

Physiotherapy Theory and Practice

An International Journal of Physical Therapy


ISSN: 0959-3985 (Print) 1532-5040 (Online) Journal homepage: <http://www.tandfonline.com/loi/iptp20>

Reliability, validity and minimal detectable change of the Mini-BESTest in Greek participants with chronic stroke

Sofia I. Lampropoulou, Evdokia Billis, Ingrid A. Gedikoglou, Christina Michailidou, Alexander V. Nowicky, Dimitra Skrinou, Fotini Michailidi, Danae Chandrinou & Margarita Meligkoni

To cite this article: Sofia I. Lampropoulou, Evdokia Billis, Ingrid A. Gedikoglou, Christina Michailidou, Alexander V. Nowicky, Dimitra Skrinou, Fotini Michailidi, Danae Chandrinou & Margarita Meligkoni (2018): Reliability, validity and minimal detectable change of the Mini-BESTest in Greek participants with chronic stroke, *Physiotherapy Theory and Practice*, DOI: [10.1080/09593985.2018.1441931](https://doi.org/10.1080/09593985.2018.1441931)

To link to this article: <https://doi.org/10.1080/09593985.2018.1441931>

 Published online: 23 Feb 2018.

 Submit your article to this journal [↗](#)

 Article views: 4

 View related articles [↗](#)

 View Crossmark data [↗](#)

REPORT



Reliability, validity and minimal detectable change of the Mini-BESTest in Greek participants with chronic stroke

Sofia I. Lampropoulou, PT, MSc, PhD^{a,b}, Evdokia Billis, PT, MSc, PhD^a, Ingrid A. Gedikoglou, PT, MSc^c, Christina Michailidou, PT, MSc, PhD^d, Alexander V. Nowicky, BA, PhD^e, Dimitra Skrinou, PT^a, Fotini Michailidi, PT^a, Danae Chandrinou, PT^a, and Margarita Meligkoni, PT^a

^aDepartment of Physical Therapy, School of Health and Welfare Professions, TEI of Western Greece, Aigio, Greece; ^bPhysiotherapy Program, Department of Life and Health Sciences, Nicosia University, Nicosia, USA; ^cPhysio Point, Athens, Greece; ^dChronic Fatigue Research and Treatment Unit, South London and Maudsley NHS Trust, London, UK; ^eDepartment of Clinical Sciences, College of Health and Life Sciences, Brunel University London, Uxbridge, UK

ABSTRACT

Objectives: This study aimed to investigate the psychometric characteristics of reliability, validity and ability to detect change of a newly developed balance assessment tool, the Mini-BESTest, in Greek patients with stroke. **Design:** A prospective, observational design study with test-retest measures was conducted. **Methods:** A convenience sample of 21 Greek patients with chronic stroke (14 male, 7 female; age of 63 ± 16 years) was recruited. Two independent examiners administered the scale, for the inter-rater reliability, twice within 10 days for the test-retest reliability. Bland Altman Analysis for repeated measures assessed the absolute reliability and the Standard Error of Measurement (SEM) and the Minimum Detectable Change at 95% confidence interval ($MDC_{95\%}$) were established. The Greek Mini-BESTest (Mini-BESTest_{GR}) was correlated with the Greek Berg Balance Scale (BBS_{GR}) for assessing the concurrent validity and with the Timed Up and Go (TUG), the Functional Reach Test (FRT) and the Greek Falls Efficacy Scale-International (FES-I_{GR}) for the convergent validity. **Results:** The Mini-BESTest_{GR} demonstrated excellent inter-rater reliability (ICC (95%CI) = 0.997 (0.995–0.999, SEM = 0.46) with the scores of two raters within the limits of agreement ($mean_{dif} = -0.143 \pm 0.727, p > 0.05$) and test-retest reliability (ICC (95% CI) = 0.966 (0.926–0.988), SEM = 1.53). Additionally, the Mini-BESTest_{GR} yielded very strong to moderate correlations with BBS_{GR} ($r = 0.924, p < 0.001$), TUG ($r = -0.823, p < 0.001$), FES-I_{GR} ($r = -0.734, p < 0.001$) and FRT ($r = 0.689, p < 0.001$). MDC_{95} was 4.25 points. **Conclusion:** The exceptionally high reliability and the equally good validity of the Mini-BESTest_{GR}, strongly support its utility in Greek people with chronic stroke. Its ability to identify clinically meaningful changes and falls risk need further investigation.

ARTICLE HISTORY

Received 28 July 2016
Revised 18 May 2017
Accepted 18 July 2017

KEYWORDS

Balance; Mini-BESTest; reliability; validity; stroke; Greek

Introduction

Balance is an important functional ability to maintain static and dynamic posture safely during activities of everyday life and it is commonly impaired in people with neurological disorders, such as stroke (Mancini and Horak, 2010). The clinical evaluation of balance is, therefore, of great importance in identifying risks of falling as well as in organizing appropriate rehabilitation protocols for those patients (Duncan et al., 2015). The literature identifies various ways of balance assessment either via observation of the patient's balance performance or via patients' self-reports in regards their balance confidence, or their fear of falling (Yelnik and Bonan, 2008). The Activities-Specific Balance Confidence Scale (ABC) (Botner, Miller, and

Eng, 2005), and the Falls Efficacy Scale (Yardley et al., 2005) are examples of the most widely used self-reported measures. The Short Physical Performance Battery (SPPB) of the lower limbs for the elderly, via examination of strength, balance and mobility as well as the Dynamic Gait Index (DGI) are examples of high reliability and validity tests of measurements based on performance (Fisher et al., 2009; Jonsdottir and Cattaneo, 2007; Shumway-Cook et al., 2013). In the same category, the Performance Oriented Mobility Assessment (POMA) is a widely used tool for mobility examination and falls prediction, especially in the elderly (Faber, Bosscher, and Van Wieringen, 2006). The Berg Balance Scale (BBS) constitutes one of the very well-characterized and established clinical

CONTACT Sofia I. Lampropoulou ✉ lampropoulou.s@unic.ac.cy 📧 Department of Physical Therapy, School of Health and Welfare Professions, TEI of Western Greece, Aigio, Greece.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/iptp.

assessment measures for balance in the elderly and patients with balance deficiencies due to various neurological conditions (e.g. stroke, multiple sclerosis and Parkinson's disease) (La Porta et al., 2012; Toomey and Coote, 2013). Isolated functional balance tests include the Romberg Test, which has the advantage of differentiating between lesions in the central and peripheral nervous system, the Functional Reach Test and the Timed Up and Go test, which have all been used in combination with several functional scales in clinical and research settings (Barry et al., 2014; Garcia et al., 2012; Tyson and Connell, 2009; Yelnik and Bonan, 2008).

In order to identify the underlying postural control systems responsible for the balance deficits, a new balance tool, the Balance Evaluation Systems Test (BESTest) was developed (Horak, Wrisley, and Frank, 2009). However, prolonged time needed for administration (from 30 to 45 min) has limited its utility in clinical practice. Thus, an abbreviated version of the BESTest, the mini-Balance Evaluation Systems Test (Mini-BESTest), that can be administered in only 15 min, has recently been developed with the main focus being on 'dynamic balance' (Franchignoni et al., 2010). The Mini-BESTest is the only measure among 26 common balance measures for community-dwelling people that incorporates almost all contents of balance, such as static and dynamic body stability, transfers, gait, variation of support surfaces, variation of visual conditions, obstacle negotiation, external forces and dual-tasking (Di Carlo et al., 2016; Franchignoni et al., 2010; Pardasanev et al., 2013). The Mini-BESTest appears to gain ground in neurological assessment as it reveals a lower ceiling effect compared to BBS (Godi et al., 2013; King et al., 2012); robust psychometric properties especially for Parkinson and stroke populations (Dahl and Jørgensen, 2014; Duncan et al., 2013; King et al., 2012; Leddy, Crouner, and Earhart, 2011; Mak and Auyeung, 2013; Tsang, Liao, Chung, and Pang, 2013) and a broad range of activities along the entire spectrum of task and environment complexity (Pardasanev et al., 2013). A further advantage of the Mini-BESTest is that its total score can now be transformed from ordinal summed raw scores into internal-level measurements using a specific nomogram, created recently by Franchignoni et al. (2015).

Being characterized as one of the most comprehensive balance measures for community-dwelling adults and elderly individuals (O'Hoski, Sibley, Brooks, and Beauchamp, 2015; Pardasanev et al., 2013); together with its excellent reliability (Godi et al., 2013; Leddy, Crouner, and Earhart, 2011; Padgett, Jacobs, and Kasser, 2012; Tsang, Liao, Chung, and Pang, 2013) and validity properties (Bergström, Lenholm, and Franzen,

2012; Franchignoni et al., 2010; King et al., 2012; Tsang, Liao, Chung, and Pang, 2013), the Mini-BESTest could be one of the preferred means of international balance assessments for stroke patients who present balance impairments in any of the following domains: static and dynamic body stability; transfers and gait. Its utility into the Greek clinical environment is now achievable as an official adaptation of the Mini-BESTest into Greek has been performed (Lampropoulou et al., 2016a). The Greek version of the scale (Mini-BESTest_{GR}) can now be found on the official BESTest website (www.bestest.us). However, a thorough evaluation of the reliability and validity of the scale on specific clinical population is needed in order to further establish psychometric characteristics for the scale, such as the minimum detectable change, the distribution of the scores and the floor and ceiling effects for this specific disease in Greek population. Stroke was chosen because of its increased incidence internationally and its risk that roughly doubles with each decade of age during adulthood (Hirtz et al., 2007). Stroke leads to chronic balance impairments and functional, sociological, economical and psychological negative effects that necessitate the health care professionals' attendance (Ma, Chang, and Carruthers, 2014).

The aim of this study, therefore, is to evaluate the psychometric characteristics of the Mini-BESTest_{GR} on Greek neurological patients who have suffered a stroke. Specifically, the objectives of the study are: the investigation of the distribution of the scores, the internal consistency, the test-retest and inter-rater reliability, the concurrent and convergent validity and the ability to detect change of the Greek version of the Mini-BESTest on patients with chronic stroke. This will further establish Mini-BESTest_{GR} as a valid means for balance assessment on specific population.

Methodology

Participants

Participants consisted of a convenience sample of Greek adult (18 or older years old) neurological patients who had stroke for more than 6 months (chronic stage). All the patients had adequate cognitive status (at least 24 at the Mini Mental State Examination) (Chinsongkram et al., 2014; Solias et al., 2014), in order to ensure good comprehension of the instructions provided by the assessor. The exclusion criteria were pregnancy, dementia or other cognitive impairment and recent surgery to the lower limbs that could restrict the performance of the standing and walking activities included in the scale. The patients had been informed about the study and had signed an

informed consent form prior to their participation. The study was approved by ethics review board of the Scientific Committee of the Technological Educational Institute (TEI) of Western Greece.

Materials and assessment tools

The diagnostic tools that were used in this study were: Greek version of the Mini-BESTest (Mini-BESTest_{GR}) (Lampropoulou et al., 2016a); Greek version of the Berg Balance Scale (BBS_{GR}) (Lampropoulou et al., 2016a); Greek version of the international Falls Efficacy Scale International (FES-I_{GR}) (Billis et al., 2011); Functional Reach Test (FRT) and the Timed Up and Go test (TUG).

Mini-balance evaluation systems test (Mini-BESTest)

The Mini-BESTest scale consists of 14 tasks – elements of the original BESTest, that represent four systems of balance control; the anticipatory postural adjustments, the compensatory postural adjustments, the sensory orientation and gait stability (Mancini and Horak, 2010). Some of the tasks include sit to stand, standing on toes, single leg stance, compensatory reactions forwards, backwards and sideways, tasks with the eyes closed, on a foamy surface, on an inclined surface and tasks involving gait with change of speed, head rotations, walking over obstacles and timing (Franchignoni et al., 2010). Each item is rated on a three-point ordinal scale, from 0 to 2, yielding to a total maximum score of 28 points (King and Horak, 2013).

Berg Balance Scale

This scale consists also of 14 balance tasks, like sit to stand and back, standing, transfers, reaching forwards, rotations and single leg stance. The grading varies from 0 to 4 for each task and the maximum total score is 56 (Berg, Wood-Dauphinee, and Williams, 1995). BBS was chosen because similarly to Mini-BESTest assesses balance through functional tasks, is easy to be performed, has been validated in Greek population yielding excellent inter-rater reliability (ICC = 0.998) and test-retest reliability (ICC = 0.976) as well as good validity properties (Lampropoulou et al., 2016a) and constitutes the best measurement outcome to evaluate the concurrent validity of the Mini-BESTest.

Falls Efficacy Scale - Internationa

The FES-I, is a questionnaire with 16 questions that assesses the risk of falling during daily life activities, at home and outdoors. The questions are graded from 1 to 4, where 1 indicates that the patient has no concerns about falling, and 4 indicates high patient concerns

about the possibility of a fall. Its excellent psychometric characteristics with very good validity, excellent test-retest reliability (ICC = 0.951) and internal consistency (Cronbach' alpha = 0.925), its large effect size (0.89) in distinguishing across fallers and non-fallers and its availability in Greek language (Billis et al., 2011) make it a suitable tool for convergent validity analysis.

Timed Up and Go test

This single balance test measures the time that a patient needs to stand up from a chair, walk a 3-m distance, come back and sit back on the chair (Bohannon, 2006). It is included in the Mini-BESTest and it has showed good reliability and validity properties in individuals with stroke (Knorr, Brouwer, and Garland, 2010).

Functional Reach Test

The FRT assesses balance by measuring the limits while the patient reaches forwards as far as possible, having the arms in 90° flexion and without lifting the heels off the floor (Yelnik and Bonan, 2008). It has yielded excellent reliability properties (ICC = 0.99) and high correlation to BBS (r = 0.92) proving its convergent validity in stroke population (Wolf et al., 1999).

Equipment

The equipment that was used for the application of the Mini-BESTest was: a foamy material (Temper[®] foam, also known as T-foam™, 10 cm. thick, medium density, density scale T41); a chair without armrests or wheels, a step of average height, a 10-degree incline ramp (at least 60 cm x 100 cm) to stand on, stopwatch, a box (23 cm height) and a 3 meter distance measured out and marked on the floor with tape (from chair). For the BBS application, a chair with armrests and a chair without armrests, a timer, a step of average height, and a ruler of 5, 12 and 25 cm were used.

Procedure

The assessment took place at participants' home; first obtaining a short case history and basic demographics documentation followed by the application of the Mini-BESTest_{GR} and the BBS_{GR}. Participants had been advised to wear comfortable clothing and footwear. The room was quiet, so as to minimize any distractions to the participants' performance. Analytical information and demonstration by the assessor was offered for each activity prior to testing. After the completion of each scale assessment, there was a 10-min break before the commencement of the next scale assessment. Raters were newly qualified

physiotherapists who had received specific training for at least 2 months prior to physically administering the scale. The training consisted of watching videos and studying materials and instructions on performing and grading the scale, provided by the official site of the BEST scale (www.bestest.us) and from Horak, Wrisley, and Frank (2009) and King and Horak (2013). They also received 1-day observational training of the use of the scale on patients as well as practical training by performing the scale on healthy co-workers under the supervision of the principal investigator of this research (SL).

Distribution of the scores, ceiling and floor effects

Ceiling and floor effects as well as skewness of the Mini-BESTest_{GR} were examined to assure that the sample distribution of scores would not affect the ability of the scale to detect changes in performance, and hence limit sensitivity (Pardasaney et al., 2012).

Reliability assessment

Reliability is defined as the stability of a tool in presenting the same result through repeating a series of measurements (Bowling, 2002; Ouzouni and Nakakis, 2011). To assess the test-retest reliability (in repetitive measurements), an initial assessment and a re-assessment, approximately at the same time and under the same conditions took place in a time frame of 7 to 10 days as similar research had followed (Godi et al., 2013; Billis et al., 2011). The inter-rater reliability was used to assess the agreement between the grades given by the two raters (Ouzouni and Nakakis, 2011). The absolute reliability was assessed via Bland Altman plot of the mean differences between test-retest measurements (Myles and Cui, 2007). The internal consistency was also evaluated to demonstrate how the different items of the Greek version of the Mini-BESTest scale measured the same thing in accordance with similar studies for ordinal data (Godi et al., 2013; Löfgren et al., 2014; Tsang, Liao, Chung, and Pang, 2013). In addition, the Minimum Detectable Change at 95% of confidence interval (MDC_{95%}) and the Standard Error of Measurement (SEM) based on the reliability of the scale and the standard deviation of the population were tested to evaluate the smallest change in score that reflects a true change in the balance ability of the patient rather than an error in measurement (De Vet et al., 2006).

Validity assessment

The validity of a measurement tool relates to the tool's capability to measure the variable that it was initially designed to measure (Golafshani, 2003). For the purposes of the present study the concurrent validity (part of the criterion validity) and the convergent validity (part of the construct validity) were used. Aiming to assess the *concurrent validity*, which expresses the grade of agreement of the results of two similar measurement tools, measuring the same concept, after having been used in the same research and at the same time period (Bowling, 2002; Ouzouni and Nakakis, 2011; Scholtes, Terwee, and Poolman, 2011), the Mini-BESTest_{GR} was assessed simultaneously to the established balance measure of BBS_{GR}. Aiming to assess the *convergent validity* referring to association of measurement tools which assess related variables, the Mini-BESTest_{GR} was correlated to clinical tests of balance performance (TUG and FRT) and balance confidence (FES-I_{GR}). FRT and TUG have shown to have moderate to high correlation with Mini-BESTest (Bergström, Lenholm, and Franzen, 2012; Tsang, Liao, Chung, and Pang, 2013), thus a moderate to strong correlation was expected when the Mini-BESTest_{GR} was correlated with FRT and TUG, accordingly. FES-I, being correlated with Mini-BESTest, presented a low correlation in patients with Parkinson's disease (Bergström, Lenholm, and Franzen, 2012) but this has not been tested on people with stroke.

Sample size calculations and data analysis

Sample size calculations

Based on sample sizes suggested by Bonnet (2002) for reliability analysis between two sets of data (i.e. two judges or two measurements in time), a sample of 15 patients is adequate at alpha level of 0.05 and correlation coefficient above 0.9. The studies of Tsang, Liao, Chung, and Pang (2013) and Dahl and Jørgensen (2014) have shown intra-class correlation coefficient above 0.94 for intra- and inter-rater reliability and thus, a correlation coefficient at least 0.9 was expected in the present study. Furthermore, sample size calculations for validity analysis were based on alpha level of 0.05 (2-tailed) and a power of 0.8 (Faul, Buchner, Erdfelder, and Lang, 2014). The study of Tsang, Liao, Chung, and Pang (2013) presented a very strong correlation of Mini-BESTest with BBS ($\rho = 0.83$) and TUG ($\rho = -0.82$) in patients with stroke, thus a large effect size was expected for the concurrent and

convergent validity. Using the value of a large effect size of $r = 0.6$ for the calculation of the sample size, the number of 19 patients for the validity analysis was yielded.

Data presentation

For the quantitative analysis of the data, non-parametric tests were used due to data not normally distributed (Kolmogorov-Smirnov analysis). Data are presented as mean \pm standard deviation (mean \pm SD), or as otherwise stated and the level of significance was set at 0.05 ($p \leq 0.05$). Statistical analysis was completed using SPSS statistical software (Statistical Package for the Social Sciences) (version 24.0, for Windows).

Distribution of the scores

Ceiling and floor effects were determined as more than 20% of the participants at the highest and lowest score, accordingly. Skewness of scores distribution, as further estimator of ceiling and floor effect, was presented at total scores.

Reliability analysis

The Internal Consistency of the Mini-BESTest_{GR} was assessed by use of the coefficient alpha for ordinal data. Due to ordinal response scales used in the present study a polychoric correlation matrix was used to calculate the ordinal alpha (Gadermann, Guhn, and Zumbo, 2012). The formula used for its calculation based on a 1-factor model (Gadermann, Guhn, and Zumbo, 2012) was:

$$\alpha = \frac{[k/(k-1)] * [(k * (\lambda_{\text{average}}^2 - h_{\text{average}}^2)) / (k * (\lambda_{\text{average}}^2 - u_{\text{average}}^2))]}{1}$$

where α : coefficient alpha for ordinal data, k : the number of items of the scale, λ : the factor loadings from the polychoric correlation matrix analysis, h^2 : Communality (for a 1-factor model $h^2 = \lambda^2$), u^2 : Uniqueness ($u^2 = 1 - h^2$). The difference between this coefficient and the conventionally used reliability coefficients, such as Cronbach's alpha, is that Cronbach's alpha coefficient is calculated using a Pearson correlation matrix. However, in that way its important assumption for its utility when data are continuous, is violated and the Pearson covariance matrix can be substantively distorted (Gadermann, Guhn, and Zumbo, 2012). Internal consistency was considered as acceptable at values of coefficient alpha above 0.7, at values between 0.7–0.8 as good and at values above 0.8 as excellent (Munro, 2005).

Relative reliability of the Mini-BESTest_{GR} total score was assessed with the Intraclass Correlation Coefficient

(ICC) between raters and between repeated measures, and values over 0.75 defined as excellent, 0.40–0.75 as moderate and below 0.4 as poor reliability (Roach, 2006; Toomey and Coote, 2013).

The Bland Altman Analysis for absolute reliability of total scores was also used to plot the differences between the two measurements against the means for each subject and to show the 'bias' (mean difference) of the measurements and the 95% limits of agreement (LoA) (Atkinson and Nevill, 1998; Myles and Cui, 2007). One Sample t-test for the differences was used to find whether these measurements differed significantly from 0.

Single Mini-BESTest_{GR} item scores agreement between raters and between test-retest measurements was calculated with the Weighted Kappa coefficient, which gives the agreement between pairs of ratings on an ordinal scale (Hallgren, 2012). Kappa statistics is interpreted as: 'poor agreement' for values < 0.00 ; 'slight agreement' for values 0.00–0.20; 'fair agreement' for values 0.21–0.40; 'moderate agreement' for values 0.41–0.60; 'substantial agreement' for values 0.61–0.80; and 'almost perfect' or perfect agreement for values 0.81–1.00 (Landis and Koch, 1977).

The item-total correlation was also assessed for Mini-BESTest_{GR} single items to check their consistency with the averaged behaviour of the others.

Validity analysis

The Spearman's correlation law (r_{ho}) was used for the correlation of Mini-BESTest_{GR} to other balance scales (BBS_{GR}) (concurrent validity) and other tests that supposedly are related to balance function (TUG, FRT and FES-I_{GR}) (convergent validity). The values for the correlation coefficient (r) are classified with correlation values of: 0.00–0.25 (no or minimal correlation); of 0.26–0.49 (poor correlation); of 0.50–0.69 (moderate correlation); 0.70–0.89 (strong correlation) and of 0.90–1.00 (very strong correlation) (Munro, 2005).

The MDC_{95%}

The MDC_{95%} was computed according to formula: $MDC_{95\%} = 1.96 * SEM * \sqrt{2}$. The SEM of the Mini-BESTest_{GR} total score was calculated according to formula $SEM = SD \sqrt{(1-ICC)}$, where ICC was the coefficient of the test-retest reliability and SD the standard deviation of the Mini-BESTest_{GR} total score (Godi et al., 2013; Portney and Watkins, 2009; Tsang, Liao, Chung, and Pang, 2013).

Results

Twenty-one Greek patients with chronic stroke participated in this study (14 men, 7 women), aged 63 ± 16 years old (range 27–88 years) during the period January 2013 to December 2014. Twenty-nine percent ($n = 6$) of the participants had experienced one fall during the last year and were identified as ‘fallers’ and 71% ($n = 15$) had no fall (‘non-fallers’). Mean Mini-BESTest_{GR} total score in fallers was 10 ± 5 and in non-fallers 18 ± 8 (mean difference = 8.0 ± 3.7 (SE), $t_{19} = -2.10$, $p = 0.049$).

Distribution of the scores, ceiling and floor effects

The distribution of the Mini-BESTest_{GR} total scores had a negative skewness (-0.427) but without presenting any ceiling or floor effect as only 10% of the participants (2/21) scored the best score (27/28) and only 5% (1/21) showed the lowest possible score (0/28) (Table 1). Comparisons with the other balance scales in regards the ceiling and floor effects are presented in Table 1.

Reliability

The Mini-BESTest_{GR} revealed excellent internal consistency (polychoric ordinal $\alpha = 0.942$. In addition, the Mini-BESTest_{GR} total score demonstrated excellent inter-rater reliability (ICC (95%CI) = 0.997 (0.995 – 0.999), SEM = 0.46) and test-retest reliability (ICC (95%CI) = 0.966 (0.926 – 0.988), SEM = 1.53). Reliability of the other balance tests in correspondence with Mini-BESTest_{GR} is showed in Table 1. Bland Altman analysis for the mean differences between repeated measurements is shown in Figure 1 (mean difference of the Mini-BESTest_{GR} total score between the two raters was -0.143 ± 0.727) and presented a

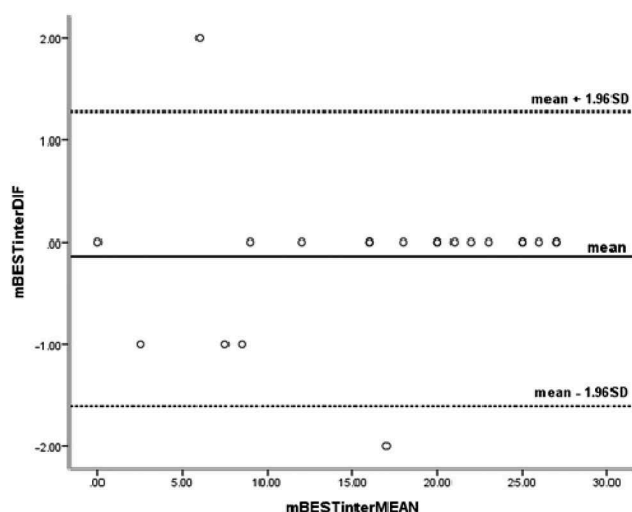


Figure 1. Bland-Altman analysis for graphical representation of differences between raters. Solid line in the middle represents the mean difference value of the sample ($n = 21$) between repeated measurements (mean difference = -0.143 ± 0.73). The dashed lines represent the upper and lower limits of agreement between the two raters (mean $\pm 1.96SD$). Points at values 0–16, 0–20, 0–25 and 0–27 of the mean to difference correlations represent two measurements (patients) each, due to the same value (17 dots are presented in the graph). Abbreviations: mBESTinterDif: the difference of the Mini-BESTest_{GR} total scores between the two raters for every patient; mBESTinterMEAN: the mean of the Mini-BESTest_{GR} total scores between the two raters for every patient.

distribution of the differences within the limits of agreement and without statistically significant difference from 0 ($t_{20} = -0.9$, $p > 0.05$).

Weighted Kappa values for each item between raters ranged from the lowest value of 0.85 for item 11 (walk with head turns) to highest 1.00 for items 6 (compensatory stepping correction-lateral), 7 (stance-firm surface), 8 (stance-eyes closed, foam), 12 (walk with pivot turns), 14 (timed up & go-dual task) (Table 2),

Table 1. Comparison of the Greek Mini-Balance Evaluation Systems Test (Mini-BESTest_{GR}) with other tests from $n = 21$.

Measurement outcome (score)	Mini-BESTest _{GR} (0–28)	BBS _{GR} (0–56)	TUG (sec)	FRT (cm)	FES-I _{GR} (16–64)
Mean score \pm SD	16 ± 8	45 ± 12	21 ± 16	21 ± 7	33 ± 14
Floor effect (% of participants with lowest score) (n of patients at min score)	4.8% (1 at 0/28)	4.8% (1 at 8/56)	–*	4.8% (1 at 7cm)	14.3% (3 at 16/64)
Ceiling effect (% of participants with highest score) (n of patients at max score)	9.5% (2) at 27/28	4.8% (1) at 56/56	–*	–*	4.8% (1 at 60/64)
Polychoric Ordinal Alpha	0.942	0.929	–*	–*	–*
Test-retest reliability (95% CI)	0.966 (0.926–0.988)	0.989 (0.973–0.995)	0.987 (0.968–0.995)	0.910 (0.778–0.963)	–*
Inter-rater reliability (95% CI)	0.998 (0.995–0.999)	0.999 (0.998–1.00)	–*	–*	–*
MDC _{95%}	4.25	3.84	5.00	6.01	–*

Abbreviations: Mini-BESTest_{GR}: Greek Mini-Balance Evaluation Systems Test; BBS_{GR}: Greek Berg Balance Scale; TUG: Timed Up & Go; FRT: Functional Reach Test; FES-I_{GR}: Greek Falls Efficacy Scale-International;

MDC_{95%}: Minimum Detectable Change at 95% of Confidence Interval

* = not applicable

Table 2. Intra- and inter-rater agreement for single item of the Greek Mini-Balance Evaluation Systems Test (Mini-BESTest_{GR}) and item-total correlation from n = 21.

Item	Single items agreement		Item/total
	Inter-rater	Test-retest (weighted Kappa)	
1	0.908	0.811	0.866
2	*0.940	*0.701	*0.876
3	*0.935	*0.814	*0.675
4	*0.950	*0.697	*0.649
5	*0.891	*0.639	*0.635
6	*1.000	*0.733	*0.589
7	*1.000	*0.788	*0.556
8	*1.000	*0.772	*0.801
9	*0.945	*0.836	*0.873
10	*0.940	*0.814	*0.853
11	*0.846	*0.787	*0.764
12	*1.000	*0.706	*0.763
13	*0.945	*0.690	*0.789
14	*1.000*	*0.556*	*0.446*

* $p < 0.05$

indicating very good agreement between raters. Test-retest agreement coefficients for individual Mini-BESTest_{GR} items presented moderate to almost perfect agreement with weighted kappa values ranging from 0.56 for item 14 to 0.84 for item 9 (Table 2). The item-total correlations were significant for all items with the item 14 to present the lowest correlation to total score of Mini-BESTest_{GR} (Table 2).

Validity

The correlation of the Mini-BESTest_{GR} was found to be very strong with the BBS_{GR} ($r = 0.924$, $p < 0.001$). In addition, Mini-BESTest_{GR} was strongly correlated with TUG ($r = -0.823$, $p < 0.001$) and FES-I ($r = -0.734$, $p < 0.001$) and moderately correlated with FRT ($r = 0.689$, $p < 0.001$) (Figure 2).

Ability to detect changes

A MDC_{95%} of 4.25 points on the scale was yielded. Table 1 presents test-retest and inter-rater reliability for total scores as well as Cronbach's α , and MDC_{95%} for all scales.

Discussion

Main findings

The present study aimed to assess the clinical application of the Mini-BESTest_{GR} scale in individuals with chronic stroke, through the evaluation of its reliability, validity and its ability to detect changes. The main findings revealed high levels of inter-rater and test-retest reliability, which suggests that the scale possesses

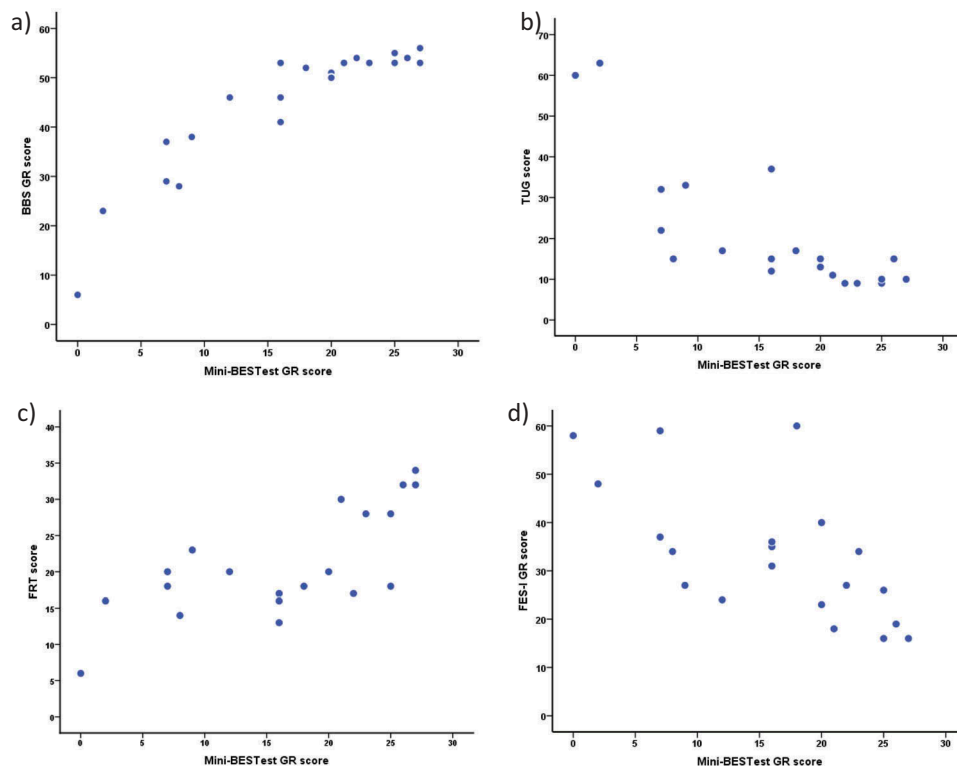


Figure 2. Scatterplot showing the relationship between the Mini-BESTest_{GR} (Mini-Balance Evaluation Systems Test-Greek) (total score of 14 activities) and the: a) BBS_{GR} (Berg Balance Scale-Greek) (total score of 14 items); b) TUG (Timed Up and Go test) (scores in seconds); c) FRT (Functional Reach Test) (scores in centimeters) and d) FES-I_{GR} (Falls Efficacy Scale International-Greek). All data are from measurements taken during the first assessment and from the first rater. Data from n = 21.

stability in the measurements. Moderate to almost perfect agreement was also revealed between raters and test-retest measurements for single items of the scale. Additionally, the Mini-BESTest_{GR} demonstrated a very strong concurrent validity with the BBS_{GR}, and its convergent validity was moderate to strong as it was initially hypothesized through correlations with TUG, FRT and FES-I_{GR}. Furthermore, the Mini-BESTest_{GR} presented the ability to discriminate between fallers and non-fallers and a minimum change of 4.25 points on the scale has been found to represent true change rather than measurement error.

In addition, the Mini-BESTest_{GR} showed a good distribution of the scores without ceiling or floor effects and less skewness than BBS_{GR}, which is in agreement with most of the studies on reported skewness (King et al., 2012; Lampropoulou, Gedikoglou, Michailidou, and Billis, 2016a; Tsang, Liao, Chung, and Pang, 2013). This may be due to the less demanding tasks that BBS consists compared to more challenging Mini-BESTest items such as those involving balance perturbations, stance on inclined or foam surface and gait with simultaneous head movements, pivot turns and cognitive tasks (Franchignoni et al., 2010).

The high relative reliability of the Mini-BESTest_{GR} with regards to both the test-retest and the inter-rater reliability is in agreement with the studies of: Godi et al. (2013) undertaken in patients of various balance deficits; Tsang, Liao, Chung, and Pang (2013) undertaken in patients with chronic stroke; and with the studies of Leddy, Crouner, and Earhart (2011) and Maia, Rodrigues-de-Paula, Magalhães, and Teixeira (2013) in patients with Parkinson's disease. Furthermore, the Mini-BESTest_{GR} showed excellent internal consistency which is in agreement with other studies as well (Godi et al., 2013; Löfgren et al., 2014; Tsang, Liao, Chung, and Pang, 2013). Overall, these results highlight the stability, consistency and reproducibility of the scale in measurements taken by different raters and at different time points under the same conditions. Reliability in total scores also agrees with the study of Dahl and Jørgensen (2014) in stroke patients, even though in their study each patient was assessed by 3 examiners, which provides an advantage compared to our study; and the reassessment took place after 4 weeks compared to the 7–10 days timeframe of our study. The use of two examiners and assessment based on live observation instead of observation of video recordings in our study was chosen because of being a commonly used methodological procedure for this kind of research, and of being a common way of assessment in clinical settings (Löfgren et al., 2014; Maia, Rodrigues-de-Paula, Magalhães, and Teixeira, 2013;

Tsang, Liao, Chung, and Pang, 2013; Billis et al., 2011). Furthermore, a relatively big interval between the two assessments was not considered as methodologically correct for our study, given that 4 weeks is enough time for alterations in the environment or health status, and eventually in the performance of the participant to occur.

Absolute reliability for Mini-BESTest_{GR} total scores regarding the evaluation of the difference in the measurements between raters is not usually assessed by researchers. Bland-Altman analysis has been conducted by only one study in individuals with Parkinson's diseases (Löfgren et al., 2014). In contrast to that study, our study presented very good agreement in the measurements between raters and their differences did not differ from zero. The different results in the two studies maybe explained by either the different neurological condition of the participants or by the training and the practice that the raters of our study had undertaken. It has been suggested that Mini-BESTest may require good training of the raters especially for items that are novel and may cause fear of falling in patients such as those involving push and release techniques to elicit compensatory stepping (Horak, Wrisley, and Frank, 2009).

Our findings regarding the reliability of individual items gave almost perfect kappa coefficient for inter-rater reliability and substantial to perfect coefficient for the test-retest reliability. One exception only was presented for item 14 (i.e. dual task walking) in test-retest reliability coefficient as well as in the item to total correlation, which presented a moderate coefficient and this may be explained by an actual change in patients' performance. During that task the participant has to count backwards by threes from a random number between 100–90 while he/she walks. This was a task that most of the participants found difficult to perform. This item has also been found to be the most difficult one among the items of BESTest in the study of Maia, Rodrigues-de-Paula, Magalhães, and Teixeira (2013). It might be possible therefore, that the participants practiced in counting backwards after the first assessment leading to a small learning effect during the second measurements. People with stroke and cognitive impairments (e.g. dementia or aphasia) may also find this task difficult to perform, because cognitively demanding tasks during walking have a destabilizing effect, thus, increasing the risk of fall (Hollman, Kovash, Kubik, and Linbo, 2007). It may therefore be useful if this dual task is transformed to a different kind of cognitive task, less demanding, such as spelling backwards a 5 letters word, when the scale is administered to patients with cognitive difficulties. Furthermore,

people who find that task challenging could incorporate dual and cognitive activities in their therapy. This will treat balance in a specific way, according to the underlying factors or systems affecting performance, rather than practicing only to a specific task (Horak, Wrisley, and Frank, 2009).

The agreement in the measurements taken either between different raters or at different time points is higher than the one reported in the other two studies which have tested similar reliability in stroke patients (Dahl and Jørgensen, 2014; Tsang, Liao, Chung, and Pang, 2013). Both of those studies reported lower coefficients in items that included complex tasks such as item 6 (i.e. compensatory stepping-lateral), and gait items. The same items in our study yielded perfect inter-rater and substantial to almost perfect test-retest reliability. The superiority of our results compared to the other two studies may be explained by multiple reasons. First, our raters had sufficient training in neurological assessment and in the utility of the functional scales tested. Additionally, they had practiced and they had watched all the training videos provided by the site of the Balance Evaluation System Test (BEST) especially for items that are considered as complicated or even 'dangerous' to test in patients with balance disorders (i.e. those with stepping postural responses) (Horak, Wrisley, and Frank, 2009). Second, the raters served as administrators of the scale, which has been suggested to increase the agreement between the measurements (Horak, Wrisley, and Frank, 2009). Third, ratings were taken by observations on real time and not by watching the video that restricts the anatomical planes that participants observed, a limitation that is stated in the study by Dahl and Jørgensen (2014). The perfect overall results suggest that the Mini-BESTest_{GR} is reliable when different raters use it and/or at repeated measures if trained physiotherapists are administering it.

The Mini-BESTest_{GR} also presented very strong correlations with the well-known and established BBS_{GR}. Similar results were reported in other studies and in various neurological conditions such as: chronic stroke (Tsang, Liao, Chung, and Pang, 2013); Parkinson's disease (King et al., 2012); and balance deficits of multi-factors aetiology (Godi et al., 2013). These comparable results suggest the concurrent validity of the Mini-BESTest_{GR} in Greek individuals with chronic stroke. The strong negative correlation of the Mini-BESTest_{GR} to the TUG and the moderate correlation to FRT are in accordance with our primary hypothesis and in line with the studies of Bergström, Lenholm, and Franzen (2012) and Tsang, Liao, Chung, and Pang (2013). Since the low levels of balance ability require more time for the execution of transfers and gait, the strong negative correlation

of the balance scale with the time to go 3 meters at TUG test is fully explained. The significant correlation of the Mini-BESTest_{GR} to the FRT could be considered as a paradox, if one thinks that FRT is a single test for balance assessment, whereas the Mini-BESTest includes the assessment of dynamic balance in more than one activities. However, this moderate but significant correlation proves the construct validity of the scale, thus both scales evaluate balance components. Less strong but significant and moderate was the correlation of the Mini-BESTest_{GR} to the FES-I_{GR} compared to the study of Bergström, Lenholm, and Franzen (2012) for Parkinson patients that presented low correlation. This discrepancy may be explained by the different population undertaken the study and together with the aforementioned results confirm the construct validity of the Mini-BESTest_{GR} for assessing balance in Greek patients with chronic stroke.

One important feature of a clinical tool is its ability to detect real changes in the patients' status and discriminate patients regarding their level of function. Literature gives inconclusive results regarding the minimum detectable change for repeated measurements which is reported as: 3.0 points in the study of Tsang, Liao, Chung, and Pang (2013); 3.5 points in the study of Godi et al. (2013); 3.4 points in the study of Löfgren et al. (2014) and 4.4 points in the study of Dahl and Jørgensen (2014). A minimum detectable change at 95% confidence interval (MDC95%) of 4.3 points was found in our study, which is not far from the other changes reported. The small variability in the results may be explained by the differences in methods and in clinical populations used in each study. Stroke patients were included in the studies of Tsang, Liao, Chung, and Pang (2013) and Dahl and Jørgensen (2014); and our MDC95% is similar to the last one reported. In addition, the Mini-BESTest_{GR} presented the ability to discriminate between fallers and non-fallers. Although at a borderline, the statistically significant difference between the two categories found in our study is in disagreement with the study of O'Hoski, Sibley, Brooks, and Beauchamp (2015), which failed to show such a discrimination (between fallers and non-fallers) in older adults. In line with the results of our study is the study of Leddy, Crowner, and Earhart (2011) undertaken in patients with Parkinson. The inconclusive results, however, regarding the ability of the scale to discriminate between fallers and non-fallers and in regards to the MDC95% or the clinical important change specifically to stroke patients necessitates further research. MDC95%, as well as clinical important change would be better assessed following an intervention or/and in combination with sensitivity and specificity assessment that could give

more conclusive results regarding the ability of the scale to detect real changes in balance improvement or in illness severity of the patient.

Clinical implications

This study is important for clinical practice as it recommends to Greek clinical environment one of the most modern as well as valid and reliable international balance assessment tools. In addition, its good distribution of the scores with less skewness compared to BBS and without ceiling or floor effects, and the significant differences presented between fallers and non-fallers, supports the ability of the Greek version of the scale to discriminate the levels of performance and balance function among patients. Furthermore, the thorough examination of the reliability, compared to other studies, give the advantage of further establishing the absolute and the relative reliability of the scale at least to people with chronic stroke. The excellent relative and absolute reliability for total scores and for each item prove therefore, the ability of the Mini-BESTest_{GR} to give stable, reproducible and consistent results over time and between raters when these are trained. The strong correlations with other balance assessment tools highlight the concurrent and convergent validity of Mini-BESTest_{GR} when this is applied on patients with chronic stroke. A minimum detectable change of 4.3 points is of great clinical use because it could help to identify any real change in the balance ability of individuals with chronic stroke following a rehabilitation program.

Limitations

This study consisted of a convenience sample of people with stroke at a chronic stage and as such its results cannot be generalized to stroke patients at other stages (i.e. acute stage). The results could be applied to this specific population of Greek patients. Furthermore, despite the clinical importance of identifying the MDC_{95%} this study did not provide a detailed investigation regarding the responsiveness of the scale. Results about sensitivity and specificity could also have given information about the cut off scores on the scale that could predict future falls. This is another clinically important issue that needs further exploration.

Future research

Future research should focus on individuals with acute and sub-acute in addition to chronic stage of stroke in order to investigate the ability of the scale to

distinguish between the various levels of seriousness of the illness. Also, psychometric evaluation of the scale on patients with balance impairments from other neurological conditions is of immediate importance in order to fully establish the scale in clinical practice. The correlation of the scale to the falls history of the patient needs to be investigated in combination with responsiveness, sensitivity and specificity assessment in order to further explore the ability of the scale to predict falls in various neurologic conditions. The meaningful clinical change will also give useful information to clinicians regarding the efficiency of a treatment protocol in improving the balance ability of patients of various conditions at different levels of recovery.

Conclusions

In conclusion, the results of the current study, account for a valid and reliable mean of balance assessment in Greek patient with chronic stroke that fulfills the requirements for stability, reproducibility and good distributions of its measurements. The excellent relative and absolute reliability as well as its very good validity establish Mini-BESTest_{GR} as an optimal tool among those that exist in the scientific community for assessing balance in chronic stroke.

Acknowledgments

Initially, we would like to thank Dr Horak for giving us the permission to adapt and use the Mini-BESTest into Greek. We also thank the physiotherapists who kindly accepted to administer the scale and all patients for their voluntary participation in the balance assessment; without them this research would not be completed. Part of the results has been orally presented at the 24th Panhellenic Scientific Conference of Physiotherapy, PSF, Athens, Greece (5-7th December 2014) and to the 5th Pancyprriot Conference of Physiotherapy, Nicosia, Cyprus (5-6th March 2016).

Declaration of Interest

Authors report no declarations of interest

References

- Atkinson G, Nevill AM 1998 Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine* 26: 217–238.
- Barry E, Galvin R, Keogh C, Frances Horgan F, Fahey T 2014 Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: A systematic review and meta- analysis. *BMC Geriatrics* 14: 14.

- Berg K, Wood-Dauphinee S, Williams JI 1995 The Berg Balance Scale: reliability assessment with elderly residents and patients with an acute stroke. *Scandinavian Journal of Rehabilitation Medicine* 27: 27–36.
- Bergström M, Lenholm E, Franzen E 2012 Translation and validation of the Swedish version of the Mini-BESTest in subjects with Parkinson's disease or stroke: A pilot study. *Physiotherapy Theory and Practice* 28: 509–514.
- Bohannon RW 2006 Reference values for the Timed Up and Go test: A descriptive meta-analysis. *Journal of Geriatric Physical Therapy* 29: 64–68.
- Bonnet DG 2002 Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in Medicine* 21: 1331–1335.
- Botner EM, Miller WC, Eng JJ 2005 Measurement properties of the activities-specific balance confidence scale among individuals with stroke. *Disability and Rehabilitation* 27: 156–163.
- Bowling A 2002 *Research Methods in Health*, 2nd edn. pp. 150–156. New York: Open University Press.
- Chinsongkram B, Chaikereee N, Saengsirisuwan V, Viriyatharakij N, Horak FB, Boonsinsukh R 2014 Reliability and validity of the Balance Evaluation Systems Test (BESTest) in people with subacute stroke. *Physical Therapy* 94: 1632–1643.
- Dahl SS, Jørgensen L 2014 Intra- and inter-rater reliability of the mini-balance evaluation systems test in individuals with stroke. *International Journal of Physical Medicine and Rehabilitation* 2: 177.
- de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM 2006 Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Quality of Life Outcomes* 4: 54.
- Di Carlo S, Bravini E, Vercelli S, Massazza G, Ferriero G 2016 The Mini-BESTest: A review of psychometric properties. *International Journal of Rehabilitation Research* 39: 97–105.
- Duncan RP, Leddy AL, Cavanaugh JT, Dibble LE, Ellis TD, Ford MP, Foreman KB, Earhart GM. 2013 Comparative utility of the BESTest, mini-BESTest, and brief-BESTest for predicting falls in individuals with Parkinson disease: a cohort study. *Physical Therapy* 93: 542–550.
- Duncan RP, Leddy AL, Cavanaugh JT, Dibble LE, Ellis TD, Ford MP, Foreman KB, Earhart GM 2015 Balance differences in people with Parkinson disease with and without freezing of gait. *Gait and Posture* 42: 306–309.
- Faber MJ, Bosscher RJ, van Wieringen PC 2006 Clinimetric properties of the performance-oriented mobility assessment. *Physical Therapy* 86: 944–954.
- Faul F, Buchner A, Erdfelder E, Lang AG 2014 G*Power Version 3.1.9.2 [computer software]. Germany: Universität Kiel. <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register>
- Fisher S, Kenneth J, Ottenbacher KJ, Goodwin JS, Graham J, Ostir GV 2009 Short physical performance battery in hospitalized older adults. *Aging Clinical and Experimental Research* 21: 445–452.
- Franchignoni F, Godi M, Guglielmetti A, Nardone A, Giordano A 2015 Enhancing the usefulness of the Mini-BESTest for measuring dynamic balance: A Rasch validation study. *European Journal of Rehabilitation Medicine* 51: 429–437.
- Franchignoni F, Horak F, Godi M, Nardone A, Giordano A 2010 Using psychometric techniques to improve the Balance Evaluation System's Test: the Mini-BESTest. *Journal of Rehabilitation Medicine* 42: 323–331.
- Gadermann AM, Guhn M, Zumbo BD 2012 Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research and Evaluation* 17: 3.
- Garcia A, Marciniak D, McCune L, Smith E, Ramsey R 2012 Promoting fall self-efficacy and fall risk awareness in older adults. *Physical and Occupational Therapy in Geriatrics* 30: 165–175.
- Godi M, Franchignoni F, Caligari M, Giordano A, Turcat AM, Nardone A 2013 Comparison of reliability, validity, and responsiveness of the Mini-BESTest and Berg Balance Scale in patients with balance disorders. *Physical Therapy* 93: 158–167.
- Golafshani N 2003 Understanding reliability and validity in qualitative research. *Qualitative Report* 8: 597–607.
- Hallgren KA 2012 Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology* 8: 23–34.
- Hirtz D, Thurman DJ, Gwinn-Hardy K, Mohamed M, Chaudhuri AR, Zalutsky R 2007 How common are the “common” neurologic disorders? *Neurology* 68: 326–337.
- Hollman JH, Kovash FM, Kubik JJ, Linbo RA 2007 Age-related differences in spatiotemporal markers of gait stability during dual task walking. *Gait and Posture* 26: 113–119.
- Horak FB, Wrisley DM, Frank J 2009 The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Physical Therapy* 89: 484–498.
- Jonsdottir J, Cattaneo D 2007 Reliability and validity of the dynamic gait index in persons with chronic stroke. *Archives of Physical Medicine and Rehabilitation* 88: 1410–1415.
- King LA, Horak F 2013 On the Mini-BESTest: scoring and the reporting of total scores. *Physical Therapy* 93: 571–575.
- King LA, Priest KC, Salarian A, Pierce D, Horak FB 2012 Comparing the Mini-BESTest with the BBS to evaluate balance in Parkinson's disease. *Parkinson's Disease* 2012: 375–419.
- Knorr S, Brouwer B, Garland SJ 2010 Validity of the community balance and mobility scale in community-dwelling persons after stroke. *Archives of Physical Medicine and Rehabilitation* 91: 890–896.
- La Porta F, Caselli S, Susassi S, Cavallini P, Tennant A, Franceschini M 2012 Is the Berg Balance Scale an internally valid and reliable measure of balance across different etiologies in neurorehabilitation? A revisited Rasch analysis study. *Archives of Physical Medicine and Rehabilitation* 93: 1209–1216.
- Lampropoulou S, Gedikoglou IA, Michailidou C, Billis E 2016a Cross cultural validation of the Mini-BESTest into Greek. *World Journal of Research and Review* 3: 61–67.
- Lampropoulou S, Gizeli A, Kalivioti C, Billis E, Gedikoglou IA, Nowicky VA 2016b Cross cultural adaptation of Berg Balance Scale in Greek for various balance impairments. *Journal of Physical Medicine, Rehabilitation and Disabilities* 2: 011.

- Landis JR, Koch GG 1977 The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.
- Leddy A, Crowner B, Earhart G 2011 Utility of the Mini-BESTest, BESTest, and BESTest Sections for balance assessments in individuals with parkinson disease. *Journal of Neurological Physical Therapy* 35: 90–97.
- Löfgren N, Lenholm E, Conderson D, Ståhle A, Franzén E 2014 The Mini-BESTest – A clinically reproducible tool for balance evaluations in mild to moderate Parkinson’s disease? *BMC Neurology* 14: 235.
- Ma VY, Chang L, Carruthers KJ 2014 The incidence, prevalence, costs and impact on disability of common conditions requiring rehabilitation in the US. *Archives of Physical Medicine and Rehabilitation* 95: 986–995.
- Maia A, Rodrigues-de-Paula F, Magalhães L, Teixeira R 2013 Cross-cultural adaptation and analysis of the psychometric properties of the Balance Evaluation Systems Test and MiniBESTest in the elderly and individuals with Parkinson’s disease: application of the Rasch model. *Brazilian Journal of Physical Therapy* 17: 195–217.
- Mak MK, Auyeung MM 2013 The mini-BESTest can predict Parkinsonian recurrent fallers: A 6-month prospective study. *Journal of Rehabilitation Medicine* 45: 565–571.
- Mancini M, Horak FB 2010 The relevance of clinical balance assessment tools to differentiate balance deficits. *European Journal of Physical and Rehabilitation Medicine* 46: 239–248.
- Munro B 2005 *Statistical Methods for Health Care Research*, 5th edn. Philadelphia: Lippincott Williams and Wilkins.
- Myles PS, Cui J 2007 Using the Bland-Altman method to measure agreement with repeated measures. *British Journal of Anaesthesia* 99: 309–311.
- O’Hoski S, Sibley KM, Brooks D, Beauchamp MK 2015 Construct validity of the BESTest, mini-BESTest and brief BESTest in adults aged 50 years and older. *Gait and Posture* 42: 301–305.
- Ouzouni C, Nakakis K 2011 Reliability and validity of measurement outcomes in quantitative research. *Nosileftiki* 50: 231–239.
- Padgett PK, Jacobs JV, Kasser SL 2012 Is the BESTest at Its Best? A suggested brief version based on interrater reliability, validity, internal consistency, and theoretical construct. *Physical Therapy* 92: 1197–1207.
- Pardasaney PK, Latham NK, Jette AM, Wagenaar RC, Pengsheng N, Slavin MD, Bean JF 2012 Sensitivity to change and responsiveness of four balance measures for community-dwelling older adults. *Physical Therapy* 92: 388–397.
- Pardasaney PK, Slavin MD, Wagenaar RC, Latham NK, Pengsheng N, Jette AM 2013 Conceptual limitations of balance measures for community-dwelling older adults. *Physical Therapy* 93: 1351–1368.
- Portney LG, Watkins MP 2009 *Foundations of Clinical Research: applications to Practice*, 3rd edn. Upper Saddle River, NJ: Pearson/Prentice Hall.
- Roach K 2006 Measurement of health outcomes: reliability, validity and responsiveness. *Journal of Prosthetics and Orthotics* 18(1S): 8–12.
- Scholtes VA, Terwee CB, Poolman RW 2011 What makes a measurement instrument valid and reliable? *Injury* 42: 236–240.
- Shumway-Cook A, Taylor CS, Matsuda P, Studer MT, Whetten BK 2013 Expanding the scoring system for the dynamic gait index. *Physical Therapy* 93: 1493–1506.
- Solias A, Skapinakis P, Degleris N, Pantoleon M, Katirtzoglou E, Politis A 2014 Mini Mental State Examination (MMSE): determination of cutoff scores according to age and educational level. *Psychiatriki* 25: 245–256.
- Toomey E, Coote S 2013 Between-rater reliability of the 6-minute walk test, Berg balance scale, and handheld dynamometry in people with multiple sclerosis. *International Journal of Multiple Sclerosis Care* 15: 1–6.
- Tsang CS, Liao LR, Chung RC, Pang MY 2013 Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. *Physical Therapy* 93: 1102–1115.
- Tyson S, Connell L 2009 The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. *Clinical Rehabilitation* 23: 1018–1033.
- Wolf SL, Catlin PA, Gage K, Gurucharri K, Robertson R, Stephen K 1999 Establishing the reliability and validity of measurements of walking time using the emory functional ambulation profile. *Physical Therapy* 79: 1122–1133.
- Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C 2005 Development and initial validation of the Falls Efficacy Scale-International (FES-I). *Age Ageing* 34: 614–619.
- Yelnik A, Bonan I 2008 Clinical tools for assessing balance disorders. *Clinical Neurophysiology* 38: 439–445.
- Billis E, Strimpakos N, Kapreli E, Sakellari V, Skelton D, Dontas I, Ioannou F, Filon G, Gioftsos G 2011 Cross-cultural validation of the Falls Efficacy Scale International (FES-I) in Greek community-dwelling older adults. *Disability and Rehabilitation* 33: 1776–1784.