ORIGINAL REPORT

UNIFIED BALANCE SCALE: AN ACTIVITY-BASED, BED TO COMMUNITY, AND AETIOLOGY-INDEPENDENT MEASURE OF BALANCE CALIBRATED WITH RASCH ANALYSIS

Fabio La Porta, MD[1], Marco Franceschini, MD[2], Serena Caselli, PT[1], Paola Cavallini, PT[1], Sonia Susassi, PT[1] and Alan Tennant, PhD[3]

From the [1]Rehabilitation Medicine Unit, AUSL Modena, Modena, [2]Neuro-Rehabilitation Units IRCCS San Raffaele Pisana, Rome, Italy and [3]Department of Rehabilitation Medicine, Faculty of Medicine and Health, University of Leeds, Leeds, UK

Objective: **To build a new activity-based, "bed to community", aetiology-independent measure of balance within the neurological rehabilitation setting by merging some existing scales.**

Methods: **Balance scales were selected using a conceptual framework and subsequently administered to a convenience sample of adult patients with balance problems due to different neurological aetiologies. Data were then processed using classical psychometric analyses and Rasch analysis in order to construct a new balance measurement tool.**

Results: **The Berg Balance Scale, the Tinetti Scales and the Fullerton Advanced Balance Scale were selected and administered to a sample of patients, giving 302 observations. Classical psychometric analyses (item and scale analysis; confirmatory factor analysis) were undertaken on the pooled 40-item set with confirmation of unidimensionality. The subsequent Rasch analysis allowed the identification of a 27-item set satisfying the Rasch Model's requirements for fundamental measurement, with further confirmation of unidimensionality by *post-hoc* confirmatory factor analysis.**

Conclusion: **The new scale (Unified Balance Scale) holds proven measurement properties and may be a candidate tool for "bed to community" balance measurement for patients with balance problems within the neuro-rehabilitation setting. Future studies are warranted to explore further its external validity and other clinical properties, as well as to improve its usability.**

Key words: postural balance; accidental falls; outcome measures; rehabilitation; neurological disorders; psychometrics.

J Rehabil Med 2011; 43: 435–444

Correspondence address: Fabio La Porta, Medicina Riabilitativa, Nuovo Ospedale Civile "S. Agostino-Estense", Via Giardini 1355, 41126, Modena, Italy. E-mail: fabiolaporta@mail.com

INTRODUCTION

Balance is a function frequently impaired in stroke (1), traumatic brain injury (2) and many other neurological conditions (3). As balance impairment may lead to falls, and as it has been advocated that fall prevention strategies should be an integral part of rehabilitation programmes (4), it follows that the identification of patients with balance problems who may be at risk for falls, and the development of fall screening tools, should be essential components of any comprehensive fall reduction plan (4–6).

A recent systematic review identified as many as 30 different functional balance assessment tools (7) and new instruments are being continually developed (8). These tools can be either generic or disease-specific, and may be used for different purposes (e.g. assessing balance problems in patients admitted to an acute rehabilitation facility, or adults living in the community) and may cover different aspects of balance. As a result, balance scales may have different operational ranges, spanning the acute-community divide, making it sometimes necessary to use them in conjunction with other instruments in order to cover the whole spectrum of balance problems (1). As such, clinicians and researchers are presented with a wide, and sometimes confusing, array of options, making the choice of the right instrument for the intended clinical or research purpose difficult, although in this respect systematic reviews might provide some help (7).

Consequently, it may be desirable for clinicians and others to have available a single tool with proven measurement properties, allowing the measurement of balance "from bed to community" (i.e. within the hospital as well as in the community setting), regardless of the aetiology of the neurological lesion causing the loss of balance. Such a goal may be achieved by either: (*i*) making a new scale that has a wider operational range, or (*ii*) in some way seeking to combine existing instruments, so that they provide comparable measurement across a wider operational range. Given the current availability of many instruments it may be appropriate to try to make use of existing scales in the first instance, rather than investing time and resources in developing new instruments. One possible approach to bringing together different tools and constructing a common frame of measurement within the healthcare setting was demonstrated recently by Elhan et al. (9), who constructed an item bank for measuring disability in patients with low back pain by calibrating items from 4 different questionnaires onto a single metric using classical psychometric methods and Rasch analysis.

Thus, the aim of the current study was to construct a new measuring tool for balance activity limitations within the neurological rehabilitation setting by merging some existing scales of balance using both classical psychometric methods and Rasch analysis. As a result, we present here the Unified Balance Scale (UBS), a 27 activity-based "from bed to community" scale, which is an aetiology-independent measure of balance for neurological patients admitted to rehabilitation.

METHODS

Selection of instruments and administration guidelines

Published balance scales were reviewed and some candidate scales selected using the following criteria: (*i*) popularity; (*ii*) possibility of being applied to patients with balance problems from various aetiologies; (*iii*) the widest possible range of measurement; (*iv*) availability of cut-off scores for predicting the risk of falls; and (*v*) coverage of as many conceptual domains of balance as possible. Regarding the latter criterion, the 4 balance sub-domains suggested by Franchignoni et al. (8) were adopted as the main conceptual framework, together with a further domain for static balance (10). Hence, the conceptual framework adopted included the following 5 subdomains: (*i*) quiet stance; (*ii*) anticipatory postural adjustments/transitions; (*iii*) responses to external perturbations; (*iv*) sensory orientation; (*v*) stability during gait.

After selection, and in order to improve the inter-rater reliability of instruments, all raters involved in their administration used written scoring guidelines, together with a video detailing the items' administration procedures. Furthermore, raters underwent a single patient-based training session with an experienced physiotherapist (PC, SS, or SC). The Functional Independence Measure (FIM™) and the Trunk Control Test (TCT) were also administered for sample selection and description.

In view of the high number of items to be administered, a protocol was devised in which items were grouped into "stations" according to the postural setting required to perform the requested activity (i.e. sitting position, transfers, static standing position, dynamic standing position, and walking). The protocol was initially trialled with 10 subjects, then items were re-ordered within each station according to their apparent difficulty (i.e. in the "static standing position" station items related to simple standing balance were administered earlier than items requiring standing with feet together). When the activity requested by 2 or more different items was the same (i.e. the 3 items related to "pivot turns") the patient was asked to perform the activity just once, whereas the rater scored the observed behaviour according to each item's specific score categories. Overall, these measures allowed both patient's and rater's acceptability to be improved by, respectively, minimizing fatigue and speeding up the administration of the entire protocol.

Patients and setting

Data were collected at the Rehabilitation Unit of Modena's Civil Hospital, Italy, from April 2007 to June 2009 as part of a larger scale study to build item banks for balance and mobility. This ward is involved mainly in early rehabilitation of patients with elective orthopaedic surgery (total hip and knee replacement: 48.6%), stroke (28.2%), traumatic brain injury and other severe brain damages (12%), other neurological conditions (e.g. peripheral neuropathies, spinal cord injuries, etc): 10.4%) and other aetiologies (e.g. burns: 0.8%). For the larger project, general inclusion criteria were: all patients with a neurological lesion consecutively admitted to the unit as inpatient or outpatients, whereas exclusion criteria were: specific contraindications to mobilization (e.g. fractures, coexisting medical complications), tetraplegia or severe tetraparesis, inability to collaborate or giving informed consent (e.g. severely confused or agitated patients) and patient's unwillingness to participate. Minimum required criteria for the assessment of balance were: a score of 25 in the item "balance in sitting position" of TCT

(able to maintain sitting balance independently without using upper limbs) or a score of at least 2 in the item "chair/wheelchair transfers" of FIM™ (requires maximal assistance or less from one caregiver only). Where possible, inpatients were assessed twice (before and after treatment), for responsiveness evaluation purposes.

All patients gave their informed consent to take part in the study that was undertaken in compliance with the ethical principles set out in the Declaration of Helsinki (11).

Statistical analyses

The statistical analyses were carried out in 4 main steps: (*i*) item and scale descriptive statistics; (*ii*) assessment of unidimensionality of the pooled item set; (*iii*) Rasch analysis; and (*iv*) confirmation of the unidimensionality of the final item set.

Classical item and scale descriptive statistics. A variety of item and scale descriptive statistics were performed (12), including the analysis of response category frequencies, analysis of missing values (both for item and persons), inter-item correlations, item-scale correlation and analysis of internal consistency reliability.

Assessment of unidimensionality of the pooled item set. As Rasch models are unidimensional measurement models, they are based on the assumption that all items measure a single underlying dimension (9). As a consequence, when the item pool is derived from a range of scales that may represent different domains, in this case related to balance, it is advisable to test acceptable unidimensionality prior to the Rasch analysis (8, 9). As such, in order to assess unidimensionality a confirmatory factor analysis (CFA) for categorical data was undertaken. As strict unidimensionality will be considered during the Rasch analysis, the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) were both set at ≤0.10, which is indicative of "mediocre fit", but is sufficient for an initial assessment. RMSEA is an estimate of the discrepancy between the covariance matrix predicted by the model and the population covariance matrix, if it were available. SRMR is the standardized difference between the observed covariance and predicted covariance where a value of zero indicates perfect fit. In addition, confirmation of a satisfactory CFA was established using the non-normed fit index (Tucker-Lewis Index; TLI) and the comparative fit index (CFI) where values above 0.95 (range: 0–1) were considered acceptable (8).

Thus, the objective at this stage is to identify a candidate set of items that will be refined further during the Rasch analysis. Should the CFA fail, an exploratory factor analysis (EFA) would be performed and model fit evaluated using the RMSEA that accounts for model parsimony (9).

Rasch analysis. Following the above analyses, data from the potential unidimensional item set were fitted to the Rasch Model (13). According to the model, a subject with a certain ability (level of balance) on the latent variable is expected to affirm (pass) items representing tasks associated with less ability, and to not affirm (fail) items representing a higher level, in this case, of balance. Where data satisfy this pattern, together with the assumptions concerning local independence and unidimensionality, the data are said to satisfy the model requirements and the raw score derived from the scale can be transformed to interval scale measurement (13).

Within the current study, a two-perspective approach was adopted. The first perspective was simply to identify and compare the operational ranges of the various tests against a single underlying metric, irrespective of the quality of fit of their respective individual items.

The second perspective was to subject the entire set of individual items to scrutiny in an iterative analytical process (widely known as Rasch analysis) to test whether these data meet the requirements of the Rasch measurement model. This process, here based upon the partial credit parameterization of the model, is reported in detail elsewhere (13–15). Briefly, various assumptions are tested:

• Local independence, which was tested by the evidence that no significant association among item responses should be found once the dominant factor (balance) influencing a person's response to those items has been conditioned out (12). This important assumption

was tested by examination of the residual correlations where values above 0.3 indicated local dependence of the item set (13).

- The stochastic ordering of the items was tested by fit to the model. It was considered achieved when: (*i*) a summary χ^2 interaction statistic was non-significant, showing no deviation from model expectation; (*ii*) where item and person summary fit statistics approached a mean of zero and standard deviation of 1; (*iii*) where individual items showed non-significant χ^2 fit statistics (Bonferroni adjusted), and (*iv*) where individual item and person residuals were within the range of ±2.5, which represents the 99% confidence interval.
- Unidimensionality was tested with a *t*-test on separate estimates for each respondent (derived from subsets of items identified by a principal component analysis of the residuals), where less than 5% of such tests should be significant, or the lower bound of the binomial confidence interval for proportions was below 5% (16).

Where assumptions failed, an iterative phase involving item modifications was undertaken, aiming at finding a solution that satisfied the model expectations and assumptions. After each cycle of modification, model fit was reassessed. This included, in order:

- Checking the ordering of response categories for each item in order to establish whether the scoring model for each item worked in the expected manner. When this is not the case, item categories can be collapsed following published guidelines (17). In particular, we established different re-scoring pattern for each item individually, aiming at finding a solution that maximized both statistical indices and clinical meaningfulness (8).
- Deletion of one or more locally-dependent items in a pair items to account for local dependency. The process started from the pair with the higher residual correlations and proceeded further until no pairs of items could be found with a residual correlation equal or above 0.3. In selecting the candidate item for deletion within a pair, the following criteria were applied: involvement in further pairs, misfit to the model, lower number of score categories, clinical meaning.
- Deletion of any further misfitting item.

After achieving a final solution fitting to the model the following further aspects were evaluated:

- Testing and accounting for differential item function (DIF) by clinically relevant key groups (such as age, gender or aetiology).
- The validity of the final item hierarchy suggested by the analysis, i.e. whether it was consistent with clinical expectations.
- The analysis of reliability, here expressed as the Person Separation Index (PSI), where values above 0.70 are regarded as the minimum requirement for group level measurement, and values above 0.85 for individual person measurement (12). On the basis of PSI, it was possible to calculate the number of statistically distinct levels

of person ability (person strata) that the scale was able reliably to distinguish (18).

- The analysis of targeting, which shows graphically how well individual item difficulty and individual person abilities can be matched on a common logit scale (13). The average person ability and spread (i.e. the standard deviation (SD)) indicates how well the scale was targeted to the sample (14). Analysis of targeting also entails the demonstration of floor and ceiling effects and can also be assessed by checking visually with a person-item distribution graph that also indicates possible areas of construct under-representation (19).

Final confirmatory factor analysis. A final CFA was performed to confirm the unidimensionality of the item set.

Statistical notes, software and sample size issues

Analyses of descriptive statistics for persons and items were undertaken using SPSS (SPSS. Version 13 for Windows). Where descriptive analyses showed skewed distributions, medians were used instead of, or along with, means.

Factor analyses for categorical data were undertaken on complete data only using Mplus software (Mplus version 6.0; Muthen & Muthen, 1998–2010; www.statmodel.com). It was estimated that 250 observations (ratio subjects to items: 6.3:1) would be a suitable sample for these analyses (20).

Rasch analysis was carried out on the whole data-set using RUMM2030 software (version 5.1 for Windows). Within the context of Rasch analysis, a sample size of 300 observations would estimate item difficulty, with an α of 0.01 to <±0.5 logits, irrespective of the targeting of persons to the items (21). A significance value of 0.05 was used throughout and corrected for the number of tests by Bonferroni correction (22).

RESULTS

Selection of instruments

After considering all the requirements for scale selection, 3 scales were selected out of an initial pool of 15 instruments: the Berg Balance Scale (BBS), the Performance-Oriented Mobility Assessment scales (Tinetti Balance (TB) and Tinetti Walking (TW)), and the Fullerton Advanced Balance Scale (FAB). Since extended descriptions for all the selected instruments are available in the literature, a synthetic description only is presented in Table I, whereas Table II summarizes the content validity

Table I. *Synthetic description of the selected instruments*

	BBS (23)	POMA (24)	FAB (25)
Number of items	14	16	10
Type of scale	ordinal, summative	ordinal, summative	ordinal, summative
Raw score range	0–56	0–28	0–40
Psychometric properties			
External validity	Yes (7)	Yes (26)	Yes (25)
Classical reliability	Yes (7)	Yes (26)	Yes (25)
Internal validity (Rasch analysis)	Yes (27)	–	–
Floor effect	Yes (1)	Expected	Expected
Ceiling effect	Yes (1)	Yes (28)	No (29)
Selection criteria			
Popularity	Yes (1)	Yes (26)	–
Various aetiologies	Yes (7)	Yes (25)	Elderly at risk of falling (29)
Setting (measurement range)	Hospital and community (1)	Hospital and community (26)	Community (25)
Cut-off scores available	Yes (27)	Yes (26)	Yes (29)
Balance subdomains covered	A, B, C	A, B, C, D, E	B, C, D, E

Appropriate references are shown in parentheses. A, B, C, D, and E represent, respectively, the balance concepts presented in Table II.
BBS: Berg Balance Scale; POMA: Performance Oriented Mobility Assessment; FAB: Fullerton Advanced Balance Scale.

of each instrument in terms of balance concepts. A referenced list of the excluded scales can be obtained on request from the corresponding author.

Patients recruited and procedures

All observations were collected by 4 raters (rater 1 and 2: 14.8% observations each; rater 3 and rater 4: 36.8% and 33.4%, respectively, of total observations) on a convenience sample of 217 patients. The mean age of the patients was 59.5 years (SD 16.3; median 64 years) and 60.8% were men. Ischaemic stroke was the most common aetiology (48.8% of cases), followed by intracerebral haemorrhage (18.0%) and traumatic brain injury (11.1%). A range of other rarer aetiologies included subarachnoid haemorrhage (6.9%), central nervous system neoplasms (5.1%), mielopathies (3.7%) and peripheral neuropathies (3.2%).

For 85 inpatients pre-treatment and post-treatment observations were available, making a total sample of 302 observations available for analysis. Pre-test observations were made, on mean, 22.4 days (SD 26; median 15 days) after the acute event, whereas post-test observations took place, on mean, 49.3 days (SD 39; median 39 days) after the acute event. Considering all cases, observations took place, on mean, 198.1 days (SD 869.8; median 33 days) after the diagnosis. Given the skewed distribution implied by these data, after excluding 26 patients with time since lesion >6 months (20 of them were outpatients), for the remaining 275 observations the assessment took place on mean 38.9 days (SD 30.5; median 31 days) after admission to the hospital. The median TCT score was 100 (mean score 87.4; scale range 0–100), whereas the median motor-FIM and cognitive-FIM scores were, respectively, 56 (mean score 54.7; scale mid-point 52) and 30 (mean score 27.5; scale mid-point 20).

Statistical analyses

Classical item and scale descriptive statistics. Analysis of the response category frequencies showed that TB01 (sitting balance) and FAB05 (tandem walk) were, respectively, the items with the highest ceiling effect (95.0%) and floor effect (79.5%).

Item missing value analysis showed that BBS, TB and TW had missing data below a median of 1%, whereas FAB showed a higher percentage at 2.4%. Considering the whole item set, the median missing item rate was 0.3%. Item FAB10 (reactive postural control) had the highest percentage of missing values (11.4%).

Person missing value analysis showed that complete data were available for 246 observations. The analysis confirmed that FAB had the highest content of missing items, accounting for 69.2% of missing total scores, whereas BBS and TW accounted for, respectively, 13.1% and 12.3% of missing total scores. TB was the scale with the least missing responses (5.4% of non-calculable total scores). The 56 observations with missing values had from 1 to 13 missing items (median 1; mean 2.3).

The mean inter-item Spearman's correlation index was 0.667 (range −0.052–0.976) and the items assessing sitting balance (BBS03 and TB01) were the only 2 items that had a mean inter-item correlation lower than 0.50 (0.33 and 0.23, respectively). The total scale score-item Spearman's correlation index was high (mean correlation 0.81; range 0.34–0.92) and, again, BBS03 and TB01 were the only 2 items with significantly lower correlations than other items (0.46 and 0.34, respectively). Finally, analysis of internal consistency reliability on complete data showed a very high Cronbach's alpha value (0.983), compatible with measurement at the individual level.

Preliminary assessment of unidimensionality of the merged item set

The CFA on the 40 items demonstrated sufficient unidimensionality for taking the item set forward with an RMSEA of 0.062, a SRMR of 0.078, a CFI of 0.996, and a

Table II. *Content validity of the selected scales*

Scales/items	A	B	C	D	E
Berg Balance Scale (BBS)					
BBS02: Standing unsupported	*				
BBS03: Sitting unsupported	*				
BBS01: From sitting to standing		*			
BBS04: From standing to sitting		*			
BBS05: Transfers		*			
BBS07: Standing with feet together		*			
BBS08: Reaching forward while standing		*			
BBS09: Retrieving object from floor		*			
BBS10: Turning trunk (feet fixed)		*			
BBS11: Turning 360°		*			
BBS13: Tandem standing		*			
BBS14: Standing on 1 leg		*			
BBS12: Placing alternate foot on stool		*			
BBS06: Standing with eyes closed			*		
Fullerton Advanced Balance Scale (FAB)					
FAB02: Reaching forward to an object		*			
FAB03: Turn in full circle		*			
FAB06: Stand 1 leg		*			
FAB08: Two-footed jump		*			
FAB01: Standing feet together, eyes closed			*		
FAB07: Stand on foam, eyes closed			*		
FAB10: Reactive postural control				*	
FAB04: Step up and over					*
FAB05: Tandem walk					*
FAB09: Walk with head turns					*
Performance Oriented Mobility Assessment (POMA)					
TB01: Sitting balance	*				
TB04: Immediate standing balance	*				
TB02: Arising from chair		*			
TB03: Attempts to arise		*			
TB05: Standing balance (feet together)		*			
TB08: Turning 360°		*			
TB09: Sitting down		*			
TB07: Standing with eyes closed			*		
TB06: Nudged (being pushed)				*	
TW10: Gait: initiation					*
TW11: Step length and height					*
TW12: Step symmetry					*
TW13: Step continuity					*
TW14: Gait path					*
TW15: Trunk stability					*
TW16: Walking stance					*

Items of BBS, FAB and POMA were linked to 5 conceptual subdomains for balance. Thus, the conceptual content of each scale could be examined separately and then compared with that of the other scales. A: quiet stance; B: anticipatory postural adjustments/transitions; C: sensory orientation; D: external perturbations; E: stability in gait.
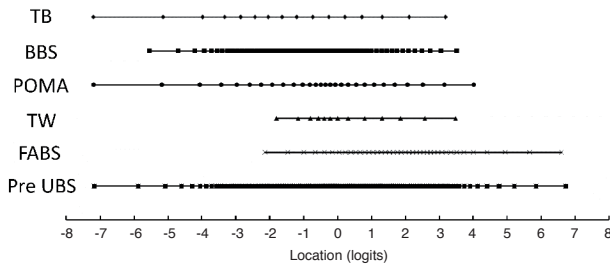
Fig. 1. Comparison of the operational ranges of the balance scales. Given the limitation of the lack of fit to the Rasch Model, the graph shows that Tinetti Balance (TB) and Fullerton Advanced Balance scale (FAB) were found to be, respectively, the easiest and the most difficult scales, whereas Performance Oriented Mobility Assessment (POMA) (TB+Tinetti Walking (TW)) had an intermediate level of difficulty. Furthermore, the total item set (Pre-analysis Unified Balance Scale, Pre-UBS) had a much wider measurement range than its originating scales.

TLI of 0.995. The standardized loadings of the items were in the range 0.680–0.969, with the only 2 items assessing static sitting balance (BBS3 and TB1) having significant lower loadings than other items.

Rasch analysis. A crude comparison of the operational range of the scales and of the total item set against a common underlying metric is shown in Fig. 1. This shows that TB and FAB had, respectively, the lowest and the highest relative difficulty, whereas BBS and TW had, respectively, the wider and the shorter range of measurement. Furthermore, the total item set had a much wider measurement range than its originating scales. There were minimal floor and ceiling effects (1% each) when all scales were considered together.

The initial Rasch analysis performed on the 40-item set showed serious misfit to the model, failing the assumptions of stochastic ordering, local independence and unidimensionality. The item analysis showed that 6 items (15%) overfitted the model (their response pattern was too predictable), 5 items

(12.5%) had highly significant χ^2 values (signalling a lack of the expected stochastic ordering) and 20 items (50% of the total item set available for analysis) had disordered thresholds.

The next stage of the analysis involved the re-scoring of items with disordered thresholds according to clinically meaningful criteria specific for any individual item. After re-scoring the 20 items with disordered thresholds, the scale still failed to meet the model's expectations in terms of uni-dimensionality and invariance. Furthermore, analysis of the inter-item residual correlations showed that 18 pairs of items had residual correlations above 0.3, indicating the presence of local dependency, with several items involved in more than one pair. The item analysis showed that 5 items (12.5%) still showed misfit to the model.

After resolving all locally-dependent pairs of items, the remaining 31-item set thus obtained still failed to meet the model's expectations in terms of unidimensionality and invariance. After elimination of 3 misfitting items and re-scoring of a further item, the remaining 27-item set finally satisfied the model's expectations in terms of strict unidimensionality, local independence and invariance. All items showed to fit the model individually. At this stage, analysis of DIF was performed by testing the following factors: *gender, age* (≤ 55; 56–68; and ≥ 69 years), *days since lesion* (≤ 20; 21–43; ≥ 44 days), *aetiology* (ischaemic stroke; haemorrhagic stroke; TBI; other) and *number of evaluations for patients* (single, pre-test, post-test evaluations). No DIF was found for any of the tested group factors. Regarding persons, only one individual had a fit residual >2.5 (+4.03), indicating a significant departure from the model's expectations.

This final set of 27 items forms a new scale, the UBS, which has a total score ranging from 0 to 65. The contribution of the originating scales to the total score was as follows: BBS: 55.4%; FAB: 24.6%; and Performance Oriented Mobility Assessment (POMA): 21.5%. The UBS' item calibrations and the re-scoring pattern, as well as all the deleted items, are
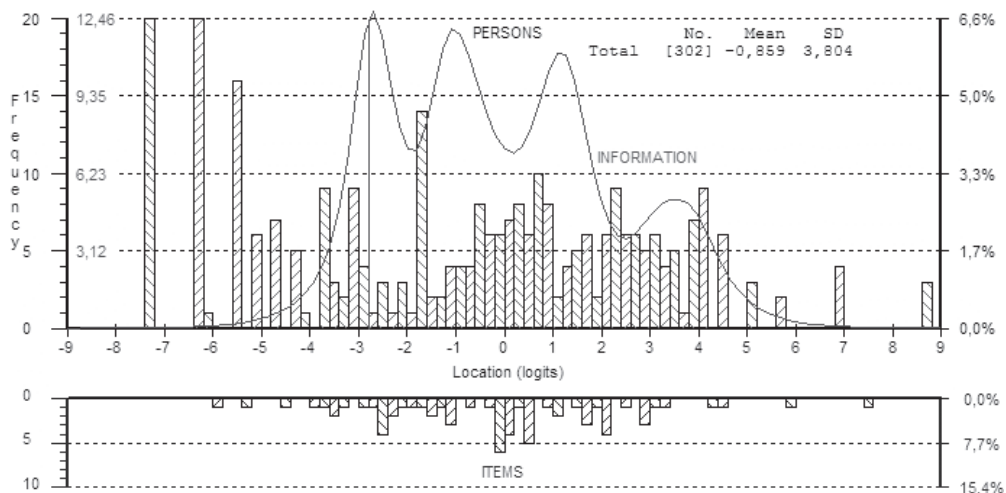

Fig. 2. Targeting of the Unified Balance Scale. Observations (*n*=302) and item thresholds are shown in the upper and the lower part of the graph, respectively, separated by the logit scale. Grouping set to interval length of 0.20 making 85 groups. Item thresholds provide 4 peaks of information, at –3.0, –1.0, +1.5 and 3.5 logits, respectively. SD: standard deviation.

Table III. *Items' parameter, fit statistics, scoring model and conceptual content for the Unified Balance Scale (UBS)*

Final UBS items	Item parameters and fit statistics					Balance concept					Scoring model				
	Loc	SE	FR	χ^2	Prob	A	B	C	D	E	0	1	2	3	4
TB04: Immediate standing balance	−3.33	0.16	−0.18	5.96	0.20	*					0	1	2	–	–
TB03: Attempts to arise	−3.03	0.16	0.39	5.49	0.24		*				0	1	2	–	–
BBS05: Transfers	−2.76	0.11	0.96	12.71	0.01		*				0	1	2	3	4
BBS04: From standing to sitting	−2.75	0.11	−0.01	1.44	0.84		*				0	1	2	3	4
BBS01: From sitting to standing	−2.43	0.13	−0.91	6.89	0.14		*				0	1	1	2	3
BBS06: Standing with eyes closed	−1.55	0.15	−1.82	12.39	0.01			*			0	1	1	1	2
TB05: Standing balance (feet together)	−1.51	0.15	−1.32	3.70	0.45		*				0	1	2	–	–
TW10: Gait: initiation	−1.18	0.21	−0.97	8.10	0.09					*	0	1	–	–	–
BBS10: Turning trunk (feet fixed)	−1.18	0.11	−0.80	12.31	0.02		*				0	1	2	3	4
BBS07: Standing with feet together	−0.95	0.12	−2.15	8.71	0.07		*				0	1	2	2	3
FAB01: Standing feet together, eyes closed	−0.86	0.15	−0.48	5.01	0.29			*			0	1	1	2	2
BBS08: Reaching forward while standing	−0.61	0.12	−1.56	3.02	0.56		*				0	1	2	2	3
BBS09: Retrieving object from floor	−0.53	0.15	−0.96	10.57	0.03		*				0	1	1	1	2
TB06: Nudged (being pushed)	−0.34	0.14	0.04	6.87	0.14				*		0	1	2	–	–
TW11: Step length and height	−0.06	0.20	−1.57	4.85	0.30					*	0	0	0	0	1
TB07: Standing with eyes closed	0.42	0.19	0.31	8.46	0.08			*			0	1	–	–	–
FAB07: Stand on foam, eyes closed	0.56	0.14	−1.09	1.86	0.76			*			0	1	1	1	2
TW15: Trunk stability	0.96	0.14	−0.34	1.11	0.89					*	0	1	2	–	–
BBS12: Placing alternate foot on stool	1.16	0.11	−0.53	3.93	0.42		*				0	1	2	2	3
BBS13: Tandem standing	1.20	0.11	−0.58	1.15	0.89		*				0	1	1	2	3
FAB10: Reactive postural control	1.26	0.15	−0.02	6.33	0.18				*		0	1	1	1	2
BBS14: Standing on 1 leg	1.27	0.11	−0.55	5.16	0.27		*				0	1	2	2	3
BBS11: Turning 360°	1.47	0.14	−1.60	9.78	0.04		*				0	0	1	1	2
FAB04: Step up and over	3.14	0.13	−0.15	0.90	0.93					*	0	1	1	2	3
FAB09: Walk with head turns	3.78	0.18	−0.51	3.48	0.48					*	0	1	1	1	2
FAB08: Two-footed jump	3.86	0.14	−0.40	1.19	0.88		*				0	1	2	2	3
FAB05: Tandem walk	3.99	0.17	−0.17	2.50	0.64					*	0	1	1	1	2

Deleted items	Reason for deletion					A	B	C	D	E					
FAB06: Stand 1 leg	Local dependency						*								
TB02: Arising from chair	Local dependency						*								
TW12: Step symmetry	Local dependency								*						
TW14: Gait path	Local dependency								*						
TB09: Sitting down	Local dependency						*								
FAB03: Turn in full circle	Local dependency						*								
FAB02: Reaching forward to an object	Local dependency						*								
TB01: Sitting balance	Local dependency					*									
TW13: Step continuity	Local dependency								*						
BBS03: Sitting unsupported	Invariance violation					*									
BBS02: Standing unsupported	Invariance violation					*									
TW16: Walking stance	Invariance violation								*						
TB08: Turning 360°	To correct overfit of BBS11						*								

In the upper part of the table, the final 27 UBS items, ordered by progressive difficulty from top to bottom, are displayed.
Loc: item difficulty expressed in logits; SE: standard error of measurement; FR: fit residual; Prob: χ^2 probability. The degrees of freedom for each χ^2 were 4 for all items. For each item the re-scoring pattern is also presented. For instance, for item BBS01, the first, fourth and fifth original categories remained unchanged, whereas the second and third ones were collapsed into 1 category (01123). A, B, C, D, E represent, respectively, the balance concepts presented in Table II: A: quiet stance; B: anticipatory postural adjustments/transitions; C: sensory orientation; D: external perturbations; E: stability in gait. In the lower part of the table the 13 deleted items are displayed (in order of deletion).

shown in Table III. The item hierarchy suggests that the easiest items are related, respectively, to quiet standing (TB04), postural change from sitting to standing (TB03 and BBS01), transfers (BBS05), and postural change standing to sitting (BBS04). On the other hand, the most difficult items were, respectively, turns (BBS11), stepping (FAB04), walking with head turns (FAB09), jumping (FAB08) and tandem walking (FAB05). The same table shows that all 5 balance concepts were represented within the UBS: quiet stance (1 item), anticipatory postural adjustment/transitions (14 items), sensory orientation (4 items), external perturbations (2 items) and stability in gait (6 items).

The person-item distribution map of the UBS (Fig. 2) shows that persons were evenly spread across 16 logits, although 20 persons were at the floor of the scale (floor effect: 6.6%). Two further groups of 20 and 16 persons were located, respectively, 1 and 2 logits further to the right in view of the relative poverty of score thresholds in this area of the measurement continuum. There was no evident ceiling effect (1%). The mean person ability of −0.859 logits indicated that, on average, the ability of the sample was relatively lower than the mean difficult of the item set. The mean standard error of the person ability estimates was 0.575, with a 95% CI of ±1.127 logits (after exclusion of extreme persons, respectively, 0.510 and ±0.999

logits) with a person reliability, expressed as PSI, of 0.971. As a consequence, persons could be separated in 9.04 ability strata.

On the basis of the item calibrations, it was possible to construct a ruler to convert the UBS total score (obtained after item re-scoring) to linear estimates of ability. A working example of the UBS ruler is shown in Fig 3. The scoring guidelines, a training video, and both paper and pencil and Excel-based version of the UBS are available on request from the corresponding author.

Final confirmatory factor analysis. A final CFA confirmed the unidimensionality of the 27-item UBS, with CFI = 0.998; TLI = 0.998, RMSEA = 0.057, and SRMR = 0.033. The standardized loadings of the items were all 0.9 and above.

DISCUSSION

In this paper, a new activity-based, bed to community and aetiology-independent measure of balance within the neurorehabilitation setting was built by merging items from 3 existing balance scales using a combination of classical psychometric methods and Rasch analysis. In this way, a new 27-item scale (the UBS) was built, and it was demonstrated that it had superior metric properties than its originating scales; that is, a wider measurement range, being independent of the aetiology of the lesion causing the loss of balance, and covering all relevant aspects of balance from a conceptual standpoint. As UBS items satisfy the strict modern psychometric criteria required by the Rasch Model, it's simple summated raw score is a sufficient statistic for an estimate of a person's ability that can easily be converted into a linear estimate of balance satisfying the rules for fundamental measurement (30).

In this study, it was decided to make use of existing scales rather than building a new instrument, considering the relative abundance of currently available assessment tools for balance (7). This approach is likely to provide significant benefits from the clinical, the service provision, and the research standpoints. From the clinical point of view, the fact that 75% of UBS items are derived from well-known scales, such as the POMA and the BBS, may facilitate its applications in routine clinical contexts. From the service standpoint, it is envisaged that there may be little need to spend extra time and money in learning how to administer it, thus enhancing its acceptability by practitioners in comparison with a totally "new instrument". Finally, from a research perspective, the UBS might facilitate pooling of historical as well as of prospective data, both within a single centre and across difference centres and within different settings (i.e. the hospital and the community), thus providing a common frame of measurement for balance.

In order to select the most appropriate instruments for this study, we used criteria aiming at facilitating the acceptability (popularity), the generalizability (possibility of being applied to patients with balance problems from various aetiologies and the widest possible range of measurement), the interpretability (availability of cut-off scores for predicting the risk of falls) and, finally, the conceptual validity (coverage of as many subdomains of balance as possible) of the new instrument. The availability of cut-off scores is an important criterion for elucidating how well the new instrument will relate to widely used existing scale and their estimates for the risk of falling. In turn, this may facilitate the acceptability of the UBS. As far as the conceptual validity is concerned, we adopted a slightly modified version of the conceptual model for balance provided by Franchignoni et al. (8). This provided a very efficient selection framework for the balance scales. For instance, a popular instrument such as TCT was excluded from the candidate scales in view of the presence of 2 items (rolling to weak and, respectively, strong side) that could not be linked to any balance subdomains, thus making explicit that the scale measured a separate construct from balance. At the same time, a relatively popular scale, the Dynamic Gait Index, was discarded in view of the fact that it appeared to cover only the gait balance subdomain, whereas it did not assess "tandem walking", which was deemed to be an important activity for balance measurement. For the opposite reason, the FAB was preferred, although less popular, in view of the coverage of 4 out of 5 balance subdomains.

When merging items from different scales, it is of paramount importance that the chosen item set is unidimensional (30). As the hypothesis that these 3 scales measured the same construct was supported by analysis of literature, of content validity, as well as by classical item analysis of the chosen scales, we used CFA instead of EFA as a first choice to test the assumption of unidimensionality prior to fitting the data to the Rasch Model (12). As data are often modified during the Rasch analysis; for example, by deleting locally dependent items, and *post-hoc* tests for unidimensionality are then undertaken on the modified item set, our purpose was to guide the Rasch analysis by avoiding clear multi-dimensional data, which breaches the assumption of the model, and which is often difficult to deal with during the analytical process. As the objective, at this stage, was to identify a candidate set of items sufficiently unidimensional for Rasch analysis, we allowed more relaxed criteria for unidimensionality (RMSEA ≤0.10). Although CFA demonstrated that the 40-item set was sufficiently unidimensional to be taken forward for the subsequent Rasch analysis, the preliminary analyses signalled that the static sitting posture items (BBS03 and TB01) behaved somehow differently from the other items. Indeed, those 2 items were both eliminated within the subsequent analysis. A possible explanation for their different behaviour may lie in the fact that static postural control in sitting is likely to be influenced, especially in low-ability hemiparetic patients, also by trunk control, which is an allied but separate construct from dynamic balance.

Within the Rasch analytical framework we needed to undertake an extensive re-scoring of most BBS and FAB items. This was entirely expected for both scales, as previous studies had already demonstrated suboptimal category functioning for BBS (8, 27) and FAB had very similar scoring criteria to BBS. Also, the discovery of several locally-dependent items was an entirely expected finding, considering the partial conceptual overlap between several items of the 3 scales (e.g. "pivot turns" were addressed by the following 3 items: BBS11, TB08, and FAB03). Local dependency may inflate reliability and leads to incorrect
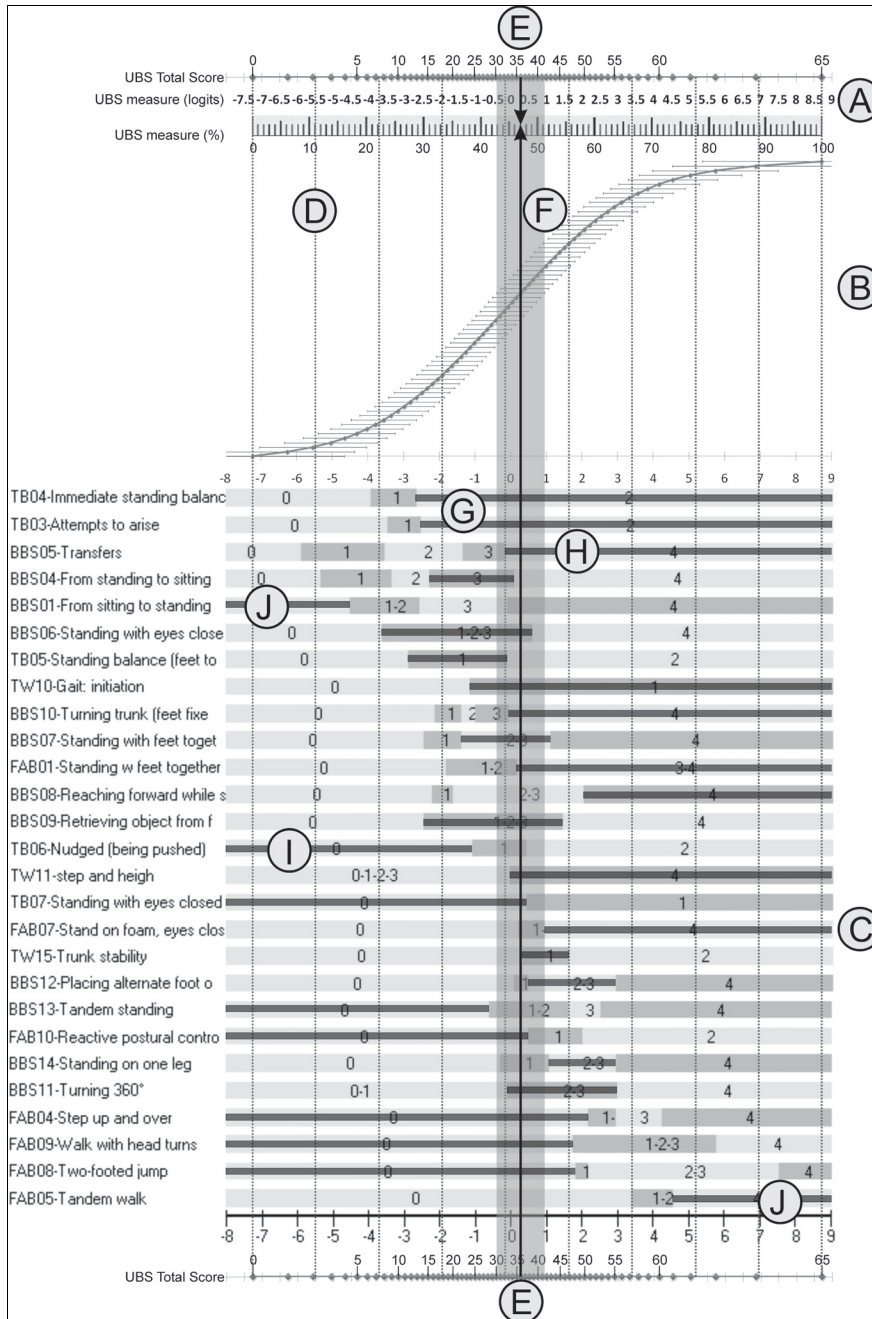
Fig. 3. Working example of the Unified Balance Scale (UBS) ruler. The UBS ruler contains 3 parts: the Rasch Nomogram (A), which allows the conversion of total score into linear measures of balance ability; a graph (B) plotting the confidence interval due to measurement error around each total score; and (C) the individual item score map. The vertical dotted lines (D) allow 9 statistically distinct strata of ability to be distinguished. For this patient, a total score of 36 was generated, transforming the item raw scores with the re-scoring key provided in Table III. A vertical arrowed line crossing the obtained UBS total score (E) was drawn. Thus, it is possible to convert the total score into measures of balance ability, expressed in both logits (0.3) and percentage (47%). Two lines parallel to the measurement line were plotted considering the lower and upper 95% confidence interval around the person's ability, thus defining a measurement area (F) that contains the true measure of the subject with a confidence of 95%. Thus, it was possible to convert the total score of 36 into a balance estimate of 47±4%, falling within the fifth stratum. The black horizontal bars on the score thresholds indicated the range of ability flagged by the given responses to each item. Most of the bars cross the measurement area as expected, either completely (G) or partially (H). However, the responses to some items do not cross the measurement area. Some of these responses are only 1 stratum away from the measurement area (I) and, therefore, are still compatible with the probabilistic expectations of the model. However, the responses to 2 items (J) are several strata away from the measurement area, thus signalling a significant departure from the model's expectation. It is notable that this individual patient appears to be unable to rise from sitting to standing, although he does appear able to perform tandem walking independently. This pattern belonged to the only subject who misfitted the model's expectations (person fit residual: +4.03). Careful inspection of his record demonstrated a transcription error of these 2 items' scores due to carelessness.

person estimates and, therefore, it has to be addressed properly (12). One way of doing so is to group items together in the form of "testlets" (a form of "super-item") (30, 31) although, in this instance, we preferred to adopt a strategy of the deletion of one item within a locally dependent pair, with the specific aim of reducing the item set to improve the usability of the scale.

The analysis of DIF demonstrated that the scale was also invariant across key demographic and clinical groups of patients. So, for instance, it was possible to demonstrate that the linear estimates of balance provided by the UBS were invariant across patients' age, the various aetiologies of the lesion causing the loss of balance, as well as across time. The lack of DIF for aetiology is particularly interesting from a clinical and research standpoint, as it ensures comparability of the assessment of patients and of the efficacy of treatments across a variety of balance problems with different neurological causes. On the other hand, regarding the lack of DIF across time, two relevant aspects were examined, i.e. the time elapsed from the onset and the possible effect of treatment on the estimation of item difficulty. The lack of DIF for all these factors suggests that the UBS may be used at various stages of the rehabilitation process, with younger or older patients, affected by acute or chronic balance problems.

The internal construct validity of the UBS was supported by several pieces of evidence. First, it contained items linking to all 5 recognized subdomains of balance, although some of these are less represented than others (such as "static balance" and "external pertubations"). Secondly, the hierarchical ordering of the items was consistent with clinical observations, as the easiest activities are the earliest motor tasks requiring balance (e.g. quiet standing, transfers, and postural changes from standing to sitting and vice versa), whereas the opposite end of the scale includes the most challenging dynamic activities requiring balance (turning, jumping, walking). Between those extremes the scale includes a variety of standing activities requiring balance, thus ensuring ample coverage of all balance subdomains. Finally, its items covered all activities that are associated with an increased risk of falls at various levels of functional independence, from the easier motor activities (transfers (32), postural changes from standing to sitting and vice versa (33) and standing (33) to the more challenging dynamic one (1-leg stance (34), reach forward task (34), pivot turns (32), and walking (32)).

The final UBS has an item content that makes it suitable for the assessment of both hospitalized and community-dwelling adults, supported by its wide measurement range. The possibility of measuring balance from bed to community with a single instrument with proven measurement properties may facilitate the follow-up of patients after discharge from the hospital or the assessment of community-dwelling older adults at risk of falls, thus avoiding the well-known ceiling effect problems of popular instruments such as the BBS (1) or the POMA (28). Notwithstanding the wide measurement range, the UBS was demonstrated to have an excellent level of reliability (0.971), as such that the scale was able to separate the sample into 9 statistically different groups of abilities that could offer the potential of identifying as many groups of patients with a progressively decreasing risk of falling.

Although not directly tested, some considerations can also be made on inter-rater reliability, considering that data were collected by 4 raters. Indeed, it is likely that any possible inconsistencies in ratings negatively influenced the scoring function of individual items and therefore were "absorbed" by the re-scoring procedure. As a consequence, the achievement of a stable scoring structure after re-scoring for all items across 4 different raters may indirectly suggest a high level of inter-rater reliability. This may be explained considering that the UBS is largely based on very well-known balance scales and that a detailed manual with scoring guidelines and a video were available. Also raters underwent only one session of patient-based training. This further suggests that the UBS may be easy to learn and to administer and, therefore, may easily be acceptable by both more- and less-experienced practitioners.

The UBS clinical ruler presented in this paper can also contribute to the acceptability of this new scale, considering that it does not only provide a simple method to transform the raw scores into linear measures of balance, but also can inform the practitioner on additional important aspects that can influence the individual patient's assessment and treatment. First, it provides a simple quality-control method for the measurement process of individual patients. By examining the item response pattern it may be possible to discover inconsistencies in ratings and a search for their most likely causes. In case of a rater's error or inconsistencies, this approach may provide instant feedback on how to correct the error and to improve future ratings. Secondly, the appraisal of individual response pattern may uncover areas of both strength and weakness (respectively, rating above and below the patient's measurement area) for that individual patient. This may constitute the basis for individualized treatment plans. Finally, the ruler provides a simple method correctly to interpret change scores occurring over time.

This study has a number of limitations. First, not all possible balance scales published in the literature (7, 8) or items could be evaluated and/or selected for the study. A second limitation relates to the main selection criteria adopted to enrol the sample. This was a convenience sample representing a cross-section of adults drawn from a single rehabilitation centre. This may limit the possibility of generalization of these findings to other samples, This, in turn, may have affected the analysis of DIF for aetiology as not all possible aetiologies encountered in all rehabilitation settings were tested. A third limitation is related to the sample size that, although sufficient for Rasch analysis, was too small to have a "set aside" sample which would have enabled us to validate the final scale further. There is thus a risk that the solution we have obtained has capitalized on chance with respect to fit to the model. As such, the raw-score interval scale transformation should be considered provisional at this time. Consequently, these findings require replication. A final limitation is related to the suboptimal targeting of the UBS, in view of the fact that approximately 20 % of the total observations were located near the floor of the scale. This is easily explained considering the specific admission criteria for the assessment of balance applied to a population of patients admitted to early

rehabilitation. We could have used more restrictive admission criteria, although we preferred not to do so, considering that this low-ability group of patients is especially at risk for falls (32, 33) and therefore may take the greatest advantage from measurement of their balance ability.

Future studies are required to explore the classic psychometric and clinimetric profile for the UBS in order to evaluate, beyond internal construct validity, whether the UBS effectively measures balance according to external validation criteria. Furthermore, analysis of responsiveness (sensitivity to change) and interpretability (identification of cut-off scores with specific clinical meaning) are warranted. Another issue that needs to be addressed in future studies regards the usability of the UBS in routine clinical setting. The administration of 27 items may impose an important administration burden that may limit its usability in routine clinical contexts. As a consequence, methods to improve its usability are being developed.

In conclusion, the UBS has proven measurement properties and may be a candidate tool for balance measurement and fall risk assessment from bed to community for patients with balance problems with different neurological causes. Future studies are warranted to explore its external validity and other clinimetric properties, as well as to improve its usability.

ACKNOWLEDGEMENTS

REFERENCES

1. Blum L, Korner-Bitensky N. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. Phys Ther 2008; 5: 559–566.
2. Campbell M, Parry A. Balance disorder and traumatic brain injury: preliminary findings of a multi-factorial observational study. Brain Inj 2005; 19: 1095–10104.
3. Bradley WG, Daroff RB, Fenichel G, Jankovic J. Neurology in clinical practice. 5th ed. Philadelphia: Butterworth-Heinemann; 2008.
4. Teasell R, McRae M, Foley N, Bhardwaj A. The incidence and consequences of falls in stroke patients during inpatient rehabilitation: factors associated with high risk. Arch Phys Med Rehabil 2002; 83: 329–333.
5. Beninato M, Portney LG, Sullivan PE. Using the International Classification of Functioning, Disability and Health as a framework to examine the association between falls and clinical assessment tools in people with stroke. Phys Ther 2009; 8: 816–825.
6. Muir SW, Berg K, Chesworth B, Klar N, Speechley M. Quantifying the magnitude of risk for balance impairment on falls in community-dwelling older adults: a systematic review and meta-analysis. J Clin Epidemiol 2009; 4: 389-406.
7. Tyson SF, Connell LA. How to measure balance in clinical practice. A systematic review of the psychometrics and clinical utility of measures of balance activity for neurological conditions. Clin Rehabil 2009; 23: 824–840.
8. Franchignoni F, Horak FB, Godi M, Nardone A, Giordano A. Using pyschometric techniques to improve the balance evaluation systems test: the mini-Bestest. J Rehab Med 2010; 42: 316–324.
9. Elhan AH, Öztuna D, Kutlay S, Küçükdeveci AA, Tennant A. An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. BMC

10. Musculoskelet Disord 2008; 9: 166.
10. Ragnarsdottir M. The concept of balance. Physiotherapy 1996; 82: 6: 368–375.
11. World Medical Association. Declaration of Helsinki: ethical principles for medical research involving human subjects. Ferney-Voltaire: WMA; 2002.
12. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007; 45 Suppl 1: S22–S31.
13. Andrich D. Rasch models for measurement. London: Sage Publications; 1988.
14. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004; 42 Suppl 1: I37–I48.
15. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007; 46: 1–18.
16. Smith E. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3: 205–231.
17. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002; 3: 85–106.
18. Wright BD, Masters GN. Rating Scale Analysis. Chicago: MESA Press; 1982.
19. Baghaei P. The Rasch Model as a construct validation tool. Rasch Meas Transact 2008; 12: 1145–1146.
20. Tabachnick BG, Fidell LS. Using multivariate statistics. 4th ed. New York: Harper Collins; 2001.
21. Linacre JM. Sample size and item calibration stability. Rasch Meas Transact 1994; 7: 328.
22. Bland J, Altman D. Multiple significance tests: the Bonferroni method. Br Med J 1995; 310: 170.
23. Berg KO, Wood-Dauphinee SL, Williams JI, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health 1992; 83 Suppl 2: S7–S11.
24. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. J Am Geriatr Soc 1986; 34: 119–126.
25. Rose DJ, Lucchese N, Wiersma LD. Development of a multidimensional balance scale for use with functionally independent older adults. Arch Phys Med Rehabil 2006; 87: 1478–1485.
26. Faber MJ, Bosscher RJ, van Wieringen PC. Clinimetric properties of the performance-oriented mobility assessment. Phys Ther 2006; 86: 944–954.
27. Kornetti DL, Fritz SL, Chiu YP, Light KE, Velozo CA. Rating scale analysis of the Berg Balance Scale. Arch Phys Med Rehabil 2004; 85: 1128–1135.
28. Raiche M, Hebert R, Prince F, Corriveau H. Screening older adults at risk of falling with the Tinetti balance scale. Lancet 2000; 356: 1001–1002.
29. Hernandez D, Rose DJ. Predicting which older adults will or will not fall using the Fullerton Advanced Balance scale. Arch Phys Med Rehabil 2008; 89: 2309–2315.
30. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human science. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2007.
31. Wainer H, Kiely G. Item clusters and computer adaptive testing: a case for testlets. J Educ Meas 1987; 24: 185–202.
32. <Perell KL, Nelson A, Goldman RL, Luther SL, Prieto-Lewis N, Rubenstein LZ. Fall risk assessment measures: an analytic review. J Gerontol A Biol Sci Med Sci 2001; 56: M761–M766.
33. Muir SW, Berg K, Chesworth B, Klar N, Speechley M. Balance impairment as a risk factor for falls in community-dwelling older adults who are high functioning: a prospective study. Phys Ther 2010; 90: 338–347.