The Balance Intensity Scales for Therapists and Exercisers Measure Balance Exercise Intensity in Older Adults: Initial Validation Using Rasch Analysis

Melanie K. Farlie, Jennifer L. Keating, Elizabeth Molloy, Kelly-Ann Bowles, Becky Neave, Jessica Yamin, Jussyan Weightman, Kelly Saber, Terry P. Haines

Background. The Balance Intensity Scales (BIS) have been developed to measure the intensity of balance exercise in older adults.

Objective. The objective was to determine whether the BIS for therapists (BIS-T) and for exercisers (BIS-E) are unidimensional measures of balance exercise intensity, able to be refined using the Rasch model into a hierarchical item order, and appropriately targeted for the older adult population with a variety of diagnoses in a range of exercise testing settings.

Design. This was a scale development study using a pragmatic mixed-methods approach.

Methods. Older adult exercisers (n = 108) and their therapists (n = 33) were recruited from a large metropolitan health service and rated balance exercise tasks on the BIS-T and BIS-E in a single session.

Results. Scores on both the BIS items and global effort ratings for therapists and exercisers had good correlation and demonstrated unidimensionality. The BIS-T and BIS-E demonstrated a hierarchical distribution of items that fit the Rasch model. The Person Separation Index was moderate (0.62) for the BIS-T but poor (0.33) for the BIS-E.

Limitations. The limitations were that therapists in this study underprescribed high-intensity balance tasks.

Conclusions. Initial validation of the BIS-T and the BIS-E demonstrated that these scales can be used for the measurement of balance exercise intensity in older adult populations. The BIS-T items and global effort ratings are recommended for use by therapists, and the global effort ratings are recommended for use by exercisers. Ongoing validation of both scales using high-intensity balance task ratings and different populations of older adults is recommended.
The optimal dosage of balance exercise intensity is yet to be described despite more than 4 decades of research into falls, mainly on interventions that may modify or ameliorate risk factors for falls and related injuries. A number of balance exercise variables are routinely reported, including frequency, type, time, and duration, but explain little of the variance in the results of balance exercises.

Guidelines are near universal in stating that balance exercise programs need to be challenging to be effective. In the absence of a validated balance exercise intensity measure, health professionals face a number of dilemmas when trying to interpret and apply these guidelines. How can the degree of challenge experienced by an exerciser be measured? Must a balance exercise challenge to the point of near-falling be effective or is a program of low or moderate intensity balance exercise as effective? Given the tension that exists between setting balance exercise training at levels that are challenging and concurrently protecting exercisers from the risk of falls, a measure of balance exercise intensity may assist therapists and exercisers to better delineate the markers of balance exercise training intensity that then informs their discussions about how to exercise safely. In addition, a measure of balance exercise intensity could also be used in the conduct of dose response studies of balance exercise to determine the most effective balance exercise intensity to retrain balance and prevent falls.

Balance exercise intensity is a unidimensional, linear construct that may rate higher or lower, analogous to the intensity of pain measured on a visual analog scale. The aim of this study was to validate 2 measures of balance exercise intensity, the Balance Intensity Scales (BIS) for therapists (BIS-T) and exercisers (BIS-E) using Rasch analysis. The specific research questions were as follows: Are the BIS-T and BIS-E unidimensional measures of balance exercise intensity? Can items on the BIS-T and BIS-E be refined using the Rasch model into a hierarchical item order? Are items on the BIS-T and BIS-E appropriately targeted—that is, matched to a range of task difficulty—in an older adult population with a variety of diagnoses in a range of exercise testing settings?

**Methods**

The process used to develop the preliminary BIS tested in this study is detailed in the Supplementary Figure (available at https://academic.oup.com/ptj). The BIS were developed using an observational study to determine all the verbal and nonverbal markers of balance exercise intensity that were then refined through key stakeholder consultation (therapists, exercisers, and researchers) to determine the scale items and global effort ratings that were then tested in this study. Key drivers that underpinned the development of these scales were to create 2 clinically relevant measures that used language and terms that were meaningful and relevant to end users and were able to be applied in a clinical setting with minimal training, equipment, or resources. Development of the scales in parallel was done to create scales that exercisers and therapists could use to rate the same task, where if an exerciser was unable to self-rate (ie, in the presence of cognitive impairment), therapists could still rate a balance task performance by an exerciser. The BIS use an objective item checklist paired with a subjective global effort rating to rate how challenging balance tasks are from the perspective of therapists and exercisers. This scale testing and validation study examined the 18-item BIS-T and the 13-item BIS-E that were the result of the initial scale development phases.

**Balance Intensity Scales**

Balance intensity is defined as “the degree of challenge to the balance control system relative to the capacity of the individual to maintain balance.” The BIS are designed for routine tests of balance to guide balance exercise prescription. Items on both scales are indicators of verbal and nonverbal markers of balance exercise intensity and are scored dichotomously (no = not observed = 0; yes = observed = 1). A higher overall score for a given balance task indicates a higher intensity balance exercise. The maximum possible raw scores are 18 for the BIS-T and 13 for the BIS-E. A 5-point subjective global rating score (GRS) of balance effort (1 [no effort at all]–5 [maximal effort]) provides a separate perspective on exercise intensity to assist validating BIS raw scores. The scales are shown in the Supplementary Appendix (available at https://academic.oup.com/ptj).

**Participants and Setting**

This study was conducted in an Australian metropolitan health care network with diverse treatment settings including acute, inpatient rehabilitation, and community services and a variety of diagnostic streams (ie, orthopedic, neurological, cardiorespiratory). Therapists with at least 1 year of experience in balance exercise prescription for adults more than 50 years old were recruited to the study. Therapists identified older adult exercisers who were currently performing balance exercises in a therapy program and were willing to participate in a testing session using the BIS-T and BIS-E. Balance exercise rating sessions were incorporated into usual care time. Approval for this study was granted by the Monash Health and Monash University ethics committees.

**Scale Testing Procedure**

Written informed consent was provided by participants. The lead researcher (M.K.F.) educated all participants in the use of the BIS-T and BIS-E with a brief training session that oriented users to the structure and scale scoring system using minimal resources reflecting realistic resources available in the clinical environment. For therapists, this included a brief orientation to the BIS-T and BIS-E and viewing and discussing video-taped Exercise Taped (BIST and BISE)
BIS for Therapists and Exercisers

performances of an older adult performing various balance challenges to explain how to use the scales. Therapists were instructed that if exercisers required assistance to complete the BIS-E rating form, the therapists were to record the exerciser’s perception/response without leading, coaching, or challenging the exerciser on their response, even if they did not agree with the exerciser self-rating. For exercisers, the education session was a conversation to orient exercisers in how to read and complete the BIS-E form. It included comparisons with other self-report exercise ratings that participants may have had experience with prior to this study (eg, use of the Borg Scale to rate perceived exertion during cardiorespiratory exercise) or self-rating scales used in other domains of therapy, such as the visual analog scale for rating pain. Exercisers were encouraged to ask questions and were asked to visualize how they might respond to each item on the scale, based on recall of the last time they completed a balance exercise, to check understanding of each item. They were also given a copy of the BIS-E at the time of recruitment and encouraged to familiarize themselves further with the items on the form between the time of recruitment and testing. Exercisers were advised that their therapist would be able to assist them with completing the BIS-E at the time of testing if requested but that therapists had been instructed not to “help” exercisers choose their responses to any item or the global rating scale. Key instructions on how to use the scales were printed on the back of each rating form (Suppl. Appendix, available at https://academic.oup.com/ptj).

Balance Task Selection and Administration

The treating therapist was asked to test the exerciser on 3 standing balance tasks taken from the de Morton Mobility Index (DEMMI), an outcome measure routinely used across the health service, and following the procedures reported previously in studies related to the development of these scales. Three key features of the DEMMI supported the decision to use this tool. First, the widespread clinical use and familiarity of this tool to the participating therapists due to its validation in all the populations treated across the health service in this study. Second, the psychometric properties of the DEMMI are well established and confirm a hierarchical order of task difficulty. Lastly, the scale does not show floor (too hard to start) or ceiling (too easy to complete) effects, indicating that the tasks listed in the DEMMI were likely to be adequate to challenge most exercisers wherever they were across the health service.

The low challenge tasks for all exercisers was “stand feet apart for 10 seconds, eyes open.” Therapists determined the medium and high challenge tasks for each exerciser depending on the exerciser performance on the DEMMI on the day of testing. The 7 standing balance task options in order of difficulty were: stand feet apart, stand feet together, pick up a pen, walk 4 steps backward, jump, stand on toes, and tandem stand with eyes closed. Therapists could seek guidance from the lead researcher (M.K.F.) on alternative tasks if the hardest item on the DEMMI “tandem stand eyes closed, 10 seconds” task was not adequately challenging. The treating therapist administered the 3 balance tests in order of increasing difficulty in a single test session, yielding 3 independent activity ratings from each exerciser participating in the study. After each balance test, the therapist rated the exerciser’s performance on the BIS-T, and the exerciser rated their own performance on the BIS-E. If the exerciser requested assistance completing the BIS-E form, the therapist assisted the exerciser with the BIS-E ratings first, prior to completing their own rating. Approximate time to complete ratings was reported to the lead researcher on collection of the rating forms after each testing session.

Data Entry and Analysis

The lead researcher (M.K.F.) entered data into a spreadsheet. Two research assistants independently audited data, with discrepancies resolved through discussion. Descriptive statistics summarized participant demographics. Analyses were conducted independently for each scale. Evaluation and refinement of the BIS-T and BIS-E items was achieved through exploratory factor analysis and Rasch analysis (Tab. 1). Exploratory factor analysis using principle components analysis examined dimensionality of each scale to identify subgroups of items, or multi-dimensionality, which could be used to guide subgroup considerations in subsequent Rasch analysis. Suitability for factor analysis, and decisions to retain factors, were based on recommendations by Pallant. These included the Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett test of sphericity to determine suitability for factor analysis. Retention of factors was predicated on eigenvalues being greater than 1.0, scree plot inspection, and parallel analysis using computer-generated eigenvalues. Analysis resulting in the retention of a single factor would indicate scale unidimensionality.

Data for both scales were analyzed separately using the Rasch unidimensional measurement model for dichotomous data with the RUMM2030 program. Each activity performed contributed to the data set as an independent observation, with each exerciser contributing 3 independent activities. If scales “fit” the Rasch model, this would indicate a hierarchical distribution of item difficulty on the scale being examined. Model fit as explained by the Rasch model for the scales in this study would be an indication that persons experiencing low levels of balance task intensity would only be expected to display the lower-level items, whereas those experiencing high levels of balance task intensity would be expected to display the higher level items as well as lower level items. Rasch analysis included examination of overall item and person fit, individual item and person fit, differential item functioning for person factors, assessments of local
dependency between items, and confirmation of the unidimensionality of the scale. Scale refinement, through removal of items, was planned based on the outcome of the Rasch analysis and alignment with the aim to develop a clinically useful scale. Following refinement of each scale, Person Separation Index (PSI) was reviewed to determine reliability, and a visual examination of person-item threshold distribution charts was used to assess targeting. If adequate fit to the Rasch model was achieved, conversion of the ordinal raw score to an interval scale score out of 100 was also planned. Correlations between the BIS-T and BIS-E total scores and global effort ratings for therapists and exercisers were examined using Spearman rho.

**Sample Size Considerations**

Various recommendations have been made regarding sample size for factor analysis. On the basis of Tinsley and Tinsley's recommendation of 5 to 10 observations per item, the minimum numbers of observations for factor analysis would be 90 to 180 for the BIS-T and 65 to 130 for the BIS-E. For Rasch analysis to achieve item calibration stability within ±0.5 logit, on the basis of a 99% confidence interval (CI), Linacre considers that 108 observations are required for well-targeted items (40%–60% success on items) and that as many as 243 observations are required for poorly targeted items (<15% or >85% success on items). The planned collection of ratings for 330 activities (3 balance tasks performed by each of the 110 exercisers) was therefore more than adequate for factor and Rasch analyses of both scales.

**Role of the Funding Source**

M.K. Farlie was supported by the Monash University Maxwell King PhD Scholarship. T.P. Haines was supported by a National Health and Medical Research Council Career Development Fellowship. The funders played no role in the design, conduct, or reporting of this study.
Results
Fifty-nine therapists (53 physical therapists and 6 exercise physiologists) were recruited across the study period, with 1 withdrawal. Thirty-three of the 59 therapists went on to identify the 110 exercisers who were recruited to the study. One exerciser withdrew because of medical illness, and 1 set of data was lost because of an administrative error. A total of 324 balance tasks were rated by the 108 exercisers and 33 therapists over a 20-week period. In 5 instances, selection of balance tasks did not adhere to the protocol and were removed, leaving 319 tasks available for analysis. Item 18 on the BIS-T (“Did the person fall during the task?”) was not rated “yes” for any task. Two exercisers rated item 13 on the BIS-E (“Did you fall during this task?”) “yes” on 1 task each; however, follow-up with therapists indicated that the more appropriate items to describe the task were item 9 (“Did you need to grab or hold on to anything to keep your balance?”) and item 12 (“Did someone else need to stop you losing your balance during this task?”). Despite confirmation of no fall during testing, exercisers’ responses to item 13 were unchanged in the dataset.

Exercise rating sessions occurred in acute (n = 3) and subacute (n = 48) inpatient units, community-based services (n = 54), and outpatient clinics (n = 3). Exerciser treatment streams included orthopedic (n = 40), neurologic (n = 31), cardiorespiratory (n = 10), and other (eg, ulcers, falls) (n = 27). Therapist, exerciser, and BIS rating characteristics are presented in Tables 2 and 3. Exercisers independently completed one-third of BIS-E forms. Therapists completed all other BIS-E forms under instruction from exercisers. There were no reports of difficulty in understanding or using any items to rate any activities. Testing sessions took 5 to 10 minutes to complete all activities and ratings. The number of items that were marked as observed increased on the BIS-T and BIS-E as task intensity increased from low to high. “Yes” (indicating observation of a marker of balance exercise intensity) was selected for at least 1 item on the BIS-T for 248 activities (78%) and at least 1 item on the BIS-E for 189 activities (59%).

Exploratory Factor Analysis
BIS-T analysis excluded item 18 (“Did you fall?”), because this item was never observed. The 17 items of the BIS-T and the 13 items of the BIS-E were analyzed with principal components analysis. Suitability for factor analysis was confirmed by acceptable Kaiser-Meyer-Olkin and Bartlett test values. Eigen values, scree plot, and parallel analysis supported a single-factor solution, indicating unidimensionality of both scales.

Rasch Analysis
BIS-T assessment of unidimensionality and item hierarchy. Initial analysis of the 17 items of the BIS-T showed poor fit to the Rasch model. Four items recorded fit residual values over the recommended level of ±2.5

Table 2.
Characteristics of 58 Therapists Recruited to the Study

| Self-Rated Proficiency in Balance Exercise Prescription | No. (%) of Clinicians in the Following Group |
|--------------------------------------------------------|--------------------------------------------|
|                                                        | Junior (n = 19)                             |
|                                                        | Intermediate (n = 31)                       |
|                                                        | Senior (n = 8)                             |
| Novice                                                 | 3 (15.8)                                   |
|                                                       | 0 (0)                                      |
|                                                       | 0 (0)                                      |
| Proficient                                             | 16 (84.2)                                  |
|                                                       | 21 (67.7)                                  |
|                                                       | 3 (37.5)                                   |
| Advanced                                               | 0 (0)                                      |
|                                                       | 10 (32.3)                                  |
|                                                       | 5 (62.5)                                   |

*The mean age of the therapists was 30.9 years (SD = 7.9 years). Of the 58 therapists, 47 (81.0%) were women. Junior = <5 y of experience; intermediate = 5-9 y of experience; senior = >9 y of experience.

Table 3.
Characteristics of 110 Exercisers Recruited to Study

| Characteristic                     | Value            |
|-----------------------------------|-----------------|
| Age, y, mean (SD)                 | 70.8 (9.9)      |
| No. (%) of women                  | 61 (55.4)       |
| Highest education level           |                 |
| Primary school                    | 3 (2.7)         |
| Secondary school                  | 85 (77.2)       |
| Undergraduate                     | 14 (12.7)       |
| Postgraduate                      | 8 (7.2)         |
| History of falls                  | 95 (86.3)       |
| 12-mo fall rate, median (IQR)     | 1 (0.3); range = 0–21 |
| Primary diagnosis                 |                 |
| Falls                             | 17 (15)         |
| Parkinsonism                      | 14 (12.7)       |
| Joint arthroplasty                | 13 (11.8)       |
| Stroke                            | 10 (9)          |
| Upper/lower limb fracture         | 9 (8.1)         |
| Cardiovascular disease            | 7 (6.3)         |
| Diabetes                          | 2 (1.8)         |
| No. of comorbidities, median (IQR)| 4 (2.5)         |
| Self-rated impact of balance on ADL performance |       |
| Nil                               | 11 (10)         |
| Minor                             | 42 (38.1)       |
| Moderate                          | 35 (31.8)       |
| Major                             | 22 (20)         |

*Values are reported as number (%) of exercisers unless otherwise indicated. ADL = activities of daily living; IQR = interquartile range.
### Table 4.
Summary of Fit Statistics for Rasch Sequential Analyses of Items on the BIS for Therapists

| Item Set | Analysis | No. of Cases | Adjusted \(\alpha\) | Overall Model Fit | Item Fit Residual Mean (SD) | Person Fit Residual Mean (SD) | Person Separation Index | Cronbach \(\alpha\) | % Significant \(t\) Tests |
|----------|----------|--------------|---------------------|-------------------|-----------------------------|-----------------------------|--------------------------|----------------|--------------------------|
| Items 1–17 | 1        | 248          | .002                | \(x^2\) 121.2 df 34 \(P < .0000\) | \(-0.09 (2.00)\) | \(-0.218 (0.44)\) | 0.73 | 0.86 | 1.14 |
| Items 1, 6, 8–14, 16, 17 | 2        | 229          | .004                | \(x^2\) 46.63 df 22 \(P < .0016\) | \(0.220 (1.55)\) | \(-0.151 (0.42)\) | 0.60 | 0.79 | 0.00 |
| Items 1, 6–10, 12, 15 | 3        | 203          | .004                | \(x^2\) 53.59 df 16 \(P < .0000\) | \(0.068 (1.98)\) | \(-0.103 (0.60)\) | 0.62 | 0.85 | 0.00 |
| Items 1, 2, 6, 7, 11–17 | 4        | 228          | .004                | \(x^2\) 40.99 df 22 \(P < .0082\) | \(-0.290 (1.31)\) | \(-0.187 (0.44)\) | 0.62 | 0.79 | 0.00 |
| Items 1, 2, 6, 8–17 | 5        | 230          | .003                | \(x^2\) 68.72 df 26 \(P < .0000\) | \(-0.325 (1.49)\) | \(-0.162 (0.47)\) | 0.66 | 0.82 | 0.31 |
| Items 1, 2, 6, 7, 13–17 | 6        | 222          | .005                | \(x^2\) 20.11 df 18 \(P < .326\) | \(-0.244 (1.137)\) | \(-0.216 (0.425)\) | 0.56 | 0.75 | 1.25 |

*Bold type indicates a marker of a misfit. BIS = Balance Intensity Scale.

### Table 5.
Summary of Fit Statistics for Rasch Analyses of Items on the BIS for Exercisers

| Item Set | Analysis | No. of Cases | Adjusted \(\alpha\) | Overall Model Fit | Item Fit Residual Mean (SD) | Person Fit Residual Mean (SD) | Person Separation Index | Cronbach \(\alpha\) | % Significant \(t\) Tests |
|----------|----------|--------------|---------------------|-------------------|-----------------------------|-----------------------------|--------------------------|----------------|--------------------------|
| Items 1–13 | 1        | 189          | 0.003               | \(x^2\) 74.54 df 26 \(P < .0000\) | \(-0.680 (1.66)\) | \(-0.219 (0.52)\) | 0.43 | 0.81 | 0.94 |
| Items 2–7, 10–13 | 2        | 162          | 0.005               | \(x^2\) 38.24 df 20 \(P < .0082\) | \(0.262 (1.29)\) | \(-0.222 (0.50)\) | 0.16 | 0.73 | 0.30 |
| Items 2–7, 9–13 | 3        | 162          | 0.004               | \(x^2\) 45.38 df 22 \(P < .0023\) | \(0.0330 (1.41)\) | \(-0.219 (0.51)\) | 0.23 | 0.76 | 0.00 |
| Items 2–8, 10–13 | 4        | 162          | 0.004               | \(x^2\) 53.54 df 22 \(P < .0001\) | \(-0.339 (1.58)\) | \(-0.208 (0.51)\) | 0.26 | 0.78 | 0.00 |
| Items 1–7, 10–13 | 5        | 189          | 0.004               | \(x^2\) 42.56 df 22 \(P < .0053\) | \(-0.506 (1.23)\) | \(-0.227 (0.50)\) | 0.33 | 0.75 | 0.00 |
| Items 4, 6–11 | 6        | 140          | 0.007               | \(x^2\) 29.28 df 14 \(P < .0095\) | \(-0.133 (1.295)\) | \(-0.212 (0.749)\) | 0.25 | 0.81 | 0.00 |

*Bold type indicates a marker of a misfit. BIS = Balance Intensity Scale.
BIS for Therapists and Exercisers

and/or probabilities below the adjusted alpha value,17 as shown in Supplementary Table 1 (available at https://academic.oup.com/ptj). Key results of analysis 1 are shown in Table 4. There was no evidence of misfit of persons.

Systematic investigation of item misfit examined differential item functioning, local dependency, and unidimensionality. Differential item functioning examined potential uniform or nonuniform bias attributable to exerciser age, sex, or therapist’s years of experience. Systematic refinement of scale items reflected statistical, theoretical, and clinical considerations. Table 4 presents analyses sequentially.

The optimal solution was analysis 4, with a moderate PSI (0.62). Other psychometrics of this item set were subsequently explored. Some misfit of item 7 (any postural reaction) persisted with a fit residual of −0.992 (P = .004), but this improved on the initial analysis where the fit residual was −2.46 (P < .001). Given the overall model fit and detrimental effect of removing item 7, this item was retained. Fit to the Rasch model indicates a hierarchical ordering of this refined item set. There was no local dependency evident in this item set and unidimensionality was preserved. On the basis of the outcome of the Rasch analysis, the refined BIS-T item set is presented in the Figure along with the conversion from raw score to interval Rasch score.

BIS-E assessment of unidimensionality and item hierarchy. Initial analysis of the 13 items of BIS-E showed poor fit to the Rasch model. Three items recorded fit residual values over the recommended level of ±2.5 and/or probabilities below the adjusted alpha value,17 as shown in Supplementary Table 2 (available at https://academic.oup.com/ptj). Key results of analysis 1 are shown in Table 5. There was no evidence of misfit of persons. Item misfit was investigated in the same way as that described for BIS-T (Tab. 5). The optimal solution arising from analysis 5 indicated model fit and item hierarchy but a poor PSI (0.33). Exploration of other psychometrics of this item set showed no item or person misfit, or local dependency, and unidimensionality was preserved.

Targeting of BIS-T and BIS-E

The person-item threshold maps shown in the Supplementary Figure 1 (available at https://academic.oup.com/ptj) were visually inspected to assess targeting of the scales. The refined versions of the BIS-T and BIS-E are moderately targeted, with most persons clustered in the lower half of the scale and > 0.5 logit between items at the higher end of the scale. The higher end of the scale (greater exercise intensity) is the area of clinical interest, but fewer items targeted this end of the scale spectrum. Reliability of the BIS-T was fair, whereas that of the BIS-E was poor. With a PSI of 0.62, the BIS-T can statistically distinguish between 2 strata of balance intensity (ie, higher/lower) at best, whereas with a PSI of 0.33 the BIS-E cannot separate out true differences in scores from measurement error at this stage.

Global Effort Ratings

Tasks were subjectively rated at “a lot” or “maximal” effort for 20% (66/319) of the activities by therapists and 15% (48/319) of the activities by exercisers using the 5-point global effort rating. If one-third of prescribed activities were of a high balance challenge, then the frequency of activities scoring 4 or 5 on the global rating of effort should have been closer to 33% (107/319). This indicates that therapist prescriptions were too conservative for exercisers to experience a high level of balance challenge, limiting higher end tests of these scales. There were strong correlations between BIS-T totals and the GRS for therapist (0.86; P = .000), BIS-E totals and the GRS for exerciser (0.70; P = .000), and the GRS for therapist and GRS for exerciser (0.70; P = .000) (Suppl. Tab. 3 and Suppl. Tab. 4, available at https://academic.oup.com/ptj). The good correlation between BIS-T and BIS-E scores and global effort ratings provides evidence of validity for both scales.

Discussion

Two new measures of balance exercise intensity, the BIS-T for therapists and BIS-E for exerciser self-rating, have been developed and refined in this validation study. Both scales tested comprised of an objective item checklist and a subjective global rating of effort to maintain balance. Both the BIS-T and BIS-E were unidimensional and demonstrated a hierarchical distribution of items that fit the Rasch model. Whereas the BIS-T items were reflective of a range of task difficulty from lower to higher intensity, the exerciser ratings for the BIS-E items were not. However, it is promising that the simple 5-point global rating scale on the BIS-E correlated well with the BIS-T scores. At this time, it is recommended that this BIS-T can be used in full (item checklist score converted to a Rasch score out of 100 plus the GRS) and the simple global rating of effort (“no effort at all” to “maximal effort”) be used by exercisers until further validation of the BIS-E items can be conducted.

The recognition that measurement of balance exercise intensity is important is growing32 and with it the importance of the development of a method to measure this complex variable. A recent study by Espy et al33 tested a 10-point subjective rating of perceived exertion, modeled on the Borg Scale for the rating of perceived exertion.33 Espy et al assigned the terms “completely stable” and “about to fall” to scores of 1 and 10, respectively, and reported poor associations between heart rate and
**STEP 1: BIS-T item checklist**

For any item ‘unsure’ or unable to be observed, score NO (+/-comments)  

| ITEMS                          | Task 1 | Task 2 | Task 3 |
|-------------------------------|--------|--------|--------|
| **PRE-TASK**                  |        |        |        |
| Hesitation to Start           |        |        |        |
| 1. Did the person hesitate, for any reason, before attempting the task? | No | Yes | No | Yes | No | Yes |
| 2. Did the person hesitate more than 5 seconds before attempting the task? | No | Yes | No | Yes | No | Yes |
| **Starting Position**         |        |        |        |
| 3. Did the person require more than 1 attempt to get into the starting position? | No | Yes | No | Yes | No | Yes |
| **Postural Reactions**        |        |        |        |
| 4. Did you see any postural reaction associated with the task? | No | Yes | No | Yes | No | Yes |
| 5. Did you see any initiation of leg movement or actual step/s to control postural sway? | No | Yes | No | Yes | No | Yes |
| 6. Did you see any initiation of arm movement or actual reach or grab to control postural sway? | No | Yes | No | Yes | No | Yes |
| **Bracing and Breathing**     |        |        |        |
| 7. Did the person hold any part of their body stiff or rigid during this task? | No | Yes | No | Yes | No | Yes |
| i.e. holding limb/s stiff or rigid, making fist, clenching jaw, pulling on own clothing, or propping limbs e.g. hand on thigh, increased tone and associated reactions, shoulder elevation |        |        |        |        |        |        |
| 8. Did the person change their breathing pattern during the task? | No | Yes | No | Yes | No | Yes |
| i.e. increased or decreased depth of breathing, sighing, breath holding, or faster rate of breathing. More shallow breathing may be characterized by increased upper chest breathing, shoulder girdle elevation, or abdominal movement. |        |        |        |        |        |        |
| **Balance Threshold**         |        |        |        |
| 9. Did the person appear unsteady at any time while preparing for or performing tasks? | No | Yes | No | Yes | No | Yes |
| 10. Did you have to say anything to prevent the person losing balance? | No | Yes | No | Yes | No | Yes |
| 11. Did you have to provide any physical assistance to prevent the person losing balance? | No | Yes | No | Yes | No | Yes |
| 12. Did the person fall* during the task? – NOT SCORED | No | Yes | No | Yes | No | Yes |

**Total raw score:**

*WHO definition: A fall is an event which results in a person coming to rest inadvertently on the ground or floor or other lower level

**STEP 2: Global Rating Scale of Balance Effort – circle for each task**

1 2 3 4 5 1 2 3 4 5 1 2 3 4 5

How hard did the person work to maintain their balance?

1 – No effort 2 – A little effort 3 – Some effort 4 – A lot of effort 5 – Maximal effort

**Conversion of BIS-T raw score to BIS-T Rasch score**

| Raw score | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|-----------|---|---|---|---|---|---|---|---|---|---|----|----|
| Rasch score | 0 | 10 | 19 | 26 | 32 | 38 | 44 | 51 | 60 | 71 | 84 | 100 |

**Balance Intensity Scale – Exerciser (BIS-E)**

| Global Rating Scale of Balance Effort | TASK 1: | TASK 2: | TASK 3: |
|--------------------------------------|---------|---------|---------|
| 1 – No effort at all                  | 1 2 3 4 5 | 1 2 3 4 5 | 1 2 3 4 5 |
| 2 – A little effort                    |         |         |         |
| 3 – Some effort                       |         |         |         |
| 4 – A lot of effort                   |         |         |         |
| 5 – Maximal effort                    |         |         |         |

**Figure.**

Balance Intensity Scale for Therapists (BIS-T) with refined item list and Rasch conversion scores and Balance Intensity Scale for Exercisers (BIS-E) global rating.
balance intensity challenge in 30 young participants who were 19 to 43 years old and were participating in balance exercises on a gaming platform. Their study does not detail the relationship between perceived balance challenge and actual task difficulty, nor was any observer rating of balance challenge intensity used for validation purposes. The authors contend that “the degree to which a person finds an activity challenging cannot be measured externally,” but our earlier observational study details a number of observable markers of balance challenge intensity that were used in the development of both the therapist- and exerciser-rated scales tested in the current study.

This study fills an important gap in the science of balance exercise prescription by reporting the initial validation of the BIS-T and BIS-E. These 2 measures were constructed using a systematic process of observational analysis of balance task performance, followed by key stakeholder consultation that guided the research team to construct therapist-reported and exerciser–self-reported scores using clinically relevant items and language meaningful to end users. This study has now resulted in refinement and validation of the BIS-T and BIS-E in a sample of older adult exercisers in a variety of settings and treatment streams. It is important to note that therapists and exercisers completed ratings during clinical consultations in <10 minutes with no equipment beyond what would normally be available in these settings. Training prior to using these scales appears to have been effective, despite minimal resource utilization. This is encouraging, given the broad range of clinical settings, therapist years of experience (1–30 years), age range of exercisers (50–91 years), and diversity of treatment programs. In addition, the psychometric assessment of the 2 scales indicates unidimensionality. With refinement, both scales fit a Rasch model and demonstrate a hierarchy of item difficulty, with certain items only scored “yes” at higher levels of balance exercise intensity (eg, extended hesitation time or need for intervention to prevent loss of balance). Item hierarchy was similar for the BIS-T and BIS-E, and there is good correlation between global ratings of balance effort and BIS-T and BIS-E item scores.

Several limitations to this study have been identified. Two-thirds of exercisers required the assistance of their therapist to complete their rating form, which may have influenced the exerciser ratings despite therapists and exercisers being instructed not to modify ratings under these circumstances. Previous comparisons of therapist and consumer measurement of outcomes have shown that therapists are prepared to recognize and disagree with patient ratings, and there is evidence that health care providers and consumers can agree when rating health outcomes. The results of this study showed correlation between exerciser and therapist ratings. This indicates a level of agreement between therapist and exerciser ratings, but the high proportion of proxy ratings may explain the large number of exercisers not selecting any items for 40% of the balance tasks tested (no indicator of balance challenge noted), which is effectively an underrating of balance task intensity when compared with only 12% of tasks having no items selected on the therapists ratings.

A related issue has also limited the validation of the BIS-T and BIS-E. Activities intended to provide high-intensity challenge were too infrequently prescribed for this sample of exercisers. This limited the number of high-intensity tasks experienced and subsequently the number of cases available for Rasch analysis. Had exercisers been more highly challenged, better discrimination between lower and higher intensity tasks might have been seen. At this stage, the ability of the scale to discriminate at the higher end of the spectrum of balance challenge requires further validation.

The reluctance of therapists in this trial to push exercisers to high levels of balance challenge for one-third of tasks, even in controlled 1:1 conditions with a set protocol, suggests that exercisers are not working at or near the limits of postural stability during routine exercise programs even when supervised. This demonstrated “underchallenging” of balance is instructive because it highlights an issue that has not previously been measured in balance exercise trials. Even though therapists know that “challenging” balance exercises are recommended and perceive they can and do prescribe exercise at this level, this does not appear to be routine practice. More research is needed to understand how to enable therapists to test exercisers at high levels of balance exercise intensity. When this is understood, further testing of BIS-T and BIS-E will be conducted.

**Conclusion**

These scales result from the first attempt to systematically develop scales to assess the intensity of balance challenge in older adults based on observations in clinical practice. It is recommended that the refined BIS-T comprising the objective item checklist in conjunction with the global effort rating be used clinically to distinguish balance exercises of increasing intensity in older adult populations. Second, it is recommended that the use of the subjective global effort rating provides the most useful differentiation of balance exercise intensity from the exercisers perspective. Furthermore, this work establishes a platform for progressing the further validation of these BIS on larger samples of high-intensity tasks and in specific older adult populations. It is anticipated that the use of these scales will give therapists and exercisers enhanced capability to accurately measure the level of balance exercise challenges for exercisers, therefore contributing to the prescription of optimal balance training.
References

1. Farlie M, Robins L, Keating J, Molloy E, Haines T. An absence of intensity reporting in the prescription of balance exercises in randomised controlled trials: a systematic review. J Physiother. 2013;59:227–235.

2. Farlie MK, Robins L, Haas R, Keating JL, Molloy E, Haines TP. Programme frequency, type, time, and duration do not explain the effects of balance exercise in older adults: a systematic review with a meta-regression analysis. Br J Sports Med. Published Online January 25, 2018. doi: 10.1136/bjsports-2016-090674. Accessed March 8, 2019.

3. Sherrington C, MichaleffZA, Fairhall N, et al. Exercise to prevent falls in older adults: an updated systematic review and meta-analysis. Br J Sports Med. 2017;51:1750–1758.

4. American College Sports Medicine. ACSM’s Resource Manual for Exercise Testing and Prescription. 7th ed. Philadelphia, PA: Wolters Kluwer Lippincott Williams and Wilkins; 2014.

5. Tiedemann A, Sherrington C, Close JCT, Lord SR. Exercise and sports science. Australia position statement on exercise and falls prevention in older people. J Sci Med Sport. 2011;14:489–495.

6. Farlie MK, Molloy E, Keating JL, Haines TP. Clinical markers of the intensity of balance challenge: observational study of older adult responses to balance tasks. Phys Ther. 2016;96:313–325.

7. Farlie MK. Development and Validation of Scales to Rate the Intensity of Balance Exercise (BIS). Melbourne, Victoria, Australia: Department of Physiotherapy, Monash University; 2017.

8. de Morton NA, Davidson M, Keating JL. The de Morton Mobility Index (DEMMI): an essential health index for an ageing world. Health Qual Life Outcomes. 2008;6:65.

9. de Morton NA, Davidson M, Keating J, Validity, responsiveness, and the minimal clinically important difference for the de Morton Mobility Index (DEMMI) in an older acute medical population. BMC Geriatr. 2010;10:72.

10. de Morton NA, Lane K. Validity and reliability of the de Morton Mobility index in the subacute hospital setting in a geriatric evaluation and management population. J Rehabil Med. 2010;42:929–4296.

11. de Morton NA, Meyer C, Moore K, Dow B, Jones C, Hill K. Validation of the de Morton Mobility Index (DEMMI) with older community care recipients. Australas J Ageing 2011;30:220–225.

12. de Morton NA, Brusco NK, Wood L, Lawler K, Taylor NF. The de Morton Mobility Index (DEMMI) provides a valid method for measuring and monitoring the mobility of patients making the transition from hospital to the community: an observational study. J Physiother. 2011;57:109–116.

13. Davenport SJ, de Morton NA. Clinimetric properties of the de Morton Mobility Index in healthy, community-dwelling older adults. Arch Phys Med Rehabil. 2011;92:51–58.

14. de Morton NA, Harding KE, Taylor NF, Harrison G. Validity of the de Morton Mobility Index (DEMMI) for measuring the mobility of patients with hip fracture during rehabilitation. Disabil Rehabil. 2013;35:325–333.

15. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th ed. Boston, MA: Pearson; 2013.

16. Pallant J. SPSS Survival Manual. 6th ed. Sydney, New South Wales, Australia: Allen & Unwin; 2016.

17. 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.

18. Bond T, Fox C. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 2nd ed. New York, NY: Routledge Taylor Francis Group; 2007.

19. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Care Res (Hoboken). 2007;57:1358–1362.

20. Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoecon Outcomes Res. 2011;11:571–585.

21. Kaiser H. A second generation Little Jiffy. Psychometrika. 1970;35:401–415.

22. Kaiser H. An index of factorial simplicity. Psychometrika. 1974;39:51–56.

23. Bartlett MS. A note on the multiplying factors for various chi square approximations. J R Stat Soc. 1954;16:296–298.
BIS for Therapists and Exercisers

24 Cattell RB. The Scree test for the number of factors. *Multivariate Behav Res.* 1966;1:245–276.
25 Watkins MW. Monte Carlo PCA for parallel analysis. http://edpsychassociates.com/Watkins3.html. Accessed June 5, 2019.
26 RUMM2030 [computer program]. Version 5.52 for Windows. Perth, Western Australia, Australia: RUMM Laboratory Pty Ltd; 2016.
27 Rasch G. An item analysis which takes individual differences into account. *Br J Math Stat Psychol.* 1966;19:49–57.
28 Wright BD, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989;70:857–860.
29 Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *J Rehabil Med.* 2003;35:105.
30 Tinsley HEA, Tinsley DJ. Uses of factor analysis in counseling psychology research. *J Couns Psychol.* 1987;34:414–424.
31 Linacre J. Sample size and item calibration stability. *Rasch Measurement Transactions.* 1994;7:328.
32 Reinthal A. Getting the dosage right in balance exercise prescription: the intensity problem. *J Neuropsychol.* 2017;7:e147.
33 Espy D, Reinthal A, Meisel S. Intensity of balance task intensity, as measured by the rate of perceived stability, is independent of physical exertion as measured by heart rate. *J Neuropsychol.* 2017;7:343.
34 Borg G. Ratings of perceived exertion and heart rates during short-term cycle exercise and their use in a new cycling strength test. *Int J Sports Med.* 1982;3:153–158.
35 McPhail S, Comans T, Haines T. Evidence of disagreement between patient-perceived change and conventional longitudinal evaluation of change in health-related quality of life among older adults. *Clin Rehabil.* 2010;24:1056–1044.
36 McPhail S, Nalder E, Hill A, Haines T. Physiotherapists have accurate expectations of their patients’ future health-related quality of life after first assessment in a subacute rehabilitation setting. *Biomed Res Int.* 2013;2013:8.
37 Wu S, Wang R, Zhao Y, et al. The relationship between self-rated health and objective health status: a population-based study. *BMC Public Health.* 2013;13:1–9.