ARTICLE ONLINE FIRST

This provisional PDF corresponds to the article as it appeared upon acceptance.

A copyedited and fully formatted version will be made available soon.

The final version may contain major or minor changes.

Characteristics that affect score reliability in the Berg Balance Scale: A meta-analytic reliability generalization study.

Ana-Belén MESEGUER-HENAREJOS, María RUBIO-APARICIO, José-Antonio LÓPEZ-PINA, Ricardo CARLES-HERNÁNDEZ, Anonia GÓMEZ-CONESA

European Journal of Physical and Rehabilitation Medicine 2019 Apr 04

DOI: 10.23736/S1973-9087.19.05363-2

Article type: Systematic reviews and meta-analyses

© 2019 EDIZIONI MINERVA MEDICA

Article first published online: April 4, 2019

Manuscript accepted: April 4, 2019

Manuscript revised: December 6, 2018

Manuscript received: May 23, 2018

Subscription: Information about subscribing to Minerva Medica journals is online at:

<http://www.minervamedica.it/en/how-to-order-journals.php>

Reprints and permissions: For information about reprints and permissions send an email to:

journals.dept@minervamedica.it - journals2.dept@minervamedica.it - journals6.dept@minervamedica.it

Characteristics that affect score reliability in the Berg Balance Scale: A meta-analytic reliability generalization study.

A meta-analytic reliability generalization study of the Berg Balance Scale.

Ana-Belén Meseguer-Henarejos¹, María Rubio-Aparicio^{2*}, José-Antonio López-Pina²,
Ricardo Carles-Hernández¹, Antonia Gómez-Conesa¹.

¹ Department of Physiotherapy, University of Murcia, Murcia, Spain; ² Department of Basic Psychology and Methodology, University of Murcia, Murcia, Spain.

* Corresponding author

María Rubio-Aparicio

Department of Basic Psychology and Methodology

The University of Murcia

30100 Campus de Espinardo

Murcia, Spain.

E-mail: maria.rubio1@um.es

ABSTRACT

INTRODUCTION: The Berg Balance Scale is the most widely used clinical scale used to assess balance performance in neurological conditions. Reliability is not an unalterable value of a scale across different applications, but is a property referring to the results obtained with the scale, not the instrument in itself. The results of a scale can vary depending on the context where it is applied and the characteristics of the participants, therefore the reliability coefficient should be reported in each study. A systematic review and meta-analysis of the reliability coefficients obtained in different applications of the test with the data at hand is the best method to examine how the reliability of a test scores varies. The objectives of this systematic revision are: To determine the mean of internal consistency, intra and interrater reliability of the Berg balance scale in the clinical, nonclinical and mixed populations, to determine the methodological and substantive characteristics and to propose a predictive model enabling researchers and clinicians to use it in the future to estimate the expected reliability based on the characteristics of the most relevant studies.

EVIDENCE ACQUISITION: The MEDLINE (Pubmed), EMBASE and CINAHL databases were searched from 1989 to 2015. Two reviewers independently selected empirical studies published in English or in Spanish that applied the Berg Balance Scale and reported any reliability coefficient.

EVIDENCE SYNTHESIS: The 80 samples in the 65 studies reported any the Berg Balance Scale reliability estimate. Coefficient alpha ranged from .62 to .98, with a mean of .92. For intra-rater agreement, the mean intraclass correlation was $ICC_+ = .957$, and for inter-rater agreement $ICC_+ = .97$. The SD of the Berg Balance Scale scores presented statistically significant relationships with the coefficient alpha and with ICC (intra-rater). In addition, the clinical population and the institutionalized population presented statistically significant

relationships with the coefficient alpha. The sample size and mean scale results were also statistically related to ICCs (intra-rater).

CONCLUSIONS: The alpha coefficient and intra- and inter-rater agreement for Berg Balance Scale scores was very satisfactory. Several characteristics of the studies were statistically associated to the alpha coefficient and with intra-rater reliability.

Keywords: Reliability, Meta-analysis, Berg Balance Scale.

Introduction

Postural balance or postural control has been defined as the ability to control the position of the body in space for stability and orientation¹, and can be classified as static and dynamic. Static equilibrium is the ability to maintain the center of body mass within the base of support, while dynamic equilibrium is the ability to move the center of body mass in relation to the support base in a controlled manner during movements².

Several factors can contribute to different balance deficits such as decreased muscle strength, limited range of movement, abnormal muscle tone, lack of motor coordination, poor sensory organization and cognitive deficit^{3,4}, factors commonly present in patients who have suffered a neurological pathology, such as a cerebrovascular accident, multiple sclerosis, Parkinson's disease, among others. With age, a series of physiological changes occur, such as reductions in muscle strength, proprioception, range of joint movement, reaction time, changes in sensory systems, as well as cognitive deficiency⁵⁻⁷.

These potentially negative factors affect the control of balance and impact on the functional capacity and activities of the elderly person's daily life. Good balance and mobility are essential for the proper development of most daily activities and recreational or leisure activities. In addition, the reduction in ability to maintain balance may be associated with an increased risk of falls^{6,8}.

Preventing and treating balance problems are common activities in the physiotherapist's clinic, and their evaluation is necessary in prescribing interventions that improve people's balance. Despite its limitations, the Berg Balance Scale is the most widely used measurement instrument for assessing balance performance in neurological conditions⁹ such as the elderly^{6,10-12}, stroke patients¹³⁻¹⁶, Parkinson's disease¹⁷⁻¹⁸, traumatic brain injury¹⁹, or multiple sclerosis²⁰, as well as a measure of the risk of falls in the elderly^{6,9}. It is also used to assess the effectiveness of interventions to improve balance²⁰⁻²⁴. The Berg Balance Scale has been adapted to the Italian²⁵, Brazilian²⁶, Norwegian²⁷, Turkish²⁸ and Iranian populations²⁹.

For a measuring instrument to be useful it must have good psychometric properties such as reliability and validity³⁰. Reliability can be defined as the consistency of measurement in test conditions³¹. Such conditions include time-sampling (e.g., test-retest correlation), content sampling (e.g., alpha coefficient), and inter-subject differences (e.g., intraclass correlation). Adequate reliability of test results is indispensable for the clinician in order to reach an accurate diagnosis²⁹. Low reliability may decrease the statistical power of the significance tests used by the research team³².

To date, more than 60 studies have been published that report some reliability coefficients of the Berg scale in adult subjects^{9,15,19,33-36}. These studies highlight a clear variability in

reliability estimates. This is due to the fact that the reliability of test scores may change depending on the composition of the sample to which it is administered (e.g., clinical vs. nonclinical population, age, gender and ethnic distribution, language, test adaptation, etc.)³⁰.

Reliability is not an unalterable value of a scale across different applications, but is a property referring to the results obtained with the scale, not the instrument in itself³⁷. The results of a scale can vary depending on the context where it is applied and the characteristics of the participants, therefore the reliability coefficient should be reported in each study.

Nevertheless, studies often do not report the reliability of scores, but rather refer to estimates obtained in other studies or to those obtained in the original psychometric study of the scale itself. This phenomenon known as reliability induction³⁸ does not consider that reliability in a particular sample is only comparable to that obtained in another sample when both have the same composition and variability. Consequently, both the statistical power and the obtained effect-size estimates can be wrongly interpreted.

As reliability can vary from one test application to the next, the best method to examine how the reliability of a test score varies is by performing a systematic review and meta-analysis of the reliability coefficients obtained in different applications of the test with the data at hand. This kind of systematic review is usually named reliability generalization meta-analysis³⁹.

The objectives of this systematic review are:

1. To determine the mean of internal consistency, intra and interrater reliability of the Berg balance scale in the clinical, nonclinical and mixed populations.
2. To determine the methodological and substantive characteristics of the studies that can be statistically associated to test score reliability coefficients.

3. To propose a predictive model enabling researchers and clinicians to use it in the future to estimate the expected reliability based on the characteristics of the most relevant studies.

Materials and methods

Identification and selection of studies

A reliability generalization study requires selecting studies that have applied the test and that report a reliability estimate from the data at hand. To be included, studies were required to meet five criteria: A) empirical studies (observational and experimental) which had applied Berg Balance Scale to a sample of participants, b) they had to report the reliability coefficient computed from their own sample data such internal consistency, intra-raters or inter-raters, c) they had to be published since November 1989, date when first article on Berg Balance Scale was published, until December 2015, and d) due to language limitations were required to be written in English or Spanish.

To locate the studies, the following electronic databases were consulted: Medline (PubMed), EMBASE and CINHALL. The search strategy was established combining key words (S1 Table). The electronic search was complemented by checking the reference lists of included studies.

SI Table. Search strategy.

The study selection process was conducted by two assessors who independently decided eligibility by titles and abstracts based on predetermined criteria. Where eligibility was unclear from the title and abstract, the full-text version was obtained and examined by both

assessors. Disagreements were resolved by consensus, with a third assessor consulted when necessary.

Assessment of characteristics of studies

Methodological and substantive characteristics of the studies were extracted in order to examine the potential influence of moderator variables on the reliability estimates. The following methodological variables were coded: test version, type of design, study focus, sample size, previous training of raters, intra- rater internal, inter- rater interval and number of raters. The following substantive variables were coded: age of the sample coded as the mean in years), standard deviation of age, continent where study was conducted, gender distribution, history of illness (mean and standard deviation), year of study, profession of raters, target population 1 (children and adolescents vs. adults), target population 2 (clinical, nonclinical, mixed) and target population 3 (institutionalized, non-institutionalized, mixed).

Data extraction

The 80 samples present in the 65 studies found were coded following a manual and registration protocol. To examine their influence on the Berg Balance Scale reliability, moderator variables related to the instrument and the subject samples were coded.

Together with these moderator variables, intra-rater reliability (Intraclass Correlation Coefficient), inter-rater reliability (Intraclass Correlation Coefficient), and alpha coefficients were obtained for the Berg Balance Scale when reported in the studies.

The reliability of the coding process of the study characteristics was checked by selecting a random sample of 50% of the studies included. This sample of studies was double-coded by

two independent coding teams and the intercoder agreement was satisfactory, with ICC ranging between 0.899 and 1.00.

The reliability coefficient was a source of heterogeneity, as one or more internal consistency coefficients, intra-rater or inter-rater appeared in the articles. ICC intra-rater reliability (48 studies, N = 1608) and ICC inter-rater reliability (39 studies, N = 1352) were the most reported, followed by coefficient alpha (23 studies, N = 1502). Since these coefficients are based on different assumptions, a reliability generalization meta-analysis has been performed for each separately^{40,41}.

Reliability estimates

In this reliability generalization study, three types of reliability coefficients were taken into account: interrater reliability (intraclass correlation coefficients), intrarrater reliability (intraclass correlation coefficients), and internal consistency of the measures (coefficients alpha). To carry out the meta-analysis, reliability coefficients were transformed in order to normalize its distribution and stabilize their variances. On the one hand, the coefficients alpha were transformed with formula proposed by Bonett (2002): $L_i = Ln(1 - |\hat{\alpha}_i|)$, with Ln being the natural logarithm⁴². On the other hand, intraclass correlation coefficients to assess interrater and intrarrater reliability were transformed into the Fisher's Z : $Z_i = .5 Ln[(1 + \hat{\alpha}_i)/(1 - \hat{\alpha}_i)]$.

Statistical analysis

Separate meta-analyses were conducted for each type of reliability coefficient. In all cases, a random-effects model was assumed and the 95% confidence limits around the weighted average reliability coefficient were calculated with the improved method proposed by

Hartung^{43,44}. The between-studies variance was estimated by restricted maximum likelihood⁴⁵.

To facilitate the interpretation of the results from each meta-analysis, the average reliability coefficients and their confidence limits obtained with the Bonett's or Fisher's Z transformations, were back-transformed into the coefficient alpha intraclass correlation metrics, respectively.

The heterogeneity exhibited by the reliability coefficients in each meta-analysis was investigated by calculating the Q statistic and the I^2 index, and by constructing a forest plot. If studies exhibited heterogeneity, then moderator analyses were carried out in order to identify the study characteristics that could explain at least part of the heterogeneity. Weighted ANOVAs and meta-regressions assuming a mixed-effects model were conducted for qualitative and quantitative moderators, respectively. Mixed-effects models were assumed, using the improved method proposed by Knapp and Hartung to test the statistical significance of moderator variables⁴⁴⁻⁴⁶. The statistical analyses were carried out with the *metafor* package in R ⁴⁷.

Results

Flow of studies through the review

Figure 1 presents a flowchart describing the selection process of the studies that met selection criteria. Electronic databases provided a total of 2,388 references screened for potential inclusion in the systematic review. In total, 1,240 references were empirical studies that applied the Berg Balance Scale. The remaining references were systematic reviews and meta-analyses, posters or abstracts published in scientific journals, studies not written in English or

Spanish, study protocols, letters, studies without persons, and a study which includes the surname Berg in the abstract and not the Berg Balance Scale. S2 Table presents the excluded studies.

Fig 1. Flow of studies through the systematic review.

SII Table. Excluded studies.

Out of 1,234 references that applied the Berg Balance Scale only 65 studies^{6,9,15,18,20,22,25-29,33-36,48-98} (5.3%) reported some reliability estimate with the data at hand, the remaining 1,169 inducing it from previous references. These figures implied a dramatically large reliability induction rate for the Berg Balance Scale of 94.7%. The 65 studies that reported some reliability estimate constituted the database for this systematic review and meta-analysis.

Characteristics of included studies

S3 Table presents the characteristics of 80 samples in 65 included studies. Some had consistency coefficients such as internal consistency, intra- and inter-rater reliability^{6,9,25,26,28,48-50}, internal consistency and intra-rater reliability^{31,51}, internal consistency and inter-rater reliability^{27,29,52-54} or intra- and inter-rater reliability^{12,20,34,55-65}, whereas others reported a single reliability coefficient such as internal consistency^{18,66-71}, intra-rater reliability^{36,72-88}, or inter-rater reliability^{15,89-94,97}. The sample size ranged from 5^{62,73,84} to 120¹⁵.

SIII Table. Summary of included studies.

The samples contained participants with or without illnesses. Some studies had samples of participants who had no pathology^{6,12,34,56,59,60,66,75,77,90,92}, while others included samples with and without pathology⁹. Most studies had samples from persons with a single illness such as stroke^{15,36,55,64,70,78,80,82,91,95,97,98}, multiple sclerosis^{20,29,58,71,81,83}, Parkinson's disease^{18, 34,53,57,61,64,88}, traumatic brain injury⁷³, peripheral arthritis⁸⁹, Huntington's disease⁸⁵, knee osteoarthritis⁸⁶, spinal cord injury^{87,94}, Alzheimer's disease⁶³, total knee arthroplasty⁵¹ or cerebral palsy⁶⁵. The most common pathology was stroke. Few studies included participants with different pathologies in their samples^{25-28,48,50,52,67,68,72,79}.

Study 1: Alpha coefficients meta-analysis

Mean Reliability and Heterogeneity

The 23 samples that reported a coefficient alpha ranged from 0.62 to 0.98, with a mean of 0.92 (95% CI = 0.89 and 0.95). Significant heterogeneity was found among the coefficients alpha, $Q(22) = 486.58$, $p < 0.0001$, with an I^2 value of 94.69%. This considerable amount of heterogeneity is also reflected in the forest plot (Figure 2). The large variability exhibited by the reliability coefficients was examined by analyzing the influence of moderator variables.

Fig 2. Forest plot displaying the coefficients alpha (and 95% confidence intervals).

Moderator Analyses

Table 1 shows the results of the simple regression analyses conducted for the continuous moderators. Of the different moderators analyzed, only the standard deviation of the test scores yielded a positive and statistically significant relationship with the coefficients alpha ($p < 0.0001$) with a 99% of variance accounted for. The remaining continuous moderator variables tested did not reach statistical significance.

Table I. Results of the simple meta-regression applied on coefficients alpha taking continuous moderator variables as predictors.

Predictor variable	<i>K</i>	<i>b_i</i>	<i>F</i>	<i>p</i>	<i>Q_E</i>	<i>R²</i>
Mean scores	10	-0.0931	2.71	0.138	64.64***	0.18
<i>SD</i> of scores	9	0.1349	87.51	<0.0001	6.39	0.99
Sample size	23	-0.0001	0.00	0.983	481.64***	0.00
Mean age (years)	18	-0.0027	0.03	0.863	274.09***	0.00
<i>SD</i> of age (years)	16	0.0465	0.84	0.375	194.89***	0.00
Gender (% male)	21	-0.0033	0.18	0.679	352.35***	0.00
Mean of disorder history (years)	5	-0.0299	0.91	0.412	15.88**	0.00
<i>SD</i> of disorder history (years)	4	-0.0389	3.34	0.209	2.04	0.00
Year of the study	23	-0.0101	0.15	0.707	474.61***	0.00

k = number of studies. *b_i* = regression coefficient of each predictor. *F* = Knapp-Hartung's statistic for testing the significance of the predictor (the degrees of freedom for this statistic are 1 for the numerator and *k* – 2 for the denominator). *p* = probability level for the *F* statistic. *Q_E* = statistic for testing the model misspecification. *R²* = proportion of variance accounted for by the predictor. In order to facilitate the interpretation, the regression coefficients were back-transformed to the metric of the original coefficients.

***p* < 0.001; *** *p* < 0.0001.

Regarding the qualitative moderators, Table 2 presents the results of ANOVAs applied on the coefficients alpha. Target population 2 showed a statistically significant association with the coefficients alpha (*p* = 0.035), higher for the clinical samples (mean = 0.94) than for nonclinical samples (mean = 0.82). Furthermore, statistically significant differences were found (*p* = 0.018) when comparing the average coefficients alpha among the institutionalized population, non-institutionalized population, and mixed category (institutionalized and non-institutionalized populations), explaining this moderator 28% of the variance, and showing the studies with non-institutionalized populations the largest coefficient alpha (mean= 0.93).

Table II. Results of the weighted ANOVAs applied on coefficients alpha, taking qualitative moderator variables as independent variables.

Variable	<i>K</i>	α_+	95% CI		ANOVA results
			LL	LU	
Test version					
Original	8	0.91	0.83	0.95	<i>F</i> (1,21) = 0.43, <i>p</i> = 0.5203
Other	15	0.93	0.89	0.95	<i>R²</i> = 0.00 <i>Q_w</i> (21)=423.18, <i>p</i> <0.0001
Design					
Observational	21	0.92	0.89	0.95	<i>F</i> (1,21) = 0.01, <i>p</i> = 0.928
Experimental	2	0.93	0.75	0.98	<i>R²</i> = 0.00 <i>Q_w</i> (21)=484.09, <i>p</i> <0.0001
Study focus					
Psychometric	19	0.92	0.88	0.95	<i>F</i> (1,21) = 0.01, <i>p</i> = 0.842

Applied	4	0.92	0.81	0.97	$R^2 = 0.00$ $Q_w(21)=458.34, p<0.0001$
Continent					
Europe	6	0.90	0.81	0.95	$F(4,18) = 1.43, p = 0.264$
N. America	7	0.95	0.90	0.97	$R^2 = 0.09$
S. America	2	0.94	0.82	0.98	$Q_w(18)=302.13, p<0.0001$
Asia	7	0.89	0.80	0.94	
Africa	1	0.97	0.87	0.99	
Target population 1					
Adults (>65 years)	7	0.92	0.86	0.95	$F(1,18) = 1.30, p = 0.569$
Adults (19-65 + >65 years)	13	0.94	0.91	0.96	$R^2 = 0.03$ $Q_w(18)=213.71, p<0.0001$
Target population 2					
Clinical	12	0.94	0.91	0.96	$F(2,20) = 3.99, p = 0.035$
Nonclinical	5	0.82	0.66	0.91	$R^2 = 0.27$
Mixed	6	0.92	0.86	0.96	$Q_w(20)=217.15, p<0.0001$
Target population 3					
Institutionalized	1	0.83	0.30	0.96	$F(2,20) = 4.96, p = 0.018$
Non-institutionalized	20	0.93	0.91	0.95	$R^2 = 0.28$
Mixed	2	0.71	0.18	0.89	$Q_w(20)=336.64, p<0.0001$
Rater training					
Psychotherapist	5	0.92	0.77	0.97	$F(2,6) = 0.36, p = 0.714$
Occupational therapist	2	0.85	0.18	0.97	$R^2 = 0.00$
Other	2	0.89	0.40	0.98	$Q_w(6)=121.38, p<0.0001$
Previous training					
Yes, all of them	9	0.90	0.83	0.94	$F(2,9) = 2.54, p = 0.133$
Yes, some of them	2	0.74	0.25	0.91	$R^2 = 0.24$
No	1	0.96	0.77	0.99	$Q_w(9)=97.24, p<0.0001$
Disease					$F(4,14) = 1.45, p = 0.269$
Ictus	2	0.97	0.92	.990	$R^2 = 0.09$
Multiple sclerosis	2	0.85	0.56	0.95	$Q_w(14)=216.99, p<0.0001$
Parkinson	3	0.92	0.80	0.97	
Musculoskeletal affections	3	0.93	0.82	0.94	
Various	9	0.94	0.89	0.97	

k = number of studies. α_+ = mean coefficient alpha. LL and LU = lower and upper 95% confidence limits for α_+ . F = Knapp-Hartung's statistic for testing the significance of the moderator variable. Q_w = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the moderator. In order to facilitate the interpretation, the average reliability coefficients and their respective confidence limits were back-transformed to the metric of the original coefficients.

An Explanatory Model

Although several moderators showed a statistically significant association with the coefficients alpha, most heterogeneity statistics presented in Tables 1 and 2 (Q_E for meta-regressions and Q_w for ANOVAs) suggested the presence of residual heterogeneity among

the reliability coefficients after including the moderator. Thus, a final step of this RG meta-analysis was to apply a multiple meta-regression analysis in order to identify a subset of study characteristics that could explain at least part of the variability. The predictors in the model were selected according to the results of previously performed ANOVAs and simple meta-regressions, with conceptual and statistical basis. Therefore, the explanatory model included three predictors: standard deviation of scores and a cluster of target population coded as two dummy variables for the “clinical” and “nonclinical” categories. Due to missing data in some variables, the number of studies included in the model was $k = 9$ (Table 3). The full model was statistically significant with the coefficients alpha ($p = 0.0012$), with a percentage of variance explained of 99%. Of the three predictors included in the model only the standard deviation of scores exhibited a statistically significant relationship with coefficient alpha once the influence of the other variable was controlled ($p = 0.0007$). In addition, the increase in R^2 when the target population cluster was included in the model was 0%, although statistically significant ($p = 0.0347$). It indicates the large influence of the standard deviation of the scores on the reliability estimates. Thus, as the psychometric theory predicts, coefficients alpha obtained in studies were larger as standard deviation increased.

Table III. Results of the multiple meta-regression applied on coefficients alpha, taking as predictors the *SD* of scores and the target population ($k = 9$).

Predictor variable	b_i	t	p	Model fit
Intercept	1.059	4.80	0.0049	$F(3, 5) = 31.07, p = 0.0012$
<i>SD</i> of scores	0.132	7.39	0.0007	$R^2 = .99$
Population (Clinical)	0.199	1.31	0.2463	$Q_E(5) = 4.34, p = 0.5015$
Population (Nonclinical)	-0.006	-0.03	0.9739	

Model	F	p	R^2	ΔR^2
Full model	31.07	0.0012	0.99	-
<i>SD</i> of scores	87.51	< 0.0001	0.99	0.74
Population	3.995	0.0347	0.25	0.00

b_j = regression coefficient of each predictor. t = statistic for testing the significance of the predictor (with 5 degrees of freedom). p = probability level for the t statistic. F = Knapp-Hartung's statistic for testing the significance of the full model. Q_E = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the predictors. ΔR^2 = increase in R^2 as consequence of including in the model a predictor once the other predictors had already been introduced. In order to facilitate the interpretation, the regression coefficients were back-transformed to the metric of the original coefficients.

Study 2: Intraclass (intrar-rater) coefficients meta-analysis

Mean Reliability and Heterogeneity

The 48 samples reported an intraclass correlation in order to estimate the intrar-rater reliability yielded a mean of 0.957 (95% CI = 0.94 and 0.97) and ranged from 0.59 to 0.99. Intraclass coefficients presented large heterogeneity, $Q(47) = 364.138$, $p < 0.0001$; $I^2 = 88.9\%$, as also seen in Figure 2. The next step consisted of analyzing the influence of potential moderators that can explain at least part of the variability found.

Fig 3. Forest plot displaying the intraclass correlation coefficients (and 95% confidence intervals) for intrarrater agreement.

Moderator Analyses

Tables 4 and 5 present the results of the moderator analyses for quantitative and categorical variables respectively, on intraclass correlations in terms of intrar-rater agreement.

Table IV. Results of the simple meta-regression applied on intraclass correlations (intrarrater agreement) taking continuous moderator variables as predictors.

Predictor variable	<i>K</i>	<i>b_j</i>	<i>F</i>	<i>p</i>	<i>Q_E</i>	<i>R²</i>
Mean scores	21	-0.0198	3.31	0.085	114.709***	0.10
<i>SD</i> of scores	19	0.0500	21.64	0.000	41.23**	0.61
Sample size	48	0.0068	3.63	0.063	312.49***	0.11
Mean age (years)	38	-0.0063	1.03	0.318	299.60***	0.05
<i>SD</i> of age (years)	37	0.0096	0.11	0.747	313.64***	0.00
Gender (% male)	32	-0.0072	2.65	0.114	263.48***	0.08
Mean of disorder history (years)	13	-0.0321	0.33	0.580	99.36***	0.00
<i>SD</i> of disorder history (years)	13	0.0060	0.01	0.948	100.31***	0.00
Year of the study	48	-0.0196	1.88	0.177	357.19***	0.03
Number of raters	38	-0.0208	0.46	0.504	261.38***	0.00
Time interval (in weeks)	35	-0.1119	0.76	0.389	217.17***	0.00

k = number of studies. *b_j* = regression coefficient of each predictor. *F* = Knapp-Hartung's statistic for testing the significance of the predictor (the degrees of freedom for this statistic are 1 for the numerator and *k* – 2 for the denominator). *p* = probability level for the *F* statistic. *Q_E* = statistic for testing the model misspecification. *R²* = proportion of variance accounted for by the predictor. In order to facilitate the interpretation, the regression coefficients were back-transformed to the metric of the original coefficients.

***p* < .001; *** *p* < .0001.

Table V. Results of the weighted ANOVAs applied on intraclass correlations (intrarrater agreement), taking qualitative moderator variables as independent variables.

Variable	K	α_+	95% CI		ANOVA results
			LL	LU	
Test versión					
Original	21	0.96	0.92	0.97	$F(2,45) = 0.67, p = 0.517$
Other	22	0.96	0.92	0.97	$R^2 = 0.00$
Mixed	5	0.93	0.83	0.98	$Q_w(45)=346.34, p<0.0001$
Design					
Observational	42	0.96	0.94	0.97	$F(1,46) = 0.13, p = 0.719$
Experimental	6	0.96	0.91	0.99	$R^2 = 0.00$
					$Q_w(46)=362.44, p<0.0001$
Study focus					
Psychometric	38	0.95	0.93	0.97	$F(1,46) = 1.69, p = 0.199$
Applied	10	0.97	0.94	0.99	$R^2 = 0.02$
					$Q_w(46)=345.17, p<0.0001$
Continent					
Europe	15	0.94	0.90	0.97	$F(5,42) = 1.21, p = 0.320$
N. America	15	0.96	0.92	0.98	$R^2 = 0.04$
S. America	2	0.99	0.92	0.99	$Q_w(42)=316.72, p<0.0001$
Asia	11	0.97	0.95	0.99	
Oceania	1	0.96	0.71	0.99	
Multicenter	5	0.93	0.83	0.98	
Target population 1					
Children and Adolescents	1	0.99	0.97	0.99	$F(3,38) = 2.25, p = 0.099$
Adults (19-65 years)	7	0.94	0.85	0.97	$R^2 = 0.12$
Adults (>65 years)	10	0.95	0.91	0.98	$Q_w(38)=263.39, p<0.0001$
Adults (19-65 + >65 years)	24	0.96	0.94	0.98	
Target population 2					
Clinical	30	0.96	0.93	0.97	$F(2,45) = 0.33, p = 0.721$
Nonclinical	10	0.97	0.93	0.98	$R^2 = 0.00$
Mixed	8	0.95	0.88	0.98	$Q_w(45)=353.72, p<0.0001$
Target population 3					
Institutionalized	4	0.94	0.82	0.98	$F(2,45) = 0.29, p = 0.744$
Non-institutionalized	40	0.96	0.94	0.97	$R^2 = 0.00$
Mixed	4	0.96	0.88	0.99	$Q_w(45)=360.11, p<0.0001$
Rater training					
Psychotherapist	29	0.95	0.93	0.97	$F(1,29) = 1.37, p = 0.252$
Other	2	0.98	0.92	0.99	$R^2 = 0.03$
					$Q_w(29)=205.06, p<0.0001$
Previous training					
Yes, all of them	20	0.96	0.93	0.98	$F(2,21) = 0.73, p = 0.496$
Yes, some of them	3	0.96	0.86	0.99	$R^2 = 0.00$
No	1	0.99	0.89	0.99	$Q_w(21)=147.12, p<0.0001$
Berg Scale					
BBC	33	0.95	0.94	0.97	$F(1,36) = 0.76, p = 0.388$
Other scale	5	0.93	0.84	0.97	$R^2 = 0.00$
					$Q_w(36)=195.28, p<0.0001$
Disease					
Ictus	9	0.97	0.93	0.99	
Multiple sclerosis	4	0.91	0.75	0.97	$F(7,30) = 1.61, p = 0.172$
Parkinson	6	0.92	0.81	0.97	$R^2 = 0.20$
Huntington's disease	5	0.94	0.83	0.98	$Q_w(30)=191.61, p<0.0001$
Visual problems	2	0.93	0.71	0.98	

Musculoskeletal affections	3	0.93	0.72	0.98
Other neurological diseases	4	0.98	0.95	0.99
Various	5	0.98	0.94	0.99

k = number of studies. α_+ = mean coefficient alpha. LL and LU = lower and upper 95% confidence limits for α_+ . F = Knapp-Hartung's statistic for testing the significance of the moderator variable. Q_W = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the moderator. In order to facilitate the interpretation, the average reliability coefficients and their respective confidence limits were back-transformed to the metric of the original coefficients.

As seen in Table 4, of the different quantitative moderator variables analyzed, only the standard deviation of the test scores exhibited statistical significance with intraclass correlations ($p < 0.001$) with a large amount of variance accounted for of 61%. In addition, the mean of scores and sample size achieve statistically significant results ($p = 0.085$ and $p = 0.063$, respectively) with a percentage of variance explained of approximately 10% in both cases. The positive sign of the regression slope for these three moderators indicated larger intraclass correlations as the standard deviation and mean of scores, and sample size increased.

Regarding qualitative moderators, marginally significant differences were found when comparing the average intraclass correlations grouped by the target population 1 ($p = .099$) with a 12% of variance explained. The only study focused on children and adolescents showed the largest intraclass correlation, whereas studies focused on adults aged from 19 to 65 years showed the lowest (mean = 0.94).

An Explanatory Model

The moderator analyses showed that some characteristics of the studies were influenced on intraclass correlations in terms of intra-rater agreement. Nevertheless, the Q_E and Q_W statistics revealed that none of the models was correctly specified (see Tables 4 and 5). Therefore, a

multiple meta-regression analysis was carried out aiming to propose a set of moderators explaining a large amount of variance.

Of the moderators analyzed, only the standard deviation of the scores showed a statistically significant result, although others such as the mean of the scores and the sample size also exhibited a marginally significant relationship with intraclass correlation with an important percentage of variance accounted for. Thus, these three moderator variables were included in the meta-regression model: standard deviation, mean of scores, and sample size. Due to missing data in some variables, the final number of studies in this meta-regression model was $k = 19$. Table 6 presents the results of the multiple meta-regression. The full model reached statistical significance ($p = 0.0001$) and its percentage of variance explained was equal to 81%.

Table VI. Results of the multiple meta-regression applied on intraclass correlations (intrarrater agreement), taking as predictors the *SD* and mean of scores and the sample size ($k = 19$).

Predictor variable	b_j	t	p	Model fit
Intercept	-1.7986	-1.84	0.0859	$F(3, 15) = 14.09, p = 0.0001$
<i>SD</i> of scores	0.1105	4.68	0.0003	$R^2 = 0.81$
Mean scores	0.0528	3.22	0.0057	$Q_E(15) = 23.78, p = 0.0689$
Sample size	0.0058	2.14	0.0495	

Model	F	p	R^2	ΔR^2
Full model	14.09	0.0001	0.81	-
<i>SD</i> of scores	21.64	0.0002	0.61	0.60
Mean scores	3.31	0.0845	0.10	0.09
Sample size	3.63	0.0629	0.11	0.10

b_j = regression coefficient of each predictor. t = statistic for testing the significance of the predictor (with 15 degrees of freedom). p = probability level for the t statistic. F = Knapp-Hartung's statistic for testing the significance of the full model. Q_E = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the predictors. ΔR^2 = increase in R^2 as consequence of including in the model a predictor once the other predictors had already been introduced. In order to facilitate the interpretation, the regression coefficients were back-transformed to the metric of the original coefficients.

All three predictors included in the model (once the influence of the other predictors was controlled), show a statistically significant and positive relationship with intraclass

correlation: standard deviation of scores ($p = 0.0003$), mean of scores ($p = 0.0057$), and sample size ($p = 0.0495$). Table 6 also shows the increase in the percentage of variance accounted for by each predictor once the others had already been introduced in the model. As expected, the most notable increment in R^2 was found when the standard deviation of scores was introduced in the model ($\Delta R^2 = 60\%$).

Study 3: Intraclass (inter-rater) coefficients meta-analysis

Mean Reliability and Heterogeneity

The 39 intraclass correlations reported to estimate interrater reliability yielded a mean of 0.97 (95% CI = 0.95 and 0.98) with a minimum intraclass correlation of 0.72 and a maximum intraclass correlation of 0.99. Significant heterogeneity was found among intraclass correlations, $Q(38) = 367.56$, $p < 0.0001$, with an I^2 value of 89.83%. This large heterogeneity is also reflected in Figure 3.

Fig 3. Forest plot displaying the intraclass correlation coefficients (and 95% confidence intervals) for interrater agreement.

Moderator Analyses

Tables 7 and 8 show the results of the moderator analyses for quantitative and qualitative variables respectively, on intraclass correlations in terms of inter-rater agreement. None of the moderators included yielded a statistically significant relationship with the intraclass correlations, not even the standard deviation of scores, as the psychometric theory predicts. In addition, the percentage of variance accounted for was 0% in most moderator variables. No marginally significant results were found.

Table VII. Results of the simple meta-regression applied on intraclass correlations (interrater agreement) taking continuous moderator variables as predictors.

Predictor variable	<i>K</i>	<i>b_j</i>	<i>F</i>	<i>p</i>	<i>Q_E</i>	<i>R²</i>
Mean scores	14	-0.0205	1.15	0.305	155.53***	0.01
<i>SD</i> of scores	12	0.0336	0.96	0.351	146.41***	0.00
Sample size	39	-0.0040	1.15	0.292	367.54***	0.00
Mean age (years)	27	-0.0079	1.56	0.223	206.79***	0.00
<i>SD</i> of age (years)	25	0.0243	0.57	0.458	207.01***	0.00
Gender (% male)	29	-0.0067	1.50	0.231	224.47***	0.03
Mean of disorder history (years)	12	-0.0037	0.01	0.922	62.26***	0.00
<i>SD</i> of disorder history (years)	12	0.0132	0.08	0.784	59.27***	0.00
Year of the study	39	0.0101	0.44	0.509	362.79***	0.00
Number of raters	35	-0.0091	0.47	0.497	332.51***	0.00
Time interval (in weeks)	9	-0.0534	0.46	0.521	43.33***	0.00

k = number of studies. *b_j* = regression coefficient of each predictor. *F* = Knapp-Hartung's statistic for testing the significance of the predictor (the degrees of freedom for this statistic are 1 for the numerator and *k* - 2 for the denominator). *p* = probability level for the *F* statistic. *Q_E* = statistic for testing the model misspecification. *R²* = proportion of variance accounted for by the predictor. In order to facilitate the interpretation, the regression coefficients were back-transformed to the metric of the original coefficients.

***p* < 0.001; *** *p* < 0.0001.

Table VIII. Results of the weighted ANOVAs applied on intraclass correlations (interrater agreement), taking qualitative moderator variables as independent variables.

Variable	<i>K</i>	α_+	95% CI		ANOVA results
			LL	LU	
Test versión					
Original	20	0.96	0.93	0.98	$F(1,37) = 0.41, p = 0.526$
Other	19	0.97	0.95	0.98	$R^2 = 0.00$
					$Q_w(37)=352.51, p<0.0001$
Design					
Observational	34	0.97	0.95	0.98	$F(1,37) = 0.11, p = 0.741$
Experimental	5	0.97	0.92	0.99	$R^2 = 0.00$
					$Q_w(37)=362.84, p<0.0001$
Study focus					
Psychometric	32	0.96	0.94	0.97	$F(1,37) = 2.20, p = 0.146$
Applied	7	0.98	0.95	0.99	$R^2 = 0.03$
					$Q_w(37)=337.07, p<0.0001$
Continent					
Europe	12	0.98	0.96	0.99	$F(3,35) = 1.27, p = 0.300$
North America	16	0.95	0.91	0.97	$R^2 = 0.05$
South America	2	0.94	0.74	0.99	$Q_w(35)=292.39, p<0.0001$
Asia	9	0.97	0.94	0.99	
Target population 1					
Children and Adolescents	1	0.99	0.92	0.99	$F(3,26) = 0.99, p = 0.413$
Adults (19-65 years)	4	0.97	0.91	0.99	$R^2 = 0.00$
Adults (>65 years)	9	0.95	0.90	0.98	$Q_w(26)=232.22, p<0.0001$
Adults (19-65 + >65 years)	16	0.97	0.95	0.99	
Target population 2					
Clinical	23	0.97	0.95	0.98	$F(2,36) = 0.49, p = 0.619$
Nonclinical	9	0.96	0.90	0.98	$R^2 = 0.00$
Mixed	7	0.97	0.94	0.99	$Q_w(36)=338.66, p<0.0001$
Target population 3	2	.902	0.59	0.98	$F(2,36) = 1.42, p = 0.744$

Institutionalized	34	.969	0.95	0.98	$R^2 = 0.01$
Non-institutionalized	3	.943	0.81	0.98	$Q_w(36)=348.84, p<0.0001$
Mixed					
Rater training					
Psychotherapist	22	0.97	0.94	0.98	$F(2,25) = 0.55, p = 0.581$
Occupational therapist	3	0.93	0.74	0.98	$R^2 = 0.00$
Other	3	0.97	0.87	0.99	$Q_w(25)=256.07, p<0.0001$
Previous training					
Yes, all of them	19	0.96	0.93	0.98	$F(2,20) = 0.15, p = 0.860$
Yes, some of them	3	0.95	0.76	0.99	$R^2 = 0.00$
No	1	0.98	0.64	0.99	$Q_w(20)=282.55, p<0.0001$
Berg Scale					
BBC	30	0.96	0.94	0.98	$F(1,30) = 0.03, p = 0.871$
Other scale	2	0.96	0.77	0.99	$R^2 = 0.00$
					$Q_w(30)=299.33, p<0.0001$
Disease					
Ictus	7	0.97	.91	0.99	$F(5,24) = 0.33, p = 0.887$
Multiple sclerosis	4	0.97	.87	0.99	$R^2 = 0.00$
Parkinson	5	0.95	.82	0.99	$Q_w(24)=230.936, p<0.0001$
Musculoskeletal affections	4	0.97	.89	0.99	
Other neurological diseases.	3	0.98	.93	0.99	
Various	7	0.97	.92	0.99	

k = number of studies. α_+ = mean coefficient alpha. LL and LU = lower and upper 95% confidence limits for α_+ . F = Knapp-Hartung's statistic for testing the significance of the moderator variable. Q_w = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the moderator. In order to facilitate the interpretation, the average reliability coefficients and their respective confidence limits were back-transformed to the metric of the original coefficients.

Discussion

As reliability is not a property of the test itself, but of the test scores obtained in each application, reliability generalization studies allow examining how reliability fluctuates through different test administrations, and to guide reliability expectations for future test applications³⁰. In this study we have examined the average reliability in addition to the sources of variability that may have influenced the reliability generalization of Berg Balance Scale in its threefold aspect: alpha coefficient, inter-rater and intra-rater agreement, the latter two evaluated through the intraclass correlation coefficient, with a low reliability reporting rate (5.3%), so that the rest of the studies chose to induce reliability from previous studies or did not make any reference to the reliability of the Berg Balance Scale scores.

The most commonly reported reliability coefficient was the intraclass coefficient (intra-raters, $n = 48$), with a weighted average for the total score of 0.96. The intraclass coefficient (inter-raters) and the alpha coefficient exhibited averages of 0.97 and 0.92, respectively. In general, the three types of reliability coefficients denoted that Berg Balance Scale produces reliable scores to evaluate balance, regardless of whether internal consistency is studied or applied, either by a rater on more than one occasion (intra-raters) or by two or more raters (inter-raters), and clearly exceed the cut-off score of 0.80 generally considered as the recommended minimum reliability on a clinical application scale³⁰.

Our results also show a wide variability between reliability estimates. Some characteristics of the studies showed a statistically significant relationship with the coefficient alpha (internal consistency) and with the intraclass coefficient (intra-rater agreement), but not with the intraclass coefficient (inter-raters agreement).

The moderator variable exhibiting the strongest relationship to the coefficient alpha was the standard deviation of the total scores as predicted by the psychometric theory³⁰. The explanatory model determined the large influence of standard deviation on reliability estimation through the internal consistency of the Berg Balance Scale, indicating that the physiotherapist must be aware that high group homogeneity can determine a low reliability coefficient in the application of Berg Balance Scale.

Other moderating variables that exhibited a statistically significant relationship with the coefficient alpha were population type (clinical vs. nonclinical) and whether the population was institutionalized. Reliability is higher in the clinical population than in the mixed or non-

clinical population and non-institutionalized population than in the institutionalized or mixed population.

In the case of the intraclass coefficient (intra-raters), again, the SD of scores had a high influence on reliability, as predicted by the psychometric theory. In addition, the sample size and the mean of the Berg Balance Scale results were highly relevant in evaluating reliability of scores, denoting that a greater number of individuals to evaluate caused an improvement in the Berg Balance Scale intra-rater reliability surely owing to an improvement in the practice of evaluating individuals. Moreover, although with only one sample, the reliability coefficient was higher when the balance was evaluated in children/adolescents than in other groups of adults, although in all cases reliability was above 0.90.

Finally, the inter-rater agreement had no significant influence on any moderating variables, revealing that reliability scores were independent of the characteristics used in this meta-analysis hence, practitioners could rely on the reliability coefficient (inter-rater) and will not vary in the different conditions where the scale can be applied.

Limitations

Although this meta-analysis of reliability allowed identifying some characteristics of the studies that were statistically related to the reliability of the Berg Balance Scale scores, there is clear evidence of residual variability that remains unexplained regarding the coefficient alpha and the intra-rater agreement, however not as regards the inter-rater. Another difficulty in this study was the presence of categories in the moderator variables with only one estimate, since in these conditions statistical analyzes are very unstable.

Finally, it is important to point out that only 5.3% of studies that applied Berg Balance Scale reported a reliability coefficient (alpha or intraclass). Most studies applying Berg Balance Scale induced reliability from previous applications or did not mention it in the scientific report. Therefore, our conclusions should be taken with caution, as the number of studies included in the meta-analysis was very low compared to the total number of studies found in the literature that have applied Berg Balance Scale. This incorrect practice of inducing reliability not only affects Berg Balance Scale but also any outcome measure³⁰. Health professionals must be aware that reliability is not a stable scale property, but should always be estimated when applying a psychometric instrument.

The Berg Balance Scale has some limitations documented to assess balance performance, such as the lack of assessing dynamic balance (eg, reacting to a perturbation, gait), a ceiling effect and a low responsiveness^{13,49,64,98}. Currently there are other tools for assessing balance impairments such as the BESTest and the Mini-BESTest, ideal as measures of balance deficits^{49,99}. However, the BBS is the most widely used instrument to quantitatively assess balance in neurological conditions from 2009 when BESTest was published to the present, according to a search performed by in PubMed to this date.

Conclusions

The alpha coefficient and intra- and inter-rater agreement for BBS scores was very satisfactory. Several characteristics of the studies were statistically associated to the alpha coefficient and with intra-rater reliability. The SD of the Berg Balance Scale scores presented statistically significant relationships with the coefficient alpha and with *ICC* (intra-rater). In addition, the clinical population and the institutionalized population presented statistically

significant relationships with the coefficient alpha. The sample size and mean scale results were also statistically related to ICCs (intra-rater).

References

1. Shumway-Cook A, Woollacott M. Motor Control: Translating Research into Clinical Practice. Lippincott Williams & Wilkins; 2007.
2. Chen CL, Yeung KT, Wang CH, Chu HT, Yeh CY. Anterior ankle-foot orthosis effects on postural stability in hemiplegic patients. *Arch Phys Med Rehabil* 1999; 80(12): 1587-92.
3. Wing MA, Goodrich S, Virji-Babul N, Jenner JR, Clap S. Balance evaluation in hemiparetic stroke patients using lateral forces applied to the hip. *Arch Phys Med Rehabil* 1993; 74(3): 292-9.
4. De Heart M, Geurts AC, Huidekoper SC, Fasotti L, Van Limbeek J. Recovery of standing balance in postacute stroke patients: A rehabilitation cohort study. *Arch Phys Med Rehabil* 2004; 85(6): 886-89.
5. Daubney ME, Culham EG. Lower-extremity muscle force and balance performance in adults aged 65 years and older. *Physical Therapy* 1999;79(12):1177–85.
6. Berg K, Wood-Dauphinee S, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. *Physiother Can* 1989; 41:304-11.
7. Mills EM. The effect of low-intensity aerobic exercise on muscle strength, flexibility, and balance among sedentary elderly persons. *Nursing Research* 1994;43(4):207–11.
8. Rossat A, Fantino B, Nitenberg C, Annweiler C, Poujol L, Herrmann FR, et al. Risk factors for falling in community-dwelling older adults: which of them are associated with the recurrence of falls?. *J Nutr Health Aging* 14(9): 787-91.

9. Berg KO, Wood-Dauphinee S, Williams JI. The Balance Scale: reliability assessment with elderly residents and patients with acute stroke. *Scand J Rehabil Med* 1995; 27:27-36.
10. Bhatt T, Espy D, Yang F, Pai YC. Dynamic gait stability, clinical correlates, and prognosis of falls among community- dwelling older adults. *Arch Phys Med Rehabil* 2011; 92:799-805.
11. Desai A, Goodman V, Kapadia N, Sahay BL, Szturm T. Relationship between dynamic balance measures and functional performance in community dwelling elderly people. *Phys Ther* 2010; 90:748-70.
12. 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;28:248–53.
13. Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. *Phys Ther* 2008; 88:559-66.
14. Wang CH, Hsueh IP, Sheu CF, Yao G, Hsieh CL. Psychometric properties of 2 simplified 3-level balance scales used for patients with stroke. *Phys Ther* 2004; 84:430-8.
15. Mao HF, Hsueh IP, Tang PF, Sheu CF, Hsieh CL. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke* 2002; 33:1022-7.
16. Hong S-J, Goh EY, Chua SY, Ng SS. Reliability and validity of step test scores in subjects with chronic stroke. *Arch Phys Med Rehabil* 2012; 93: 1065-71.
17. Brusse KJ, Zimdars S, Zalewski KR, Steffen TM. Testing functional performance in people with Parkinson disease. *Phys Ther* 2005; 85:134-41.

18. Franchignoni F, Martignoni E, Ferriero G, Pasetti C. Balance and fear of falling in Parkinson's disease. *Parkinsonism Relat Disord* 2005; 11:427-33.
19. Newstead AH, Hinman MK, Tomberlin JA. Reliability of the Berg balance scale and balance master limits of stability tests for individuals with brain injury. *J Neurol Phys Ther* 2005; 29:18-23.
20. Cattaneo D, Jonsdottir J, Zocchi M, Regola A. Effects of balance exercises on people with multiple sclerosis: a pilot study. *Clin Rehabil* 2007; 21:771-81.
21. Saiz-Llamas JR, Casado-Vicente V, Martos-Álvarez HC. Impacto de un programa de fisioterapia en atención primaria en las personas mayores, con antecedentes de caídas. *Fisioterapia* 2014; 36:103-9.
22. Conradsson M, Lundin-Olsson L, Lindelöf N, Littbrand H, Malmqvist L, Gustafson Y, et al. Berg Balance Scale: intrarater test-retest reliability among older people dependent in activities of daily living and living in residential care facilities. *Phys Ther* 2007; 87:1155-63.
23. Hwang S, Woo Y. Assessment of the Influence of Balance on Gait of Persons with Stroke. *J Phys Ther Sci* 2012; 24:249-52.
24. Cakit BD, Saracoglu M, Genc H, Erdem HR, Inan L. The effects of incremental speed-dependent treadmill training on postural instability and fear of falling in Parkinson's disease. *Clin Rehabil* 2007; 21:698.
25. Ottonello M, Ferriero G, Benevolo E, Sessarego P, Dughi D. Psychometric evaluation of the italian version of the Berg Balance Scale in rehabilitation inpatients. *Eur Med Phys* 2003; 39:181-9.
26. Miyamoto ST, Lombardi Junior I, Berg KO, Ramos LR, Natour J. Brazilian version of the Berg balance scale. *Braz J Med Biol Res* 2004; 37:1411-21.

27. Halsaa KE, Brovold T, Graver V, Sandvik L, Bergland A. Assessments of interrater reliability and internal consistency of the Norwegian version of the Berg Balance Scale. *Arch Phys Med Rehabil* 2007; 88:94-8.
28. Sahin F, Yilmaz F, Ozmaden A, Kotevolu N, Sahin T, Kuran B. Reliability and validity of the Turkish version of the Berg Balance Scale. *J Geriatr Phys Ther* 2008; 31:32-7.
29. Azad A, Taghizadeh G, Khaneghini A. Assessments of the reliability of the Iranian version of the Berg Balance Scale in patients with multiple sclerosis. *Acta Neurol Taiwan* 2011;20:22-8.
30. Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use. New York: Oxford University Press; 2008.
31. Anastasi A, Urbina S. Psychological testing. 7th ed. Upper Saddle River (NJ): Prentice Hall; 1997.
32. Nunally JC. Reliability of measurement. In: Mitzel HE, editor. *Encyclopedia of educational research* (NJ): FreePress; 1982. p. 1589-1601.
33. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. *Can J Public Health* 1992;83(2): S7-11.
34. 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; 91:102-13.
35. 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):733-46.
36. Flansbjer UB, Blom J, Brogårdh C. The reproducibility of Berg Balance Scale and the Single-leg Stance in chronic stroke and the relationship between the two tests. *PMR* 2012;4:165-70.

37. Rowley GL. The reliability of observational measures. *Am Educ Res J* 1976;13(1): 51–9.
38. Vacha-Haase T, Kogan LR, Thompson B. Sample compositions and variabilities in published studies versus those in test manuals. *Educ Psychol Meas* 2000;60(4):509–522.
39. Vacha-Haase T. Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educ Psychol Meas* 1998; 58(1): 6-20.
40. Yin P, Fan X. Assessing the reliability of Beck Depression Inventory scores: Reliability generalization across studies. *Educ Psychol Meas* 2000;60(2):201–23.
41. Barnes LLB, Harp D, Jung WS. Reliability generalization of scores on the Spielberger State-Trait Anxiety Inventory. *Educ Psychol Meas* 2002;62(4):603–18.
42. Bonett DG. Sample size requirements for testing and estimating coefficient alpha. *J Educ Behav Stat* 2002;27(4):335-40.
43. Hartung J. An alternative method for meta-analysis. *Biom J* 1999; 41:901-16.
44. Sánchez-Meca J, Marín-Martínez F. Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychol Methods* 2008; 13:31-48.
45. López-López JA, Botella J, Sánchez-Meca J, Marín-Martínez F. Alternatives for mixed-effects meta-regression models in the reliability generalization approach: A simulation study. *J Educ Behav Stat* 2013; 38:443-69.
46. Knapp G, Hartung J. Improved test for a random effects meta-regression with a single covariate. *Stat Med* 2003; 22:2693-710.
47. Viechtbauer W. Conducting meta-analysis in R with the metaphor package. *J Stat Softw* 2010; 36:1-48.
48. Salavati M, Negahban H, Mazaheri M, Soleimanifar M, Hadadi M, Sefiddashti L, et al. The Persian version of the Berg Balance Scale: inter and intra-rater reliability and construct validity in elderly adults. *Disabil Rehabil* 2012; 34 (20): 1695-8.

49. Godi M, Franchignoni F, Caligari M, Giordano A, Turcato AM, Nardone A. Comparison of reliability, validity, and responsiveness of the Min-BESTest and Berg Balance Scale in patients with balance disorders. *Phys Ther* 2013; 93: 158-17.
50. Matsushima M, Yabe I, Uwatoko H, Shirai S, Hirotani M, Sasaki H. Reliability of the Japanese Version of the Berg Balance Scale. *Intern Med* 2014; 53(15): 1621-24.
51. Chan ACM, Pang MYC. Assessing balance function in patients with total knee arthroplasty. *Phys Ther* 2015; 95: 1397-407.
52. Wang CY, Hsieh CL, Olson SL, Wang CH, Sheu CF, Liang CC. 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.
53. Scalzo PL, Nova IC, Perracini MR, Sacramento DR, Cardoso F, Ferraz HB, et al. Validation of the brazilian version of the Berg Balance Scale for patients with parkinson's disease. *Arq Neuropsiquiatr* 2009;67(3-B):831-5.
54. Major MJ, Fatone S, Roth EJ. Validity and reliability of the Beg Balance Scale for community-dwelling persons with lower-limb amputation. *Arch Phys Med Rehabil* 2013; 94: 2194-202.
55. Baer HR, Wolf SL. Modified Emory Functional Ambulation Profile: an outcome measure for the rehabilitation of poststroke gait dysfunction. *Stroke* 2001; 32: 973-9.
56. Wolf B, Feys H, De Weerd W, van der Meer J, Noom M, Aufdemkampe G. Effect of a physical therapeutic intervention for balance problems in the elderly: a single-blind randomized, controlled multicentre trial. *Clin Rehabil*. 2001; 15: 624-36.
57. Lim LI, van Wegen EE, de Goede CJ, Jones D, Rochester L, Hetherington V, et al. Measuring gait and gait-related activities in parkinson's patients own home environment: a reliability, responsiveness and feasibility study. *Parkinsonism Relat Disord* 2005; 11: 19-24.

58. Paltamaa J, West H, Sarasoja T, Wikström J, Mälkiä E. Reliability of physical functioning measures in ambulatory subjects with MS. *Physiother Res Int* 2005; 10 (2): 93-109.
59. Sackley C, Richarson P, McDonnell K, Ratib S, Dewey M, Hill HJ. The reliability of balance, mobility and self-care measures in a population of adults with a learning disability known to a physiotherapy service. *Clin Rehabil* 2005; 19: 216-23.
60. Southard V, Dave M, Davis MG, Blanco J, Hofferber A. The multiple tasks test as a predictor of falls in older adults. *Gait Posture* 2005; 22: 351-5.
61. Combs SA, Dugan EL, Passmore M, Riesner C, Whipker D, Yingling E, et al. Balance, balance confidence, and health-related quality of life in persons with chronic stroke after body weight-supported treadmill training. *Arch Phys Med Rehabil* 2010; 91: 1914-9.
62. Wong CK. Interrater reliability of the Berg Balance Scale when used by clinicians of various experience levels to assess people with lower limb amputations. *Phys Ther* 2014; 4: 94: 371-8.
63. Muir-Hunter SW, Graham L, Odasso MM. Reliability of the Berg Balance Scale as a clinical measure of balance in community-dwelling older adults with mild to moderate Alzheimer disease: a pilot study. *Physiother Can* 2015; 67(3): 255-62.
64. Schlenstedt C, Brombacher S, Hartwigsen G, Weiser B, Möller B, Deuschl G. 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; 96:218-25.
65. Jantakat C, Ramrit S, Emasithi A, Siritaratiwat W. Capacity of adolescents with cerebral palsy on paediatric balance scale and Berg balance scale. *Res Dev Disab* 2015; 36; 72-7.
66. Nemmers TM, Miller JW. Factors influencing balance in healthy community-dwelling women age 60 and older. *J Geriatr Phys Ther* 2008; 31(3): 93-100.

67. Heaney JLJ, Philips AC, Carroll D. Ageing, physical function, and the diurnal rhythms of cortisol and dehydroepiandrosterone. *Psychoneuroendocrinology* 2012; 37: 341-49.
68. La Porta F, Caselli S, Susassi S, Cavallini P, Tennant A, Franceschini M. Is the Berg Balance Scale and internally valid and reliable measure of balance across different etiologies in neurorehabilitation?. A revisited Rasch analysis study. *Arch Phys Med Rehabil* 2012; 93: 1209-16.
69. Wong CK, Chen CC, Welsh J. Preliminary assessment of balance with the Berg Balance Scale in adults who have a leg amputation and dwell in the community: Rasch Rating Scale Analysis. *Phys Ther* 2013; 93: 1520-9.
70. Puckree T, Naidoo P. Balance and stability-focused exercise program improves stability and balance in patients after acute stroke in a resource-poor setting. *PM R* 2014; 6 (12): 1081-7.
71. Azimzadeh E, Hosseini MA, Nourozi K, Davidson PM. Effect of Tai Chi Chuan on balance in women with multiple sclerosis. *Complement Ther Clin Pract* 2015; 21; 57-60.
72. Lee KKM, Scudds RJ. Comparison of balance in older people with and without visual impairment. *Age Ageing* 2003; 32: 643-9.
73. Newstead AH, Hinman MK, Tomberlin JA. Reliability of the Berg balance scale and balance master limits of stability tests for individuals with brain injury. *J Neurol Phys Ther* 2005; 29:18-23.
74. Chou CY, Chien CW, Hsueh IP, Sheu CF, Wang CH, Hsieh CL. Developing a Short Form of the Berg Balance Scale for people with stroke. *Phys Ther* 2006; 86:195-204.
75. Conradsson M, Lundin-Olsson L, Lindelöf N, Littbrand H, Malmqvist L, Gustafson Y, et al. Berg Balance Scale: intrarater test-retest reliability among older people dependent in activities of daily living and living in residential care facilities. *Phys Ther* 2007; 87:1155-63.

76. English CK, Hillier SL, Stiller KR, Warden-Flood A. Circuit class therapy versus individual's physiotherapy sessions during inpatient stroke rehabilitation: a controlled trial. *Arch Phys Med Rehabil* 2007; 88: 955-63.
77. Swanenburg J, de Bruin ED, Stauffacher M, Mulder T, Uebelhart D. Effects of exercise and nutrition on postural balance and risk of falling in elderly people with decreased bone mineral density: randomized controlled trial pilot study. *Clin Rehabil* 2007; 21:523-34.
78. Liaw LJ, Hsieh CL, Lo SK, Chen HM, Lee S, Lin JH. The relative and absolute reliability of two balance performance measures in chronic stroke patients. *Disabil Rehabil* 2008; 30 (9): 656-61.
79. Cheung KKW, Au KY, Lam WWS, Jones AYM. Effects of a structured exercise programme on functional balance in visually impaired elderly living in a residential setting. *Hong Kong Physiother J* 2008; 26:45-50.
80. Combs SA, Diehl MD, Staples WH, Conn L, Davis K, Lewis N, et al. Boxing training for patients with Parkinson disease: a case series. *Phys Ther* 2011; 91:132-42.
81. Learmonth YC, Paul L, McFadyen AK, Mattison P, Miller L. Reliability and clinical significance of mobility and balance assessments in multiple sclerosis. *Int J Rehabil Res* 2012; 35 (1): 69-74.
82. Hiengkaew V, Jitaree K, Chaiyawat P. Minimal detectable changes of the Berg Balance Scale, Fugl-Meyer Assessment Scale, Timed "Up & Go" Test, gait speeds, and 2-minute walk test in individuals with chronic stroke with different degrees of ankle plantarflexor tone. *Arch Phys Med Rehabil* 2012; 93: 1201-8.
83. Rasova K, Martinkova P, Vyskotova J, Sedova M. Assessment set for evaluation of clinical outcomes in multiple sclerosis: psychometric properties. *Patient Relat Outcome Meas* 2012; 3: 59-70.

84. Tsaih PL, Shih YL, Hu MM. Low intensity task-oriented exercise for ambulation-challenged residents in long-term care facilities: a randomized, controlled trial. *Am J Phys Med Rehabil* 2012; 91; 616-24.
85. Quinn L, Khalil H, Dawes H, Fritz NE, Kegelmeyer D, Kloos AD, et al. Reliability and minimal detectable change of physical performance measures in individuals with pre-manifest and manifest Huntington disease. *Phys Ther* 2013; 93: 942-56.
86. Takacs J, Garland SJ, Carpenter MG, Hunt MA. Validity and reliability of the Community Balance and Mobility Scale in individuals with knee osteoarthritis. *Phys Ther* 2014; 94: 866-74.
87. Tamburella F, Scivoletto G, Iosa M, Molinari M. Reliability, validity, and effectiveness of center of pressure parameters in assessing stabilometric platform in subjects with incomplete spinal cord injury: a serial cross-sectional study. *J Neuroeng Rehabil* 2014; 11: 86.
88. Geroi C, Smania N, Schena F, Dimitrova E, Verzini E, Bombieri F, et al. Does the Pisa syndrome affect postural control, balance, and gait in patients with Parkinson's disease? An observational cross-sectional study. *Parkinsonism Relat Disord* 2015; 21; 736-41.
89. Norén AM, Bogren U, Bolin J, Stenström C. Balance assessment in patients with peripheral arthritis: applicability and reliability of some clinical assessments. *Physiother Res Int* 2001; 6(4): 193-204.
90. Mihay L, Iltzsche E, Tribby A, Rushing K, Spears J, Wiltfong H, et al. Balance and perceived confidence with performance of instrumental activities of daily living: a pilot study of Tai Chi inspired exercise with elderly retirement-community dwellers. *Phys Occup Ther Geriatr* 2003; 21 (3): 75-86.

91. Smith PS, Hembree JA, Thompson ME. Berg balance scale and functional reach: determining the best clinical tool for individuals post acute stroke. *Clin Rehabil* 2004; 18: 811-8.
92. Williams B, Doherty NL, Bender A, Mattox H, Tibbs JR. The effect of Nintendo wii on balance: a pilot study supporting the use of the wii in occupational therapy for the well elderly. *Occup Ther Health Care* 2011; 25 (2-3): 131-9.
93. Toomey E, Coote S. Between-rater reliability of the 6-Minute Walk Test, Berg Balance Scale, and Handheld Dynamometry in people with multiple sclerosis. *Int J MS Care* 2013; 15(1): 1-6.
94. Srisim K, Saengsuwan J, Amatachaya S. Functional assessments for predicting a risk of multiple falls in independent ambulatory patients with spinal cord injury. *J Spinal Cord Med* 2015; 38: 4; 439-45.
95. Liston RAL, Brouwer BJ. Reliability and validity of measures obtained from stroke patients using the Balance Master. *Arch Phys Med Rehabil* 1996; 77: 425-30.
96. Schachten T, Jansen P. The effects of golf training in patients with stroke: a pilot study. *Int Psychogeriatr* 2015; 27(5): 865-73.
97. Stevenson TJ. Detecting change in patients with stroke using the Berg Balance Scale. *Aust J Physiother* 2001; 47: 29-38.
98. Portnoy S, Frechtel A, Raveh E, Schwartz I. Prevention of genu recurvatum in poststroke patients using a hinged soft knee orthosis. *PM R* 2015; 7(10): 1042-51.
99. King LA, Priest KC, Salarian A, Pierce D, Horak FB. Comparing the Mini-BESTest with the Berg Balance Scale to Evaluate Balance Disorders in Parkinson's Disease. *Parkinsons Dis* 2012; 2012:375419.
100. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Phys Ther* 2009; 89: 484-98.

Supporting information

SI Table. Search strategy.

MEDLINE (PubMed) search strategy

("Berg test"[Title/Abstract] OR "Berg balance measure"[Title/Abstract] OR "Berg balance"[Title/Abstract] OR "Berg scale"[Title/Abstract] OR "Berg balance scale"[Title/Abstract])

Limit:

1. Publication date from 1989/11/01 to 2015/12/31

EMBASE search strategy

"Berg test":ti,ab OR "Berg balance measure":ti,ab OR "Berg balance":ti,ab OR "Berg scale":ti,ab OR "Berg balance scale":ti,ab

Limits:

1. Publication date from 1989/11/01 to 2015/12/31
2. To exclude papers from Medline

CINHAL search strategy

TI("Berg test" OR "Berg balance measure" OR "Berg balance" OR "Berg scale" OR "Berg balance scale") OR AB ("Berg test" OR "Berg balance measure" OR "Berg balance" OR "Berg scale" OR "Berg balance scale")

Limits:

1. Publication date from 1989/11/01 to 2015/12/31
2. To exclude papers from Medline

SII Table. Excluded studies.

Study	Reason for Exclusion
Li F, Harmer P, Fisher KJ, Mcauley E. Tai Chi: Improving functional balance and predicting subsequent falls in older persons. <i>Med Sci Sports Exerc</i> 2004; 36 (12): 2046-52.	BBS scale adapted to children
Iatridou G, Dionyssiotis Y. Reliability of balance evaluation in children with cerebral palsy. <i>Hippokratia</i> 2013; 17; 4: 303-6.	BBS scale with 13 items
Russell T, Hoffmann TC, Nelson M, Thompson L, Vincent A. Internet-based physical assessment of people with Parkinson disease is accurate and reliable: A pilot study. <i>JRRD</i> 2013; 50 (5): 643-50.	Article provides two values intra-rater of BBS scale
La Porta F, Giordano A, Caselli S, Foti C, Franchignoni F. Is the Berg Balance Scale an effective tool for the measurement of early postural control impairments in patients with Parkinson's disease?. Evidence from Rasch analysis. <i>Eur J Phys Rehabil Med</i> 2015; 51:705-16.	Article provides reliability of 4 evaluations of BBS scale
Joa KL, Kwon SY, Choi JW, Hong SSE, Kim CH, Jung HY. Classification of walking ability of household walkers versus community walkers based on K-BBS, gait velocity and upright motor control. <i>Eur J Phys Rehabil Med</i> 2015; 51:619-25.	Article provides Spearman's rank correlation of BBS scale
Seong-Gil K, Myoung-Kwon K. The intra- and inter-rater reliabilities of the Short Form Berg Balance Scale in institutionalized elderly people. <i>J Phys Ther Sci</i> 2016; 27: 2733-34.	Article provides only data from the Short Form Berg Balance Scale

SIH Table. Characteristics of the included studies.

Study Year Country	Sample/s	BBS (mean, SD)	Target population 2/ and 3	Psychometric study (BBS and/or other/s scale/s)	Rater/s	Type /s of reliability	Time interval (Inter-rater and/or intra- rater reliability)
Berg ⁶ 1989 Canada	N=14 Age (mean/SD): ? Gender (% male): ? Disease: No Disorder history (mean/SD): ?	?	Non-clinical/ Non- institutionalized	BBS: Yes Other scale: No	Number: 6 inter-rater; 4 intra-rater Profession: Physical therapist Previous training: No Experience (Yes or No/time): Yes/?	α de Cronbach: 0.96 ICC inter-rater: 0.98 ICC intra-rater: 0.99	Inter-rater: ? Intra-rater: 1 w
Berg ⁹ 1995 USA	Sample a (Berg et al, 1995a) N=113 Age (mean/SD): 83.5/5.3 Gender (% male): 17.69 Disease: No Disorder history (mean/SD): ?-	?	Non-clinical/ Institutionalized	BBS: Yes Other scale: No	Number: ? Profession: ? Previous training: ? Experience (Yes or No /time): ?/?	α de Cronbach: 0.83	NA
	Sample b (Berg et al, 1995 b) N=28 Age (mean/SD): ? Gender (% male): ? Disease: No Disorder history (mean/SD): ?	?	Clinical/ Institutionalized	BBS: Yes Other scale: No	Number: 32 Profession: -Physical therapist -Occupational therapist -Others Previous training: No Experience (Yes or No/time): Yes/?	ICC inter-rater: 0.92	0 w
	Sample c (Berg et al, 1995 c) N=18 Age (mean/SD): ? Gender (% male): ? Disease: No	?	Clinical/ Institutionalized	BBS: Yes Other scale: No	Number: 7 Profession: -Physical therapist -Occupational therapist -Others	ICC intra-rater: 0.91	1 w

Disorder history (mean/SD):?						
				Previous training: No Experience (Yes or No/time): Yes/?		
Sample d (Berg et al, 1995 d)	?	Clinical/ Non-institutionalized	BBS: Yes Other scale: No	Number: ? Profession: ?	α de Cronbach: 0.97	NA
N=69 Age (mean/SD): 71.6/10.1 Gender (% male): 51.43 Disease: stroke Disorder history (mean/SD): ?				Previous training: ? Experience (Yes or No/time): Yes/?		
Sample e (Berg et al, 1995 e)	?	Clinical Non-institutionalized	BBS: Yes Other scale: No	Number: 32 Profession: -Physical therapist -Occupational therapist -Others	ICC inter-rater: 0.98	0 w
N=35 Age (mean/SD): ? Gender (% male): ? Disease: stroke Disorder history (mean/SD): ?				Previous training: No Experience (Yes or No/time): Yes/?		
Sample f (Berg et al, 1995 f)	?	Clinical Non-institutionalized	BBS: Yes Other scale: No	Number: 7 Profession: -Physical therapist -Occupational therapist -Others	ICC intra-rater: 0.99	1 w
N=6 Age (mean/SD): ? Gender (% male): ? Disease: stroke Disorder history (mean/SD): ?				Previous training: No Experience (Yes or /time): Yes/?		
Liston ⁹⁵ 1996 Canada	N=20 Age (mean/SD): 64.0/8.5 Gender (% male): 25 Disease: stroke Disorder history (mean/SD): 3.6y/3.8y	?	Clinical Non-institutionalized	BBS: Yes Other scale: No	Number: ? Profession: ?	ICC intra-rater: 0.98 1 w
				Previous training: ? Experience (Yes or No/time): ?/?		

Baer ⁵⁵ 2001 USA	Sample a (Baer et al, 2001a) N=6 Age (mean/SD):? Gender (% male): ? Disease: stroke Disorder history (mean/SD):32.2d/13.7d	?	Clinical Non- institutionalized	BBS: Yes Other scale: -MEFAP	Number: ? Profession: ? Previous training: ? Experience (Yes or No/time):?/?	ICC intra-rater:0.605	?
	Sample b (Baer et al, 2001b) N=4 Age (mean/SD):? Gender (% male): ? Disease: stroke Disorder history (mean/SD):?	?	Clinical Non- institutionalized	BBS: Yes Other scale: -MEFAP	Number: ? Profession: ? Previous training: ? Experience: (Yes or No/time):?/?	ICC inter-rater:0.992	?
Norén ⁸⁹ 2001 Sweden	N=19 Age (mean/SD):55 Gender (% male): 10.53 Disease: peripheral arthritis Disorder history (mean/SD):?		Clinical Non- institutionalized	BBS: Yes Other scale: -ABC tests -TUG	Number: 3 Profession: Physical therapist Previous training:? Experience (Yes or No/time):?/?	ICC inter-rater:0.97	?
Wolf ⁵⁶ 2001 Holland	N=19 Age (mean/SD):84.03/5.62 Gender (% male): 27.27 Disease: No Disorder history (mean/SD):----		Non- clinical/Mixed	NA	Number: ? Profession: Physical therapist Previous training: ? Experience (Yes or No/time):Yes/?	ICC inter-rater:0.98 ICC intra-rater:0.98	?
Stevenson ⁹⁷ 2001 Canada	N=48 Age (mean/SD):73.5/7.0 Gender (% male): 50 Disease: Stroke Disorder history (mean/SD):30.3/23.3 days	43.0/?	Clinical Non- institutionalized	BBS: Yes Other scale: No	Number: 2 Profession: Physical therapist Previous training: ? Experience: (Yes or No/time):Yes/5y	ICC inter-rater:0.92	?

Mao ¹⁵ 2002 Taiwan	N=120 Age (mean/SD):69.0/11.3 Gender (% male): 60 Disease: Stroke Disorder history (mean/SD):?	22.3/22.2	Clinical/ Non- institutionalized	BBS: Yes Other scale: -Balance subscale Fulg- Meyer test. -PASS	Number: 2 Profession: Occupational therapist Previous training:? Experience (Yes or No/time):?/?	ICC inter-rater:0.97	?
Lee ⁷² 2003 China	N=66 Age (mean/SD):? Gender (% male): 34.85 Disease: Visual problems Disorder history (mean/SD):?	?	Clinical/ Institutionalized	BBS: NA Other scale: NA	Number: 1 Profession: Physical therapist Previous training:? Experience (Yes or No/time):Yes/?	ICC intra-rater:0.97	1 w
Ottonello ²⁵ 2003 Italy	Sample a (Ottonello et al, 2003 a) N=85 Age (mean/SD): 67.1/13.8 Gender (% male): 32.9 Disease: stroke, parkinson's disease, orthopedic diseases Disorder history (mean/SD):?	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: No	Number: ? Profession: ? Previous training: ? Experience (Yes or No/time):?/?	α de Cronbach:0.95	NA
	Sample b (Ottonello et al, 2003 b) N=13 Age (mean/SD): 68.9/9.5 Gender (% male): 30.77 Disease: stroke, parkinson's disease, orthopedic diseases Disorder history (mean/SD):?		Clinical/ Non- institutionalized	BBS: Yes Other scale: No	Number: 2 Profession: ? Previous training: Yes, all of them Experience (Yes or No/time):?/?	ICC inter-rater:0.99 ICC intra-rater:0.99	?

Mihay ⁹⁰ 2003 USA	N=35 Age (mean/SD): ? Gender (% male): 40 Disease: No Disorder history (mean/SD):?	?	Non-clinical/ Non- institutionalized	NA	Number: 8 Profession: ? Previous training: ? Experience (Yes or No/time):?/?	ICC inter-rater:0.92	6-8 week
Miyamoto ²⁶ 2004 Brazil	N=36 Age (mean/SD): 72 (?) Gender (% male): 25 Disease: rheumatic disorders, cardiovascular diseases, metabolic diseases, neurological disturbances, lung diseases, gastrointestinal disorders. Disorder history (mean/SD):?	49.9 (?)	Clinical/ Non- institutionalized	BBS: Yes Other scale: No	Number: 2 (inter-rater reliability; 1 intra-rater reliability) Profession: ? Previous training: ? Experience (Yes or No/time):?/?	α de Cronbach:0.96 ICC inter-rater:0.98 ICC intra-rater:0.99	Inter-rater: 0 w Intra-rater: 1 w
Smith ⁹¹ 2004 USA	N=75 Age (mean/SD): ? Gender (% male): ? Disease: stroke Disorder history (mean/SD):?	42.12/10.39	Clinical/ Non- institutionalized	BBS: Yes Other scale: No	Number: 20 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time):?/?	ICC inter-rater: 0.80	?
Franchignoni ¹⁸ 2005 Italy	N=70 Age (mean/SD): 71/? Gender (% male): 52.86 Disease: parkinson's disease Disorder history (mean/SD):7y/?	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: -FFM -PCS	Number: ? Profession: ? Previous training: ? Experience (Yes or No/time):?/?	α de Cronbach:0.95	NA

Holbein-Jenny ¹² 2005 USA	N=26 Age (mean/SD): 85.3/4.9 Gender (% male): 19.23 Disease: No Disorder history (mean/SD):	?	Non-clinical/ Institutionalized	BBS: Yes Other scale: -MRT -ABC scale	Number: 6 inter-rater; ? intra-rater Profession: Physical therapist Previous training: ? Experience (Yes or No/time):?/?	ICC inter-rater:0.88 ICC intra-rater:0.77	Inter-rater:1-2w
Lim ⁵⁷ 2005 Holland	N=26 Age (mean/SD): 62.5/8.2 Gender (% male): 57.7 Disease: parkinson's disease Disorder history (mean/SD): 6.5y/4.2	53.77/1.99	Clinical/ Non-institutionalized	BBS: Yes Other scale: No	Number: 3 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time):?/?	ICC inter-rater:0.74 ICC intra-rater:0.87	Inter-rater:1 h Intra-rater:1 w
Newstead ¹⁹ 2005 USA	N=5 Age (mean/SD):24.4/5.3 Gender (% male): 80 Disease: traumatic brain injury Disorder history (mean/SD):?	?	Clinical/ Non-institutionalized	BBS: Yes Other scale: -BMLOST	Number: 2 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time): No/?	ICC intra-rater:0.986	1 w
Paltamaa ⁵⁸ 2005 Finland	Sample a (Paltamaa et al, 2005a) N=19 Age (mean/SD):42.7/9.2 Gender (% male): 52.63 Disease: multiple sclerosis Disorder history (mean/SD):5.8y/5.7	?	Clinical/ Non-institutionalized	BBS: Yes Other scale: -Physical functioning measures	Number: 1 Profession: Physical therapist Previous training: No Experience (Yes or No/time):Yes/?	ICC intra-rater:0.85	1 w
	Sample b (Paltamaa et al, 2005b) N=9 Age (mean/SD):48.9/8.8	?	Clinical/ Non-institutionalized	BBS: Yes Other scale: No	Number: 2 Profession: Physical therapist	ICC inter-rater:0.99	0 w

	Gender (% male): 33.33 Disease: multiple sclerosis Disorder history (mean/SD):9.3y/8.3				Previous training: No Experience (Yes or No/time):Yes/?		
Sackley ⁵⁹ 2005 UK	N=47 Age (mean/SD):43.1/14.3 Gender (% male): 48.94 Disease: No Disorder history (mean/SD):?/?	?	Non-clinical/ Non- institutionalized	BBS: Yes Other scale: -RMI -BI	Number: 2 Profession: ? Previous training: No Experience (Yes or No/time):?/?	ICC intra-rater:0.98 ICC inter-rater:0.99	Intra-rater:1 w
Southard ⁶⁰ 2005 USA	N=22 Age (mean/SD):78.7/6.9 Gender (% male): 4.54 Disease: No Disorder history (mean/SD):?/?	?	Non-clinical/ Non- institutionalized	BBS: No Other scale: -MTT	Number: 6 inter-rater; ? intra-rater Profession: ? Previous training: No Experience (Yes or No/time):?/?	ICC inter-rater:0.96 r (intra-rater):0.97	Intra-rater:1 w
Chou ⁷⁴ 2006 Taiwan	N=113 Age (mean/SD):78.7/6.9 Gender (% male): 4.54 Disease: No Disorder history (mean/SD):?/?	23.3/21.4	Non-clinical/ Non- institutionalized	BBS: No Other scale: -Short Form of The Berg Balance Scale	Number: 1 Profession: Occupational therapist Previous training: No Experience (Yes or No/time):?/?	ICC intra-rater:0.98	?
Wang ⁵² 2006 Taiwan	N=68 Age (mean/SD):?/? Gender (% male): 47.06 Disease: high blood pressure, diabetes, heart disease, arthritis. Disorder history (mean/SD):?/?	?	Clinical/Mixed	BBS: Yes Other scale: No	Number: 2 Profession: -Physical therapist -Occupational therapist Previous training: Yes, all	α de Cronbach: 0.77 ICC inter-rater:0.87	?

					of them Experience (Yes or No/time):?/?		
Conradsson ²² 2007 Sweden	N=45 Age (mean/SD):82.3/6.6 Gender (% male): 20 Disease: No Disorder history (mean/SD):?/?	30.35/15.7	Non-clinical/ Institutionalized	BBS: Yes Other scale: No	Number: 4 Profession: Physical Therapist Previous training: Yes, all of them Experience (Yes or No/time):?/?	ICC intra-rater:0.97	1-3 d
Cattaneo ²⁰ 2007 Italy	N=25 Age (mean/SD):41.7/12.5 Gender (% male): 32 Disease: multiple sclerosis Disorder history (mean/SD):8.7y/8.8	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: -DGI -DHI -ABC scale	Number: 1 intra-rater , 2 inter-rater Profession: Physical therapist Previous training: No Experience (Yes or No/time):Yes/?	ICC intra-rater:0.96 ICC inter-rater:0.96	Intra-rater:3 d
English ⁷⁶ 2007 Australia	N=68 Age (mean/SD):65.57/12.61 Gender (% male): 60.29 Disease: stroke Disorder history (mean/SD):?/?	30.07/16.8	Clinical/ Non- institutionalized	NA	Number: 1 Profession:? Previous training: ? Experience (Yes or No/time):?/?	ICC intra-rater:0.96	4 w
Swanenburg ⁷⁷ 2007 Switzerland	N=24 Age (mean/SD):71.2/6.8 Gender (% male): 100 Disease: No Disorder history (mean/SD):?/?	52.45/3.56	Non-clinical/ Non- institutionalized	NA	Number: 1 Profession:? Previous training: ? Experience (Yes or No/time):?/?	ICC intra-rater:0.90	?
Halsaa ²⁷ 2007 Norway	N=83 Age (mean/SD):82.0/5.5 Gender (% male): 30.12 Disease: stroke, parkinson's	44.3/8.6	Mixed/ Non- institutionalized	BBS: Yes Other scale:No	Number: 2 Profession:? Previous training: Yes, all	α de Cronbach: 0.87 ICC inter-rater:0.988	?

	disease, low back pain, pneumonia, heart failure, osteoarthritis hip, rheumatoid arthritis, diabetes. Disorder history (mean/SD):?/?				of them Experience (Yes or No/time):Yes/?		
Nemmers ⁶⁶ 2008 USA	N=203 Age (mean/SD):77.33/7.59 Gender (% male): 0 Disease: No Disorder history (mean/SD):?/?	51.21/5.05	Non-clinical/ Non-institutionalized	NA	Number: ? Profession:? Previous training: No Experience (Yes or No/time):Yes/?	α de Cronbach:0.821	NA
Steffen ³⁵ 2008 USA	N=37 Age (mean/SD):77.33/7.59 Gender (% male): 0 Disease: No Disorder history (mean/SD):?/?	50.0/7.0	Non-clinical/ Non-institutionalized	BBS: Yes Other scale: -FRT -RT -SRT -ABC scale -6MWT -TUG -Gait speed -SF-36 -UPDRS	Number: 3 (intra-rater reliability) Profession:? Previous training: Yes, all of them Experience: (Yes or No/time):Yes/?	α de Cronbach:0.86 ICC intra-rater:0.94	?
Liaw ⁷⁸ 2008 Taiwan	N=52 Age (mean/SD):60.4/13.4 Gender (% male): 65.4 Disease: stroke Disorder history (mean/SD):?/?	?	Clinical/ Non-institutionalized	BBS: Yes Other scale:No	Number:1 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time):Yes/?	ICC intra-rater:0.98	1 w

Sahin ²⁸ 2008 Turkey	N=60 (20 intra-rater, 20 inter-rater) Age (mean/SD):77.0/5.67 Gender (% male): 33.33 Disease: hypertension, osteoarthritis, cardiac disease, osteoporosis, diabetes mellitus, stroke, thyroid disease, vertigo, pulmonary disease, romatoid arthritis, hypercholesterolemia, ischemic heart disease, renal failure, pulmonary disease, parkinson´ disease, polyneuropathy, epilepsy. Disorder history (mean/SD):?/?	47.63/9.88	Mixed/ Non-institutionalized	BBS: Yes Other scale:No	Number:2 Profession: Psychiatrist Previous training: ? Experience (Yes or No/time):?/?	α de Cronbach:0.93 ICC intra-rater:0.98 ICC inter-rater:0.97	Intra-rater:0 w Inter-rater:0 w
Cheung ⁷⁹ 2008 China	N=50 Age (mean/SD):83.64/5.64 Gender (% male): 100 Disease: visual problems Disorder history (mean/SD):?/?	43.3/5.1	Mixed/Mixed	NA	Number:? Profession:? Previous training: ? Experience (Yes or No/time):?/?	ICC intra-rater:0.821	?
Scalzo ⁵³ 2009 Brazil	N=53 Age (mean/SD):62.0/7.9 Gender (% male): 69.8 Disease: parkinson´ disease Disorder history (mean/SD):7.8y/4.4	47.2/8.2	Clinical/ Non-institutionalized	BBS: Yes Other scale:No	Number:2 Profession:? Previous training: Yes, all of them Experience (Yes or No/time):?/?	α de Cronbach:0.92 ICC intrer-rater:0.84	?
Combs ⁶¹ 2010 USA	N=19 Age (mean/SD):61.1/11.7 Gender (% male): 31 Disease: stroke Disorder history (mean/SD):3.8y/3.2	43.4/?	Mixed/ Non-institutionalized	NA	Number:1 Profession:? Previous training: ? Experience: (Yes or No/time):?/?	ICC intra-rater:0.99	?
Azad ²⁹ 2011 Iran	N=60 Age (mean/SD):36.6/9.5 Gender (% male): 34.0 Disease: multiple sclerosis	?	Clinical/ Non-institutionalized	BBS: Yes Other scale:No	Number:2 Profession: Occupational therapist	α de Cronbach:0.9 ICC intrer-rater:0.9	?

	Disorder history (mean/SD):0.1y/0.3				Previous training: Yes, all of them Experience (Yes or No/time):?/?		
Combs ⁷³ 2011 USA	N=6 Age (mean/SD):60.17/10.26 Gender (% male): ? Disease: parkinson´disease Disorder history (mean/SD):28.67m/24.34	?	Clinical/ Non- institutionalized	NA	Number:3 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time):?/?	ICC intra-rater:0.998 ICC inter-rater:0.998	?
Leddy ³⁴ 2011 USA	Sample a (Leddy et al, 2011a) N=24 Age (mean/SD):?/? Gender (% male): ? Disease: parkinson´disease Disorder history (mean/SD):6y/3.38	?	Mixed/ Non- institutionalized	BBS: No Other scale: -FGA -BESTest	Number:2 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time):Yes/?	ICC intra-rater:0.79	2 w
	Sample b (Leddy et al, 2011b) N=15 Age (mean/SD):?/? Gender (% male): ? Disease: parkinson´disease Disorder history (mean/SD):6.8y/3.26	?	Mixed/ Non- institutionalized	BBS: No Other scale: -FGA -BESTest	Number:3 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time):Yes/13 and 21y	ICC inter-rater:0.95	2 w
Williams ⁹² 2011 USA	N=22 Age (mean/SD):83.86/5.47 Gender (% male): 18.18 Disease: No Disorder history (mean/SD):?/?		Non-clinical/ Non- institutionalized	NA	Number:2 Profession: Physical Therapist Previous training: Yes, all of them Experience (Yes or No/time):?/?	ICC inter-rater:0.948	?

Learmonth ⁸¹ 2012 UK	N=24 Age (mean/SD):51.8/7.9 Gender (% male): ? Disease: multiple sclerosis Disorder history (mean/SD):13.7y/6.5	45.92/12.4	Clinical/ Non- institutionalized	BBS: Yes Other scale: -T25FW -6MWT -TUG	Number:1 Profession: Physical Therapist Previous training: Yes Experience (Yes or No/time):Yes/?	ICC intra-rater:0.95	1 w
Heaney ⁶⁷ 2012 UK	N=36 Age (mean/SD):72.5/6.47 Gender (% male): 50 Disease: hypertension, arthritis, osteoarthritis, renal disease, glaucoma. Disorder history (mean/SD):?/?	?	Mixed/ Non- institutionalized	NA	Number:? Profession:? Previous training: ? Experience (Yes or No/time):?/?	α de Cronbach:0.96	NA
Flansbjer ³⁶ 2012 Sweden	N=50 Age (mean/SD):58.0/6.0 Gender (% male): 76 Disease: stroke Disorder history (mean/SD):17m/9	?	Mixed/ Non- institutionalized	BBS: Yes Other scale: -Single leg stance	Number:? Profession: Physical therapist Previous training: ? Experience (Yes or No/time):Yes/?	ICC intra-rater:0.88	1 w
La Porta ⁶⁸ 2012 Italy	N=217 Age (mean/SD):59.50/16.30 Gender (% male): 60.80 Disease: ischemic stroke, intracerebral hemorrhage, traumatic brain injury, subarachnoid hemorrhage, central nervous system neoplasms, myelopathies, peripheral neuropathies. Disorder history (mean/SD):?/?	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: No	Number: 4 raters Profession: ? Previous training: ? Experience (Yes or No/time):Yes/?	α de Cronbach:0.974	NA
Hiengkaew ⁸² 2012	N=61 Age (mean/SD):63.5/10.0	46.2/7.64	Mixed/ Non-	BBS: Yes Other scale:	Number: 2 Profession: Physical	ICC intra-rater: 0.95	5-10 d

Thailand	Gender (% male): 70 Disease: stroke Disorder history (mean/SD):?/?		institutionalized	-FMA-LE -TUG -CGS -FGS -2MWT	therapist Previous training:? Experience (Yes or No/time):Yes/?		
Rasova ⁸³ 2012 Czech Republic	N=17 Age (mean/SD):43.30/9.0 Gender (% male): 41.0 Disease: multiple sclerosis Disorder history (mean/SD):10.10y/5.80	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: No	Number: 1 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time):Yes/?	ICC intra-rater:0.780	?
Tsaih ⁸⁴ 2012 Taiwan	N=5 Age (mean/SD):?/? Gender (% male): ? Disease: No Disorder history (mean/SD):?/?	22.1/14.02	Non- clinical/Mixed	NA	Number: 1 Profession: Physical therapist Previous training:Yes Experience (Yes or No/time):?/?	ICC intra-rater:0.98	1 w
Salavati ⁴⁸ 2012 Iran	N=58 Age (mean/SD):?/? Gender (% male): ? Disease: No Disorder history (mean/SD):?/?	?	Non-clinical/ Mixed	BBS: Yes Other scale: No	Number: 2 Profession: -Physical therapist -Orthostic and prosthetist Previous training: No Experience: (Yes or No/time):Yes/?	α de Cronbach:0.62 ICC inter-rater:0.93 ICC intra-rater:0.95	Inter-rater:? Intra-rater:1 w
Wong ⁶⁹ 2013 USA	N=40 Age (mean/SD):?/? Gender (% male): ? Disease: Leg amputation Disorder history (mean/SD):?/?		Mixed/ Non- institutionalized	BBS: Yes Other scale: No	Number: 8 Profession: Physical therapist Previous training:Yes, all of them	α de Cronbach:0.88	NA

					Experience (Yes or No/time):?/?		
Toomey ⁹³ 2013 Ireland	N=8 Age (mean/SD):?/? Gender (% male): 37.5 Disease: Leg amputation Disorder history (mean/SD):?/?	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: -6MWT -HHD	Number: 4 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time):?/?	ICC inter-rater: 0.982	?
Quinn ⁸⁵ 2013 UK, Holland, Portugal, USA	Sample a (Quinn et al, 2013a) N=11 Age (mean/SD): 43.2/13.1 Gender (% male): ? Disease: Huntington disease Disorder history (mean/SD):?/?	55/1	Clinical/ Non- institutionalized	BBS: Yes Other scale: -6MWT -10MWT -TUG -PPT -BI -RMI -TMT	Number: 10 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or No/time): Yes/?	ICC intra-rater: 0.86	1 w
	Sample b (Quinn et al, 2013b) N=64 Age (mean/SD): 53.7/11.0 Gender (% male): ? Disease: Huntington disease Disorder history (mean/SD):?/?	47/8	Clinical/ Non- institutionalized	BBS: Yes Other scale: -6MWT -10MWT -TUG -PPT -BI -RMI -TMT	Number: 10 Profession: Physical therapist Previous training: Yes, all of them Experience: (Yes or No/time): Yes/?	ICC intra-rater: 0.96	1 w
	Sample c (Quinn et al, 2013c) N=20 Age (mean/SD): 57.3/9.5 Gender (% male): ? Disease: Huntington disease Disorder history (mean/SD):?/?	51/4	Clinical/ Non- institutionalized	BBS: Yes Other scale: -6MWT -10MWT -TUG -PPT -BI	Number: 10 Profession: Physical therapist Previous training: Yes, all of them Experience (Yes or	ICC intra-rater: 0.90	1 week

			-RMI -TMT	No(time):Yes/?		
Sample d (Quinn et al, 2013d)	47/6	Clinical/ Non- institutionalized	BBS: Yes Other scale: -6MWT -10MWT -TUG -PPT -BI -RMI -TMT	Number: 10 Profession: Physical therapist Previous training:Yes, all of them Experience (Yes or No(time):Yes/?	ICC intra-rater: 0.91	1 week
Sample e (Quinn et al, 2013e)	45/12	Clinical/ Non- institutionalized	BBS: Yes Other scale: -6MWT -10MWT -TUG -PPT -BI -RMI -TMT	Number: 10 Profession: Physical therapist Previous training:Yes, all of them Experience (Yes or No(time):Yes/?	ICC intra-rater: 0.97	1 w
Major ⁵⁴ 2013 USA	N=30 Age (mean/SD): 54.0/12.0 Gender (% male): 66.6 Disease: lower-limb amputation Disorder history (mean/SD):21y/15	51/5	Mixed/ Non- institutionalized	BBS: Yes Other scale: No -Biomedical engineer -Prosthetist/orthotist Previous training:Yes, all of them Experience: (Yes or No(time):Yes/4y	α de Cronbach:0.827 ICC inter-rater:0.945	Inter-rater:20 m
Godi ⁴⁹ 2013 Italy	N=32 Age (mean/SD): 67.3/13.5 Gender (% male): 40.62 Disease: stroke, parkinson's disease, vestibular disorders. Disorder history (mean/SD):21y/15	42/11.2	Clinical/ Non- institutionalized	BBS: Yes Other scale: -Mini-BESTest Previous training:Yes, all of them Experience: (Yes or No(time):?/?	Number: 3 inter-rater, 1 intra-rater Profession: physical therapist α de Cronbach:0.93 ICC inter-rater:0.97 ICC intra-rater:0.92	Inter-rater:? Intra-rater:1-3 d
Matsushima ⁵⁰	N=33	42.8/11.7	Clinical/	BBS: Yes	Number: 2 inter-rater, ? α de Cronbach:0.941	?

2014 Japan	Age (mean/SD): 67.3/13.5 Gender (% male): 40.62 Disease: neurological diseases Disorder history (mean/SD):21y/15		Non- institutionalized	Other scale: No	intra-rater Profession: ? Previous training:? Experience (Yes or No/time):?/?	ICC inter-rater:0.933 ICC intra-rater:0.977	
Puckree ⁷⁰ 2014 South Africa	N=50 Age (mean/SD): ?/? Gender (% male): 52 Disease: stroke Disorder history (mean/SD):?/?	44.07/19.85	Mixed/ Non- institutionalized	NA	Number: ? Profession: ? Previous training:? Experience (Yes or No/time):?/?	α de Cronbach:0.974	NA
Takacs ⁸⁶ 2014 Canada	N=20 Age (mean/SD): ?/? Gender (% male): ? Disease: Knee osteoarthritis Disorder history (mean/SD):?/?	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: -CB&M	Number:1 Profession: ? Previous training:? Experience: (Yes or No/time):?/?	ICC intra-rater:0.59	2 w
Tamburella ⁸⁷ 2014 Italy	N=17 Age (mean/SD): ?/? Gender (% male): ? Disease: incomplete spinal cord injury Disorder history (mean/SD):?/?	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: -stabilometric parameters	Number:1 Profession: ? Previous training:? Experience: (Yes or No/time):?/?	ICC intra-rater:0.97	15 d
Wong ⁶² 2014 USA	N=5 Age (mean/SD): 53.0/15.7 Gender (% male): 80 Disease: lower limb amputation Disorder history (mean/SD):8.2y/7.9	?	Mixed/ Non- institutionalized	BBS: Yes Other scale: No	Number:16 Profession: -Physical therapist -Occupational therapist Previous training:Yes, all of them Experience: (Yes or No/time):Yes/8y	ICC inter-rater:0.99 ICC intra-rater: 0.99	?
Srisim ⁹⁴ 2015 Thailand	N=25 Age (mean/SD): ?/? Gender (% male): ? Disease: spinal cord injury		Mixed/ Non- institutionalized	BBS: Yes Other scale: -TUG -10MWT	Number:2 Profession: Physical therapist	ICC inter-rater: 0.998	?

	Disorder history (mean/SD):?/?			-FRT -Step test -FTSST	Previous training:Yes, all of them Experience (Yes or No/time):Yes/>5y		
Schlenstedt ⁶⁴ 2015 Germany	Sample a (Schlenstedt et al, 2015a) N=15 Age (mean/SD): ?/? Gender (% male): ? Disease: parkinson´ disease Disorder history (mean/SD):?/?	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: -FAB -MiniBESTest -BBS	Number:3 Profession: -Sport scientists -Psychology student Previous training:Yes, all of them Experience (Yes or No/time):Yes/?	ICC inter-rater:0.98	?
	Sample b (Schlenstedt et al, 2015b) N=17 Age (mean/SD): ?/? Gender (% male): ? Disease: parkinson´ disease Disorder history (mean/SD):?/?	?	Clinical/ Non- institutionalized	BBS: Yes Other scale: -FAB Scale -MiniBESTest -BBS -TUG	Number:1 Profession: ? Previous training:Yes, all of them Experience (Yes or No/time):Yes/?	ICC intra-rater:0.95	3±1 d
Schachten ⁹⁶ 2015 Germany	N=14 Age (mean/SD): 54.14/15.75 Gender (% male): ? Disease: stroke Disorder history (mean/SD):3.6y/1.7	?	Clinical/ Non- institutionalized	NA	Number:? Profession: ? Previous training:? Experience (Yes or No/time):?/?	ICC inter-rater:0.99	?
Portnoy ⁹⁸ 2015 USA	N=31 Age (mean/SD): 54.14/15.75 Gender (% male): 74.19 Disease: stroke Disorder history (mean/SD):6.1y/6.7	36.8/11.4	Clinical/ Non- institutionalized	NA	Number:? Profession: ? Previous training:? Experience (Yes or No/time):?/?	ICC inter-rater:0.977	
Muir-Hunter ⁶³ 2015 Canada	N=15 Age (mean/SD): 80.20/5.03. Gender (% male): 73.33 Disease: Alzheimer´ disease Disorder history (mean/SD):?/?	41.67/14.2	Mixed/ Non- institutionalized	BBS: Yes Other scale: No	Number:2 inter-rater, 1 intra-rater Profession: Physical therapist Previous training:Yes, all of them	ICC inter-rater:0.72 ICC intra-rater:0.95	Inter-rater:? Intra-rater:1w

					Experience (Yes or No/time):Yes/?		
Geroin ⁸⁸ 2015 Italy	N=20 Age (mean/SD): 68.25/9.89 Gender (% male): 60 Disease: parkinson´ disease Disorder history (mean/SD):2.63y/2.6	?	Clinical/ Non-institutionalized	NA	Number:1 Profession: ? Previous training:? Experience (Yes or No/time):Yes/?	ICC intra-rater:0.80	?
Chan ⁵¹ 2015 China	N=46 Age (mean/SD): 69.1/6.1 Gender (% male): 26 Disease: total knee arthroplasty Disorder history (mean/SD):?/?	?	Clinical/ Non-institutionalized	BBS: Yes Other scale: No	Number:? Profession: ? Previous training:Yes, all of them Experience (Yes or No/time):Yes/10y	α de Cronbach:0.98 ICC intra-rater:0.97	NA
	N=25 Age (mean/SD): 69.7/6.8 Gender (% male): 32 Disease: total knee arthroplasty Disorder history (mean/SD):?/?		Clinical/ Non-institutionalized	BBS: Yes Other scale: No	Number:3 Profession:Physical therapist Previous training:Yes, all of them Experience (Yes or No/time):Yes/10y	ICC inter-rater:0.98	
Azimzadeh ⁷¹ 2015 Iran	N=34 Age (mean/SD): ?/? Gender (% male): 100 Disease: multiple sclerosis Disorder history (mean/SD):?/?	?	Clinical/ Non-institutionalized	NA	Number:? Profession: ? Previous training:? Experience (Yes or No/time):?/?	α de Cronbach:0.78	NA
Jantakat ⁶⁵ 2015 Thailand	N=58 Age (mean/SD): 15.09/1.84? Gender (% male): 55.17	?	Clinical/ Non-institutionalized	NA	Number:2 inter-rater, 1 intra-rater Profession: Physical	ICC inter-rater:0.992 ICC intra-rater: 0.996	Inter-rater:? Intra-rater:1 w

Disease: cerebral palsy
Disorder history (mean/SD):?/?

therapist

Previous training:Yes, all
of them
Experience (Yes or
No/time):Yes/2y

Abbreviations: SD, Standard Deviation; Target population 2, clinical, non-clinical, mixed; Target population 3, Institutionalized, non-institutionalized, mixed; y,years; d, days; m, month; w; week/s; h; hour/s; m; minute/s; MEFAS, Modified Emory Functional Ambulation Profile; NA, not applicable; PASS, Postural Assessment Scale for Stroke Patients; FFM, Fear of Falling Measure; PCS, Postural Changes Scale; MRT, Multi-directional Reach Test; ABC scale, The Activities-specific Balance Confidence Scale; BMLOST, Balance Master Limits of Stability Test; RMI, Rivermead Mobility Index; BI, Barthel Activities of Daily Living Index; MTT, Multiple Taks Test; DGI, Dynamic Gait Index; DHI, Dizziness Handicap Inventory; FRT, Forward and Backward Functional Reach; RT, Romberg Test; SRT, Sharpened Romberg Test; 6MWT, Six-Minute Walk Test; TUG, Timed “Up & Go” Test; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; UPDRS, Unified Parkinson Disease Rating Scale; FGA, Functional Gait Assessment; BESTest, Balance Evaluation Systems Test; FMA-LE; Lower Limb Subscale of Fugl-Meyer Assessment; CGS, Comfortable Gait Speed; FGS, Fast Gait Speed, 2MWT, 2-minute walk test; HHD, Handheld Dynamometry; 10MWT, 10-Meter Walk Test; PPT; Physical Performance Test; RMI, Rivermead Mobility Index; TMT, Tinetti Mobility Test; CB&M, Community Balance and Mobility Scale; MiniBESTest, Mini-Balance Evaluation Systems Test; FTSST, Five Times Sit-to-Stand Test; FAB Scale, Fullerton Advanced Balance Scale.