doi: 10.2522/ptj.20110063.  
37  
38  
39 434 [8] Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate  
40  
41 435 balance deficits. *Phys Ther* 2009;89:484-498. doi: 10.2522/ptj.20080071.  
42  
43  
44  
45 436 [9] Jacobs JV, Horak FB, Tran VK, Nutt JG. Multiple balance tests improve the assessment of  
46  
47 437 postural stability in subjects with Parkinson's disease. *J Neurol Neurosurg Psychiatry*  
48  
49  
50 438 2006;1;77(3):322-6. PubMed PMID: 16484639.  
51  
52  
53 439 [10] Jacobs JV, Earhart GM, McNeely ME. Can postural instability tests improve the prediction of  
54  
55 440 future falls in people with Parkinson's disease beyond knowing existing fall history? *J Neurol*  
56  
57  
58 441 2016;263(1):133-9. doi: 10.1007/s00415-015-7950-x.  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 442 [11] Tsang CS, Liao LR, Chung RC, Pang MY. Psychometric properties of the Mini-Balance  
443 Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic  
444 stroke. *Phys Ther* 2013;93:1102-15. doi: 10.2522/ptj.20120454.
- 445 [12] Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to  
446 improve the Balance Evaluation Systems Test: the Mini-BESTest. *J Rehabil Med* 2010;42:323-  
447 331. doi: 10.2340/16501977-0537.
- 448 [13] Padgett PK, Jacobs JV and Kasser SL. Is the BESTest at its best? A suggested brief version  
449 based on interrater reliability, validity, internal consistency, and theoretical construct. *Phys*  
450 *Ther* 2012;92:1197-1207. doi: 10.2522/ptj.20120056.
- 451 [14] Bravini E, Nardone A, Godi M, Guglielmetti S, Franchignoni F, Giordano A. Does the Brief-  
452 BESTest meet classical test theory and Rasch analysis requirements for balance assessment in  
453 people with neurological disorders? *Phys Ther* 2016;96:1610-1619. PubMed PMID:  
454 27103223.
- 455 [15] Jácome C, Cruz J, Oliveira A, Marques A. Validity, reliability, and ability to identify fall status  
456 of the Berg Balance Scale, BESTest, Mini-BESTest, and Brief-BESTest in patients with COPD.  
457 *Phys Ther* 2016;96(11):1807-1815. doi: 10.2522/ptj.20150391.
- 458 [16] Leddy AL, Crowner BE, Earhart GM. Utility of the Mini-BESTest, BESTest, and BESTest  
459 sections for balance assessments in individuals with Parkinson disease. *J Neurol Phys Ther*  
460 2011;35:90-97. doi: 10.1097/NPT.0b013e31821a620c.
- 461 [17] Duncan RP, Leddy AL, Cavanaugh JT, Dibble LE, Ellis TD, Fors MP, Foreman KB, Earhart GM.  
462 Comparative utility of the BESTest, Mini-BESTest, and Brief-BESTest for predicting falls in

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 463 individuals with Parkinson disease: a cohort study. *Phys Ther* 2013;93:542-550. doi:  
464 10.2522/ptj.20120302.
- [18] O'Hoski S, Winship B, Herridge L, Agha T, Brooks D, Beauchamp MK, Sibley KM. Increasing  
465 the clinical utility of the BESTest, Mini-BESTest, and Brief-BESTest: normative values in  
466 Canadian adults who are healthy and aged 50 years or older. *Phys Ther* 2014;94:334-342. doi:  
467 10.2522/ptj.20130104.
- [19] Franchignoni F, Giordano A. On "Is the BESTest at its best?...." Padgett PK, Jacobs JV, Kasser  
469 SL. *Phys Ther* 2012;92:1197-1207. *Phys Ther* 2012;92(9):1236-7. doi:  
470 10.2522/ptj.2012.92.9.1236.
- [20] Marques A, Almeida S, Carvalho J, Cruz J, Oliveira A, Jácome C. Reliability, Validity, and  
472 Ability to Identify Fall Status of the Balance Evaluation Systems Test, Mini-Balance Evaluation  
473 Systems Test, and Brief-Balance Evaluation Systems Test in Older People Living in the  
474 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  
478 developed for data collection from medical records for use in retrospective analyses. *J Clin  
479 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  
481 tools in people with diabetic peripheral neuropathy. *Phys Ther* 2012;92:1461-1470. doi:  
482 10.2522/ptj.20120070.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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.

1

2

3 506 [33] Atkinson TM, Rosenfeld BD, Sit L, Mendoza TR, Fruscione M, Lavene D, Shaw M, Li Y, Hay J,  
4  
5  
6 507 Cleeland CS, Scher HI, Breitbart WS, Basch E. Using confirmatory factor analysis to evaluate  
7  
8 508 construct validity of the Brief Pain Inventory (BPI). *J Pain Symptom Manage* 2011;41(3):558-  
9  
10 509 565. doi: 10.1016/j.jpainsymman.2010.05.008.

12

13

14 510 [34] Bentler PM. Comparative fit indexes in structural models. *Psychol Bull* 1990;107(2):238-46.  
15  
16 511 PubMed PMID: 2320703.

17

18

19 512 [35] Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS,  
20  
21  
22 513 editors. *Testing Structural Equation Models*. Newbury Park: Sage 1993;136–62.

23

24

25 514 [36] Hu LT, Bentler P. Cutoff criteria for fit indices in covariance structure analysis: conventional  
26  
27  
28 515 criteria versus new alternatives. *Struct Equ Model Multidiscip J* 1999;6:1-55. doi:  
29  
30 516 10.1080/10705519909540118.

31

32

33 517 [37] DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: Considerations for  
34  
35  
36 518 the applied researcher. *PARE* 2009;14(20):1-11.

37

38

39 519 [38] Schönberger M, Ponsford J, Olver J, Ponsford, M. A longitudinal study of family functioning  
40  
41  
42 520 after TBI and relatives' emotional status. *Neuropsychol Rehabil* 2010;20:813-829. doi:  
43  
44 521 10.1080/09602011003620077.

45

46

47 522 [39] Hancock GR, Mueller RO. Rethinking construct reliability within latent variable systems.  
48  
49  
50 523 *Structural equation modeling: Present and future-A Festschrift in honor of Karl Jöreskog*.  
51  
52 524 Lincolnwood, IL: Scientific Software International 2001;195-216.

53

54

55 525 [40] Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*  
56  
57  
58 526 1951;16(3):297–334. PubMed PMID: 8449658.

59

60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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.

1

2

3 548 [49] Grasso R, Zago M, Lacquaniti F. Interactions between posture and locomotion: motor

4

5 549 patterns in humans walking with bent posture versus erect posture. *J Neurophysiol*

6

7 550 2000;83:288-300. doi: 10.1152/jn.2000.83.1.288.

8

9

10

11 551 [50] Massion J. Movement, posture and equilibrium: interaction and coordination. *Prog*

12

13 552 *Neurobiol* 1992;38(1):35-56. Review. PubMed PMID: 1736324.

14

15

16 553 [51] Giardini M, Nardone A, Godi M, Guglielmetti S, Arcolin I, Pisano F, Schieppati M.

17

18 554 Instrumental or physical-exercise rehabilitation of balance improves both balance and gait in

19

20 555 Parkinson's disease. *Neural Plast* 2018;2018:5614242. doi: 10.1155/2018/5614242.

21

22

23 556 [52] Nardone A, Godi M, Grasso M, Guglielmetti S, Schieppati M. Stabilometry is a predictor of

24

25 557 gait performance in chronic hemiparetic stroke patients. *Gait Posture* 2009;30(1):5-10. doi:

26

27 558 10.1016/j.gaitpost.2009.02.006.

28

29

30 559 [53] Morlino P, Balbi B, Guglielmetti S, Giardini M, Grasso M, Giordano C, Schieppati M, Nardone

31

32 560 A. Gait abnormalities of COPD are not directly related to respiratory function. *Gait Posture*

33

34 561 2017;58:352-357. doi: 10.1016/j.gaitpost.2017.08.020.

35

36

37 562 [54] Bloem BR, Marinus J, Almeida Q, Dibble L, Nieuwboer A, Post B, Ruzicka E, Goetz C, Stebbins

38

39 563 G, Martinez-Martin P, Schrag A; Movement Disorders Society Rating Scales Committee.

40

41 564 Measurement instruments to assess posture, gait, and balance in Parkinson's disease:

42

43 565 Critique and recommendations. *Mov Disord* 2016;31(9):1342-55. doi: 10.1002/mds.26572.

44

45

46 566 [55] Chan AC, Pang MY. Assessing balance function in patients with total knee arthroplasty. *Phys*

47

48 567 *Ther* 2015;95(10):1397-407. doi: 10.2522/ptj.20140486.

49

50

51

52

53

54

55

56

57

58

59

60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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 Sijsma. *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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
607**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)	<i>p</i>
	N %	N %	N %	
Sex (M/F)	55/45	51/49	58/42	.23 <sup>a</sup>
Walking Aid (no/yes)	54/46	47/53	68/32	<.001 <sup>a</sup>
	Mean (SD)	Mean (SD)	Mean (SD)	
Age (y)	66.5 (14.2)	66.3 (13.4)	64.4 (14.8)	.28 <sup>b</sup>
OLS mean (s)	4.2 (6.3)	3.3 (5.0)	6.9 (7.5)	<.001 <sup>b</sup>
TUG test (s)	17.6 (15.3)	20.6 (20.3)	13.3 (9.5)	<.001 <sup>b</sup>
Functional Reach (cm)	18.6 (9.7)	16.5 (9.6)	20.5 (9.8)	<.01 <sup>b</sup>
Total score BBT	8.9 (6.4)	7.3 (5.1)	11.8 (6.9)	<.001 <sup>b</sup>
Total score ABC 5-levels	29.6 (16.8)	26.0 (15.0)	36.2 (17.1)	<.001 <sup>b</sup>

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 12 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.

<sup>a</sup> =  $\chi^2$  test.

<sup>b</sup> = 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	$\chi^2$	df	RMSEA (90% C.I.)	CFI	TLI	SRMR
Model 1 ( <i>Padgett et al., 2012</i> )	134.0	20	0.12 (0.10-0.14)	0.78	0.70	0.09
Model 2 ( <i>adapted from Franchignoni &amp; Giordano, 2012</i> )	60.3	18	0.08 (0.06-0.10)	0.92	0.88	0.05
Model 3 ( <i>adapted from Bravini et al., 2016</i> )	26.2	12	0.05 (0.03-0.08)	0.97	0.95	0.03
Comparison of factor models		Significant difference between models <sup>a</sup>				
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.

<sup>a</sup> Calculated as a  $\chi^2 = \chi^2_2 - \chi^2_1$  with  $df = df_2 - df_1$ .



627 **Table 4**

628 Ability of the Brief-BESTest Models 1, 2 and 3 to discriminate fallers from non-fallers compared to  
 629 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%

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

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

1  
2  
3 635 **Appendix 1**  
4  
5

6 636 Items and factors of the three models of Brief-BESTest.  
7  
8  
9

Model item (Scoring 0-3 per item)	Model 1	Model 2	Model 3	
Item 1: Hip/Trunk Lateral Strength	Dynamic Balance	Static Balance		
Item 2: Functional Reach Forward		Dynamic Balance	Dynamic Balance	Dynamic Balance
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	0-24	0-24 (Static 0-5; Dynamic 0-19)	0-24	

28 637  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 1 **Is the Brief-BESTest brief enough? Suggested modifications based on structural validity and**  
6  
7  
8 2 **internal consistency**  
9

10  
11  
12 3 **Marco Godi <sup>a</sup>, Marica Giardini <sup>a,\*</sup>, Ilaria Arcolin <sup>a</sup>, Simona Ferrante <sup>b</sup>, Antonio Nardone <sup>a,c</sup>,**  
13  
14 4 **Stefano Corna <sup>a</sup>, Roberto Colombo <sup>a</sup>**  
15  
16

17 5  
18  
19 6 Author affiliation:

20  
21 7 <sup>a</sup> Istituti Clinici Scientifici Maugeri IRCCS, Italy

22  
23 8 <sup>b</sup> Neuroengineering and Medical Robotics Laboratory, Politecnico di Milano, Italy

24  
25 9 <sup>c</sup> Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Italy  
26  
27

28 10  
29  
30 11 \* Corresponding author: Marica Giardini.

31  
32 12 Istituti Clinici Scientifici Maugeri IRCCS,

33  
34 13 Via per Revislate 13, Veruno 28010, Italy.

35  
36 14 E-mail address: marica.giardini@icsmaugeri.it

37  
38 15 Ph: +39 0322 884905  
39  
40  
41  
42 16  
43  
44 17  
45  
46 18  
47  
48 19  
49  
50  
51  
52 20  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 21 **ABSTRACT**  
4  
5  
6 22  
7

8 23 **Background:** The Brief Balance Evaluation Systems Test (Brief-BESTest) could be a useful tool for  
9  
10 24 balance assessment. Although some psychometric characteristics have been examined, others  
11  
12  
13 25 still need to be clarified.  
14

15 26 **Objectives:** To assess the structural validity, convergent validity, discriminant validity and  
16  
17  
18 27 internal consistency of the Brief-BESTest in neurological patients.  
19

20 28 **Design:** Cross-sectional.  
21

22 29 **Methods:** Data were from 416 patients with neurological disease and related balance disorders.  
23  
24

25 30 Patients were assessed with the 5-levels Activities-Specific Balance Confidence Scale (ABC 5-  
26

27 31 levels), Brief-BESTest, **some simple balance tests, i.e. One-Leg Stance (OLS), Timed Up and Go**  
28

29 32 **(TUG) test, Functional Reach (FR), simple balance tests** and a fall history questionnaire. Three  
30  
31

32 33 **models of Brief-BESTest models** were examined through confirmatory factor analysis (CFA) and  
33  
34

35 34 the following indices calculated: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean  
36

37 35 Square Error of Approximation (RMSEA). Convergent validity was assessed by calculating the  
38

39 36 correlation between Brief-BESTest and ABC 5-levels total scores. Receiver operating  
40

41 37 characteristics (ROC) assessed the ability of each model to differentiate between people with vs.  
42  
43

44 38 without falls. Internal consistency was measured by Cronbach's alpha and coefficient omega.  
45  
46

47 39 **Results:** CFA showed Model 3 (CFI=0.97, TLI=0.95, RMSEA=0.05), **with item 1 removed and error**  
48

49 40 **covariance between items 3-4 and between items 5-6**, to have a significantly better structure  
50

51 41 than Models 1 and 2 ( $p < 0.001$ ). The correlation between Brief-BESTest and ABC 5-levels was 0.61  
52  
53

54 42 (Spearman's rho) for all three models. Area Under the Curve (AUC) of ROC **was showed an**  
55

56 43 **acceptable accuracy (0.72) in distinguishing patients with vs. without history of falls** (95%  
57  
58

59 44 C.I.=0.66–0.78) for all models, and superior to AUCs of other simple balance tests (**OLS, TUG test,**  
60

1  
2  
3 45 **FR).** Cronbach's alpha was good for Brief-BESTest Models 1 (0.92) and 3 (0.92), but omega was  
4  
5  
6 46 >0.80 only for Model 3.

7  
8 47 **Limitations:** Heterogeneous sample size was a heterogeneous population.

9  
10 48 **Conclusions:** The Brief-BESTest, after some changes, shows good validity and internal  
11  
12  
13 49 consistency in patients affected by different balance disorders, after applying some changes.  
14

15 50

16  
17  
18 51 **Contribution of the Paper:**

19  
20 52 • Although some psychometric characteristics of the Brief-BESTest have been examined in  
21  
22  
23 53 previous studies, other properties such as validity still need to be clarified.

24  
25 54 • This study shows that the Brief-BESTest has good validity and internal consistency in  
26  
27  
28 55 patients affected by different balance disorders, after applying some changes: removal of item 1  
29  
30 56 and using an appropriate weighting method for the calculation of the total score.

31  
32 57 • This study confirms the ability of the Brief-BESTest to distinguish between people with vs.  
33  
34  
35 58 without history of falls, in contrast to other simple balance tests. Moreover, it highlights once  
36  
37 59 again the superiority of a clinical scale composed of several items compared to single-item  
38  
39  
40 60 measures such as the TUG test and OLS.

41  
42 61

43  
44 62

45  
46  
47 63 **Keywords:** Brief-Balance Evaluation Systems Test, balance assessment, confirmatory factor  
48  
49 64 analysis, structural validity, internal consistency.  
50

51  
52 65

53  
54 66

55  
56  
57  
58  
59  
60

## 67 INTRODUCTION

68

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

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

82 One of the most recent balance scales is the BESTest, a 36-item scale developed to identify  
83 impairments in six balance control subsections, which **it has been** shown to be a valid and  
84 reliable tool [8]. However, one of its drawbacks is that it is time-consuming **to administer** [12].

85 For this reason, shorter versions have been proposed such as the Mini-BESTest [12] and the  
86 Brief-BESTest [13]. In particular, the Brief-BESTest, an 8-item version of the original scale, has  
87 demonstrated good to excellent psychometric properties [8,13,14,15]. It is less time-consuming  
88 [16] and more feasible than its parent scale in clinical settings [15], while it encompasses and  
89 should adequately evaluate all subsections of balance endorsed by the BESTest [13,17,18].

1  
2  
3 90 However, the latter hypothesis was rejected through exploratory factor analysis **dismantled the**  
4  
5  
6 91 **latter hypothesis**, demonstrating **that the Brief-BESTest actually has at most two subsections,**  
7  
8 92 **or dimensions** [19]. Furthermore, Bravini et al. [14] showed by Rasch analysis that all items of  
9  
10 93 the Brief-BESTest except for item 1 account for the same underlying theoretical construct and  
11  
12  
13 94 indicated that the Brief-BESTest should in fact be considered as unidimensional. Therefore, the  
14  
15 95 authors suggested the adoption of a 7-item version of the test.

16  
17  
18 96 Although some psychometric characteristics of the Brief-BESTest, such as the internal  
19  
20  
21 97 consistency [13,14], reliability [14,17,20] and sensitivity to change [14,20], have been  
22  
23  
24 98 investigated in previous studies, other properties still need to be clarified. In particular, the Brief-  
25  
26 99 BESTest structure has not yet **been investigated with undergone** confirmatory factor analysis  
27  
28  
29 100 (CFA). This statistical tool provides information on possible independent factors and can be very  
30  
31 101 useful for developing shortened forms of an **evaluating** instrument [21,22]. Finally, the Brief-  
32  
33  
34 102 BESTest seems to have good sensitivity and accuracy in identifying retrospectively people who  
35  
36 103 have had at least one fall [13,17]. However, these findings are based only on small samples of  
37  
38 104 patients with multiple sclerosis [13] or Parkinson's disease [17].

39  
40  
41  
42 105 **The We aimed in of** the present study **was** to fill the existing knowledge gap by examining the  
43  
44 106 structural validity, convergent validity and discriminant validity of the Brief-BESTest in a large  
45  
46  
47 107 group of patients with a variety of balance disorders. In particular, we hypothesized that:

48  
49  
50 108 1) among the three models of the Brief-BESTest presented in the literature [13,14,19], the 7-item  
51  
52 109 version would be the one with the best structural validity;  
53  
54  
55  
56 110 2) in spite of its conciseness, the 7-item model [14] would have the same ability as the other two  
57  
58 111 Brief-BESTest models to **predict patients at risk of falls; discriminate between people with vs.**  
59  
60 112 **without a history of falls;**

1  
2  
3 113 3) in discriminating between people with vs. without a history of falls, the Brief-BESTest would be  
4  
5  
6 114 superior to other simple balance tests such as One Leg Stance, Timed Up and Go test and  
7  
8 115 Functional Reach.

9  
10  
11 116

## 12 13 14 15 117 **METHODS**

16  
17 118

### 18 19 20 21 119 **Participants**

22  
23  
24 120 This was an observational retrospective study conducted in a group of 416 patients affected by  
25  
26  
27 121 different neurological diseases: 186 females and 230 males; mean age  $66.5 \pm 16.0$  years (**mean  $\pm$**   
28  
29 122 standard deviation) consecutively admitted for in-patient rehabilitation at the Istituti Clinici  
30  
31 123 Scientifici Maugeri IRCCS, Institute of Veruno (Novara, Italy) between February 2014 and April  
32  
33  
34 124 2017. **Patients' clinical and treatment data were extracted from the electronic medical record**  
35  
36 125 **system and transferred to a specific database (Microsoft Excel).** Patients were stratified into  
37  
38  
39 126 different groups according to their diagnosis: 118 **with** Parkinson's disease, 79 **with** acute stroke,  
40  
41 127 43 **with** sensorimotor polyneuropathy, 32 **with** cerebellar ataxia, 32 **with** diffuse  
42  
43  
44 128 encephalopathy, 31 **with** chronic stroke, 21 **with** multiple sclerosis, 19 **with** traumatic brain  
45  
46 129 injury, 16 **with** vestibular disorder, 13 **with** neuromuscular disorders, 12 **with** central nervous  
47  
48  
49 130 system neoplasm. Inclusion criteria were: a) ability to maintain an upright position without  
50  
51 131 support for at least 5 seconds; b) ability to understand the required motor tasks; c) no hip or  
52  
53  
54 132 knee replacement surgery within the previous 6 months. Exclusion criteria were: a)  
55  
56 133 musculoskeletal injury limiting the ability to walk; b) any other serious cardio-respiratory  
57  
58  
59 134 problem. The study was carried out in conformity with the Declaration of Helsinki of the World  
60



1  
2  
3 135 Medical Association and the guidelines for retrospective studies [23]. The local scientific and  
4  
5  
6 136 ethics committee approved the study.  
7  
8

9 137

### 12 138 **Assessment tools**

16 139 Patients' demographic and clinical characteristics were gathered by **Aa** team of trained physical  
17  
18 140 therapists engaged in clinical practice collected patients' demographic and clinical  
19  
20  
21 141 characteristics. During the routine clinical assessment following admission to the rehabilitation  
22  
23 142 department, patients underwent the following assessments:

#### 26 143 **- Fall history**

30 144 A history of falls over the past 6 months was obtained **from patients** at admission through  
31  
32 145 **patient** interview. **A fall was defined as an unintentional event in which any part of the body**  
33  
34  
35 146 **came into contact with the ground [10]**. Patients who reported two or more falls in the defined  
36  
37 147 period were classified as 'fallers' [24]. A fall history was not **recorded taken in the case of for**  
38  
39  
40 148 patients with acute stroke at the time of admission.  
41  
42

#### 43 149 **- Brief-BESTest**

47 150 The Brief-BESTest is an 8-item scale with each item scored on a 4-level rating scale from 0 (severe  
48  
49 151 balance impairment) to 3 (no balance impairment). Its items cover the six subsections of the  
50  
51 152 original BESTest (biomechanical constraints, stability limits/verticality, anticipatory postural  
52  
53  
54 153 responses, postural responses, sensory orientation, stability in gait); the maximum total score is  
55  
56 154 **24 [13]. The Brief-BESTest requires less time and equipment to administer than the BESTest**  
57  
58  
59 155 **and the Mini-BESTest; thus, the Brief-BESTest seems to be more feasible for clinical use [17].**  
60

1  
2  
3 156 - **Simple balance tests**

4  
5  
6  
7 157 During administration of the Brief-BESTest, we recorded the time required by patients to  
8  
9 158 complete item 3 (left One Leg Stance), 4 (right One Leg Stance) and 8 (Timed Up and Go test) and  
10  
11 159 the distance covered by patients during item 2 (Functional Reach). This allowed us to obtain the  
12  
13  
14 160 scores of three additional simple balance tests: One Leg Stance (OLS) [25,26], Timed Up and Go  
15  
16 161 (TUG) test [27,28] and Functional Reach (FR) [29,30].

17  
18  
19  
20 162 - **Activities-specific Balance Confidence Scale**

21  
22 163 The short version of the Activities-specific Balance Confidence Scale (ABC) is a self-reported 16-  
23  
24  
25 164 item questionnaire that scores the perceived level of balance confidence when performing  
26  
27 165 common activities of daily living [31]. We used the 5-levels **rating** version of rating (ABC 5-levels)  
28  
29  
30 166 [32] in which each item is scored from 0 (no confidence) to 4 (fully confident), giving a total score  
31  
32 167 range 0-64.

33  
34  
35  
36 168  
37  
38  
39 169 **Data analysis**

40  
41  
42 170 Descriptive statistics were used to describe mean demographic and balance performance  
43  
44  
45 171 characteristics of the entire sample and of the two **smaller** subgroups, classified as fallers and  
46  
47 172 non-fallers. **The analysis of discriminant validity was conducted only in these two subgroups,**  
48  
49  
50 173 **i.e. those patients who had a history of falls available.** These values were also determined  
51  
52 174 separately both for fallers and non-fallers. For each item of the Brief-BESTest we calculated:  
53  
54  
55 175 median value, spread (25<sup>th</sup>–75<sup>th</sup> percentiles), skewness and kurtosis. Floor and ceiling effects  
56  
57 176 were analyzed by calculating the percentage of individuals obtaining the lowest and the highest  
58  
59  
60 177 score for each scale item. In order to detect differences in clinical characteristics between fallers

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

178 and non-fallers, the Chi-square ( $\chi^2$ ) test was used for two parameters, sex and use of walking  
179 aids, while the Mann-Whitney U-test was used for age, OLS, TUG test, FR, total score of Brief-  
180 BESTest and total score of ABC 5-levels. Significance was set at  $p < 0.05$ . All analyses were  
181 performed using STATA 13.0 software (StataCorp LLC, College Station, Texas, USA).

### 183 **Structural validity**

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

188 The original model (Model 1) of the Brief-BESTest [13] comprises one factor with 8 independent  
189 items that contribute with the same weights to the total score. In the recent literature, two  
190 additional models of Brief-BESTest have been presented. Model 2 was designed based on [19]. It  
191 includes two factors: one named "static balance" that comprises items 1 and 2, and another  
192 called "dynamic balance" that contains items 3 to 8. As demonstrated by the authors, items 5  
193 and 6 showed local dependence, so Model 2 was designed allowing correlation between the  
194 errors of these items. In Model 3, was drawn up without item 1 was dropped, as suggested by  
195 [14]. In this 7-item model, the error of item 3 was allowed to correlate with that of item 4 and  
196 the error of item 5 with that of item 6.

197 For all models, the score of each item ranges from 0 to 3. Then the total score was is obtained by  
198 multiplying the rated score by the coefficient fitted for each model (see below formula and  
199 supplementary data). In order to allow comparison of the score models, we adjusted the

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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**he 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.  $\chi^2$  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  $\geq 0.95$ ; b) RMSEA value  $< 0.06$ ; and c) SRMR value  $\leq 0.08$  [36].

The goodness of fit parameters of the three models were compared by computing the  $\chi^2$  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

1  
2  
3 222 and the coefficient omega for congeneric models [39]. Cronbach's alpha measures the extent to  
4  
5  
6 223 which the items consistently measure the same construct, with the value  $\geq 0.80$  indicating good  
7  
8 224 internal consistency [40]. The coefficient omega is the ratio of the true score variance and the  
9  
10  
11 225 total variance of the scale. Interpretation of coefficient omega is similar to that of Cronbach's  
12  
13 226 alpha [41].  
14  
15  
16  
17 227

### 20 228 **Convergent validity**

21  
22  
23 229 We used the correlation between the three models of Brief-BESTest and the ABC 5-levels total  
24  
25  
26 230 score to assess the convergent validity. The ABC scale led to rate rates the balance self-efficacy  
27  
28 231 of patients [31]. This is associated with measures of balance [32], walking capacity [42],  
29  
30  
31 232 functional mobility [43], Activities of Daily Living performance [44], and perceived health status  
32  
33 233 [45]. The choice to use the ABC as a competitor an external criterion was also based on the need  
34  
35  
36 234 to avoid the comparisons of the three models with a scale that had items similar to the Brief-  
37  
38 235 BESTest. The correlation was assessed by means of Spearman's rho: coefficients  $< 0.30$  were  
39  
40 236 interpreted as weak, those between 0.30 and 0.49 as moderate, and those  $\geq 0.50$  as strong  
41  
42  
43 237 correlations [46].  
44  
45  
46 238

### 49 239 **Discriminant validity**

50  
51  
52  
53 240 To assess the ability of the three Brief-BESTest models to distinguish between 'fallers' and 'non-  
54  
55  
56 241 fallers', receiver operating characteristic (ROC) curves were computed, by plotting sensitivity on  
57  
58 242 the x-axis against 1 – specificity on the y-axis. In our study, sensitivity was calculated as the  
59  
60 243 number of patients correctly identified as 'fallers' and specificity as the number of patients

1  
2  
3 244 correctly identified as 'non-fallers'. The optimal cut-off value was chosen on the ROC curve at the  
4  
5  
6 245 point that jointly maximized sensitivity and specificity [47]. For each ROC, the area under the  
7  
8 246 curve (AUC) and the positive and negative likelihood ratios (LR+ and LR-) were computed to  
9  
10 247 **maximize the cut off scores**. Low, moderate and high accuracy of discrimination were defined  
11  
12  
13 248 respectively as  $AUC < 0.70$ ,  $0.70 < AUC < 0.90$ , and  $AUC > 0.90$  [47]. In addition, the predictive  
14  
15 249 performance of the three models was compared to that of the OLS, TUG test and FR tests by  
16  
17  
18 250 reporting the above described parameters and the percentage of correctly classified patients.  
19  
20  
21 251

## 25 252 **RESULTS**

26  
27  
28 253 Fall history data, **for the analysis of discriminant validity, was available** could be collected only  
29  
30 254 **from** in 295 subjects: 135 fallers (45%) and 160 non-fallers (55%). Table 1 reports mean **scores**  
31  
32  
33 255 and standard deviation for each balance measure as well as information on the use of walking  
34  
35 256 aids in the overall sample and in fallers vs. non-fallers. We found a significant difference between  
36  
37  
38 257 fallers and non-fallers in the mean score of all clinical evaluations, while mean age and sex did  
39  
40 258 not differ significantly. Table 2 reports descriptive statistics for each item score and total score of  
41  
42  
43 259 the original 8-item Brief-BESTest in the whole sample.  
44  
45

### 50 261 **Structural validity**

#### 57 263 ***Analysis of the Brief-BESTest models***

58  
59  
60 264 Figure 1A shows the standardized solution of the CFA for Model 1 that was fitted using all 8

1  
2  
3 265 items. We found  $\chi^2$  value of 134.0 (df = 20,  $p < 0.001$ ), with CFI of 0.78, TLI of 0.70, RMSEA of 0.12  
4  
5  
6 266 (90% C.I. = 0.10–0.14) and SRMR above 0.09. Both CFI and TLI were below the cut-off value of  
7  
8 267 0.95 defined for a well-fitting model. In addition, the RMSEA value suggested that this model  
9  
10 268 exhibited a poor fit of the data.

11  
12  
13  
14 269 Model 2 (figure 1B) showed better values of goodness of fit with respect to the original model. In  
15  
16 270 fact, Model 2 had  $\chi^2$  of 60.3 (df = 18,  $p < 0.001$ ), with CFI of 0.92, TLI 0.88, RMSEA 0.08 (90% C.I. =  
17  
18 0.06–0.10) and SRMR 0.05. However, only SRMR exhibited a value lower than the preselected  
19 271 well-fitting index value.  
20  
21 272

22  
23  
24  
25 273 Figure 1C shows the CFA solution for Model 3. The results show a significantly better goodness of  
26  
27 274 fit for this model ( $\chi^2 = 26.2$ , df = 12, CFI = 0.97, TLI = 0.95, RMSEA = 0.05 (90% C.I. = 0.03–0.08),  
28  
29 275 SRMR = 0.03) than for Models 1 and 2. Comparison of Model 3 ( $\chi^2 = 26.2$ ) to Model 1 ( $\chi^2 = 134.0$ )  
30  
31 and Model 2 ( $\chi^2 = 60.3$ ) yielded a difference in  $\chi^2$  value of 107.8 and 34.1 respectively and a  
32 276 difference of 6 degrees of freedom, suggesting that Model 3 performed better than the original  
33  
34 277 Brief-BESTest (Model 1) and Model 2. Table 3 summarizes the goodness of fit indices of each  
35  
36 278 model and the significance level of the comparison between each model pair.  
37  
38 279

39  
40  
41  
42  
43 280 The factor loadings of each item were significant ( $p < 0.001$ ) and higher than 0.6 for all three  
44  
45 281 models. Item 1 (Hip/Trunk Lateral Strength), when present, and item 2 (FR Forward) had the  
46  
47 282 lowest factor loading.  
48  
49  
50

#### 51 283

#### 52

#### 53

#### 54

#### 55 284 **Internal consistency**

#### 56

#### 57

58 285 Cronbach's alpha was good for Brief-BESTest Models 1 (0.92) and 3 (0.92), but not for Model 2  
59  
60 286 (0.56 and 0.88, respectively for static and dynamic balance factor). Coefficient omega was higher

1  
2  
3 287 than 0.80 only for Model 3, while Model 1 (0.75) and Model 2 (0.71 and 0.61, respectively for  
4  
5  
6 288 static and dynamic balance factor) showed a lower coefficient, thus suggesting that only Model 3  
7  
8 289 had good internal consistency.  
9

10  
11  
12 290 The total score of the Brief-BESTest needs to be weighted considering the loading coefficient of  
13  
14 291 each item. Therefore, scores were adjusted to yield a uniform score range 0-24 for all three  
15  
16 292 models, where 0 represents severe impairment and 24 no balance impairment. Below are the  
17  
18  
19 293 weighted total score equations used to estimate Models 2 and 3:  
20

21  
22  
23 Brief-BESTest (Model 2) =  
24  $(0.860504 \cdot \text{item1}) + (0.914286 \cdot \text{item2}) + (1.089076 \cdot \text{item3}) + (1.048739 \cdot \text{item4}) + (1.021849 \cdot \text{item5}) +$   
25  $(0.994958 \cdot \text{item6}) + (1.008438 \cdot \text{item7}) + (1.062185 \cdot \text{item8})$   
26 294

27  
28  
29 Brief-BESTest (Model 3) =  
30  $(0.963107 \cdot \text{item2}) + (1.133981 \cdot \text{item3}) + (1.118447 \cdot \text{item4}) + (1.21165 \cdot \text{item5}) + (1.165049 \cdot \text{item6}) +$   
31  $(1.165049 \cdot \text{item7}) + (1.242718 \cdot \text{item8})$   
32 295  
33

### 34 35 36 296 **Convergent validity**

37  
38  
39 297 The relationship of the total score estimated by the three different models of Brief-BESTest and  
40  
41  
42 298 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. =  
43  
44 299 0.54-0.69) for Model 2 and  $\rho = 0.61$  (95% C.I. = 0.54-0.69) for Model 3. No significant ( $p = 0.85$ )  
45  
46 300 difference was found between the three Spearman correlation coefficients.  
47  
48

### 49 50 301 51 52 53 302 **Discriminant validity**

54  
55  
56  
57 303 Figure 2 shows the ROC curve plotted to assess the ability of the three models of Brief-BESTest to  
58  
59 304 discriminate between patients with vs. without a history of falls. Table 4 reports the  
60



1  
2  
3 305 discrimination parameters (**cut-off scores**) for Models 1, 2 and 3 and for the simple balance tests  
4  
5  
6 306 (OLS, TUG test and FR). The AUC was 0.72 (95% C.I. = 0.66–0.78) for all three models. **Model 3**  
7  
8 307 **showed superior sensitivity, specificity and likelihood ratios compared to the other two**  
9  
10 308 **models.** The other simple balance tests (OLS, TUG test and FR) did not reach the AUC value of  
11  
12  
13 309 0.70, i.e. the cut-off value required to distinguish between fallers and non-fallers. In addition,  
14  
15 310 their AUCs were lower than those of the Brief-BESTest models.  
16  
17  
18  
19 311  
20  
21

## 22 312 **DISCUSSION**

23  
24  
25  
26 313  
27  
28  
29 314 The main purpose of this study was to compare, in a large group of patients with balance  
30  
31 315 disorders, the structural, convergent and discriminant validity of three different models of the  
32  
33  
34 316 Brief-BESTest. The Brief-BESTest, in particular Model 3, was found to be unidimensional and to  
35  
36 317 have a good convergent validity with other measures of balance confidence. In addition, the  
37  
38  
39 318 Brief-BESTest confirmed its ability to distinguish subjects with a history of falling from those  
40  
41 319 without a history of falling, in contrast to other simple balance tests.  
42  
43  
44

### 45 320 ***Structural validity***

46  
47  
48 321 CFA showed that measurement properties of the original Brief-BESTest scale (Model 1) [13] little  
49  
50 322 fitted our data; none of CFA performance indices reached satisfactory values. As suggested in a  
51  
52  
53 323 previous study [14], the original Brief-BESTest can be improved by making the following  
54  
55 324 modifications: a) removing item 1; b) covarying errors between items 3 and 4 and between items  
56  
57  
58 325 5 and 6; c) using an appropriate weighting method for the calculation of the total score. The  
59  
60 326 internal structure of the scale can, in fact, be improved by removing item 1 rather than by

1  
2  
3 327 increasing the number of factors as proposed earlier [19]. The analysis of structural validity  
4  
5  
6 328 prompted two main considerations. First, item 1 does not belong to the same construct as the  
7  
8 329 other items. This finding, in line with previous studies [14,19], is also supported by the fact that  
9  
10  
11 330 “lift a leg to the side of the body” (item 1) could reflect a general reduction in strength rather  
12  
13 331 than a decreased ability to maintain static balance. Second, the Brief-BESTest is unidimensional  
14  
15 332 because no advantage in terms of fitting performance was found when more than one factor was  
16  
17  
18 333 taken into consideration. In other words, our study confirmed that the Brief-BESTest is not able  
19  
20 334 to measure multiple dimensions of balance as claimed by [13]. The Brief-BESTest items include all  
21  
22  
23 335 subsections covered by the BESTest, but this does not mean that the two scales have the same  
24  
25 336 capability to measure the different aspects underlying postural control. In accordance with other  
26  
27  
28 337 authors, one could object that a unidimensional tool is a poor representation of the balance  
29  
30 338 concept, which by definition is multidimensional. We agree with this objection and believe that  
31  
32  
33 339 the Brief-BESTest can assess only certain aspects of dynamic balance. It neglects static  
34  
35 340 components of balance such as those measured by the “Romberg test”. On the other hand, some  
36  
37 341 aspects of balance considered independent in animal models [48], such as walking and  
38  
39  
40 342 maintenance of upright position, could be actually considered as well belonging to the same  
41  
42 343 construct in humans. In fact, balance control and locomotion are interdependent at many  
43  
44  
45 344 different levels of the central nervous system and these functions share some common principles  
46  
47 345 of organization [49,50]. Furthermore, previous studies have reported that the ability to maintain  
48  
49  
50 346 upright stance could be related to walking [51,52,53] in patients affected by balance disorders.  
51  
52 347 Hence, in these patients it might be misleading for a clinical scale to investigate balance and  
53  
54  
55 348 walking as two separate factors, as also suggested by recent recommendations [54].

56  
57  
58 349 ***Internal consistency***

59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

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

### 367 ***Convergent validity***

368 The three models compared in this study exhibit an equivalent moderate convergent validity  
369 (Spearman's rho = 0.61) with the ABC 5-levels scale. This finding confirms previous studies in  
370 which the Brief-BESTest showed a moderate correlation with ABC in specific populations, such as  
371 chronic obstructive pulmonary disease [15], stroke [11] and cancer survivors patients [58]. This is

1  
2  
3 372 not surprising since it is well known that balance confidence is a consequence of balance  
4  
5  
6 373 impairments [59].

### 9 374 ***Discriminant validity***

10  
11  
12 375 The three Brief-BESTest models exhibited an equally **ivalent and** acceptable accuracy (AUC = 0.72)  
13  
14  
15 376 **of the Brief-BESTest** in distinguishing patients with vs. without history of falls. However, our AUCs  
16  
17 377 are lower than those of [17], who found in patients with Parkinson's disease AUC values of 0.82,  
18  
19  
20 378 0.86 and 0.84 respectively for the Brief-BESTest, Mini-BESTest and BESTest, **thus** indicating a  
21  
22 379 moderate accuracy of the three scales in recognising a history of falls. Likely, the small difference  
23  
24  
25 380 is due to the heterogeneity of our patients. **It We cannot be** excluded that studying separate  
26  
27 381 disease populations might improve the discriminant validity.

28  
29  
30 382 In **accordance line** with previous studies, **we confirm that balance scales can discriminate**  
31  
32  
33 383 **between fallers and non-fallers, in contrast to single balance tests which have a** **we found a** low  
34  
35 384 level of accuracy in identifying fallers **using only a single balance test** [11,60,61]. Due to the  
36  
37  
38 385 heterogeneity of patients in our study, it is unlikely that a single test could accurately identify  
39  
40 386 patients at risk of falls. This is in line with recommendations of [62] **who suggested** to use two or  
41  
42  
43 387 more clinical tests for an accurate evaluation of the various components of risk of falls in patients  
44  
45 388 with Parkinson's disease. The limited association between single measures of balance and fall  
46  
47  
48 389 history could be partly **explained by** **due to** the multifactorial nature of causes of fall.

### 51 390 52 53 54 391 **Study limitations**

55  
56  
57  
58 392 Our study was based on a heterogeneous **population sample** of patients affected by different  
59  
60 393 balance disorders **that are representative of the spectrum of neurological diseases routinely**

1  
2  
3 394 **observed in departments of rehabilitation, but may not be representative of other clinical**  
4  
5  
6 395 **settings. Although the Brief-BESTest assesses balance performance independently of the**  
7  
8 396 **pathology that causes the problem, future studies should assess psychometric characteristics**  
9  
10  
11 397 **such as responsiveness, the minimal clinically important difference and sensitivity to change of**  
12  
13 398 **the Brief-BESTest in specific groups of disease.**

14  
15  
16 399 The history of falls should be considered useful only for discriminative purposes (i.e. to  
17  
18 distinguish fallers from non-fallers) and not to predict patients who will fall in the future. In fact,  
19 400  
20 falls change people's behavior and the cutoff scores for prospective prediction of falls may be  
21 401  
22 very different from those reported in our study.  
23  
24 402

### 25 26 27 403 **Implications for clinical practice and conclusions**

28  
29  
30 404 The comparison of the three structural models of Brief-BESTest proposed in the literature clearly  
31  
32 highlights that Model 3, i.e. the shortest version, has psychometric characteristics equal or  
33 405  
34 superior to the other two. This makes Model 3 the best of the three versions to recommend for  
35 406  
36 use in clinical practice, given that it is also the fastest to perform. For this reason, in the  
37 407  
38 supplementary material we provide a simple calculator for the three models tested in this study.  
39 408  
40 Finally, this study highlights once again the superiority of a clinical scale composed of several  
41 409  
42 items compared to single-item measures.  
43 410  
44  
45 411  
46  
47  
48  
49 411

### 50 51 52 412 **Authors' contributions**

53  
54  
55 413 MGo and AN contributed to the concept/idea/research design. MGo, MGi and IA contributed to  
56  
57 the writing and data analysis. MGi and IA provided data collection. MGo and SC provided project  
58 414  
59 management and study participants. AN and SC provided facilities/equipment and institutional  
60 415

1  
2  
3 416 liaisons. SC and RC contributed consultation in different phases of the study (including review of  
4  
5  
6 417 manuscript before submission).  
7  
8

9 418 **Ethical Approval**  
10  
11

12 419 The Scientific Technical Committee of the Istituti Clinici Scientifici Maugeri approved this study  
13  
14  
15 420 with the following identification number: rrf41.  
16  
17

18 421 **Funding**  
19  
20

21  
22 422 This research did not receive any specific grant from funding agencies in the public, commercial,  
23  
24 423 or not-for-profit sectors.  
25  
26

27  
28 424 **Conflict of Interest**  
29

30  
31 425 None declared.  
32  
33

34 426  
35  
36  
37  
38 427  
39  
40  
41 428 **Bibliography**  
42

- 43 429 [1] Geiger RA, Allen JB, O'Keefe J, Hicks RR. Balance and mobility following stroke: effects of  
44  
45  
46 430 physical therapy interventions with and without biofeedback/forceplate training. Phys Ther  
47  
48 431 2001;81(4):995-1005. PubMed PMID: 11276182.  
49  
50  
51 432 [2] Hyndman D, Ashburn A. People with stroke living in the community: Attention deficits,  
52  
53  
54 433 balance, ADL ability and falls. Disabil Rehabil 2003;25(15):817-22. PubMed PMID: 12851091.  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 476 individuals with Parkinson disease: a cohort study. *Phys Ther* 2013;93:542-550. doi:  
477 10.2522/ptj.20120302.
- [18] O'Hoski S, Winship B, Herridge L, Agha T, Brooks D, Beauchamp MK, Sibley KM. Increasing  
479 the clinical utility of the BESTest, Mini-BESTest, and Brief-BESTest: normative values in  
480 Canadian adults who are healthy and aged 50 years or older. *Phys Ther* 2014;94:334-342. doi:  
481 10.2522/ptj.20130104.
- [19] Franchignoni F, Giordano A. On "Is the BESTest at its best?...." Padgett PK, Jacobs JV, Kasser  
482 SL. *Phys Ther* 2012;92:1197-1207. *Phys Ther* 2012;92(9):1236-7. doi:  
483 10.2522/ptj.2012.92.9.1236.
- [20] Marques A, Almeida S, Carvalho J, Cruz J, Oliveira A, Jácome C. Reliability, Validity, and  
485 Ability to Identify Fall Status of the Balance Evaluation Systems Test, Mini-Balance Evaluation  
486 Systems Test, and Brief-Balance Evaluation Systems Test in Older People Living in the  
487 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  
491 developed for data collection from medical records for use in retrospective analyses. *J Clin  
492 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  
494 tools in people with diabetic peripheral neuropathy. *Phys Ther* 2012;92:1461-1470. doi:  
495 10.2522/ptj.20120070.

- 1  
2  
3 497 [25] Lin MR, Hwang HF, Hu MH, Wu HDI, Wang YW, Huang FC. Psychometric comparisons of the  
4  
5  
6 498 timed up and go, one-leg stand, functional reach, and Tinetti balance measures in  
7  
8 499 community-dwelling older people. *J Am Geriatr Soc* 2004;52:1343-1348. doi: 10.1111/j.1532-  
9  
10 500 5415.2004.52366.x.
- 11  
12  
13  
14 501 [26] Vellas BJ, Wayne SJ, Romero L, Baumgartner RN, Rubenstein LZ, Garry PJ. One-leg balance is  
15  
16 502 an important predictor of injurious falls in older persons. *J Am Geriatr Soc* 1997;45:735–738.  
17  
18  
19 503 PubMed PMID: 9180669.
- 20  
21  
22 504 [27] Sebastião E, Sandroff BM, Learmonth YC, Motl RW. Validity of the timed up and go test as a  
23  
24 505 measure of functional mobility in persons with multiple sclerosis. *Arch Phys Med Rehabil*  
25  
26  
27 506 2016;97:1072-1077. doi: 10.1016/j.apmr.2015.12.031.
- 28  
29  
30 507 [28] Podsiadlo D, Richardson S. The timed 'Up & Go': A test of basic functional mobility for frail  
31  
32 508 elderly persons. *J Am Geriatr Soc* 1991;39:142–148. PubMed PMID: 1991946.
- 33  
34  
35  
36 509 [29] Franchignoni F, Tesio L, Martino M, Ricupero C. Reliability of four simple, quantitative tests  
37  
38 510 of balance and mobility in healthy elderly females. *Aging Clin Exp Res* 1998;10:26-31.  
39  
40  
41 511 PubMed PMID: 9589748.
- 42  
43  
44 512 [30] Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of  
45  
46 513 balance. *J Gerontol* 1990;45(6):M192-7. PubMed PMID: 2229941.
- 47  
48  
49  
50 514 [31] Powell LE, Myers AM. The activities-specific balance confidence (ABC) scale. *J Gerontol A Biol*  
51  
52 515 *Sci Med Sci* 1995;50(1):M28-M34. PubMed PMID: 7814786.
- 53  
54  
55 516 [32] Franchignoni F, Giordano A, Ronconi G, Rabini A, Ferriero G. Rasch validation of the  
56  
57  
58 517 Activities-specific Balance Confidence Scale and its short versions in patients with Parkinson's  
59  
60 518 disease. *J Rehabil Med* 2014;46:532-539. doi: 10.2340/16501977-1808.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [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 Sijsma. *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.

602 **Table 1**

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

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

605  
 606 **The total sample comprises all patients, i.e. those in whom a fall history was collected (the two**  
 607 **subgroups fallers/non-fallers) as well as those in whom a fall history was not collected due to**  
 608 **acute stroke event (n=79) or failure to complete the questionnaire (n=42), as this is a**  
 609 **retrospective study.** Participants were classified as fallers if they reported 1 2 or more falls in the  
 610 last 6 months.

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

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

615 <sup>a</sup> =  $\chi^2$  test.

616 <sup>b</sup> = Mann-Whitney U-test.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

621  
622  
623

**Table 2**

624 Details on each single item and total score of the original Brief-BESTest (Model 1) in the whole sample (N  
625 = 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	

626

627 Item 1 - Hip/Trunk Lateral Strength; item 2 - Functional Reach Forward; item 3 and 4 - One Leg

628 Stance, Left and Right; item 5 and 6 - Compensatory Stepping- Lateral, Left and Right; item 7 -

629 stand on a foam surface; 8 - Timed Up and Go test.

630



631 **Table 3**  
632 Summary of fit statistics of the specified models (N = 416)

Models	$\chi^2$	df	RMSEA (90% C.I.)	CFI	TLI	SRMR
Model 1 ( <i>Padgett et al., 2012</i> )	134.0	20	0.12 (0.10-0.14)	0.78	0.70	0.09
Model 2 ( <i>adapted from Franchignoni &amp; Giordano, 2012</i> )	60.3	18	0.08 (0.06-0.10)	0.92	0.88	0.05
Model 3 ( <i>adapted from Bravini et al., 2016</i> )	26.2	12	0.05 (0.03-0.08)	0.97	0.95	0.03
Comparison of factor models		Significant difference between models <sup>a</sup>				
Model 1 vs. Model 2		p<0.001				
Model 2 vs. Model 3		p<0.001				
Model 3 vs. Model 1		p<0.001				

634 df, degrees of freedom; RMSEA, Root Mean-Square Error of Approximation; C.I., Confidence

635 Interval; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; SRMSR, Standardized Root Mean

636 Square Residual.

637 <sup>a</sup> Calculated as a  $\chi^2 = \chi^2_2 - \chi^2_1$  with  $df = df_2 - df_1$ .

640 **Table 4**

641 Ability of the Brief-BESTest Models 1, 2 and 3 to discriminate fallers from non-fallers compared to  
 642 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%

643

644

644 OLS, One Leg Stance; TUG test, Timed Up and Go test; FR, Functional Reach; AUC, Area Under the

645

645 Curve; C.I., Confidence Interval; LR+, positive Likelihood Ratio; LR-, negative Likelihood Ratio.

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

1  
2  
3 648 **Appendix 1**  
4  
5

6 649 Items and factors of the three models of Brief-BESTest.  
7  
8  
9

Model item (Scoring 0-3 per item)	Model 1	Model 2	Model 3	
Item 1: Hip/Trunk Lateral Strength	Dynamic Balance	Static Balance		
Item 2: Functional Reach Forward		Dynamic Balance	Dynamic Balance	Dynamic Balance
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	0-24	0-24 (Static 0-5; Dynamic 0-19)	0-24	

28 650  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

<b>Items</b>
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

<b>Total score (range 0-24)</b>
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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only



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

Model 1	Model 2	Model 3*
Dynamic Balance	Static Balance	
	Dynamic Balance	Dynamic Balance

	<b>14.00</b>	<b>13.95</b>	<b>13.01</b>
		2.46	
		11.52	

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

. .  
. .  
. .

. .  
. .  
. .  
. .

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

.  
. .  
.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Peer Review Only



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Peer Review Only

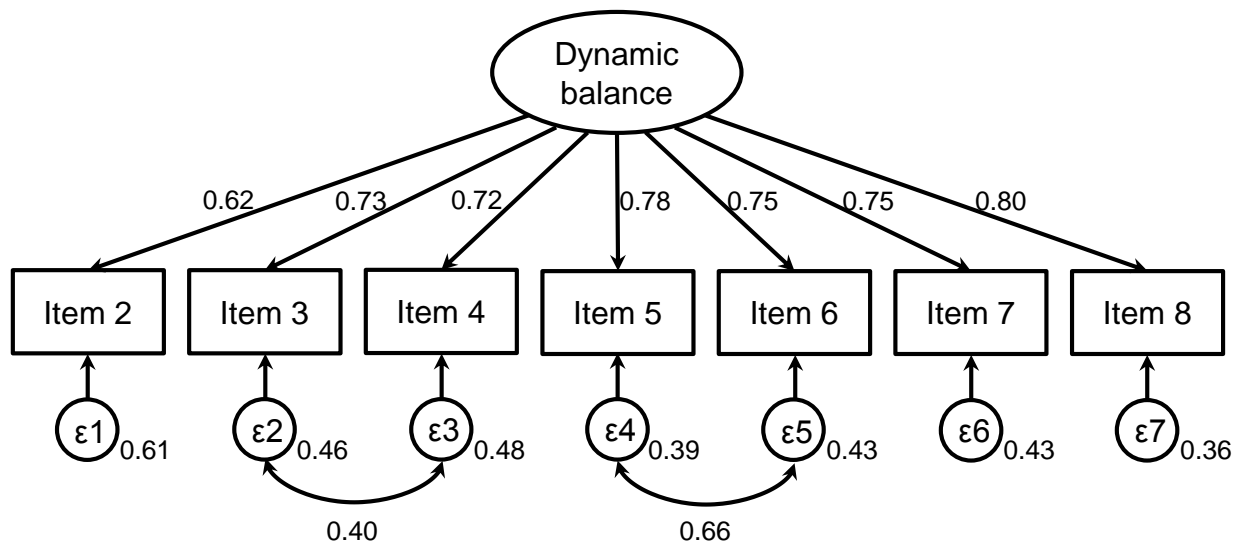
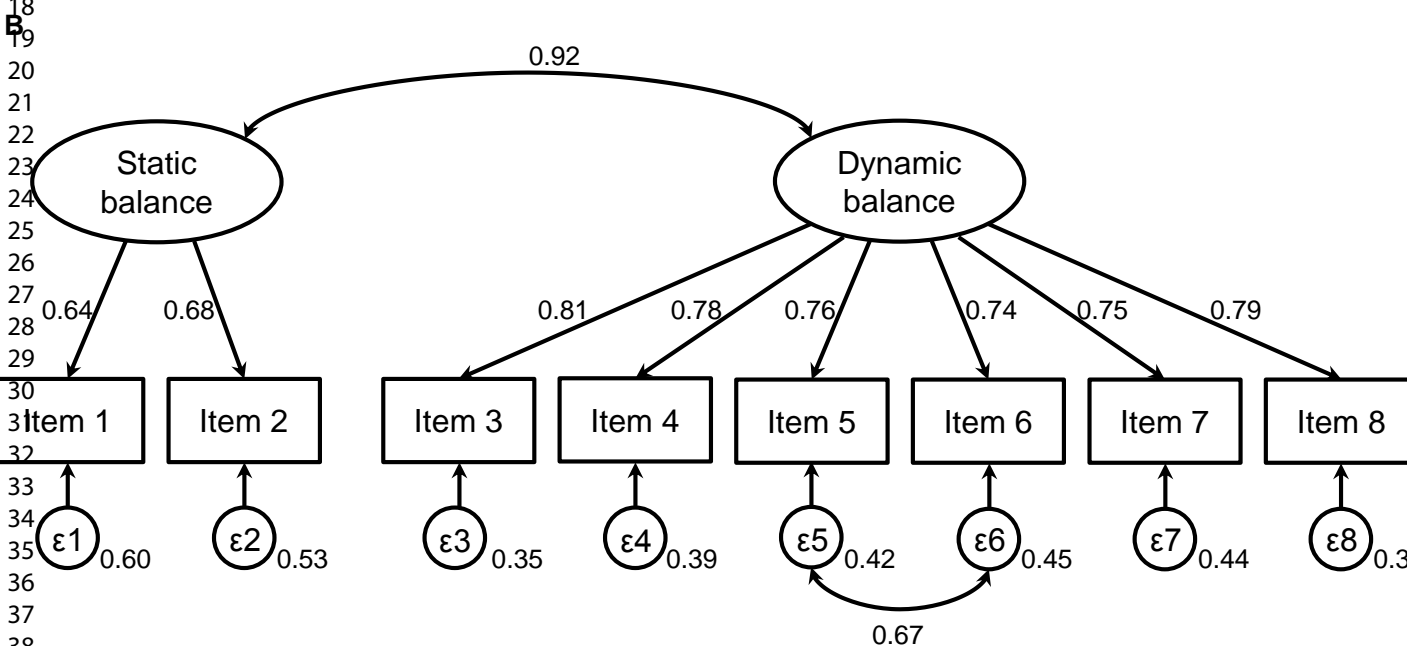
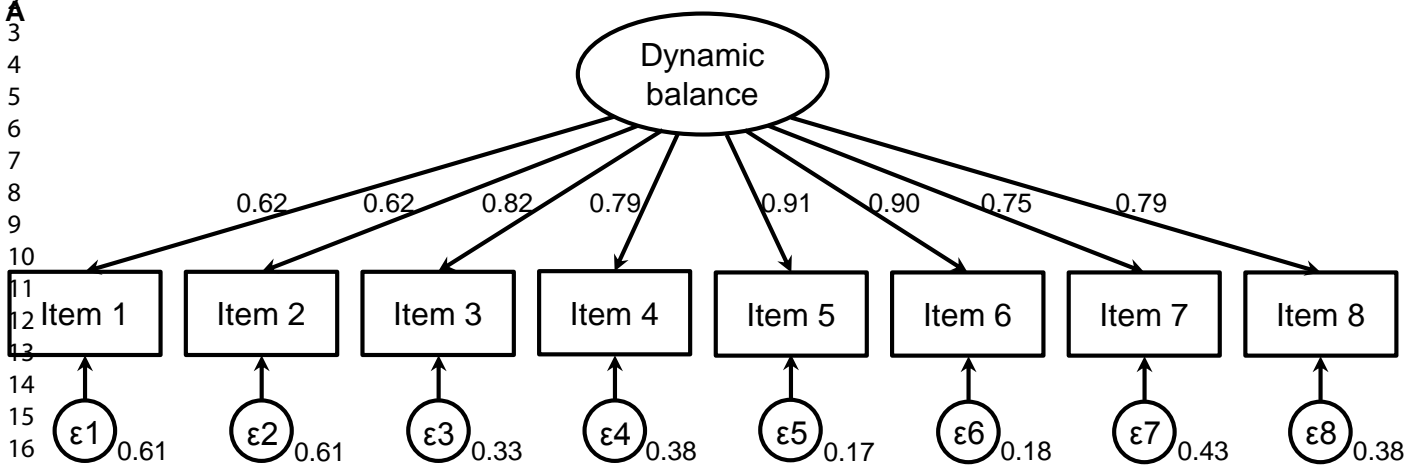
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

