Reliability, validity and ability to identify fall status of the BESTest, Mini-BESTest and Brief-BESTest in older people living in the community

Abstract

Objective: The reliability, validity and ability to identify fall status of the Balance Evaluation Systems Test (BESTest), Mini-BESTest and Brief-BESTest with the Berg Balance Scale (BBS) in older people living in the community was assessed.

Design: Cross-sectional.

Setting: Community centers.

Participants: 122 older adults (76±9 years) participated.

Interventions: Not applicable.

Main Outcome Measures: Participants reported on falls history in the preceding year and completed the Activities-specific Balance Confidence (ABC) Scale. BBS, BESTest and the Five times sit-to-stand test were administered. Interrater (two physiotherapists) and test-retest relative (48-72h) and absolute reliability were explored with the Intraclass correlation coefficient (ICC) equation (2,1) and the Bland and Altman method. Minimal detectable changes at the 95% confidence level (MDC_{95}) were established. Validity was assessed by correlating the balance tests with each other and with the ABC Scale (Spearman correlation coefficients - *rho*). Receiver operating characteristics assessed the ability of each balance test to differentiate between people with and without a history of falls.

Results: All balance tests presented good to excellent interrater (ICC 0.71-0.93) and test-retest (ICC 0.50-0.82) relative reliability, with no evidence of bias. MDC_{95} values were 4.6, 9, 3.8 and 4.1 points for the BBS, BESTest, Mini-BESTest and Brief-BESTest, respectively. All tests were significantly correlated with each other (*rho*=0.83-0.96) and with the ABC Scale (*rho*=0.46-0.61). Acceptable ability to identify fall status (areas under the curve 0.71-0.78) was found for all tests. Cut-off points were 48.5, 82, 19.5 and 12.5 points for BBS, BESTest, Mini-BESTest and Brief-BESTest, respectively.

Conclusions: All balance tests are reliable, valid and able to identify fall status in older people living in the community therefore, the choice of which test to use will depend on the level of balance impairment, purpose and time availability.

Key-words: Postural Balance; geriatric assessment; risk assessment; ROC curve.

List of abbreviations

ABC	Activities-specific Balance Confidence Scale
AUC	Area under the curve
BBS	Berg Balance Scale
BESTest	Balance Evaluation Systems Test
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
GRRAS	Guidelines for reliability and agreement studies
ICC	Intraclass correlation coefficient
MDC	Minimal detectable changes
rho	Spearman correlation coefficients
SEM	Standard error of measurement
STROBE	STrengthening the Reporting of OBservational studies in Epidemiology

Falls are a major public health problem.[1] About a third of community-living older people fall each year.[2] For older people living in the community, the human cost of falling includes distress, pain, injury, loss of confidence, and ultimately loss of independence and mortality.[3] Therefore, falling has an enormous impact on quality of life and health of community-living older people. Fall prevention programs are currently a worldwide priority[1] and to develop such programs it is essential to tackle modifiable fall risk factors. Balance impairment is one of the major modifiable risk factor for falls.[4]

Several clinical balance tests are directed at screening for balance impairments and predicting falls. One of the most commonly used tool in clinical situations is the Berg Balance Scale (BBS), which has shown to be reliable, valid,[5,6] sensitive and specific in predicting falls in community-based older populations.[7] Despite being well-studied, this scale has considerable ceiling effect in various populations,[5,8-10] namely in older people living in the community who often represent a population with a higher level of functioning, transmitting the idea that no balance impairment is present, when in fact specific balance training is often warrant.

Development of balance tests able to identify the underlying systems responsible for balance deficits are critical for health professionals to diagnose specific impairments and develop individualized treatment plans in community-living older people.[11] The Balance Evaluation Systems Test (BESTest) was developed to assess functioning of six balance control systems.[12] This test has been widely used[13-17] however, its administration takes a considerable amount of time (20-30 minutes), which may not be feasible and practical for routine clinical use. Thus, two abbreviated versions of the BESTest, i.e., Mini-BESTest and Brief-BESTest, which take approximately half of the time to be administrated (10-15 minutes), have been developed.[18,19] The measurement properties of BESTest,[12,13] Mini-BESTest[20,21] and Brief-BESTest[19] have been explored in people with balance disorders and with specific conditions (e.g., Parkinson disease, chronic stoke), however little is known in healthy older people. Recently, construct validity and discriminative ability for older people living in the community have been explored for the BESTest, Mini-BESTest and Brief-BESTest[22] however, a comparison with BBS was not performed and reliability and cut-offs to differentiate between fallers and non-fallers were not established.

Therefore, this study aimed to compare the reliability, validity and the ability to identify fall status of the BESTest, the Mini-BESTest and the Brief-BESTest with the BBS in older people living in the community. We hypothesized that all four balance tests will be i) moderately reliable, ii) highly correlated with each other and with self-reported balance confidence and iii) able to differentiate between fallers and non-fallers. Our secondary aims were to determine the minimal detectable change and propose cut-offs to differentiate between fallers and non-fallers for the same balance tests.

Methods

Study design and ethics

A cross-sectional study was conducted from November 2014 to February 2015 in the central region of Portugal. Ethical approval was obtained from the Ethics Committee

(238/10-2014). This study is reported following the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN),[23] the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement[24] and the guidelines for reliability and agreement studies (GRRAS).[25]

Participants

Fifteen community centers, which main purpose is to promote physical activity and active cognitive functioning of older adults were contacted. Nine agreed to participate, i.e., 2 community activity “clubs”, offering exercises classes, and 7 day care centers, offering a range of recreational, cultural, educational, health and social support services. Staff from the community-centers identified eligible participants according to the following criteria: (i) aged 60 years or older; (ii) living independently in the community and iii) able to ambulate with or without a walking aid (but without the assistance of another person). Individuals were excluded if they i) had comorbidities that interfered with independent ambulation (e.g., hip fracture, lower limb amputation and hemiplegia); ii) were taking medication that could cause dizziness or affect their balance (e.g., psychotropic medications); iii) had self-reported cardiorespiratory diseases (e.g., chronic obstructive pulmonary disease, heart failure) or neurological (e.g., Parkinson disease, multiple sclerosis, dementia), musculoskeletal (e.g., severe osteoarthritis) or psychiatric impairments (e.g., psychosis, schizophrenia), that could interfere with the measurements. Individuals were contacted by the researchers, who explained the purpose of the study and asked about their willingness to participate. When individuals agreed to participate, an appointment with the researchers was scheduled at their institution. Written informed consent was obtained prior to data collection.

Data collection procedures

Socio-demographic (age and gender), anthropometric (height, weight, body mass index) and clinical (self-reported medication and common age-related comorbidities - e.g., hypertension and hyperlipidemia) data were collected. Individuals were provided with a clear definition of falls (an event when you find yourself unintentionally on the ground, floor or lower level)[26] and required to report their history of falls using two standardized questions (1. "*Have you had any falls in the last 12 months?*" and, if yes, 2.*"How many times did you fall down in the last 12 months?").*[27] The self-reported Activities-Specific Balance Confidence (ABC) scale was used to assess balance confidence.[28] The ABC Scale quantifies an individual's perceived ability to maintain his/her balance under different circumstances, using a scale of 0% (no confidence) to 100% (total confidence).[29] Then, the BBS, the BESTest and the Five times sit-to-stand test (FTSTS) were performed and participants were encouraged to rest, as needed. Two trained physiotherapists, collected the data.

For interrater reliability, each of the two physiotherapists performed a simultaneous independent balance assessment at baseline.[30,31] For each item of the BBS and of BESTest, one physiotherapist read the instructions to the participant while the other physiotherapist performed the task. Then, the participant performed the task with close supervision. Each task was scored immediately after completion by the two physiotherapists and scoring was not discussed. For test-retest reliability, participants were reassessed by 1 of the 2 physiotherapists after a 48–72h interval. The order of testing was the same.

The Mini-BESTest and Brief-BESTest scores were computed based on the performance of the BESTest tasks. A custom designed worksheet was used by the physiotherapists to simultaneously record the BESTest and the Mini-BESTest item scores. Brief-BESTest scores were extracted from the relevant subset of BESTest items.

Tests

Berg Balance scale. The BBS is composed of 14 items to assess performance of specific functional tasks. Each item is scored from 0 (unable to perform) to 4 (normal performance) and the maximum score is 56 points. BBS has been shown to have high interrater and test-retest reliability in personal care home residents[32] and in patients with Parkinson disease and stroke.[13,33,34]

BESTest. The BESTest contains 36 items organized into 6 subsections: biomechanical constraints, stability limits and verticality, anticipatory postural adjustments, postural responses to external perturbations, sensory orientation during stance and stability in gait.[12] Each item is scored from 0 (severe balance impairment) to 3 (no balance impairment) and the maximum possible total score is 108 points. BESTest has high interrater reliability (ICC=0.91-0.99) and test-retest reliability (ICC=0.80-0.99) in subjects with and without balance disorders, in patients with Parkinson disease and in older cancer survivors.[12,13,16] Moreover, the BESTest has been found to be accurate in identifying future recurrent fallers in patients with Parkinson disease.[35]

Mini-BESTest. The Mini-BESTest includes 14 items from sections of the BESTest related to anticipatory postural adjustments, reactive postural responses, sensory orientation and stability in gait.[18] Two of the 14 items are scored bilaterally. Each item is scored from 0 (severe balance impairment) to 2 (no balance impairment) and the maximum possible score is 28 points. Higher scores indicate better balance performance. In patients with Parkinson disease, the Mini-BESTest test showed high sensitivity (89%) and specificity (81%) in identifying people with abnormal postural responses.[36] High interrater (ICC=0.86-0.98) and test-retest reliability (ICC=0.88-0.97) have also been found in patients with balance disorders, Parkinson disease, chronic stroke and also in older cancer survivors.[16,20,21,31]

Brief-BESTest. The Brief-BESTest is a 6-item balance test, which contains 1 item from each of the 6 subsections of the BESTest.[19] Two of the 6 items, are scored bilaterally, resulting in an 8-item balance test. Each item is scored from 0 (severe balance impairment) to 3 (no balance impairment) and the maximum possible score is 24 points. Higher scores indicate better balance performance.[19] The Brief-BESTest was accurate in identifying future recurrent fallers in patients with Parkinson disease.[35] It has also showed high interrater (ICC=0.86-0.96) and test-retest (ICC=0.90-0.98) reliability in individuals with and without neurological diagnoses[19] and older cancer survivors.[16]

A summary of the core components of each balance test, based on the information provided in the Rehabilitation Measures Database,[37] is presented in Supplementary Table 1.

(Supplementary table 1)

FTSTS

The FTSTS was used as a measure of functional mobility.[38] Participants started the test in a seated position with their arms folded across their chests and their back against the chair backrest. The following verbal instructions were given: “Please stand up and sit down 5 times as quickly as possible”. Timing (seconds) with a digital stopwatch was initiated manually at the “Go” instruction, and stopped when the subject’s back touched the backrest after the fifth stand. Two trials were performed and the mean of the trials was used in the analysis. This test has shown good intrarater, interrater and test-retest reliability in healthy older adults.[39]

Statistical analysis

Descriptive statistics were used to describe the sample. A z-test was applied for normality test using skewness and kurtosis.[40] Individuals with a history of falls were defined as those who reported at least one fall during the past year; individuals without a history of falls were defined as those who reported no falls during the past year. Participants’ characteristics were compared between those with and without a history of falls, using independent t-tests for normally distributed data, Mann-Whitney U-tests for non-normally distributed and ordinal data and Chi-square tests for categorical data. When significant differences on the performance of balance tests were found between those with and without a history of falls, effect sizes were computed. The *Cohen’s d* was used as this is the effect size estimate recommended for Independent t-tests.[41] Cohen’s *d* was calculated using the G*Power 3.1 software (University Düsseldorf, Düsseldorf, Germany) and was interpreted as a small (≥0.2), medium (≥0.5), or large (≥0.8) effect.[42] The skewness of the distribution of scores was also assessed for each balance test to verify the occurrence of ceiling and floor effects. A positive skewness value greater than 1 denoted a substantial floor effect and a negative value lower than -1.0 indicated a substantial ceiling effect.[43]

Reliability

Interrater and test-retest reliability were analyzed in a subsample of the first consecutive 28 participants. This sample size was determined according to the study from Bonnet,[44] where it has been established that a sample size of 21 individuals is sufficient to estimate an ICC of 0.9 with a 95% confidence interval width of 0.2 (α=0.05 and k=2).[44] A 30% attrition rate was estimated, yielding a sample of 28 individuals.

Interrater reliability was computed using the scores obtained from the 2 physiotherapists in session 1 and test-retest reliability using the scores from 1 physiotherapist in sessions 1 and 2. For relative reliability, the ICC equation (2,1) was used as it was intended to generalize the present results to a variety of raters.[45] ICC was interpreted as excellent (>0.75), moderate to good (0.4-0.75) or poor (<0.4).[46] Absolute reliability was explored with the Bland and Altman method and standard error of measurement (SEM). The Bland and Altman method was used for visual judgement of how well the measurements between raters or sessions agreed.[47]

The SEM indicates the extent to which a score varies on repeated measurements. It provides a value for measurement error in the same units as the measurement itself and thus, is easier to use in clinical practice.[48] It was calculated using equation 1:

$$SEM = SD\sqrt{(1 - ICC)} \qquad (1)$$

where SD is the standard deviation of the scores obtained from all individuals and ICC is the test-retest reliability coefficient.

Minimal Detectable Change

The MDC at the 95% level of confidence (MDC_{95}), is also reported in the same units as the measurement itself and represents the smallest change that can be interpreted as a real difference. It was calculated as follows (equation 2):

$$MDC_{95} = SEM \times 1.96 \times \sqrt{2} \qquad (2)$$

The MDC was also expressed as a percentage ($MDC_{\%}$), defined as (equation 3):

$$MDC_{\%} = (MDC_{95}/mean) \times 100 \qquad (3)$$

where mean is the mean of the scores obtained in the 2 testing sessions. The $MDC_{\%}$ is independent of the units of measurement and facilitates the comparison of the random measurement error among different measures.[49] A $MDC_{\%}$ below 30% is considered acceptable and below 10% excellent.[50]

Validity

Spearman's correlation coefficient (*rho*) was used to examine the relationship among balance tests (concurrent validity) and between each balance test and the ABC Scale (convergent validity).

Ability to identify fall status

Receiver operating characteristic analysis and area under the curves (AUC) and the 95% confidence interval were used to determine the cut-off points of each balance test differentiating individuals with and without a history of falls.[41] Receiver operating characteristic analysis is a plot of a true-positive rate (sensitivity) and false-positive rate (1-specificity) across the full range of cut-off scores. The AUC is an index of the diagnostic accuracy of the test. Interpretation of the AUC values was as follow: AUC=0.5 no discrimination; 0.7≤AUC<0.8 acceptable discrimination; 0.8≤AUC<0.9 excellent discrimination and AUC≥0.9 outstanding discrimination.[51] The optimal cut-off value for each balance test to differentiate between individuals with and without a history of falls was chosen as the intersection point which maximized both, sensitivity and specificity values.

All statistical analyses were performed using IBM SPSS Statistics version 20.0 (IBM Corporation, Armonk, NY, USA) and plots created using GraphPad Prism version 5.01 (GraphPad Software, Inc., La Jolla, CA, USA). The level of significance was set at 0.05.

Results

Participants

A total of 130 individuals were invited to participate. However, 8 refused to complete the assessment. Therefore, 122 participants were included. Participants were mainly female (n=86; 70.5%), with a mean age of 76±9 years old. Thirteen participants used walking aids, eleven used a cane and two used a walker. Participants' characteristics and balance scores are presented in Table 1.

(table 1)

Skewness of BESTest (-0.62), Mini-BESTest (-0.67) and Brief-BESTest (-0.40) were lower than the BBS (-1.46).

Reliability and Minimal Detectable Change

Table 2 presents the relative and absolute interrater and test-retest reliability results. Moderate to good interrater relative reliability was observed for the Mini-BESTest (ICC 0.71), while the other balance tests presented excellent interrater relative reliability (ICC from 0.86 to 0.93). The Bland and Altman method also showed good agreement between the two physiotherapists, with small mean differences (from -0.1±1.1 to 0.4±3.5) for all balance tests (Table 2). BBS (ICC=0.50) and Mini-BESTest had moderate to good (ICC=0.73) whereas BESTest and Brief-BESTest had excellent (ICC of 0.77 and 0.82) relative test-retest reliability (Table 2).

(table 2)

The MDC_{95} was 4.6 (SEM=1.7 $MDC_{\%}$=8.6%), 9 (SEM=3.3; $MDC_{\%}$=10.2%), 3.8 (SEM=1.4; $MDC_{\%}$=16.3%) and 4.1 (SEM=1.5; $MDC_{\%}$=21.5%) points for BBS, BESTest, Mini-BESTest and Brief-BESTest, respectively.

Validity

All balance tests were strongly correlated with each other (*rho* ranged from 0.83-0.96; p<0.001). Scores of the ABC Scale were also significantly correlated with the scores of BBS (*rho*=0.58; p<0.001), BESTest (*rho*=0.46; p<0.001), Mini-BESTest (*rho*=0.57; p<0.001) and Brief-BESTest (*rho*=0.61; p<0.001).

Ability to identify fall status

Thirty four participants (27.9%) had a history of falls and were significantly older than those without a history of falls (Table 1). The FTSTS and balance tests were able to significantly differentiate between participants with and without a history of falls ($p<0.007$) (Table 1). Large effect sizes were found for BBS (*Cohen's d* =1.00), BESTest (*Cohen's d* =0.81), Mini-BESTest (*Cohen's d* =1.02) and Brief-BESTest (*Cohen's d* =1.00).

Table 3 presents the results from the receiver operating characteristic analysis. Balance tests had acceptable ability to differentiate participants with and without a history of falls (AUC from 0.71 to 0.78). Sensitivity to differentiate fall status was similar across balance tests (71-74%). Specificity, however, was higher for the Brief-BESTest (76%), Mini-BESTest (71%) and BBS (72%) than for BESTest (67%). The identified cut-offs are 48.5 points for BBS, 82 points for BESTest, 19.5 points for Mini-BESTest and 12.5 points for Brief-BESTest (Table 3). Figure 1 illustrates the AUC as well as the identified cut-off points.

(table 3 and Figure 1)

Discussion

This study showed that the four balance tests presented good to excellent inter-rater and test-retest reliability, were highly correlated with each other (concurrent validity) and with the self-reported balance confidence (convergent validity), and were able to differentiate between older people living in the community with and without a history of falls with similar sensitivity/specificity. Therefore, findings support the use of all tests to assess balance impairment. However, similar to previous studies the BBS presented a substantial ceiling effect.[8,36] Thus it may be more appropriate for health professionals to use the BESTest and its short versions to assess balance in older people living in the community.

Although the reliability of the BBS has been widely studied,[5,8] much less is known about the BESTest and its short versions. This study showed that the BBS, BESTest, Mini-BESTest and Brief-BESTest have good to excellent interrater (ICC 0.71-0.93) and test-retest (ICC 0.50-0.82) relative reliability in older people living in the community. Similar findings have been reported for relative reliability of BESTest and its short versions in other studied populations.[13,16,19-21]

Absolute reliability has been much less studied than relative reliability despite the GRRAS guidelines[25] recommendations to study both. In this study, balance tests showed good agreement with no systematic bias between raters and sessions, with relatively small SEM. Although, the Bland and Altman method has not been commonly used for visual inspection of BESTest and its short versions, some studies have reported the SEM for BESTest,[16] Mini-BESTest[16,20] and Brief-BESTest.[16] The SEM is largely independent of the population from which it is determined however, its calculation depends of the standard deviation of the scores from all individuals studied.[48] In this study, the standard deviations of BESTest and its short versions were higher than those already reported when studying patients with balance disorders or older cancer survivors, thus higher values of SEM have been found (BESTest - 3.3 vs 1.75;[16] Mini-BESTest – 1.4 vs 1.22[16] or 1.26;[20] Brief-BESTest – 1.5

vs 1.13[16]). Nevertheless, these SEM differences might not be of clinical relevance considering the score range of the balance tests assessed and the achieved changes (considerably higher than the SEM values here reported) in studies conducting balance training.[20,52] Nevertheless, results should be interpreted with caution as absolute reliability has been little studied in BESTest and its short versions. Future studies exploring both types of reliability are needed.

In this study, acceptable values of $MDC_{\%}$ and MDC_{95} values were found. Higher MDC_{95} (6.5[53] and 6.2[20]) have been reported for BBS however, in this study, a sample of healthy older people were studied and a substantial ceiling effect for the BBS was found, which may explain the lower MDC_{95} found. For BESTest and its short versions no ceiling or floor effects were observed and all MDC_{95} values were higher than those previously published.[15,16,20,21] However, these studies were conducted with patients with specific health conditions, such as chronic stroke, total knee arthroplasty, balance disorders and history of cancer.[15,16,20,21] Therefore, it is not surprising that higher MDC_{95} values were found in a healthy older population. Therefore, the MDC_{95} established in this study will be useful for health professionals working with older people living in the community in determining whether an intervention (e.g., fall prevention program) has caused any real change in balance ability.

Similarly to other studies,[16,20-22] all balance tests were significantly correlated with each other and with the ABC scale, indicating good concurrent and convergent validity. All balance tests were also able to significantly differentiate between participants with and without a history of falls with large effect sizes, acceptable ability and similar sensitivity and specificity. The cut-off points identified were similar to those previously published for BBS,[36] BESTest,[35] Mini-BESTest[35,36] and Brief-BESTest[35]. These cut-off scores indicate the level of balance impairment that is associated with falling. This is especially important for health professionals to detect fall risk in older adults before a fall occurs and implement preventive effective interventions.

Study Limitations

The sample included older people living in the community, which limits the generalizability of the results to the overall older population. It is unclear whether the participation in community centers activities have contributed to minimize the risk of falling. Future studies could compare the balance impairment and risk of falling between older people integrating and not integrating community centers in order to clarify these issues. Older people living in the community were classified as having or not having a history of falls based on their self-report hence, some degree of bias in their responses might have been present. Moreover, tasks that were similar among BESTest, Mini-BESTest and Brief-BESTest were only performed once and therefore, this influenced the correlation results. Future studies should assess the reliability, validity and ability to identify fall status of the Mini-BESTest and of the Brief-BESTest when performed separately from the BESTest. Finally, as this was a cross-sectional study, the ability of the balance tests to identify fall status in older people living in the community was analyzed retrospectively. Longitudinal studies should be conducted in order to assess the ability of these tests in identifying recurrent fallers prospectively.

Conclusions

BBS, BESTest, Mini-BESTest and Brief-BESTest are highly reliable, valid and able to differentiate between older people living in the community with and without a history of falls. However, as BBS presented a substantial ceiling effect, BESTest and its short versions seem more appropriate to assess specific balance impairments in older people living in the community. MDC_{95} values and cut-off points for BBS, BESTest, Mini-BESTest and Brief-BESTest, were established. It is believed that these findings will help health professionals to assess balance impairments in older people living in the community and will also provide information on whom to prioritize intervention.

References

1. World Health Organization. Falls. 2012:Fact sheet N°344.
2. Tinetti M, Speechley M, Ginter S. Risk factors for falls among elderly persons living in the community. *N Engl J Med.* 1988;319:1701–1707.
3. National Institute for Health and Care Excellence. Falls in older people: assessing risk and prevention (CG161). 2013: https://www.nice.org.uk/guidance/cg161.
4. Delbaere K, Close JC, Heim J, Sachdev PS, et al. A multifactorial approach to understanding fall risk in older people. *J Am Geriatr Soc.* 2010;58(9):1679-1685.
5. Steffen T, Seney M. Test-Retest Reliability and Minimal Detectable Change on Balance and Ambulation Tests, the 36-Item Short-Form Health Survey, and the Unified Parkinson Disease Rating Scale in People With Parkinsonism. *Phys Ther.* 2008;88(6):733-746.
6. Oliveira CC, Lee A, Granger CL, Miller KJ, et al. Postural control and fear of falling assessment in people with chronic obstructive pulmonary disease: a systematic review of instruments, international classification of functioning, disability and health linkage, and measurement properties. *Arch Phys Med Rehabil.* 2013;94(9):1784-1799.e1787.
7. Wang CY, Hsieh CL, Olson SL, Wang CH, et al. Psychometric properties of the Berg Balance Scale in a community-dwelling elderly resident population in Taiwan. *J Formos Med Assoc.* 2006;105(12):992-1000.

8. Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. *Phys Ther.* 2008;88(5):559-566.

9. Tanji H, Gruber-Baldini AL, Anderson KE, Pretzer-Aboff I, et al. A comparative study of physical performance measures in Parkinson's disease. *Mov Disord.* 2008;23(13):1897-1905.

10. Beauchamp MK, O'Hoski S, Goldstein RS, Brooks D. Effect of Pulmonary Rehabilitation on Balance in Persons With Chronic Obstructive Pulmonary Disease. *American Congress of Rehabilitation Medicine.* 2010;91:1460-1465.

11. Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? *Age Ageing.* 2006;35 Suppl 2:ii7-ii11.

12. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to Differentiate Balance Deficits. *Phys Ther.* 2009;89(5):484-498.

13. Leddy AL, Crowner BE, Earhart GM. Functional gait assessment and balance evaluation system test: reliability, validity, sensitivity, and specificity for identifying individuals with Parkinson disease who fall. *Phys Ther.* 2011;91(1):102-113.

14. Beauchamp MK, Sibley KM, Lakhani B, Romano J, et al. Impairments in systems underlying control of balance in COPD. *Chest.* 2012;141(6):1496-1503.

15. Chan AC, Pang MY. Assessing Balance Function in Patients With Total Knee Arthroplasty. *Phys Ther.* 2015;[Epub ahead of print].

16. Huang MH, Miller K, Smith K, Fredrickson K, et al. 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.* 2015;[Epub ahead of print].

17. O'Hoski S, Winship B, Herridge L, Agha T, et al. Increasing the clinical utility of the BESTest, mini-BESTest, and brief-BESTest: normative values in Canadian adults who are healthy and aged 50 years or older. *Phys Ther.* 2014;94(3):334-342.

18. Franchignoni F, Horak F, Godi M, Nardone A, et al. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. *J Rehabil Med.* 2010;42(4):323-331.

19. Padgett PK, Jacobs JV, Kasser SL. Is the BESTest at its best? A suggested brief version based on interrater reliability, validity, internal consistency, and theoretical construct. *Phys Ther.* 2012;92(9):1197-1207.

20. Godi M, Franchignoni F, Caligari M, Giordano A, et al. Comparison of reliability, validity, and responsiveness of the mini-BESTest and Berg Balance Scale in patients with balance disorders. *Phys Ther.* 2013;93(2):158-167.

21. Tsang CS, Liao LR, Chung RC, Pang MY. Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. *Phys Ther.* 2013;93(8):1102-1115.

22. O'Hoski S, Sibley KM, Brooks D, Beauchamp MK. Construct validity of the BESTest, mini-BESTest and briefBESTest in adults aged 50 years and older. *Gait Posture.* 2015;42(3):301-305.

23. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21(4):651-657.

24. von Elm E, Altman DG, Egger M, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ.* 2007;335(7624):806-808.

25. Kottner J, Audigé L, Brorson S, Donner A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96-106.

26. Biderman A, Cwikel J, Fried A, Galinsky D. Depression and falls among community dwelling elderly people: a search for common risk factors. *J Epidemiol Community Health.* 2002;56(8):631-636.

27. Whitney SL, Poole JL, Cass SP. A review of balance instruments for older adults. *Am J Occup Ther.* 1998;52(8):666-671.

28. Myers AM, Fletcher PC, Myers AH, Sherk W. Discriminative and evaluative properties of the activities-specific balance confidence (ABC) scale. *J Gerontol A Biol Sci Med Sci.* 1998;53(4):M287-294.

29. Branco PS. Validation of the Portuguese Version of the “Activities-specific Balance Confidence Scale”. *RSPMFR.* 2010;19(2):20-25.

30. Schlenstedt C, Brombacher S, Hartwigsen G, Weisser B, et al. Comparing the Fullerton Advanced Balance Scale with the Mini-BESTest and Berg Balance Scale to assess postural control in patients with Parkinson disease. *Arch Phys Med Rehabil.* 2015;96(2):218-225.

31. 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(2):90-97.

32. Holbein-Jenny MA, Billek-Sawhney B, Beckman E, Smith T. Balance in personal care home residents: a comparison of the Berg Balance Scale, the

Multi-Directional Reach Test, and the Activities-Specific Balance Confidence Scale. *J Geriatr Phys Ther.* 2005;28(2):48-53.

33. Mao HF, Hsueh IP, Tang PF, Sheu CF, et al. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke.* 2002;33(4):1022-1027.

34. Flansbjer UB, Blom J, Brogardh C. The reproducibility of Berg Balance Scale and the Single-leg Stance in chronic stroke and the relationship between the two tests. *PM R.* 2012;4(3):165-170.

35. Duncan RP, Leddy AL, Cavanaugh JT, Dibble LE, et al. Comparative utility of the BESTest, mini-BESTest, and brief-BESTest for predicting falls in individuals with Parkinson disease: a cohort study. *Phys Ther.* 2013;93(4):542-550.

36. King LA, Priest KC, Salarian A, Pierce D, et al. Comparing the Mini-BESTest with the Berg Balance Scale to Evaluate Balance Disorders in Parkinson's Disease. *Parkinsons Dis.* 2012;2012:375419.

37. Rehabilitation Measures Database. http://www.rehabmeasures.org/default.aspx. Accessed April 20th, 2016.

38. Goldberg A, Chavis M, Watkins J, Wilson T. The five-times-sit-to-stand test: validity, reliability and detectable change in older females. *Aging Clin Exp Res.* 2012;24(4):339-344.

39. Teo TW, Mong Y, Ng SS. The repetitive Five-Times-Sit-To-Stand test: its reliability in older adults. *International Journal of Therapy and Rehabilitation.* 2013;20(3):122-130.

40. Troiano RP, Berrigan D, Dodd KW, Masse LC, et al. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40(1):181-188.

41. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39(2):175-191.

42. Cohen J. *Statistical power analysis for the behavioral sciences.* 2 ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.

43. Bulmer MG. *Principles of statistics.* New York: Dover Publications; 1979.

44. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med.* 2002;21(9):1331-1335.

45. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil.* 1998;12(3):187-199.

46. Fleiss J. Reliability of measurements. In: Fleiss J, ed. *The design and analysis of clinical experiments*. 1 ed. New York: John Wiley & Sons; 1986.

47. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil.* 1998;12:187-199.

48. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005;19(1):231-240.

49. Huang S-LH, Ching-Lin; Wu, Ruey-Meei; Tai, Chun-Hwei; Lin, Chin-Hsien; Lu, Wen-Shian. Minimal Detectable Change of the Timed "Up & Go" Test and the Dynamic Gait Index in People With Parkinson Disease. *Phys Ther.* 2011;91(1):114-121.

50. Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, et al. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehabil.* 2002;83(8):1145-1150.

51. Cohen HS, Kimball KT. Usefulness of some current balance tests for identifying individuals with disequilibrium due to vestibular impairments. *J Vestib Res.* 2008;18(5-6):295-303.

52. Beauchamp MK, Janaudis-Ferreira T, Parreira V, Romano JM, et al. A randomized controlled trial of balance training during pulmonary rehabilitation for individuals with COPD. *Chest.* 2013;144(6):1803-1810.

53. Romero S, Bishop MD, Velozo CA, Light K. Minimum detectable change of the Berg Balance Scale and Dynamic Gait Index in older persons at risk for falling. *J Geriatr Phys Ther.* 2011;34(3):131-137.

Tables captions

Table 1 - Participants' characteristics and balance scores (n=122).

Characteristics	**Total (n=122)**	**Without a history of falls (n=88)**	**With a history of falls (n=34)**	**p-value**
Age (years)	75.9 ± 8.9	74.7 ± 8.8	78.9 ± 8.6	0.020
Gender				
Female	86 (70.5%)	58 (65.9%)	28 (82.4%)	0.081
Male	36 (29.5%)	30 (34.1%)	6 (17.6%)	
BMI (Kg/m^2)	26.8 ± 4	26.9 ± 4	26.6 ± 4.2	0.715
Walking aid	13 (10.7%)	7 (8%)	6 (17.6%)	0.187
Comorbidities, M[IQR]	2 [1, 3]	1.5 [1, 3]	2 [1, 3]	0.573
ABC scale	80.9 ± 22.1	86.6 ± 17.7	66.1 ± 25.4	<0.001
FTSTS (seconds)	12.9 ± 5	12.2 ± 4.4	15 ± 6.1	0.007
BBS	47.2 ± 10	49.9 ± 8.3	40.2 ± 10.9	<0.001
BESTest	84.7 ± 24.5	84 ± 13.9	71.7 ± 16.5	<0.001
Mini-BESTest	19.8 ± 6.8	21.6 ± 5.8	15 ± 7.1	<0.001

Table 2 – Interrater and test-retest reliability of the Berg Balance Scale (BBS), Balance Evaluation Systems Test (BESTest), Mini-BESTest and Brief-BESTest (n=28).

	Interrater reliability			**Test-retest reliability**		
Balance test	**ICC (95% CI)**	**MD±SD**	**95% LA**	**ICC (95% CI)**	**MD±SD**	**95% LA**
BBS	0.88 (0.77→0.94)	-0.1 ± 1.1	-2.2→2	0.50 (0.15→0.73)	-0.7 ± 2	-4.6→3.2
BESTest	0.86 (0.73→0.92)	0.4 ± 3.5	-6.5→7.2	0.77 (0.55→0.88)	-1.6 ± 4.3	-10.1→6.9
Mini-BESTest	0.71 (0.50→0.84)	-0.2 ± 2	-4.2→3.7	0.73 (0.49→0.86)	-0.4 ± 1.9	-4.2→3.3
Brief-BESTest	0.93 (0.87→0.97)	-0.2 ± 1.2	-2.5→2.1	0.82 (0.65→0.91)	-0.4 ± 2	-4.4→3.5

Abbreviations: 95% CI, 95% confidence intervals; ICC, Intraclass Correlation Coefficient; 95% LA, 95% limits of agreement; MD, mean difference; SD, standard deviation.

Table 3 – Ability of the Berg Balance Scale (BBS), Balance Evaluation Systems Test (BESTest), Mini-BESTest and Brief-BESTest to identify fall status (n=122).

Balance test	**AUC (SEM)**	**95% CI**	**Cut-off point**	**% Sensitivity / % Specificity**	**Positive / Negative Likelihood Ratios**
BBS	0.78 (0.05)	0.68→0.87	48.5	74 / 72	2.59 / 0.37
BESTest	0.71 (0.06)	0.60→0.82	82	74 / 67	2.23 / 0.40
Mini-BESTest	0.76 (0.05)	0.66→0.86	19.5	74 / 71	2.49 / 0.38
Brief-BESTest	0.76 (0.05)	0.66→0.86	12.5	71 / 76	2.92 / 0.39

Abbreviations: AUC, area under the curve; 95%CI, 95% confidence intervals; SEM, standard error.

Figure captions

Figure 1 - Receiver operator characteristics of the Berg Balance Scale (BBS), Balance Evaluation Systems Test (BESTest), Mini-BESTest and the Brief-BESTest to differentiate participants with and without a history of falls (n=122).