ORIGINAL REPORT

USING PSYCHOMETRIC TECHNIQUES TO IMPROVE THE BALANCE EVALUATION SYSTEMS TEST: THE MINI-BESTEST

Franco Franchignoni, MD[1], Fay Horak, PT, PhD[2], Marco Godi, PT[3], Antonio Nardone, MD[3,4] and Andrea Giordano, PhD[5]

From the [1]Unit of Occupational Rehabilitation and Ergonomics, Salvatore Maugeri Foundation, Clinica del Lavoro e della Riabilitazione, IRCCS, Veruno, Italy, [2]Department of Neurology, Oregon Health and Science University, Portland, OR, USA, [3]Department of Physical Medicine and Rehabilitation, Salvatore Maugeri Foundation, Clinica del Lavoro e della Riabilitazione, IRCCS, Veruno, [4]Department of Clinical and Experimental Medicine, University of Eastern Piedmont, Novara and [5]Unit of Bioengineering – Salvatore Maugeri Foundation, Clinica del Lavoro e della Riabilitazione, IRCCS, Veruno, Italy

Objective: To improve, with the aid of psychometric analysis, the Balance Evaluation Systems Test (BESTest), a tool designed to analyse several postural control systems that may contribute to poor functional balance in adults.

Methods: Performance of the BESTest was examined in a convenience sample of 115 consecutive adult patients with diverse neurological diagnoses and disease severity, referred to rehabilitation for balance disorders. Factor (both explorative and confirmatory) and Rasch analysis were used to process the data in order to produce a new, reduced and coherent balance measurement tool.

Results: Factor analysis selected 24 out of the 36 original BESTest items likely to represent the unidimensional construct of "dynamic balance". Rasch analysis was then used to: (*i*) improve the rating categories, and (*ii*) delete 10 items (misfitting or showing local dependency). The model consisting of the remaining 14 tasks was verified with confirmatory factor analysis to meet the stringent requirements of modern measurement.

Conclusion: The new 14-item scale (dubbed mini-BESTest) focuses on dynamic balance, can be conducted in 10–15 min, and contains items belonging evenly to 4 of the 6 sections from the original BESTest. Further studies are needed to confirm the usefulness of the mini-BESTest in clinical settings.

Key words: postural balance, outcome assessment, psychometrics.

J Rehabil Med 2010; 42: 323–331

Correspondence address: Franco Franchignoni, Fondazione Salvatore Maugeri, Clinica del Lavoro e della Riabilitazione, IRCCS, Via Revislate 13, IT-28010 Veruno, Italy. E-mail: franco.franchignoni@fsm.it

INTRODUCTION

Assessment of balance and mobility in clinical settings can help to determine both risk of falling (1) and the most suitable measures to reduce postural instability (2–3). Laboratory studies have shown that postural control embraces different subdomains, including stability during quiet stance, postural reactions to external disturbances, anticipatory postural adjustments to perturbations caused by self-initiated movements (e.g. lifting an object), and dynamic balance during gait (4). However, until recently clinical balance tests did not systematically evaluate all these subdomains (5–6).

Recently, a new clinical tool for assessing subdomains underlying balance deficits has been presented: the Balance Evaluation Systems Test (BESTest) (7). The BESTest is a comprehensive balance assessment tool developed to identify the postural control systems underlying poor functional balance, so that treatments can be targeted to the specific balance deficit. Since the BESTest encompasses 4–6 items for each of 6 different balance domains, it takes approximately 35 min to administer, compared with only approximately 15 min for other balance scales (e.g. the Berg Balance Scale; BBS) (8). This is an important shortcoming of the BESTest, limiting its routine use. On the other hand, the main disadvantage of other popular balance scales, including the BBS, is that they do not include important aspects of dynamic balance control, such as the capability to react to postural perturbations, to stand on a compliant or inclined surface, or to walk while performing a cognitive task. All of these features of balance control are known to be important in assessing balance disorders in different types of patients, and reflect balance challenges during activities of daily living (5, 7, 9). Therefore, there is need for a comprehensive balance assessment tool that can be administered in a short time period.

In developing and validating new clinical instruments there is a growing trend of using Rasch analysis (10). Whereas traditional psychometric approaches focus on an instrument's total score, item response theory (IRT) models, such as the Rasch measurement models, are founded on the probability that a person will make a particular response according to their level of the underlying latent variable. In this framework, it is possible to evaluate how well an item performs in terms of its relevance or contribution for measuring the underlying construct, the level of the underlying construct targeted by the question, the possible redundancy of the item relative to other items in the scale, and the appropriateness of the response categories (11). For these reasons, Rasch analysis has been recommended as a

complementary method to assess the scaling properties of new clinical instruments, in addition to the traditional psychometric criteria for disability outcomes research (12).

The purpose of this study was to use both classical psychometric techniques and Rasch analysis to evaluate the BESTest, investigating a wide range of measurement requirements (e.g. dimensionality, quality of the rating categories, construct validity, reliability indexes) in order to improve the structure and measurement qualities of the test. Based on this analysis, we present a new, mini-BESTest that focuses on dynamic balance and can be conducted in 10–15 min.

MATERIAL AND METHODS

Patients

A total of 115 patients (53 men and 62 women), mean age 62.7 years (standard deviation, (SD) 16), were studied. They represent a convenience sample of patients with balance disorders, recruited with a consecutive sampling method. Patient diagnosis was as follows: 22 hemiparesis (12 right, 10 left), 21 Parkinson's disease, 15 neuromuscular diseases, 14 hereditary ataxia, 11 multiple sclerosis, 10 unspecific age-related balance disorders, 7 peripheral vestibular disorders, 6 traumatic brain injury, 4 diffuse encephalopathy, 3 cervical myelopathy, and 2 central nervous system (CNS) neoplasm. All subjects were inpatients referred to the Scientific Institute of Veruno for rehabilitation assessment and treatment. Inclusion criteria were: ability to walk with or without a cane; absence of severe cognitive or communication impairments; ability to tolerate the balance tasks without fatigue. Prior to taking part in the study, all participants signed an informed consent that had been approved by the central ethics committee of the "Salvatore Maugeri" Foundation.

Instrument and procedure

The BESTest (7) contains 6 subscales, covering a broad spectrum of performance tasks: (*i*) biomechanical constraints, (*ii*) stability limits, (*iii*) transitions and anticipatory postural adjustments, (*iv*) postural responses to perturbation, (*v*) sensory orientation while standing on a compliant or inclined base of support, and (*vi*) dynamic stability in gait with and without a cognitive task (Table I). The BESTest consists of 27 items, some of which are subdivided into 2–4 sub-items (e.g.

for left and right sides) for a total of 36 tasks. Each item is scored on a 4-category ordinal scale from 0 (worst performance) to 3 (best performance). Specific patient and rating instructions, and stopwatch and ruler values are used to improve reliability (see www.bestest. us). Patients were rated by a physical therapist (MG) with 4 years of practice experience in balance assessment, who participated in a one week training course on the BESTest, at the Balance Disorders Laboratory, Oregon Health & Science University.

Statistical analysis

Unidimensionality, i.e. whether items are measuring a single underlying dimension or several separate dimensions, is one of the key requisites for test analysis and must be verified before applying Rasch models (13). To test the dimensionality of the BESTest, we performed the following statistical steps.

1. A *confirmatory factor analysis* for categorical data (CFA, LISREL 8.80 software, Scientific Software International, Inc., Lincolnwood, IL 60712, USA) was performed to evaluate the fit of the scale to a unidimensional model. The extent to which the model can be used to reproduce the sample data was determined by examining the following indexes: the non-normed fit index (NNFI, or Tucker-Lewis index), the comparative fit index (CFI), the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR). NNFI and CFI scores range from 0 to 1 with higher values indicating better fit: values greater than 0.95 are indicative of an acceptable model fit. A RMSEA value lower than 0.08 reflects an adequate fit and a RMSEA value equal to or less than 0.05–0.06 suggests a good fit. A SRMR value between < 0.10 and 0.05 is reflective of an acceptable fit (14–15).

2. In the event of a poor fit (i.e. multidimensionality is suspected) the following statistical steps were performed sequentially:

   a. *Horn's parallel analysis* (16) was used to estimate the number of meaningful dimensions in the response matrix: the size of eigenvalues obtained from principal component analysis (PCA) was compared with those obtained from a randomly generated data set of the same size and number of variables. Only factors with eigenvalues exceeding the values obtained from the corresponding random dataset were retained for further investigation. Parallel analysis was conducted using ViSta (17) Parallel Analysis plugin (http://www.mdp.edu.ar/psicologia/vista/).

   b. *Explanatory factor analysis* (EFA, STATA 10.1 software, StataCorp LP College Station TX 77845, USA) was performed with a principal factor analysis using the number of factors suggested

Table I. *Summary of BESTest items and subsystem categories. The 14 items forming the mini-BESTest for dynamic balance are in bold. Only the worst performance in items 11 "Stand on one leg" and 18 "Lateral stepping" have to be taken into account for the score. Moreover, the performance in item 27 "Cognitive Get Up and Go" must be compared with that in the baseline item 26*

I Biomechanical constraints	II Stability limits	III Anticipatory- transitions
1. Base of Support	6. a. Lateral Lean L	**9. Sit to Stand**
2. Alignment	b. Lateral Lean R	**10. Rise to Toes**
3. Ankle Strength	c. Sitting Verticality L	**11. Stand on One Leg (both right and left)**
4. Hip Strength	d. Sitting Verticality R	12. Alternate Stair Touch
5. Sit on Floor and Stand Up	7. Reach Forward	13. Standing Arm Raise
	8. a. Reach L	
	b. Reach R	

IV Postural responses	V Sensory orientation	VI Dynamic gait
14. In-place forward	**19. a. Stance EO (firm surface)**	21. Gait Natural
15. In-place backward	b. Stance EC (firm surface)	**22. Change Speed**
16. Stepping forward	c. Foam EO	**23. Head Turns**
17. Stepping backward	d. **Foam EC**	**24. Pivot Turns**
18. Lateral stepping (both right and left)	**20. Incline EC**	**25. Obstacles**
		26. Get Up and Go
		27. Cognitive Get Up and Go

L: left; R: right; EO: eyes open; EC: eyes closed.

by the parallel analysis. After varimax rotation, the relationships between the test items and retained factors were taken into account. For a solution that is stable and approximates the population pattern, given the sample size, only items with loading > 0.50 were considered as correlated to the factors (18).

    c. Item exclusion, based on the EFA results and expert review, was performed leading to a preliminary reduced set of test items.

Following the above analysis and item exclusion, the matrix of item responses of the 24 retained items for each subject underwent Rasch analysis using WINSTEPS software (Linacre JM, WINSTEPS Rasch measurement computer program, version 3.68. Chicago: Winsteps. com; 2009) (19).

As a first step, we investigated whether the rating scale of each BESTest item was used in the expected manner. We evaluated the rating scale categories (partial credit model) using criteria suggested by Linacre (20, 21): (*i*) at least 10 observations per response option; (*ii*) even distribution of category use; (*iii*) monotonic increase in both average measures of persons with a given score/category and thresholds (thresholds, or step calibrations, are the ability levels at which the response to either of 2 adjacent categories is equally likely); (*iv*) category outfit mean square (MnSq) values less than 2 (see below); and (*v*) threshold differences larger than 1.4 and lower than 5 logits. We collapsed categories following these guidelines, and compared different collapsing solutions, examining not only the category diagnostics, but also reliability indices. We were guided by the intention to select a solution that maximized statistical indices and clinical meaningfulness.

After this rating scale modification, a new Rasch analysis was performed, including PCA on the standardized residuals to evaluate: (*i*) the presence of sub-dimensions, as an independent confirmation of the unidimensionality of the scale, and (*ii*) the local independence of items.

1. "Unidimensionality" assumes that, after removal of the trait that the scale intended to measure (the "Rasch factor"), the residuals will be uncorrelated and normally distributed (i.e. there are no principal components) (19). The following criteria were used to determine whether additional factors were likely to be present in the residuals: (*i*) a cutoff of 50% of the variance explained by the Rasch factor; and (*ii*) eigenvalue of the first residual factor smaller than 3 (19).

2. "Local independence" between items indicates that they do not duplicate some feature of each other or they both incorporate some shared dimension. Item couples with a standardized residual correlation > 0.30 were considered as possibly dependent components (22). Based on examination of the respective item information functions and expert judgement, we progressively eliminated all dependencies, either removing one of the items, or, in the case of dependent items that were related to the same task performed in different directions (e.g. scores assessing right and left sides), collapsing the items into a new one reporting only the worst performance.

Internal validity of the scale was assessed by evaluating the fit of individual test items to determine if the pattern of item difficulty was consistent with the model predictions. We estimated the goodness-of-fit of the observed data to data predicted by the Rasch model (23, 24). Information-weighted (infit) and outlier-sensitive (outfit) mean-square statistics (MnSq) for each item were calculated to test whether there were items that did not fit the model expectancies. Both of these fit statistics are expected to approach 1 if the data fit the model. In accordance with the literature (10), we considered MnSq > 0.7 and < 1.3 as an indicator of acceptable fit.

We also estimated the level of difficulty of each item ("item difficulty") and the ability of each individual subject, and then we examined the data for floor and ceiling effects. Item difficulty and subject ability are expressed, on a common interval scale, in logit units, a logit being the natural logarithm of the ratio (odds) of mutually exclusive alternatives (e.g. pass vs fail, or higher vs lower response option) (23, 24). Logit-transformed measures represent linear measures. By convention, 0 logit was ascribed to the mean item difficulty. For Rasch analysis, a sample size of more than 100 persons will estimate item difficulty with an alpha of 0.05 within ± 0.5 logits (25).

Reliability was evaluated in terms of "separation" across test items, defined as the ratio of the true spread of the measures to their measurement error (23, 24). Two indexes were calculated: the item separation index and the person separation index, that give an estimate (in standard error units) of the spread or "separation" of items and persons along the measurement construct, respectively. A separation of 2.0 is considered good (24). Related indexes are the reliability of the item separation index and of the person separation index. These provide the degree of confidence that can be placed in the consistency of the estimates. This confidence ranges from 0 to 1, and coefficients > 0.80 and > 0.90 are considered respectively good and excellent (23).

RESULTS

The confirmatory factor analysis (CFA) gave, using all the items in the BESTest, an inadequate fit (NNFI = 0.91, CFI = 0.91, RMSEA = 0.12, SRMR = 0.15). Horn's Parallel Analysis (PA) revealed 3 factors with empirical eigenvalues exceeding those from the random data. These 3 factors explained 43%, 11% and 8% of the variance, respectively. To investigate the contribution of each item to the scale, we tested the 3-factor model suggested by PA using explanatory factor analysis for ordinal data (EFA) with a principal axis factor extraction method. After varimax rotation, 24 items loaded > 0.50 in the first factor, 4 items (6 a–d) in the second factor, and 3 items (7, 8a and 8b) in the third factor, while items 1–4 and 13 failed to load meaningfully in any factor.

Taking into account these results and expert opinion, 12 items (1–4, 6a–d, 7, 8a–b, 13) were deemed as not belonging to the main trait and therefore were dropped from subsequent analyses. The expert review judged the remaining 24 items to potentially measure a factor likely to represent "dynamic balance" in a variety of functional conditions. These 24 items underwent Rasch analysis.

Rating scale diagnostics showed that the 0–3 level rating categories did not comply with our pre-set criteria for category function. The model best meeting the criteria reduced the rating scale from 4 to 3 levels by combining categories 0 (absent) and 1 (mild) or 1 (mild) and 2 (moderate) (Table II), with different collapsing strategies used across items.

After combining these rating scale categories, 22 out of the 24 items fitted the underlying construct of dynamic balance that the scale was intended to measure (infit and outfit MnSq between 0.7 and 1.3). Item 5 "Sit on floor and stand up" was underfitting (i.e. with unexpectedly high variability) and item 26 "Get up and go" was overfitting (i.e. with an overly predictable pattern), so they were eliminated.

The PCA of standardized residuals showed several high (> 0.30) residual correlations between items.

Based on examination of the respective item information functions and expert judgment, all misfitting items and residual correlations > 0.30 were eliminated one by one, and the Rasch analysis was re-run. Correlated (redundant) items were removed either by deleting one of them, or by maintaining only the worst performance in items 11 and 18, which assessed the same task on both right and left side. At the end of these iterations, only 14 test items remained. This set of items (called the mini-Balance Evaluation Systems Test of dynamic balance; mini-BESTest) (see Table I) underwent further analyses.

Table II. *Mean difficulty estimates for each of the 14 items of the mini-BESTest with standard errors (SE) and infit and outfit mean-square statistics (MnSq). The more difficult the item estimate, the less likely it is for any subject to gain a high score. Alongside each item is its number in the original BESTest (see Table I). The rating scale column shows how the 4 scaling categories were collapsed into 3 categories, e.g. 0012 means that categories 0 and 1 have been collapsed and then the remaining 3 categories have been re-numbered accordingly. L: left; R: right; EO: eyes open; EC: eyes closed*

Item	Mean difficulty	SE	Infit MnSq	Outfit MnSq	Rating scale
11 a/b – Stand on L/R leg	2.43	0.25	0.90	1.07	0112
18 a/b – Postural Stepping L/R	1.10	0.22	0.84	0.76	0112
23 – Head turns	1.00	0.19	0.91	0.83	0012
17 – Postural Stepping backward	0.93	0.22	0.97	1.08	0112
27 – Cognitive "Get Up and Go" with dual task	0.77	0.24	1.07	1.08	0112
10 – Rise to toes	0.65	0.20	0.94	1.11	0012
19 d – Foam Surface EC	0.54	0.20	1.04	1.12	0112
25 – Obstacles	0.10	0.21	0.75	0.73	0112
16 – Postural Stepping forward	−0.03	0.21	1.14	1.23	0112
20 – Incline EC	−0.64	0.21	1.12	1.00	0112
24 – Pivot turns	−0.85	0.21	0.99	1.32	0112
22 – Change speed	−1.00	0.20	0.89	0.78	0112
9 – Sit to stand	−1.78	0.24	1.30	1.32	0012
19 a – Stance EO	−2.51	0.39	1.12	0.66	0012

All of the final 14 items showed good infit and outfit MnSq values (Table II). The variance explained by the estimated Rasch measures was 58.8%, whereas only 5.3% of the variance was explained by the first residual factor (eigenvalue 1.8). Regarding the hierarchic ordering of items, Figs 1 and 2 show, according to the Rasch model, the distribution of subject ability and item difficulty. Item difficulty showed a fairly even spread (from the most easy item "Stand with eyes open on a firm surface" to the most difficult item "Stand on one leg"), and subject ability presented a normal distribution spanning from −5 to +4.9 logits, with an average measure = +0.15 (mean SE 0.59). Only 2 subjects showed extreme maximum scores: the precision of their ability estimates was quite low, the SE being approximately 30% of the corresponding measure. No floor effect was found. Overall, these findings demonstrate an adequate sample-item distribution. The item difficulty estimates spanned from −4 to +2.5 logits. The reliability indices of mini-BESTest were as follows: item separation index = 7.35 and item separation reliability = 0.98; Person separation index = 2.50 and Person separation reliability = 0.86.

A final CFA confirmed the unidimensionality of the mini-BESTest, supporting the unidimensional model with the following indexes: NNFI = 0.98, CFI = 0.99, RMSEA = 0.064, and SRMR = 0.098.

The final version of the mini-BESTest is shown in Appendix I.

DISCUSSION

The original BESTest is composed of a comprehensive battery of 36 balance tasks, developed to analyse 6 different postural control systems that may contribute to poor functional balance in adults of any age (7). Thus, it is not surprising that this test failed to meet a unidimensionality assumption (i.e. that a single dimension underlies all item responses), when applied to 115 patients with a wide range of diagnoses and severity of disease.

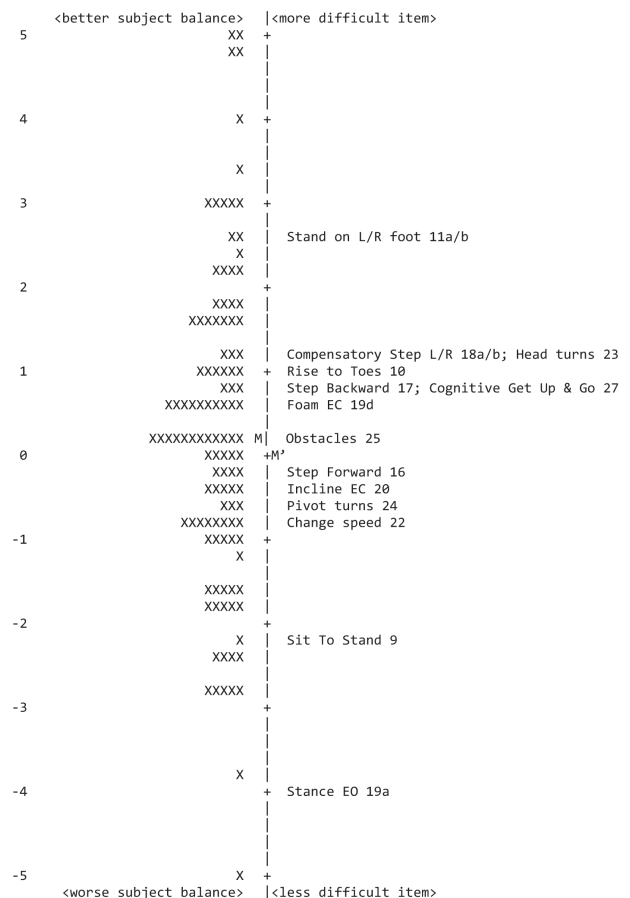Our dimensionality assessment extracted from the test battery 24 item assumed to define "dynamic balance". On these

```
          <better subject balance>  |<more difficult item>
     5                        XX    +
                              XX    |
                                    |
                                    |
     4                         X    +
                                    |
                                    |
                               X    |
     3                      XXXXX   +
                                    |
                              XX    | Stand on L/R foot 11a/b
                               X    |
                             XXXX   |
     2                              +
                             XXXX   |
                           XXXXXXX  |
                                    |
                              XXX   | Compensatory Step L/R 18a/b; Head turns 23
     1                     XXXXXX   + Rise to Toes 10
                              XXX   | Step Backward 17; Cognitive Get Up & Go 27
                         XXXXXXXXXX | Foam EC 19d
                                    |
                      XXXXXXXXXXXX M| Obstacles 25
     0                      XXXXX   +M'
                             XXXX   | Step Forward 16
                            XXXXX   | Incline EC 20
                              XXX   | Pivot turns 24
                          XXXXXXXX  | Change speed 22
    -1                     XXXXX    +
                               X    |
                                    |
                            XXXXX   |
                            XXXXX   |
    -2                              +
                               X    | Sit To Stand 9
                             XXXX   |
                                    |
                            XXXXX   |
    -3                              +
                                    |
                                    |
                                    |
                               X    |
    -4                              + Stance EO 19a
                                    |
                                    |
                                    |
                                    |
    -5                         X    +
          <worse subject balance>  |<less difficult item>
```

Fig. 1. Subject-ability and item-difficulty maps of the mini-BESTest (*n* = 115). In both maps, the vertical line represents the measure of the variable, in linear logit units. The left-hand column locates each patient's ability, from best to worst dynamic balance. The right-hand column locates each item's relative difficulty for this sample (for each item, the difficulty estimate represents the mean calibration of the threshold parameters according to the partial credit model). From bottom to top, measures indicate better balance for patients and higher difficulty for items. By convention, the average difficulty of items in the test is set at 0 logits (and indicated with M') and patients with average ability are located at M. L: left; R: right; EO: eyes open; EC: eyes closed.
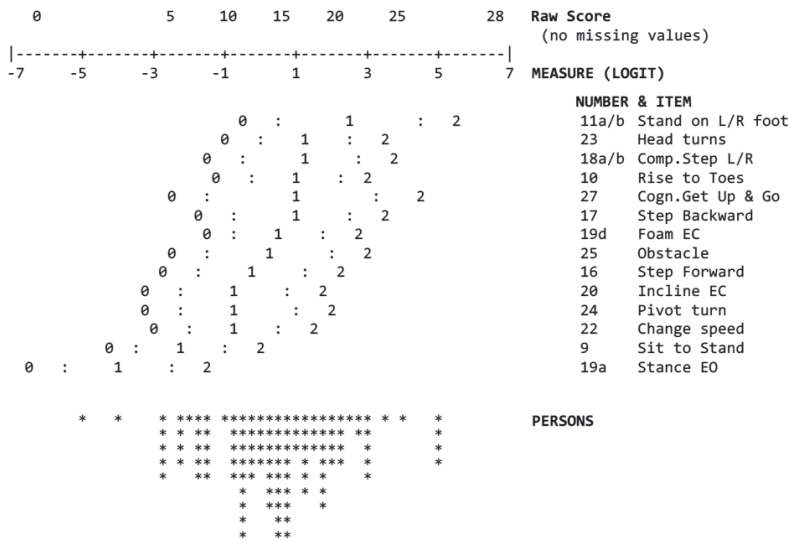
```
        0        5    10    15    20    25      28   Raw Score
                                                      (no missing values)

|-------+------+------+------+------+------+------|
 -7     -5     -3     -1     1      3      5      7   MEASURE (LOGIT)

                                                      NUMBER & ITEM
                   0    :        1        :    2      11a/b  Stand on L/R foot
                0   :      1      :    2             23     Head turns
              0   :       1     :      2            18a/b  Comp.Step L/R
                 0   :      1    :    2              10     Rise to Toes
            0   :           1         :      2      27     Cogn.Get Up & Go
              0   :        1      :    2            17     Step Backward
               0   :      1    :      2             19d    Foam EC
             0   :        1       :      2          25     Obstacle
              0   :     1    :      2               16     Step Forward
           0   :     1      :      2                20     Incline EC
           0   :     1     :       2                24     Pivot turn
            0   :     1    :     2                  22     Change speed
         0   :     1     :     2                     9     Sit to Stand
       0   :       1    :     2                      19a    Stance EO


                *    *    *  **** **************** * *    *      PERSONS
                *  **  *************** **      *       *
                *  *  **   ************* *        *
                *  *  **  ******* * ***  *        *
                *    **   *** *** * *    *
                     *   *** * *
                     *   *** *
                     *    **
                     *    **
```

Fig. 2. Expected scores for the mini-BESTest (*n* = 115). Distance between points is equal-interval. Logit measure at top of key, centred at the mean item difficulty. The rating scale is collapsed from 4 to 3 categories renumbered 0 (severely impaired), 1 (moderately impaired), and 2 (normal). The threshold between adjacent categories is marked by ":". At the bottom is the distribution of the person measures (subject ability): each marker is a single person. L: left; R: right; EO: eyes open; EC: eyes closed.

items we performed an analysis of category and item properties using Rasch psychometric methods, which led to the definition of the 14 most psychometrically useful and practical items: the refined mini-BESTest measures the unidimensional construct of "dynamic balance" without redundant items or significant ceiling/floor effects (26) and takes 10–15 min to administer.

The rating scale diagnostics (21) performed on the 24 items retained after EFA showed that the original 4 levels were redundant (23). This finding was expected, since some BESTest items were borrowed (with modifications) from the BBS and the Dynamic Gait Index. These 2 well-known balance and mobility scales have been shown to include sub-optimal category functioning (27, 28) when strict diagnostic criteria are applied (20). In addition, it has already been demonstrated that the BBS (and other balance scales) show essentially identical psychometric properties, including responsiveness, when used with a 3-category, instead of a 4- or 5-category rating scale (29). Appropriate combination of levels 0–1 or 1–2 eliminated underutilized rating categories, and ensured that each rating category was distinct from the others in representing a distinct balance ability.

After collapsing the categories to 3 distinct levels, the data from the 24-item set were reanalysed to calculate fit statistics and the PCA of the residuals. This analysis enabled us to eliminate 10 misfitting or redundant items without loss of measurement information and with the great advantage of improving test acceptability and feasibility. For the remaining 14-item (the mini-BESTest), we calculated fit statistics, extracted Rasch-modelled parameters of ability and difficulty, and then examined internal validity and test reliability. The average ability of this group of patients was very similar to the mean value of 0 logits (+0.15): this means that the test is well targeted to the sample. Moreover, the person-ability and item-difficulty mapped logit scale showed a broad range for both person-ability and item-difficulty (Fig. 1). The 1.7% of subjects (2/115) with extreme maximum scores, the 2 "×" at the top of the left-hand column in Fig. 1, constituted a minor trend toward a ceiling effect in very highly functioning subjects.

No floor effect was found. However, one should interpret the extreme results with caution, since these person measures have the least precision due to the larger errors of measurement. On the other hand, the high item separation reliability indicates that great confidence can be placed in the consistency of item difficulty estimate across future samples.

Content validity of the dynamic mini-BESTest is high, since many items included in the test are part of well-known balance batteries: (*i*) "Sit to stand" is from the Berg Balance Scale (30) and the Performance-Oriented Mobility Assessment (31); (*ii*) "Stand on one leg" is from the Ataxia Test Battery (32) and the Berg Balance Scale; (*iii*) "Stance – eyes open" and "Stance on foam – eyes closed" are from the modified Clinical Test of Sensory Integration of Balance (33, 34); (*iv*) Gait when balance is challenged by changing speed, head rotations, pivot turns, or stepping over obstacles comes from the Dynamic Gait Index (35); (*v*) the "Get Up and Go" test (36) and the "Get Up and Go with a simultaneous cognitive task" (37) are stand-alone tests. In the BESTest, Horak et al. (7) made only minor modifications to some of the above original items, in order to increase their challenge and improve their consistency and reliability. Novel items in the mini-BESTest have been adapted from laboratory tests where they were shown to distinguish different types of balance disorders: (*i*) postural reactions to external perturbations (38); (*ii*) rise to toes (39); and (*iii*) stance on an inclined surface with eyes closed (40).

As an additional demonstration of the internal construct validity of the scale, the general hierarchical arrangement found by Rasch analysis (Table II) is consistent with clinical expectations. For example, the maintenance of feet-together stance, eyes open on a firm surface ("Stance EO") is the easiest task and "Stand on one leg" the most difficult task item (28). In fact, "Stance EO" makes few sensory demands and requires low effort, whereas "Stand on one leg" is very challenging because of the narrow base of support and musculoskeletal demands. In addition, the results of Rasch analysis of the mini-BESTest show a hierarchical order of item difficulty: "Gait with hori-

zontal head turns", "Stand on one leg", and "Lateral stepping responses" were the most difficult items, whereas "Stance EO" and "Sit to Stand" were the easiest items. The high difficulty of the item "Gait with horizontal head turns" may be attributed to vestibular influences (35) and is in line with the results of the two Rasch studies on the Dynamic Gait Index (28, 41).

The mini-BESTest contains 14 items belonging evenly to 4 of the 6 sections from the original BESTest (Table I): section III "Anticipatory Postural Adjustments" (sit to stand, rise to toes, stand on 1 leg); section IV "Postural Responses" (stepping in 4 different directions); section V "Sensory Orientation" (stance – eyes open; foam surface – eyes closed; incline – eyes closed); and section VI "Balance during Gait" (gait during change speed, head turns, pivot turns, obstacles; cognitive "Get Up and Go" with dual task).

Our factor analysis procedure (42) isolated a number of items, primarily in the first 2 sections of the BESTest, that did not contribute to the dominant trait (dynamic balance), suggesting that parts I "Biomechanical constraints" and II "Stability limits" of the BESTest warrant separate psychometric studies. Biomechanical constraints (such as orthopaedic limitations on the base of foot support, postural alignment and strength) and stability limits (ability to lean to perceived limits of stability and perception of verticality) are also important facets of postural control, but appear to be independent of the construct "dynamic balance".

This study has several limitations, which restrict the generalization of our results to different groups or settings, and raters. In particular, the selection criteria of our convenience sample (recruited with a consecutive sampling method) may represent a threat to external validity. Our sample was a cross-section of adults drawn from a single rehabilitation facility and with balance disorders of very different origins and severities. Moreover, we used only one rater, but to improve the reliability of results he participated in a one week training course on BESTest, held by one of its developers (FBH).

In conclusion, the new mini-BESTest offers a unique, brief clinical rating scale for dynamic balance that has excellent psychometric characteristics. The potential interest of the mini-BESTest in clinical settings is high, but further studies are needed. They should include: (*i*) analysis of the actual performance of the new 3-level response structure; and (*ii*) a study of differential item functioning, i.e. the stability of item hierarchy across subsamples defined according to potentially relevant clinical criteria; (*iii*) relation of the scores to fall risk and to other clinical tests of balance; and (*iv*) age-related normative values.

ACKNOWLEDGEMENT

REFERENCES

1. Thurman DJ, Stevens JA, Rao JK. Practice parameter: assessing patients in a neurology practice for risk of falls (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2008; 70: 473–479.
2. Gillespie LD, Robertson MC, Gillespie WJ, Lamb SE, Gates S, Cumming RG, et al. Interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2009 Apr 15; 2: CD007146.
3. Horak FB. Clinical assessment of balance disorders. Gait Posture 1997; 6: 76–84.
4. Horak FB, Macpherson JM. Postural orientation and equilibrium. In: Shepard J, Rowell L, editors. Regulation and integration of multiple systems. Handbook of physiology: section 12, exercise. New York: Oxford University Press; 1996, p. 255–292.
5. Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? Age Ageing 2006; 35: ii7–ii11.
6. Pérennou D, Decavel P, Manckoundia P, Penven Y, Mourey F, Launay F, et al. Evaluation of balance in neurologic and geriatric disorders. Ann Readapt Med Phys 2005; 48: 317–335.
7. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther 2009; 89: 484–498.
8. Franchignoni F, Tesio L, Martino MT, Ricupero C. Reliability of four simple, quantitative tests of balance and mobility in healthy elderly females. Aging (Milano) 1998; 10: 26–31.
9. Woollacott M, Shumway-Cook A. Abnormal postural control. In: Motor control - theory and practical applications. Baltimore: Lippincott Williams & Wilkins; 2001, p. 248–270.
10. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation. J Rehabil Med 2003; 35: 105–115.
11. Reeve BB, Fayers P. Applying item response theory modeling for evaluating questionnaire item and scale properties. In: Fayers P, Hays RD, editors. Assessing quality of life in clinical trials: methods of practice. 2nd edn. Oxford, NY: Oxford University Press; 2005, p 55–73.
12. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil 2000; 81 Suppl 2: S15–S20.
13. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. PROMIS Cooperative Group. 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.
14. Schermelleh-Engel K, Moosbrugger H, H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods Psychol Res Online 2003; 8: 23–72.
15. Cook KF, Teal CR, Bjorner JB, Cella D, Chang CH, Crane PK, et al. IRT health outcomes data analysis project: an overview and summary. Qual Life Res 2007; 16 Suppl 1: 121–132.
16. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965; 30: 179–185.
17. Young FW. ViSta: the Visual Statistics System, UNC L.L. Thurstone Psychometric Laboratory Research Memorandum 94–1(c), 1996.
18. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull 1988; 103: 265–275.
19. Linacre JM. A user's guide to WINSTEPS-MINISTEP: Rasch-model computer programs. Program manual 3.68.0. Chicago, IL: WINSTEPS.com; 2009 [cited 2009 September 24]. Available from: URL: http://www.winsteps.com/a/winsteps.pdf.
20. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002; 3: 85–106.
21. Wolfe EW, Smith EV Jr. Instrument development tools and activities for measure validation using Rasch models: part II – validation activities. J Appl Meas 2007; 8: 204–234.
22. Davidson M. Rasch analysis of 24-, 18- and 11-item versions of the Roland-Morris Disability Questionnaire. Qual Life Res 2009; 18: 473–481.

23. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 2nd edn. Mahwah: Lawrence Erlbaum Associates; 2001.

24. Wright BD, Masters GN. Rating scale analysis. Chicago: Mesa Press; 1982.

25. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994; 7: 328.

26. Hobart JC, Lamping DL, Freeman JA, Langdon DW, McLellan DL, Greenwood RJ, et al. Evidence-based measurement: which disability scale for neurologic rehabilitation? Neurology 2001; 57: 639–644.

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. Chiu YP, Fritz SL, Light KE, Velozo CA. Use of item response analysis to investigate measurement properties and clinical validity of data for the dynamic gait index. Phys Ther 2006; 86: 778–787.

29. 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: 430–438.

30. 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.

31. Tinetti ME, Richman D, Powell L. Falls efficacy as a measure of fear of falling. J Gerontol 1990; 45: P239–P243.

32. Graybiel A, Fregly AR. A new quantitative ataxia test battery. Acta Otolaryngol 1966; 6: 292–312.

33. Shumway-Cook A, Horak FB. Assessing the influence of sensory interaction of balance. Suggestion from the field. Phys Ther 1986; 66: 1548–1550.

34. Cohen H, Blatchly CA, Gombash LL. A study of the clinical test of sensory interaction and balance. Phys Ther 1993; 73: 346–351.

35. Whitney SL, Hudak MT, Marchetti GF. The dynamic gait index relates to self-reported fall history in individuals with vestibular dysfunction. J Vestib Res 2000; 10: 99–105.

36. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991; 39: 142–148.

37. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. Phys Ther 2000; 80: 896–903.

38. Henry SM, Fung J, Horak FB. EMG responses to maintain stance during multidirectional surface translations. J Neurophysiol 1998; 80: 1939–1950.

39. Nardone A, Schieppati M. Postural adjustments associated with voluntary contraction of leg muscles in standing man. Exp Brain Res 1988; 69: 469–480.

40. Kluzik J, Horak FB, Peterka RJ. Differences in preferred reference frames for postural orientation shown by after-effects of stance on an inclined surface. Exp Brain Res 2005; 162: 474–489.

41. Marchetti GF, Whitney SL. Construction and validation of the 4-item dynamic gait index. Phys Ther 2006; 86: 1651–1660.

42. Coste J, Bouée S, Ecosse E, Leplège A, Pouchot J. Methodological issues in determining the dimensionality of composite health measures using principal component analysis: case illustration and suggestions for practice. Qual Life Res 2005; 14: 641–654.

APPENDIX I. Mini-BESTest of Dynamic Balance – Balance Evaluation System's Test© 2009

Subjects should be tested with flat-heeled shoes OR shoes and socks off. If subject must use an assistive device for an item, score that item one category lower. If subject requires physical assistance to perform an item, score the lowest category (0) for that item.

1. SIT TO STAND
(2) Normal: Stable for 3 s with maximum height .
(1) Moderate: Comes to stand WITH use of hands on first attempt.
(0) Severe: Impossible to stand up from chair without assistance, OR several attempts with use of hands.

2. RISE TO TOES
(2) Normal: Stable for 3 s with maximum height.
(1) Moderate: Heels up, but not full range (smaller than when holding hands), OR noticeable instability for 3 s.
(0) Severe: ≤3 s.

3. STAND ON ONE LEG
Left Time in Seconds Trial 1:_____ Trial 2:_____   *Right* Time in Seconds. Trial 1: _____ Trial 2:_____
(2) Normal: 20 s.   (2) Normal: 20 s.
(1) Moderate: <20 s.   (1) Moderate: <20 s.
(0) Severe: Unable.   (0) Severe: Unable.

4. COMPENSATORY STEPPING CORRECTION – FORWARD
(2) Normal: Recovers independently a single, large step (second realignment step is allowed)
(1) Moderate: More than one step used to recover equilibrium.
(0) Severe: No step, OR would fall if not caught, OR falls spontaneously.

5. COMPENSATORY STEPPING CORRECTION – BACKWARD
(2) Normal: Recovers independently a single, large step.
(1) Moderate: More than one step used to recover equilibrium.
(0) Severe: No step, OR would fall if not caught, OR falls spontaneously.

6. COMPENSATORY STEPPING CORRECTION – LATERAL
Left   *Right*
(2) Normal: Recovers independently with 1 step (crossover or lateral OK).   (2) Normal: Recovers independently with 1 step (crossover or lateral OK).
(1) Moderate: Several steps to recovers equilibrium.   (1) Moderate: Several steps to recovers equilibrium.
(0) Severe: Falls, or cannot step.   (0) Severe: Falls, or cannot step.

7. STANCE EYES OPEN, FIRM SURFACE (FEET TOGETHER)
Time in seconds:_____
(2) Normal: 30 s.
(1) Moderate: <30 s.
(0) Severe: Unable.

8. STANCE EYES CLOSED, FOAM SURFACE (FEET TOGETHER)
Time in seconds:_____
(3) Normal: 30 s.
(1) Moderate: <30 s.
(0) Severe: Unable.

9. INCLINE – EYES CLOSED
Time in seconds:_____
(2) Normal: Stands independently 30 s and aligns with gravity.
(1) Moderate: Stands independently <30 s, OR aligns with surface.
(0) Severe: Unable.

10. CHANGE IN GAIT SPEED
(2) Normal: Significantly changes walking speed without imbalance.
(1) Moderate: Unable to change walking speed or imbalance.
(0) Severe: Unable to achieve significant change in speed AND signs of imbalance.

11. WALK WITH HEAD TURNS – HORIZONTAL
(2) Normal: performs head turns with no change in gait speed and good balance.
(1) Moderate: performs head turns with reduction in gait speed.
(0) Severe: performs head turns with imbalance.

12. WALK WITH PIVOT TURNS
(2) Normal: Turns with feet close, FAST (≤3 steps) with good balance.
(1) Moderate: Turns with feet close SLOW (≥4 steps) with good balance.
(0) Severe: Cannot turn with feet close at any speed without imbalance.

13. STEP OVER OBSTACLES
(2) Normal: Able to step over box with minimal change of speed and with good balance.
(1) Moderate: Steps over shoe boxes but touches box, OR displays cautious behaviour by slowing gait.
(0) Severe: Cannot step over shoe boxes, OR hesitates, OR steps around box.

14. TIMED GET UP & GO (TUG) AND COGNITIVE GET UP & GO WITH DUAL TASK:
TUG:_____seconds; Dual Task TUG: _____seconds
(2) Normal: No noticeable change between sitting and standing in backward counting and no change in gait speed compared with TUG with dual task.
(1) Moderate: Dual task affects either counting OR walking.
(0) Severe: Stops counting while walking OR stops walking while counting.

INSTRUCTIONS
1. SIT TO STAND
Examiner Instructions: Note the initiation of the movement, and the use of hands on the arms of the chair or their thighs or thrusts arms forward.

Patient: Cross arms across your chest. Try not to use your hands unless you must. Do not let your legs lean against the back of the chair when you stand. Please stand up now.

2. RISE TO TOES
Examiner Instructions: Allow the patient to try it twice. Record the best score. (If you suspect that subject is using less than their full height, ask them to rise up while holding the examiners' hands.) Make sure subjects look at a non–moving target 4–12 feet away.

Patient: Place your feet shoulder width apart. Place your hands on your hips. Try to rise as high as you can onto your toes. I will count out loud to 3 seconds. Try to hold this pose for at least 3 seconds. Look straight ahead. Rise now.

3. STAND ON ONE LEG
Examiner Instructions: Allow the patient two attempts and record the best. Record the number of seconds they can hold posture up to a maximum of 30 seconds. Stop timing when subject moves their hand off hips or puts a foot down. Make sure subjects look at a non–moving target 4–12 feet ahead. Repeat other side.

Patient: Look straight ahead. Keep your hands on your hips. Bend one leg behind you. Do not touch your raised leg on your other leg. Stay standing on one leg as long as you can. Look straight ahead. Lift now.

4. COMPENSATORY STEPPING CORRECTION – FORWARD
Examiner Instructions: Stand in front to the side of patient with one hand on each shoulder and ask them to push forward. (Make sure there is room for them to step forward.) Require them to lean until their shoulders and hips are in front of their toes. The test must elicit a step. NOTE: Be prepared to catch patient.

Patient: Stand with your feet shoulder width apart, arms at your sides. Lean forward against my hands beyond your forward limits. When I let go, do whatever is necessary, including taking a step, to avoid a fall.

5. COMPENSATORY STEPPING CORRECTION – BACKWARD
Examiner Instructions: Stand in back to the side of the patient with one hand on each scapula and ask them to lean backward. (Make sure there is room for them to step backward.) Require them to lean until their shoulders and hips are in back of their heels. After you feel their body weight in your hands, very suddenly release your support. Test must elicit a step. NOTE: Be prepared to catch patient

Patient: Stand with your feet shoulder width apart, arms down at your sides. Lean backward against my hands beyond your backward limits. When I let go, do whatever is necessary, including taking a step, to avoid a fall.

6. COMPENSATORY STEPPING CORRECTION – LATERAL

Examiner Instructions: Stand behind the patient, place one hand on either the right (or left) side of the pelvis, and ask them to lean their whole body into your hand. Require them to lean until the midline of pelvis is over the right (or left) foot and then suddenly release your hold. NOTE: Be prepared to catch patient.

Patient: Stand with your feet together, arms down at your sides Lean into my hand beyond your sideways limit. When I let go, step if you need to, to avoid a fall.

7. STANCE EYES OPEN – FIRM SURFACE

Examiner Instructions: Record the time the patient was able to stand with feet together up to a maximum of 30 seconds. Make sure subjects look at non-moving target 4–12 feet away.

Patient: Place your hands on your hips. Place your feet together until almost touching. Look straight ahead. Stay as stable as possible until I say stop

8. STANCE EYES CLOSED – FOAM SURFACE

Examiner Instructions: Use medium density Temper® foam, 4 inches thick. Assist subject in stepping onto foam. Tell patient to "Close Eyes". Record the time the patient was able to stand in each condition to a maximum of 30 seconds. Have the subject step of the foam between trials.

Patient: Place your hands on your hips. Place your feet together until almost touching. Stay as stable as possible until I say stop. I will start timing when you close your eyes.

9. INCLINE – EYES CLOSED

Examiner Instructions: Aid the patient onto the ramp. Once the patient closes their eyes, begin timing and record and average both times. Note if sway is greater than when standing on firm, level surface with eyes closed (Item 7) or if there is poor alignment to vertical.

Patient: Please stand on the incline ramp with your toes toward the top. Place your feet shoulder width apart and your arms on your hips. I will start timing when you close your eyes.

10. CHANGE IN SPEED

Examiner Instructions: Allow the patient to take 3–5 steps at their normal speed, and then say "fast", after 3–5 fast steps once say "slow". Allow 3–5 slow steps before they stop walking.

Patient: Begin walking at your normal speed, when I tell you "fast" walk as fast as you can. When I say "slow", walk very slowly.

11. WALK WITH HEAD TURNS – HORIZONTAL

Examiner Instructions: Allow the patient to reach their normal speed, and give the commands "right, left" every 3–5 steps. Score if you see a problem in either direction. If patient has severe cervical restrictions allow combined head and trunk movements.

Patient: Begin walking at your normal speed, when I say "right", turn your head and look to the right. When I say "left" turn your head and look to the left. Try to keep yourself walking in a straight line.

12. WALK WITH PIVOT TURNS

Examiner Instructions: Demonstrate a pivot turn. Once the patient is walking at normal speed, say "turn and stop". Count the steps from turn" until the subject is stable. Imbalance may be indicated by wide stance width, extra stepping or trunk motion.

Patient: Begin walking at your normal speed. When I tell you to "turn and stop", turn as quickly as you can to face the opposite direction and stop. After the turn, your feet should be close together.

13. STEP OVER OBSTACLES

Examiner Instructions: Place the box (9" / 22.9 cm height) 10 feet away from where the patient will begin walking. Use a stopwatch to time gait duration to calculate average velocity by dividing the number of seconds into 20 feet.

Patient: Begin walking at your normal speed. When you come to the shoe boxes (9" / 22.9 cm height), step over them, not around them and keep walking.

14. TIMED GET UP & GO (TUG) and COGNITIVE GET UP & GO with DUAL TASK

Examiner Instructions: Use the TUG score to determine the effects of dual tasking.
TUG: Have the patient sit with their back against the chair. Time the patient from the time you say "Go" until they return to sitting in chair. Stop timing when the patient's buttocks hit the chair bottom. The chair should be firm with arms to push from if necessary.
TUG with Dual Task: While sitting, determine how fast and accurately the patient can count backwards by 3 seconds from a number between 90–100. Then, ask them to count from a different number and after a few numbers say "Go". Time the patient from the time you say "Go" until they return to the sitting position. Score dual task as affecting walking if speed slows >10% from TUG &/or new signs of imbalance.

Patient:
TUG: When I say "Go", stand up from chair, walk at your normal speed across the tape on the floor; turn around, and come back to sit in the chair. Continue counting backwards the entire time.
 TUG with Dual Task: Count backwards by 3 seconds starting at _____.
When I say "Go", stand up from chair, walk at your normal speed across the tape on the floor, turn around, and come back to sit in the chair. Continue counting backwards the entire time.