5-2014

# Inter- and Intra-rater Reliabililty of a Modified Approach to Scoring the Balance Error Scoring System BESS in Undistracted and Distracted Conditions

Jacob Ryan Pierce
*Utah State University*

5-2014

# Inter- and Intra-rater Reliabililty of a Modified Approach to Scoring the Balance Error Scoring System (BESS) in Undistracted and Distracted Conditions

Jacob Ryan Pierce
*Utah State University*

**Inter- and Intra-rater Reliability of a Modified Approach to Scoring the Balance Error Scoring System (BESS) in Undistracted and Distracted Conditions**

Jacob R. Pierce[1,2], Adam C. Raikes[1,2], and Sydney Y. Schaefer[1,3]

[1]Department of Health, Physical Education and Recreation, Utah State University; [2]Utah State University Sports Medicine, Athletics Department; [3]Department of Physical Therapy, University of Utah

Running head: Reliability of modified BESS approach

Corresponding author: Sydney Y. Schaefer, PhD
Utah State University
HPER Department
7000 Old Main Hill
Logan, UT 84322-7000
Phone: 435-797-8441
Fax: 435-797-3759
E-Mail: sydney.schaefer@usu.edu

Other authors: Jacob R. Pierce, LAT, ATC, CSCS
Utah State University
HPER Department
7000 Old Main Hill
Logan, UT 84322-7000
E-Mail: jrpierce@me.com

Adam C. Raikes, MS, LAT, ATC
Utah State University
HPER Department
7000 Old Main Hill
Logan, UT 84322-7000
E-Mail: adam.raikes@gmail.com

Number of words: 3,496

Number of figures: 1

Number of tables: 7

**ACKNOWLEDGEMENTS**

Thanks to Dr. Eadric Bressel and Dr. Richard Gordin for their assistance with study design and statistical analyses. We would also like to thank Dale Mildenberger, MS, LAT, ATC and Dr. Jonathan Finnoff for developing the modified BESS used in this study.

**ABSTRACT**

**Context:** We have developed a modified approach to scoring balance errors with the Balance Error Scoring System (BESS) by eliminating two of the six original error criteria.

**Objective:** To measure the inter- and intra-rater reliability of this modified BESS approach. Because of the potential auditory distractions that may be present in a clinical setting, we also measured reliability for distracted rating conditions.

**Design:** Cross-sectional.

**Setting**: University sports medicine clinic.

**Patients or Other Participants**: Board of Certification (BOC) certified athletic trainers as raters (n=6).

**Intervention**: Raters used a modified approach to scoring the BESS to score 32 different sets of pre-recorded videos of balancers completing the six stance positions of the BESS. They first completed this in an undistracted condition, then one week later in a distracted condition as they performed a concurrent auditory vigilance task.

**Main Outcome Measures:** Two-way, random, absolute agreement intraclass correlation coefficients (ICC) measured the inter- and intra-rater reliability of the modified approach for undistracted and distracted conditions. Repeated measures ANOVAs tested for effects of condition (undistracted vs. distracted) and also BESS stance position (e.g. single-leg, foam) on the raters' reported number of errors.

**Results:** For the undistracted condition, inter- and intra-rater reliability ICCs were 0.93 and 0.92, respectively. For the distracted condition, inter- and intra-rater reliability ICCs were 0.89 and 0.92, respectively. There was no significant effect of condition (undistracted vs. distracted) on the reported number of errors in the total error score or in the individual stance position scores.

**Conclusions**: This modified approach to scoring the BESS is reliable, even when used in a potentially distracting environment.

**Key Words**: balance; reliability; distraction; Balance Error Scoring System

## INTRODUCTION

Reliable balance assessments are necessary in clinical settings to document musculoskeletal injury recovery,[1] determine fall risk,[2] and decide post-concussion return-to-play[3]. Although there are a variety of balance assessment tools, athletic trainers may use the Balance Error Scoring System (BESS).[3-6] The BESS includes three stances (double-leg, single-leg, and tandem) on two separate surfaces (firm ground and foam), yielding six different stance positions.[7] This test requires individuals to stand in each position for 20 seconds with their eyes closed and their hands on their hips. To assess an individual's balance using the BESS, a clinician counts the number of errors the individual makes in each stance position. There are six criteria for what constitutes a "balance error" in the original version of the BESS: 1) lifting hands off of hips; 2) opening eyes; 3) stepping, stumbling, or falling out of position; 4) taking >5 seconds to return to position; 5) lifting forefoot or heel; or 6) abducting/flexing hip joint >30 degrees. The clinician records the number of errors made in each stance position, and then sums each position to yield a total BESS score. The original BESS is valid, can be performed quickly in a small space, requires little equipment, and is inexpensive.[7] It has been shown, however, to have "poor to moderate" inter- and intra-rater reliability,[8-10] according to general guidelines.[11] One possible explanation for the "poor to moderate"[11] reliability of the original BESS is the number of possible error criteria.[8]

We have modified the criteria for counting balance errors within the BESS criteria by removing two of the six original error criteria. This modified approach (mBESS) aims to increase the reliability of the original BESS, based on the assumption that clinicians are not highly reliable in visually confirming whether a balancer has lifted his/her forefoot or heel, or has abducted or flexed the hip joint more than 30 degrees. Thus, the mBESS counts only the first four criteria from the BESS as an error, and does not count the last two described above. By minimizing these possible error counts, clinicians using this modified approach may be more reliable at detecting changes in balance performance. Regardless of which approach is used, however, clinicians typically rate the BESS in environments that contain a number of potential auditory distractions, such as a busy athletic training room or a noisy sideline. Distraction can affect balance performance during the BESS,[12] but less is known about the extent to which distraction affects rating ability. If a clinician is required to attend to another task at the same time as rating the BESS, his/her ability to reliably rate balance errors may decline. Thus, the purpose of this study was to measure the inter- and intra-rater reliability of a modified approach to scoring the balance error scoring system (BESS) in undistracted and distracted conditions. We hypothesized that both the inter- and intra-rater reliability of this modified approach would be “good” for the undistracted condition, but only “poor to moderate” for the distracted condition. We also hypothesized that the reported number of errors would be different for the undistracted condition compared to the distracted condition.

## METHODS

### Participants

Six Board of Certification (BOC) certified athletic trainers participated in this study (mean±SD age: 29.04±7.87 yrs). Participants used a modified approach to scoring balance errors in the BESS to rate balance performance of 32 (out of 64 total) pre-recorded sets of videos of individuals completing all six stance positions of the BESS, and are therefore referred to as ‘raters.’ Details regarding these pre-recorded sets of videos and the modified BESS (mBESS) scoring approach are described below in Experimental Design. Mean and SD years of certification was 4.81±6.15 across raters. We first determined the reliability of the mBESS in an undistracted condition (Time 1). We then re-evaluated reliability measures in a distracted condition using an auditory vigilance task one week later (Time 2; description of auditory vigilance task below). All raters provided informed consent prior to participation in compliance with the Institutional Review Board at Utah State University.

**Experimental design**

*General methods*

All raters (n=6) sat in front of a computer screen in a quiet room and scored sets of pre-recorded balance videos using only four of the six original BESS criteria. Table 1 illustrates the differences in criteria between the original and modified error criteria. If the balancer made multiple errors at the same time, such as opening the eyes and falling out of position, only one error was recorded. Also, once the balancer made an error according to the modified approach, the rater was instructed not to record another error until the balancer had returned to the correct stance position. To ensure that the raters used the correct criteria in their scoring, they were asked to list from memory the criteria that they used to score balance errors for each video at the end of both Time 1 and Time 2. Raters viewed 32 (out of 64) pre-recorded videos within a PowerPoint (Microsoft, Redmond, WA) presentation that had been prepared previously by the experimenter. At both Time 1 and Time 2, raters completed their scoring while wearing headphones, regardless of whether they were subjected to the auditory vigilance task or not.

Prior to rater recruitment for this study, we recruited 32 'balancers' who were free from a lower extremity injury in the previous year, and had no history of any physician-diagnosed concussion. Balancers provided informed consent for their participation in this study. Balancers were then video-recorded in the frontal plane as they completed all six stance positions of the BESS: double-leg, firm (ground); single-leg, firm; tandem, firm; double-leg, foam; single-leg, foam; tandem, foam. Each stance position was held for 20 seconds. Additional details regarding these conditions and the test itself have been published previously.[7,13] All videos were archived by balancer and by stance position, such that each balancer yielded six separate 20-sec video files, which equaled one set of videos. For reasons beyond the scope of this study, the 32 balancers were video-recorded again a week later in the same fashion, yielding a total of 64 sets of videos. A random selection of 32 sets were then inserted offline by the experimenter into PowerPoint presentations for the raters to view, with each 20-sec video separated by a blank slide to allow the raters time to record the corresponding error score on a paper data collection sheet. The composition of video sets within the PowerPoint presentations enabled the testing of inter- and intra-rater reliability of the mBESS approach.

*Time 1*

Raters scored 16 sets of videos, were given a five-minute rest period, and then scored another 16 sets of videos. Raters were undistracted during Time 1. After each video, raters were instructed to write down the total number of errors they observed on their data collection sheet. These data were used to determine the inter-rater reliability of the mBESS approach before any potential distractions were introduced in the study.

*Time 2*

One week later, raters again scored 16 sets of videos, were given a five-minute rest period, and then scored another 16 sets of videos. To determine the effect of distraction on the a) reliability and b) rater performance of the mBESS approach, raters scored one of the two sets of videos (either the first or second set) while performing a concurrent auditory vigilance task that served as a potential distractor to the rater.

*Distracted rating conditions with an auditory vigilance task*

We used an auditory vigilance task to potentially distract the raters as they used the mBESS to score balance errors. Auditory vigilance has been shown previously to significantly distract healthy young adults[14] and adults with stroke[15,16] as they performed a concurrent motor task (i.e. under dual-task conditions), regardless of general intelligence level.[17,18] The auditory vigilance task in this study required raters to listen to 20-sec recordings of 35-letter sequences. Each 35-letter sequence consisted of a random series of the same four letters (A, G, M, and O). The sequence began as soon as the balancer began the condition, followed by 35 letters (at

1.75 Hz), and ran continuously for the duration of the stance position. Prior to each sequence, raters were instructed to pay attention to the number of times a target letter was heard. The target letter was A, G, M, or O, and was changed for each trial. Immediately after each trial, raters were asked to record the number of times a target letter was heard on a paper data collection sheet. Raters were asked to both score the videos to the best of their ability, as well as get the number of target letters correct. All sequences were recorded live using an external microphone (Gigaware Omnidirectional model 33-119, Ignition L.P., Dallas, TX) and played at a comfortable volume through headphones (Sony MDR-V700). All audio files were embedded in the PowerPoint that included the videos. Raters were given two familiarization trials of the auditory vigilance task prior to doing so while simultaneously rating the 16 sets of balance videos at Time 2. During two more trials, to gather a baseline auditory vigilance score, raters remained seated, with eyes closed, and focused only on the number of times a target letter was read.

**Data analysis**

Raters reported the number of balance errors using the mBESS approach for each stance position: double-leg, firm (ground); single-leg, firm; tandem, firm; double-leg, foam; single-leg, foam; tandem, foam. The reported numbers of errors from the individual stance positions were also summed to yield a total 'composite' score for each set of videos. These individual stance positions, along with the total composite score, were used to compute reliability measures and to compare rater performance.

**Statistical analysis**

To measure the inter- and intra-rater reliability of the mBESS approach for both undistracted and distracted conditions, we used two-way, random, absolute agreement intraclass correlation coefficients ($ICC_{2,1}$)[11,19] All ICC values and confidence intervals were determined using SPSS software version 21 (IBM, Armonk, NY). Inter-rater reliability for undistracted conditions was computed across video sets 1-16 in Time 1 and Time 2. Intra-rater reliability for undistracted conditions was computed within raters across Time 1 and Time 2, using only the error reports from the 16 sets of videos viewed without distraction. Inter-rater reliability for distracted conditions was computed across raters in Time 2, using only the error reports from the 16 sets of videos viewed while simultaneously performing the auditory vigilance task. Intra-rater reliability for distracted conditions was computed within raters using the 16 sets of videos from Time 1 (undistracted) that corresponded with the same video sets rated under distracted conditions from Time 2. All ICC values are reported as 0.00-1.00, with larger values indicating higher reliability. Reliabilities were computed for all six separate BESS stance positions as well as total composite scores.

To test whether the reported number of errors was different in the undistracted condition compared to the distracted condition, we compared mean total composite scores across raters using a one-way repeated measures ANOVA with condition (undistracted vs. distracted) as a within-subject factor. Based on our hypothesis, we expected no main effect of condition on the reported number of errors. To test whether the auditory vigilance task affected the raters' scoring of a particular stance position compared to another position, we used a 2x4 repeated measures ANOVA with condition (undistracted vs. distracted) and stance position (single-leg, firm vs. tandem, firm vs. single-leg, foam vs. tandem, foam) as within-subject factors. While there are six separate stance positions in the BESS protocol,[7] the two double-leg stance positions (firm and foam) yielded zero reported errors (see Results). Thus, only the four error-inducing stance positions were included in our ANOVA. We expected significant main effects of condition and stance position on the reported number of errors, yet no interaction between condition and stance position. Significant effects were tested post hoc using Tukey-Kramer Honestly Significant Different (HSD) tests.[20,21] Multiple comparisons were accounted for using the Kackar-Harville correction.[22] All ANOVAs and subsequent posthoc analyses were performed with JMP 8.0 (SAS Institute Inc., Carey, NC), with an alpha level of .05.

## RESULTS

### What is the reliability of the modified BESS approach for the undistracted condition?

*Inter-rater reliability*

Table 2 shows the inter-rater intraclass correlation coefficient (ICC) values and confidence intervals (CI) for each stance position and for the total score when the raters were undistracted. These values are reported for Time 1 (i.e. initial rating session). The inter-rater ICC value for the total score was 0.93. The values for both double-leg stance positions (firm and foam) were zero, as all raters reported no errors. Of the other four stance positions, only single-leg, foam had an inter-rater ICC value below a "good" range.[11]

Table 3 shows the inter-rater ICC values and CIs for each stance position and for the total score when the raters were undistracted. These values are reported for Time 2 (i.e. second rating session). The inter-rater ICC value for the total score was also 0.93. The values for both double-leg stance positions (firm and foam) were zero, as all raters again reported no errors. All other four stance positions had inter-rater ICC values within the "good" range.[11]

*Intra-rater reliability*

Table 4 shows the intra-rater ICC values and CIs for each stance position and for the total score when the raters were undistracted. These values are reported within rater between Time 1 and Time 2 (i.e. initial and second rating sessions). The intra-rater ICC value for the total score was 0.92. The values for both double-leg stance positions (firm and foam) were zero, as all raters reported no errors. All other four stance positions had intra-rater ICC values within the "good" range.[11]

**What is the reliability of the modified BESS approach for the distracted condition?**

*Inter-rater reliability*

Table 5 shows the inter-rater ICC values and CIs for each stance position and for the total score when the raters were distracted. These values are reported for Time 2 (i.e. second rating session w/ distractor). The inter-rater ICC value for the total score was 0.89. The values for both double-leg stance positions (firm and foam) were zero, as all raters reported no errors. Of the other four stance positions, only single-leg, foam had an inter-rater ICC value below a "good" range.[11]

*Intra-rater reliability*

Table 6 shows the intra-rater ICC values and CIs for each stance position and for the total score when the raters were distracted. These values are reported for the comparison between Time 1 (undistracted) and Time 2 (distracted) within rater. The intra-rater ICC value for the total score was 0.92. The values for both double-leg stance positions (firm and foam) were zero, as all raters reported no errors. Of the other four stance positions, only single-leg, foam had an inter-rater ICC value below a "good" range.[11]

**Was the reported number of errors different between the undistracted and distracted conditions?**

Figure 1A shows the reported number of errors for the total BESS score in undistracted and distracted conditions. Our one-way repeated measures ANOVA reported no significant effect of condition (undistracted vs. distracted) on the reported number of errors in the total error score ($F_{1,5}$ = .045; p=.83). Given that ICC values varied by BESS stance position (see Tables 2-6), it is plausible that different stance positions could be differentially affected when the rater concurrently performed the auditory vigilance task. Our 2x4 repeated-measures ANOVA reported no significant main effect of condition ($F_{1,5}$ = .101; p=.75) nor an interaction between condition and stance position ($F_{3,5}$=.057; p=.98) on the reported number of errors. There was, however, a significant main effect of stance position on reported number of errors ($F_{3,5}$=262.831; p<.0001; Fig. 1B). Post-hoc analyses revealed that the reported number of errors in all error-inducing stance positions (single leg, firm; tandem, firm; single leg, foam; tandem, foam) were significantly different from each other (p<.05).

**Were these results due to balancer and rater characteristics?**

To ensure that the results reported above were not due to the order in which the raters viewed the sets of videos, or the characteristics of the balancers or raters, we repeated the same experiment in three additional raters with a mean±SD age of 26.47±2.65 years and a mean±SD years of experience of 2.74±1.66,and also provided prior informed consent. These additional raters viewed other PowerPoint presentations containing 32 sets of videos that were different from those used in the above results. ICC values and CIs from these additional raters are shown in Table 7A-E. Note the similarities between this table and Tables 2-6, suggesting that the "good"[11] reliability of the mBESS in this study was not due to certain balancer characteristics.

## DISCUSSION

The purpose of this study was to measure the inter- and intra-rater reliability of a modified approach to scoring the balance error scoring system (BESS) for undistracted and distracted conditions, as measured by intraclass correlation coefficient (ICC) values. Although we hypothesized "poor to moderate"[11] reliability for the distracted condition, we found that the modified approach had "good"[11] inter- and intra-rater reliability, regardless of condition. General guidelines for interpreting ICC values indicate that values >0.75 as having "good"[11] reliability, and those below 0.75 as having "poor to moderate"[11] reliability. Moreover, the raters' reported number of errors was not significantly different between the undistracted and distracted conditions. Thus, this modified approach to scoring the BESS is reliable, even when used in a potentially distracting environment.

The presence of distractions has been shown to affect balance performance. Numerous studies have reported changes in postural sway,[23,24] fall risk,[25] and postural responses to perturbations[26] under dual-task conditions. Even performance on the BESS itself has been shown to decline in situations when the balancer is susceptible to distraction.[12] Typically, if a distraction is present enough to disrupt the balancer, it could also disrupt the rater. Given that the reliability and the raters' performance were comparable, however, between undistracted and distracted conditions, our findings suggest that modified approach to rating is a reliable balance assessment regardless of whether a distraction is present or not.

When rating the pre-recorded sets of videos at Time 1 and at Time 2, the raters maintained "good"[11] inter-rater ICC values for the total score and for all but one stance position: single-leg, foam at Time 1. The ICC value for this stance position and time point fell within the "poor to moderate"[11] range as 0.68, but by Time 2 was categorized as "good"[11] with a value of 0.82. This improvement in inter-rater reliability of the single-leg, foam stance position from Time 1 to Time 2 may be a) an effect of learning within raters or b) an effect of stance position difficulty within balancers, as described below.

Learning effects have been documented for the original BESS approach, such that balancers improve their total scores the more times they complete the BESS.[6,27,28] Balancing on the foam surface is particularly susceptible to such learning effects. Even in healthy adults, learning to stand on the foam surface requires several consecutive trials that are longer than 20 seconds.[29] Our results for the foam surface are likely not attributable to this balancer learning effect, however, as all balancers in this study completed the BESS only two times, separated by one week. The change in reliability may have instead been due to rater learning between Time 1 and Time 2, yet all raters in this study had two or more years of clinical experience and had used this modified approach in a clinical setting previously. Moreover, if there was a rater learning effect, then all stance positions should have shown an increase in reliability, rather than only one. Future studies are needed, however, to identify the minimum number of times a clinician needs to have administered this modified BESS approach, as well as any other rated assessment, to reliably use it.

Alternatively, the lower ICC values for the single-leg, foam stance position may have been due to a higher number of errors committed by the balancer (and subsequently reported by the rater). Findings from this study are consistent with many others that report significantly higher error rates in the single-leg, foam position compared to all other positions in healthy adults[7,10,30-33] and concussed individuals.[32] This suggests that the balancers who were video-recorded in this study completed the BESS using similar balance strategies to those in other studies. Regardless, though, if the more ‘difficult’ stance positions of the BESS (i.e. on foam or in single and tandem stances) influenced the reliability of modified BESS approach, then the ICC values should scale inversely with stance difficulty. The tables of results described earlier do not consistently show this trend, however, with the ICC values often being even higher for the foam surface relative to the firm surface.

Although our results suggest that the modified BESS approach is reliable even when the rater is performing a concurrent task that may be distracting, participants in this study who served as raters were rating pre-recorded videos of healthy adult balancers who were free of musculoskeletal injury and concussion history. Future studies are needed to determine if the “good”[11] reliability of this modified BESS approach is preserved when the balancer is injured or concussed.

**REFERENCES**


1. Linens SW, Ross SE, Arnold BL, Gayle R, Pidcoe P. Postural-Stability Tests That Identify Individuals With Chronic Ankle Instability. *J Athl Train.* 2014;49(1):15-23.
2. Leddy AL, Crowner BE, Earhart GM. Functional Gait Assessment and Balance Evaluation System Test: Reliability, Validity, Sensitivity, and Specificity for Identifying Individuals With Parkinson Disease Who Fall. *Physical Therapy.* Jan 2011;91(1):102-113.
3. McCrory P, Meeuwisse WH, Aubry M, et al. Consensus statement on concussion in sport--the 4th International Conference on Concussion in Sport held in Zurich, November 2012. *PM R.* Apr 2013;5(4):255-279.
4. McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on concussion in sport: the 3rd International Conference on Concussion in Sport held in Zurich, November 2008. *J Athl Train.* Jul-Aug 2009;44(4):434-448.
5. Guskiewicz KM, Bruce SL, Cantu RC, et al. National Athletic Trainers' Association Position Statement: Management of Sport-Related Concussion. *J Athl Train.* Sep 2004;39(3):280-297.
6. Valovich TC, Perrin DH, Gansneder BM. Repeat Administration Elicits a Practice Effect With the Balance Error Scoring System but Not With the Standardized Assessment of Concussion in High School Athletes. *J Athl Train.* Mar 2003;38(1):51-56.
7. Riemann BL, Guskiewicz KM, Shields EW. Relationship between clinical and forceplate measures of postural stability. *J Sport Rehabil.* May 1999;8(2):71-82.
8. Finnoff JT, Peterson, V.J., Hollman, J.H., & Smith, J. Intrarrater and interrater reliability of the balance error scoring system (BESS). *PM R.* 2009 2009;1:50-54.
9. Valovich McLeod TC, Barr WB, McCrea M, Guskiewicz KM. Psychometric and measurement properties of concussion assessment tools in youth sports. *J Athl Train.* Oct-Dec 2006;41(4):399-408.

10. Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error scoring system. *Sports Health.* May 2011;3(3):287-295.

11. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* 3rd ed. Upper Saddle River, New Jersey: Pearson Education, Inc.; 2009.

12. Onate JA, Beck BC, Van Lunen BL. On-field testing environment and balance error scoring system performance during preseason screening of healthy collegiate baseball players. *J Athl Train.* Oct-Dec 2007;42(4):446-451.

13. Guskiewicz KM, Register-Mihalik J, McCrory P, et al. Evidence-based approach to revising the SCAT2: introducing the SCAT3. *Br J Sports Med.* Apr 2013;47(5):289-293.

14. Schaefer SY, Lang CE. Using dual tasks to test immediate transfer of training between naturalistic movements: a proof-of-principle study. *J Mot Behav.* 2012;44(5):313-327.

15. Lang CE, Bastian AJ. Cerebellar damage impairs automaticity of a recently practiced movement. *J Neurophysiol.* Mar 2002;87(3):1336-1347.

16. Schaefer SY, Patterson CB, Lang CE. Transfer of training between distinct motor tasks after stroke: implications for task-specific approaches to upper-extremity neurorehabilitation. *Neurorehabil Neural Repair.* Sep 2013;27(7):602-612.

17. Bakan P. Extraversion-introversion and improvement in an auditory vigilance task. *Br J Psychol.* Nov 1959;50:325-332.

18. Beck LH, Bransome ED, Jr., Mirsky AF, Rosvold HE, Sarason I. A continuous performance test of brain damage. *J Consult Psychol.* Oct 1956;20(5):343-350.

19. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol.* 2012;8(1):23-34.

20. Kramer CY. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics.* 1956;12:307-312.

**21.** Stoline MR. The Status of Multiple Comparisons - Simultaneous Estimation of All Pairwise Comparisons in One-Way Anova Designs. *American Statistician.* 1981;35(3):134-141.

**22.** Kackar RN, Harville DA. Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear-Models. *Journal of the American Statistical Association.* 1984;79(388):853-862.

**23.** Shumway-Cook A, Woollacott M, Kerns KA, Baldwin M. The effects of two types of cognitive tasks on postural stability in older adults with and without a history of falls. *J Gerontol A Biol Sci Med Sci.* Jul 1997;52(4):M232-240.

**24.** Woollacott M, Shumway-Cook A. Attention and the control of posture and gait: a review of an emerging area of research. *Gait Posture.* Aug 2002;16(1):1-14.

**25.** Toulotte C, Thevenon A, Watelain E, Fabre C. Identification of healthy elderly fallers and non-fallers by gait analysis under dual-task conditions. *Clin Rehabil.* Mar 2006;20(3):269-276.

**26.** Brown LA, Shumway-Cook A, Woollacott MH. Attentional demands and postural recovery: the effects of aging. *J Gerontol A Biol Sci Med Sci.* Apr 1999;54(4):M165-171.

**27.** Mulligan IJ, Boland, M.A., McIlhenny, C.V. The balance error scoring system learned response among young adults. *Sports Health.* 2013;5:22-26.

**28.** Valovich McLeod TC, Perrin, D. H., Guskiewicz, K. M., Schultz, S. J., Diamond, R., Gansneder, B. M. Serial administration of clinical concussion assessments and learning effects in healthy young athletes. *Clin J Sport Med.* 2004;14:287-295.

**29.** Pagnacco G, Carrick FR, Pascolo PB, Rossi R, Oggero E. Learning effect of standing on foam during posturographic testing preliminary findings. *Biomed Sci Instrum.* 2012;48:332-339.

**30.** Hunt TN, Ferrara MS, Bornstein RA, Baumgartner TA. The reliability of the modified Balance Error Scoring System. *Clin J Sport Med.* Nov 2009;19(6):471-475.

**31.** Riemann BL, Schmitz R. The relationship between various modes of single leg postural control assessment. *Int J Sports Phys Ther.* Jun 2012;7(3):257-266.

**32.** Riemann BL, Guskiewicz KM. Effects of mild head injury on postural stability as measured through clinical balance testing. *J Athl Train.* Jan 2000;35(1):19-25.

**33.** Wilkins JC, Valovich McLeod TC, Perrin DH, Gansneder BM. Performance on the Balance Error Scoring System Decreases After Fatigue. *J Athl Train.* Jun 2004;39(2):156-161.

**LEGEND TO FIGURES**

Figure 1. Mean reported number of errors in undistracted and distracted conditions for (A) total score and (B) each stance position. Error bars indicate standard error.

**Table 1. Criteria for Scoring Errors using the Original and Modified BESS criteria.**

| Criteria for scoring an error | Original | Modified |
|---|---|---|
| Lifts hands off of iliac crests. | X | X |
| Opens eyes | X | X |
| Steps/Stumbles/Falls | X | X |
| Takes >5 seconds to return to the testing position | X | X |
| Lifts forefoot/heel | X | |
| Abducts/Flexes hip >30° | X | |

**Table 2. Undistracted Inter-Rater Reliability ICC's and CI's; Time 1**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.96 | 0.926-0.985 |
| Tandem Firm | 0.89 | 0.805-0.955 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.68 | 0.489-0.849 |
| Tandem Foam | 0.93 | 0.869-0.972 |
| Total Score | 0.93 | 0.845-0.964 |

[a]Both double leg stance ICCs and CIs could not be calculated due to a lack of errors.

**Table 3. Undistracted Inter-Rater Reliability ICC's and CI's; Time 2**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.95 | 0.911-0.981 |
| Tandem Firm | 0.85 | 0.736-0.936 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.82 | 0.690-0.923 |
| Tandem Foam | 0.89 | 0.797-0.953 |
| Total Score | 0.93 | 0.847-0.970 |

[a]Both double leg stance ICCs and CIs could not be calculated due to a lack of errors.

**Table 4. Undistracted Intra-Rater Reliability ICC's and CI's**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.96 | 0.923-0.982 |
| Tandem Firm | 0.88 | 0.789-0.947 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.76 | 0.614-0.886 |
| Tandem Foam | 0.89 | 0.808-0.952 |
| Total Score | 0.92 | 0.855-0.965 |

[a]Both double leg stance ICCs and CIs could not be calculated due to a lack of errors.

**Table 5. Distracted Inter-Rater Reliability ICC's and CI's**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.86 | 0.741-0.938 |
| Tandem Firm | 0.90 | 0.816-0.958 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.56 | 0.344-0.780 |
| Tandem Foam | 0.91 | 0.834-0.965 |
| Total Score | 0.89 | 0.769-0.958 |

[a]Both double leg stance ICCs and CIs could not be calculated due to a lack of errors.

**Table 6. Undistracted vs. Distracted Intra-Rater Reliability ICC's and CI's**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.89 | 0.811-0.953 |
| Tandem Firm | 0.91 | 0.832-0.959 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.64 | 0.465-0.818 |
| Tandem Foam | 0.92 | 0.862-0.968 |
| Total Score | 0.92 | 0.845-0.964 |

[a]Both double leg stance ICCs and CIs could not be calculated due to a lack of errors.

**Table 7.**

**A. Undistracted Inter-Rater Reliability ICC's and CI's; Time 1**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.88 | 0.753-0.953 |
| Tandem Firm | 0.64 | 0.368-0.837 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.81 | 0.594-0.922 |
| Tandem Foam | 0.89 | 0.729-0.959 |
| Total Score | 0.89 | 0.772-0.958 |

**B. Undistracted Inter-Rater Reliability ICC's and CI's; Time 2**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.90 | 0.772-0.961 |
| Tandem Firm | 0.92 | 0.826-0.968 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.70 | 0.424-0.874 |
| Tandem Foam | 0.76 | 0.540-0.898 |
| Total Score | 0.88 | 0.658-0.956 |

**C. Undistracted Intra-Rater Reliability ICC's and CI's**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.90 | 0.805-0.956 |
| Tandem Firm | 0.79 | 0.647-0.907 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.73 | 0.545-0.878 |
| Tandem Foam | 0.82 | 0.679-0.921 |
| Total Score | 0.89 | 0.777-0.952 |

**D. Distracted Inter-Rater Reliability ICC's and CI's**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.79 | 0.582-0.914 |
| Tandem Firm | 0.66 | 0.392-0.848 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.43 | 0.072-0.733 |
| Tandem Foam | 0.94 | 0.858-0.975 |
| Total Score | 0.80 | 0.400-0.930 |

**E. Undistracted vs. Distracted Intra-Rater Reliability ICC's and CI's**

| Stance Position | ICC | CI |
|---|---|---|
| Double Leg Firm | 0.00[a] | 0.000[a] |
| Single Leg Firm | 0.86 | 0.743-0.939 |
| Tandem Firm | 0.79 | 0.637-0.904 |
| Double Leg Foam | 0.00[a] | 0.000[a] |
| Single Leg Foam | 0.55 | 0.294-0.776 |
| Tandem Foam | 0.91 | 0.823-0.960 |
| Total Score | 0.85 | 0.685-0.939 |

[a]Both double leg stance ICCs and CIs could not be calculated due to a lack of errors.

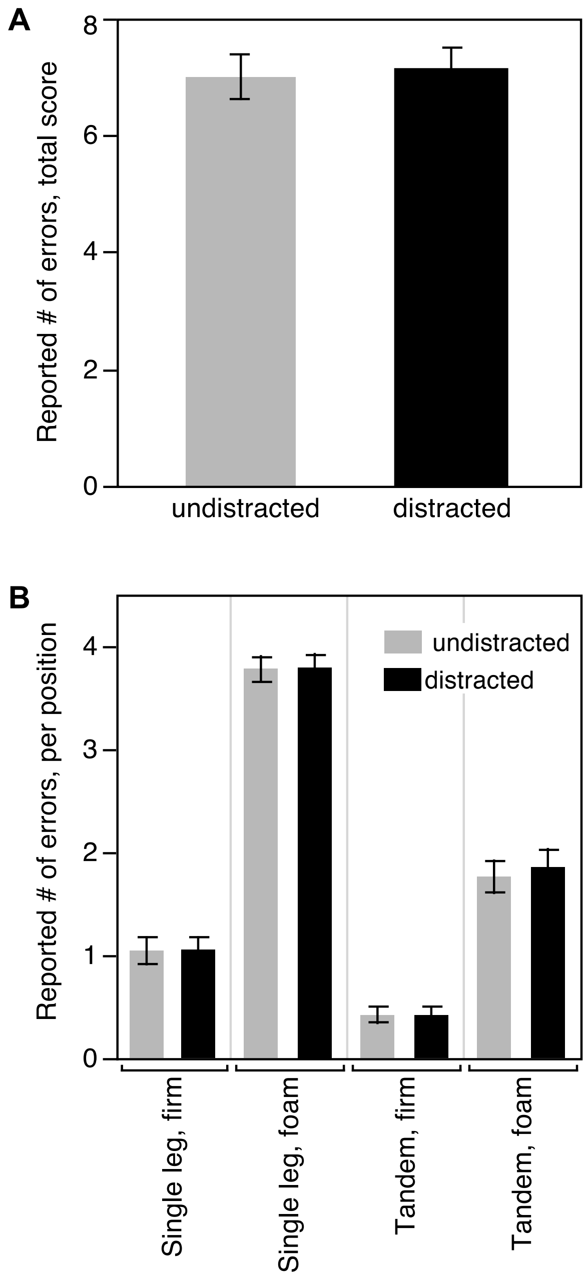


**Figure 1. Mean reported number of errors in undistracted and distracted conditions for (A) total score and (B) each stance position. Error bars indicate standard error.**