

Utah State University

DigitalCommons@USU

All Graduate Plan B and other Reports

Graduate Studies

5-2014

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

Jacob Ryan Pierce
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/gradreports>



Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Pierce, Jacob Ryan, "Inter- and Intra-rater Reliability of a Modified Approach to Scoring the Balance Error Scoring System BESS in Undistracted and Distracted Conditions" (2014). *All Graduate Plan B and other Reports*. 362.

<https://digitalcommons.usu.edu/gradreports/362>

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



5-2014

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

Jacob Ryan Pierce
Utah State University

Follow this and additional works at: <http://digitalcommons.usu.edu/gradreports>

 Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Pierce, Jacob Ryan, "Inter- and Intra-rater Reliability of a Modified Approach to Scoring the Balance Error Scoring System (BESS) in Undistracted and Distracted Conditions" (2014). *All Graduate Plan B and other Reports*. Paper 362.

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact becky.thoms@usu.edu.



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}

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

1 **ABSTRACT**

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

4 **Objective:** To measure the inter- and intra-rater reliability of this modified BESS approach.

5 Because of the potential auditory distractions that may be present in a clinical setting, we also
6 measured reliability for distracted rating conditions.

7 **Design:** Cross-sectional.

8 **Setting:** University sports medicine clinic.

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

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

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

20 **Results:** For the undistracted condition, inter- and intra-rater reliability ICCs were 0.93 and
21 0.92, respectively. For the distracted condition, inter- and intra-rater reliability ICCs were 0.89
22 and 0.92, respectively. There was no significant effect of condition (undistracted vs. distracted)
23 on the reported number of errors in the total error score or in the individual stance position
24 scores.

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

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

28

29 INTRODUCTION

30 Reliable balance assessments are necessary in clinical settings to document
31 musculoskeletal injury recovery,¹ determine fall risk,² and decide post-concussion return-to-
32 play³. Although there are a variety of balance assessment tools, athletic trainers may use the
33 Balance Error Scoring System (BESS).³⁻⁶ The BESS includes three stances (double-leg, single-
34 leg, and tandem) on two separate surfaces (firm ground and foam), yielding six different stance
35 positions.⁷ This test requires individuals to stand in each position for 20 seconds with their eyes
36 closed and their hands on their hips. To assess an individual's balance using the BESS, a
37 clinician counts the number of errors the individual makes in each stance position. There are six
38 criteria for what constitutes a "balance error" in the original version of the BESS: 1) lifting hands
39 off of hips; 2) opening eyes; 3) stepping, stumbling, or falling out of position; 4) taking >5
40 seconds to return to position; 5) lifting forefoot or heel; or 6) abducting/flexing hip joint >30
41 degrees. The clinician records the number of errors made in each stance position, and then
42 sums each position to yield a total BESS score. The original BESS is valid, can be performed
43 quickly in a small space, requires little equipment, and is inexpensive.⁷ It has been shown,
44 however, to have "poor to moderate" inter- and intra-rater reliability,⁸⁻¹⁰ according to general
45 guidelines.¹¹ One possible explanation for the "poor to moderate"¹¹ reliability of the original
46 BESS is the number of possible error criteria.⁸

47 We have modified the criteria for counting balance errors within the BESS criteria by
48 removing two of the six original error criteria. This modified approach (mBESS) aims to increase
49 the reliability of the original BESS, based on the assumption that clinicians are not highly
50 reliable in visually confirming whether a balancer has lifted his/her forefoot or heel, or has
51 abducted or flexed the hip joint more than 30 degrees. Thus, the mBESS counts only the first
52 four criteria from the BESS as an error, and does not count the last two described above. By
53 minimizing these possible error counts, clinicians using this modified approach may be more
54 reliable at detecting changes in balance performance. Regardless of which approach is used,

55 however, clinicians typically rate the BESS in environments that contain a number of potential
56 auditory distractions, such as a busy athletic training room or a noisy sideline. Distraction can
57 affect balance performance during the BESS,¹² but less is known about the extent to which
58 distraction affects rating ability. If a clinician is required to attend to another task at the same
59 time as rating the BESS, his/her ability to reliably rate balance errors may decline. Thus, the
60 purpose of this study was to measure the inter- and intra-rater reliability of a modified approach
61 to scoring the balance error scoring system (BESS) in undistracted and distracted conditions.
62 We hypothesized that both the inter- and intra-rater reliability of this modified approach would be
63 “good” for the undistracted condition, but only “poor to moderate” for the distracted condition.
64 We also hypothesized that the reported number of errors would be different for the undistracted
65 condition compared to the distracted condition.

66

67 **METHODS**

68 **Participants**

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

80

81 **Experimental design**

82 *General methods*

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

95 Prior to rater recruitment for this study, we recruited 32 'balancers' who were free from a
96 lower extremity injury in the previous year, and had no history of any physician-diagnosed
97 concussion. Balancers provided informed consent for their participation in this study. Balancers
98 were then video-recorded in the frontal plane as they completed all six stance positions of the
99 BESS: double-leg, firm (ground); single-leg, firm; tandem, firm; double-leg, foam; single-leg,
100 foam; tandem, foam. Each stance position was held for 20 seconds. Additional details regarding
101 these conditions and the test itself have been published previously.^{7,13} All videos were archived
102 by balancer and by stance position, such that each balancer yielded six separate 20-sec video
103 files, which equaled one set of videos. For reasons beyond the scope of this study, the 32
104 balancers were video-recorded again a week later in the same fashion, yielding a total of 64
105 sets of videos. A random selection of 32 sets were then inserted offline by the experimenter into
106 PowerPoint presentations for the raters to view, with each 20-sec video separated by a blank

107 slide to allow the raters time to record the corresponding error score on a paper data collection
108 sheet. The composition of video sets within the PowerPoint presentations enabled the testing of
109 inter- and intra-rater reliability of the mBESS approach.

110

111 *Time 1*

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

117

118 *Time 2*

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

124

125 *Distracted rating conditions with an auditory vigilance task*

126 We used an auditory vigilance task to potentially distract the raters as they used the
127 mBESS to score balance errors. Auditory vigilance has been shown previously to significantly
128 distract healthy young adults¹⁴ and adults with stroke^{15,16} as they performed a concurrent motor
129 task (i.e. under dual-task conditions), regardless of general intelligence level.^{17,18} The auditory
130 vigilance task in this study required raters to listen to 20-sec recordings of 35-letter sequences.
131 Each 35-letter sequence consisted of a random series of the same four letters (A, G, M, and O).
132 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.

146

147 **Data analysis**

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

154

155 **Statistical analysis**

156 To measure the inter- and intra-rater reliability of the mBESS approach for both
157 undistracted and distracted conditions, we used two-way, random, absolute agreement
158 intraclass correlation coefficients (ICC_{2,1})^{11,19} All ICC values and confidence intervals were

159 determined using SPSS software version 21 (IBM, Armonk, NY). Inter-rater reliability for
160 undistracted conditions was computed across video sets 1-16 in Time 1 and Time 2. Intra-rater
161 reliability for undistracted conditions was computed within raters across Time 1 and Time 2,
162 using only the error reports from the 16 sets of videos viewed without distraction. Inter-rater
163 reliability for distracted conditions was computed across raters in Time 2, using only the error
164 reports from the 16 sets of videos viewed while simultaneously performing the auditory vigilance
165 task. Intra-rater reliability for distracted conditions was computed within raters using the 16 sets
166 of videos from Time 1 (undistracted) that corresponded with the same video sets rated under
167 distracted conditions from Time 2. All ICC values are reported as 0.00-1.00, with larger values
168 indicating higher reliability. Reliabilities were computed for all six separate BESS stance
169 positions as well as total composite scores.

170 To test whether the reported number of errors was different in the undistracted condition
171 compared to the distracted condition, we compared mean total composite scores across raters
172 using a one-way repeated measures ANOVA with condition (undistracted vs. distracted) as a
173 within-subject factor. Based on our hypothesis, we expected no main effect of condition on the
174 reported number of errors. To test whether the auditory vigilance task affected the raters'
175 scoring of a particular stance position compared to another position, we used a 2x4 repeated
176 measures ANOVA with condition (undistracted vs. distracted) and stance position (single-leg,
177 firm vs. tandem, firm vs. single-leg, foam vs. tandem, foam) as within-subject factors. While
178 there are six separate stance positions in the BESS protocol,⁷ the two double-leg stance
179 positions (firm and foam) yielded zero reported errors (see Results). Thus, only the four error-
180 inducing stance positions were included in our ANOVA. We expected significant main effects of
181 condition and stance position on the reported number of errors, yet no interaction between
182 condition and stance position. Significant effects were tested post hoc using Tukey-Kramer
183 Honestly Significant Different (HSD) tests.^{20,21} Multiple comparisons were accounted for using

184 the Kackar-Harville correction.²² All ANOVAs and subsequent posthoc analyses were performed
185 with JMP 8.0 (SAS Institute Inc., Carey, NC), with an alpha level of .05.

186

187 **RESULTS**

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

189 *Inter-rater reliability*

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

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

201

202 *Intra-rater reliability*

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

209

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

211 *Inter-rater reliability*

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

218

219 *Intra-rater reliability*

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

226

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

229 Figure 1A shows the reported number of errors for the total BESS score in undistracted
230 and distracted conditions. Our one-way repeated measures ANOVA reported no significant
231 effect of condition (undistracted vs. distracted) on the reported number of errors in the total error
232 score ($F_{1,5} = .045$; $p=.83$). Given that ICC values varied by BESS stance position (see Tables 2-
233 6), it is plausible that different stance positions could be differentially affected when the rater
234 concurrently performed the auditory vigilance task. Our 2x4 repeated-measures ANOVA
235 reported no significant main effect of condition ($F_{1,5} = .101$; $p=.75$) nor an interaction between

236 condition and stance position ($F_{3,5}=.057$; $p=.98$) on the reported number of errors. There was,
237 however, a significant main effect of stance position on reported number of errors (F
238 $_{3,5}=262.831$; $p<.0001$; Fig. 1B). Post-hoc analyses revealed that the reported number of errors in
239 all error-inducing stance positions (single leg, firm; tandem, firm; single leg, foam; tandem,
240 foam) were significantly different from each other ($p<.05$).

241

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

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

252

253 **DISCUSSION**

254 The purpose of this study was to measure the inter- and intra-rater reliability of a
255 modified approach to scoring the balance error scoring system (BESS) for undistracted and
256 distracted conditions, as measured by intraclass correlation coefficient (ICC) values. Although
257 we hypothesized “poor to moderate”¹¹ reliability for the distracted condition, we found that the
258 modified approach had “good”¹¹ inter- and intra-rater reliability, regardless of condition. General
259 guidelines for interpreting ICC values indicate that values >0.75 as having “good”¹¹ reliability,
260 and those below 0.75 as having “poor to moderate”¹¹ reliability. Moreover, the raters’ reported
261 number of errors was not significantly different between the undistracted and distracted

262 conditions. Thus, this modified approach to scoring the BESS is reliable, even when used in a
263 potentially distracting environment.

264 The presence of distractions has been shown to affect balance performance. Numerous
265 studies have reported changes in postural sway,^{23,24} fall risk,²⁵ and postural responses to
266 perturbations²⁶ under dual-task conditions. Even performance on the BESS itself has been
267 shown to decline in situations when the balancer is susceptible to distraction.¹² Typically, if a
268 distraction is present enough to disrupt the balancer, it could also disrupt the rater. Given that
269 the reliability and the raters' performance were comparable, however, between undistracted and
270 distracted conditions, our findings suggest that modified approach to rating is a reliable balance
271 assessment regardless of whether a distraction is present or not.

272 When rating the pre-recorded sets of videos at Time 1 and at Time 2, the raters
273 maintained "good"¹¹ inter-rater ICC values for the total score and for all but one stance position:
274 single-leg, foam at Time 1. The ICC value for this stance position and time point fell within the
275 "poor to moderate"¹¹ range as 0.68, but by Time 2 was categorized as "good"¹¹ with a value of
276 0.82. This improvement in inter-rater reliability of the single-leg, foam stance position from Time
277 1 to Time 2 may be a) an effect of learning within raters or b) an effect of stance position
278 difficulty within balancers, as described below.

279 Learning effects have been documented for the original BESS approach, such that
280 balancers improve their total scores the more times they complete the BESS.^{6,27,28} Balancing on
281 the foam surface is particularly susceptible to such learning effects. Even in healthy adults,
282 learning to stand on the foam surface requires several consecutive trials that are longer than 20
283 seconds.²⁹ Our results for the foam surface are likely not attributable to this balancer learning
284 effect, however, as all balancers in this study completed the BESS only two times, separated by
285 one week. The change in reliability may have instead been due to rater learning between Time
286 1 and Time 2, yet all raters in this study had two or more years of clinical experience and had
287 used this modified approach in a clinical setting previously. Moreover, if there was a rater

288 learning effect, then all stance positions should have shown an increase in reliability, rather than
289 only one. Future studies are needed, however, to identify the minimum number of times a
290 clinician needs to have administered this modified BESS approach, as well as any other rated
291 assessment, to reliably use it.

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

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

309

310 **REFERENCES**

- 311 **1.** Linens SW, Ross SE, Arnold BL, Gayle R, Pidcoe P. Postural-Stability Tests That
312 Identify Individuals With Chronic Ankle Instability. *J Athl Train.* 2014;49(1):15-23.
- 313 **2.** Leddy AL, Crowner BE, Earhart GM. Functional Gait Assessment and Balance
314 Evaluation System Test: Reliability, Validity, Sensitivity, and Specificity for Identifying
315 Individuals With Parkinson Disease Who Fall. *Physical Therapy.* Jan 2011;91(1):102-
316 113.
- 317 **3.** McCrory P, Meeuwisse WH, Aubry M, et al. Consensus statement on concussion in
318 sport--the 4th International Conference on Concussion in Sport held in Zurich, November
319 2012. *PM R.* Apr 2013;5(4):255-279.
- 320 **4.** McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on concussion in
321 sport: the 3rd International Conference on Concussion in Sport held in Zurich, November
322 2008. *J Athl Train.* Jul-Aug 2009;44(4):434-448.
- 323 **5.** Guskiewicz KM, Bruce SL, Cantu RC, et al. National Athletic Trainers' Association
324 Position Statement: Management of Sport-Related Concussion. *J Athl Train.* Sep
325 2004;39(3):280-297.
- 326 **6.** Valovich TC, Perrin DH, Gansneder BM. Repeat Administration Elicits a Practice Effect
327 With the Balance Error Scoring System but Not With the Standardized Assessment of
328 Concussion in High School Athletes. *J Athl Train.* Mar 2003;38(1):51-56.
- 329 **7.** Riemann BL, Guskiewicz KM, Shields EW. Relationship between clinical and forceplate
330 measures of postural stability. *J Sport Rehabil.* May 1999;8(2):71-82.
- 331 **8.** Finnoff JT, Peterson, V.J., Hollman, J.H., & Smith, J. Intrarrater and interrater reliability
332 of the balance error scoring system (BESS). *PM R.* 2009 2009;1:50-54.
- 333 **9.** Valovich McLeod TC, Barr WB, McCrea M, Guskiewicz KM. Psychometric and
334 measurement properties of concussion assessment tools in youth sports. *J Athl Train.*
335 Oct-Dec 2006;41(4):399-408.

- 336 **10.** Bell DR, Guskiewicz KM, Clark MA, Padua DA. Systematic review of the balance error
337 scoring system. *Sports Health*. May 2011;3(3):287-295.
- 338 **11.** Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd
339 ed. Upper Saddle River, New Jersey: Pearson Education, Inc.; 2009.
- 340 **12.** Onate JA, Beck BC, Van Lunen BL. On-field testing environment and balance error
341 scoring system performance during preseason screening of healthy collegiate baseball
342 players. *J Athl Train*. Oct-Dec 2007;42(4):446-451.
- 343 **13.** Guskiewicz KM, Register-Mihalik J, McCrory P, et al. Evidence-based approach to
344 revising the SCAT2: introducing the SCAT3. *Br J Sports Med*. Apr 2013;47(5):289-293.
- 345 **14.** Schaefer SY, Lang CE. Using dual tasks to test immediate transfer of training between
346 naturalistic movements: a proof-of-principle study. *J Mot Behav*. 2012;44(5):313-327.
- 347 **15.** Lang CE, Bastian AJ. Cerebellar damage impairs automaticity of a recently practiced
348 movement. *J Neurophysiol*. Mar 2002;87(3):1336-1347.
- 349 **16.** Schaefer SY, Patterson CB, Lang CE. Transfer of training between distinct motor tasks
350 after stroke: implications for task-specific approaches to upper-extremity
351 neurorehabilitation. *Neurorehabil Neural Repair*. Sep 2013;27(7):602-612.
- 352 **17.** Bakan P. Extraversion-introversion and improvement in an auditory vigilance task. *Br J*
353 *Psychol*. Nov 1959;50:325-332.
- 354 **18.** Beck LH, Bransome ED, Jr., Mirsky AF, Rosvold HE, Sarason I. A continuous
355 performance test of brain damage. *J Consult Psychol*. Oct 1956;20(5):343-350.
- 356 **19.** Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and
357 Tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34.
- 358 **20.** Kramer CY. Extension of multiple range tests to group means with unequal numbers of
359 replications. *Biometrics*. 1956;12:307-312.

- 360 **21.** Stolone MR. The Status of Multiple Comparisons - Simultaneous Estimation of All
361 Pairwise Comparisons in One-Way Anova Designs. *American Statistician*.
362 1981;35(3):134-141.
- 363 **22.** Kackar RN, Harville DA. Approximations for Standard Errors of Estimators of Fixed and
364 Random Effects in Mixed Linear-Models. *Journal of the American Statistical Association*.
365 1984;79(388):853-862.
- 366 **23.** Shumway-Cook A, Woollacott M, Kerns KA, Baldwin M. The effects of two types of
367 cognitive tasks on postural stability in older adults with and without a history of falls. *J*
368 *Gerontol A Biol Sci Med Sci*. Jul 1997;52(4):M232-240.
- 369 **24.** Woollacott M, Shumway-Cook A. Attention and the control of posture and gait: a review
370 of an emerging area of research. *Gait Posture*. Aug 2002;16(1):1-14.
- 371 **25.** Toulotte C, Thevenon A, Watelain E, Fabre C. Identification of healthy elderly fallers and
372 non-fallers by gait analysis under dual-task conditions. *Clin Rehabil*. Mar
373 2006;20(3):269-276.
- 374 **26.** Brown LA, Shumway-Cook A, Woollacott MH. Attentional demands and postural
375 recovery: the effects of aging. *J Gerontol A Biol Sci Med Sci*. Apr 1999;54(4):M165-171.
- 376 **27.** Mulligan IJ, Boland, M.A., McIlhenny, C.V. The balance error scoring system learned
377 response among young adults. *Sports Health*. 2013;5:22-26.
- 378 **28.** Valovich McLeod TC, Perrin, D. H., Guskiewicz, K. M., Schultz, S. J., Diamond, R.,
379 Gansneder, B. M. Serial administration of clinical concussion assessments and learning
380 effects in healthy young athletes. *Clin J Sport Med*. 2004;14:287-295.
- 381 **29.** Pagnacco G, Carrick FR, Pascolo PB, Rossi R, Oggero E. Learning effect of standing on
382 foam during posturographic testing preliminary findings. *Biomed Sci Instrum*.
383 2012;48:332-339.
- 384 **30.** Hunt TN, Ferrara MS, Bornstein RA, Baumgartner TA. The reliability of the modified
385 Balance Error Scoring System. *Clin J Sport Med*. Nov 2009;19(6):471-475.

- 386 **31.** Riemann BL, Schmitz R. The relationship between various modes of single leg postural
387 control assessment. *Int J Sports Phys Ther.* Jun 2012;7(3):257-266.
- 388 **32.** Riemann BL, Guskiewicz KM. Effects of mild head injury on postural stability as
389 measured through clinical balance testing. *J Athl Train.* Jan 2000;35(1):19-25.
- 390 **33.** Wilkins JC, Valovich McLeod TC, Perrin DH, Gansneder BM. Performance on the
391 Balance Error Scoring System Decreases After Fatigue. *J Athl Train.* Jun
392 2004;39(2):156-161.
- 393
- 394
- 395

396 **LEGEND TO FIGURES**

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

399

400 **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	

401

402 **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

403 ^aBoth double leg stance ICCs and CIs could not be calculated due to a lack of errors.

404

405

406 **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

407 ^aBoth double leg stance ICCs and CIs could not be calculated due to a lack of errors.

408

409

410 **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

411 ^aBoth double leg stance ICCs and CIs could not be calculated due to a lack of errors.

412

413

414

415

416 **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

417 ^aBoth double leg stance ICCs and CIs could not be calculated due to a lack of errors.

418

419

420 **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

421 ^aBoth double leg stance ICCs and CIs could not be calculated due to a lack of errors.

422

423

424

425

426

427

428

429

430

431

432 **Table 7.**

433 **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

434 **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

435 **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

436 **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

437 **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

438 ^aBoth double leg stance ICCs and CIs could not be calculated due to a lack of errors.

439

440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480

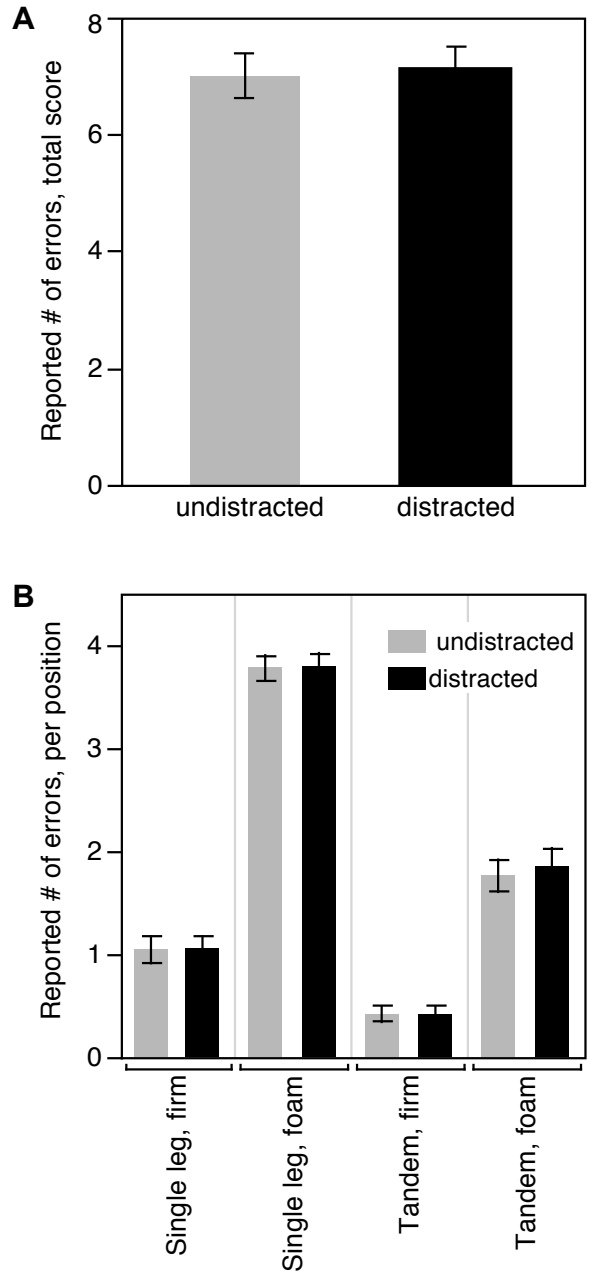


Figure 1. Mean reported number of errors in undistracted and distracted conditions for (A) total score and (B) each stance position. Error bars indicate standard error.