The Reliability of the Modified Balance Error Scoring System

Tamerah N. Hunt, PhD, ATC, CSCS, Michael S. Ferrara, PhD, ATC,† Robert A. Bornstein, PhD, ABPP,‡ and Ted A. Baumgartner, PhD†*

Objective: Study 1 investigated the intraclass reliability and percent variance associated with each component within the traditional Balance Error Scoring System (BESS) protocol. Study 2 investigated the reliability of subsequent modifications of the BESS.

Design: Prospective cross-sectional examination of the traditional and modified BESS protocols.

Setting: Schools participating in Georgia High School Athletics Association.

Intervention: The modified BESS consisted of 2 surfaces (firm and foam) and 2 stances (single-leg and tandem-leg stance) repeated for a total of three 20-second trials.

Participants: Participants consisted of 2 independent samples of high school athletes aged 13 to 19 years.

Main Outcome Measures: Percent variance for each condition of the BESS was obtained using GENOVA 3.1. An intraclass reliability coefficient and repeated measures analysis of variance were calculated using SPSS 13.0.

Results: Study 1 obtained an intraclass correlation coefficient (r = 0.60) with stance accounting for 55% of the total variance. Removing the double-leg stance increased the intraclass correlation coefficient (r = 0.71). Study 2 found a statistically significant difference between trials 1 and 2 ($F(1.65,286) = 4.890$, $P = 0.013$) and intraclass reliability coefficient of r = 0.88 for 3 trials of 4 conditions.

Conclusions: The variance associated with the double-leg stance was very small, and when removed, the intraclass reliability coefficient of the BESS increased. Removal of the double-leg stance and addition of 3 trials of 4 conditions provided an easily administered, cost-effective, time-efficient tool that provides reliable objective information for clinicians to base clinical decisions upon.

Key Words: balance, head injury, concussion, BESS, youth

(*Clin J Sport Med* 2009;19:471–475)

Submitted for publication June 5, 2009; accepted September 14, 2009.

From *School of Allied Medical Professions, The Ohio State University, Columbus, Ohio; †Department of Kinesiology, University of Georgia, Athens, Georgia; and ‡Department of Psychiatry, The Ohio State University, Columbus, Ohio.

The authors state that they have no financial interest in the products mentioned within this article.

Reprints: Tamerah N. Hunt, PhD, ATC, CSCS, Ohio State University Sports Medicine Center, The Ohio State University, 2050 Kenny Rd, Ste 3100, Columbus, OH 43221 (tamerah.hunt@osumc.edu).

INTRODUCTION

Postural stability assessment has been recognized as an important component of evaluation after concussion and can be assessed using several different methods.[1–4] Furthermore, it was confirmed by experts at the 3rd International Conference on Concussion in Sport that "postural stability testing provides a useful tool for objectively assessing the motor domain of neurologic function."[5] The Balance Error Scoring System (BESS) was initially developed as an easily administered cost-effective objective assessment tool used by clinicians for the evaluation of postural stability after concussion.

Considerable research has been conducted using the BESS.[4,6–10] The BESS has been shown to correlate well with other measures of postural stability. The first study using the BESS correlated Sensory Organization Test composite scores and sway measures to BESS scores, which resulted in moderate to high correlations.[6] In studies following concussion, subjects exhibit acute postural stability alterations up to 5 days after injury with recovery usually occurring within 4 to 7 days postinjury to preinjury baseline values.[3,4,11] Although the BESS seems to be sensitive to subtle deficits after concussion, it has several drawbacks. Recently, Valovich et al found that administration of multiple trials of the BESS results in practice effects, with the number of errors decreasing with each consecutive trial.[6] Furthermore, the effects of muscle fatigue increase the number of errors acutely, but these correct after 20 minutes of rest after an exercise session.[10]

Psychometric properties of the BESS have demonstrated limited reliability and validity data. Riemann et al[6] found that most components of the BESS correlated strongly with long force plate measures (0.78–0.96 dependent on the stance). However, the double-leg stance on the firm surface was the only condition that did not significantly correlate to long force plate measures. Intrarater reliability ranged from 0.87 to 0.98.[7] Additionally, test–retest reliability (0.673) of the BESS has been reported.[12] These test–retest reliabilities fall below commonly accepted thresholds for reliability level used for tools during clinical assessment.

Studies have not found significant difference among subjects for the double-leg stance.[3,6,8] Variance associated with the double-leg stance on both the firm and foam surfaces was low in collegiate and high school populations.[6,7,9] Riemann et al[6] did not find significant differences between injured and control participants using the double-leg stance on the foam surface. Additionally, the number of errors performed during the double-leg stance does not increase like the other stances after exertion or during fatigue.[9,10] In measures with multiple

components, it is important to understand the properties of these components and their contribution to overall reliability. This article will describe 2 studies. Study 1 investigated the intraclass reliability and the percent variance associated with each component within the standard BESS. Study 2 investigated the subsequent modifications of the BESS to improve reliability of a commonly used concussion assessment tool.

STUDY 1

Participants

Seventy-eight high school football athletes (mean age 15.78 ± 1.16 years) were tested before the fall competitive season using the standard BESS protocol. The BESS protocol and determination of errors followed protocols described in previous studies.[6–8] The University of Georgia Institutional Review Board approved study 1 and study 2. All participants and guardians read and signed university-approved informed consent and assent forms before participating in the study.

Methods

The participant was positioned 10 to 15 feet away from the tester and video camera. Each participant was videotaped and scored at a later date, which allowed for repeated and precise scoring. All tapes were scored twice by the primary investigator with an intratester reliability ($r = 0.97$).

Statistical Analysis

All analyses were completed by GENOVA, a computer program developed principally for generalizability analysis to determine the percent variance associated with each component of the BESS (GENOVA 3.1, Fortran IV, Dorchester, Massachusetts).[13,14] The components analyzed included person, surface, stance, person by stance, and person by surface by stance. The variance associated with the person by surface by stance interaction was labeled as error. Intraclass reliability coefficients were obtained using SPSS 11.3.1 (SPSS Inc, Chicago, Illinois). The Spearman–Brown prophecy formula was used to estimate reliability coefficients when the length of the test was changed. The formula is

$$\frac{r_{k,k} = k(r_{1,1})}{1 + (k-1)(r_{1,1})}$$

where $r_{k,k}$ is the theorized reliability, $r_{1,1}$ is the reliability of the present measure, and k is the number of times the measure is increased.[15]

RESULTS

As shown in Figure 1, stance accounted for more than half of the total variance within the model. The firm surface single-leg stance (29.4%) and foam surface single-leg stance (35.3%) had the highest percentage of variance, whereas the firm surface double-leg stance demonstrated few errors and had the lowest percentage of variance (0.17). The results for the other components in the model were low, whereas the error term was 35% (Table 1).

Intraclass reliability for the traditional protocol of the BESS (1 trial of 6 conditions) had a reliability of $r = 0.60$.

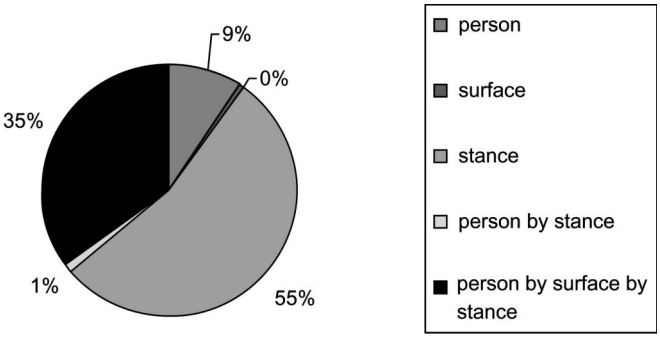

FIGURE 1. Percent variance for each component of the BESS. BESS, Balance Error Scoring System.

When the double-leg stance was removed for both the firm and foam surfaces from the model, the reliability coefficient increased to 0.71, which is still below conventionally accepted levels of reliability.

Finally, to estimate how the reliability of the BESS might be improved, we used the Spearman–Brown prophecy formula with 4 conditions (single-leg stance firm and foam surfaces and tandem-leg stance firm and foam surfaces). We found that we could increase the reliability of the measure by increasing the number of trials of these 4 conditions. As seen in Table 2, the theoretical reliability of the measure substantially increased with more trials per condition.

Study 1 again demonstrated insufficient reliability of the standard BESS protocol. Low variances associated with the double-leg stance and the poor sensitivity of the double-leg stance during baseline and postinjury evaluation in previous studies suggest that the double-leg stance could be eliminated and the number of trials per condition should be increased.[4,6] Therefore, we conducted a follow-up study to assess the reliability of a modified BESS protocol.

STUDY 2

Procedures

Participants

An independent sample of 144 high school football athletes (mean age 15.57 ± 1.15 years) were tested before the fall competitive season. Any participant who had sustained a lower extremity musculoskeletal injury and/or head injury within the 3 months before testing was excluded from the study.

TABLE 1. Percent Variance Associated With Each Stance Within the Traditional BESS

Stance	Percent Variance
Firm double leg	0.17
Foam double leg	2.25
Firm tandem	14.21
Foam tandem	18.60
Firm single leg	29.43
Foam single leg	35.31

BESS, Balance Error Scoring System.

TABLE 2. Theorized Reliabilities Using a Modified Protocol of 4 Conditions (Single-Leg Firm, Tandem-Leg Firm, Single-Leg Foam, and Tandem-Leg Foam) With Increases in the Number of Trials Administered

Model	ICC Reliability
1 trial 6 conditions	$R = 0.60$
1 trial 4 conditions	$R = 0.71$
2 trials 4 conditions	$R = 0.83$
3 trials 4 conditions	$R = 0.88$
4 trials 4 conditions	$R = 0.91$
5 trials 4 conditions	$R = 0.92$
6 trials 4 conditions	$R = 0.94$
7 trials 4 conditions	$R = 0.94$

ICC, intraclass correlation coefficient.

Revised BESS

The protocol used for BESS administration was a modification of the original BESS. The revised protocol comprised 4 conditions: single-leg and tandem-leg stance on firm and foam surfaces. Completion of 4 conditions was considered a trial. Each participant completed 3 trials of the 4 conditions. All other aspects of the revised model remained identical to previously published literature inclusive of the determination of errors.[4,6–8]

Statistical Analysis

Intraclass reliability coefficients using a 1-way analysis of variance (ANOVA) protocol were calculated for the revised protocol. Additionally, a repeated measures (SxT) ANOVA was calculated to examine differences between trials. The intraclass reliability coefficients and repeated measures ANOVA were obtained using SPSS 11.3.1.

RESULTS

The mean total number of errors was 20.08 ± 11.76, whereas the mean number of errors for each trial was 6.69 ± 0.36. The number of errors for each surface, stance, and trial can be found in Table 3. The participants made the greatest amount of errors in the single-leg stance foam surface conditions. The tandem-leg stance had slightly less errors than the single-leg stance on either condition (Table 3).

Evaluation of differences between trials within the protocol was calculated using a repeated measures ANOVA. There were statistically significant differences between the trials. Simple contrasts were calculated to determine whether the statistical significance was equal across all trials. Statistically significant differences existed between trials 1 and 2 ($F(2,142) = 3.675$, $P = 0.028$, $\eta^2 = 0.049$); however, there were no significant differences between trials 2 and 3 (Figure 2).

Finally, an intraclass reliability coefficient was calculated for the revised protocol. The intraclass reliability measures internal consistency, the extent to which the parts of the test measure the same construct.[15] A reliability coefficient for the revised protocol was calculated (0.88) using 3 trials of 4 conditions. Due to the presence of a practice effect for trial 1, it was eliminated from the analysis. This resulted in a reliability coefficient for the criterion score of trials 2 and 3 of 0.84. Consistency of scores was obtained between trials 2 and 3. This would suggest that the use of 1 trial might be appropriate for estimation of postural stability. However, the intraclass reliability coefficient for 1 trial was 0.73, which negates the change in protocol as only moderate reliability was obtained (Table 4).

TABLE 3. Errors Committed During Administration of the Modified BESS

Trial	Stance	Surface	Mean ± SD
1	Single leg	Floor	1.75 ± 1.74
1	Tandem leg	Floor	0.44 ± 0.88
1	Single leg	Foam	3.73 ± 2.39
1	Tandem leg	Foam	1.27 ± 1.35
2	Single leg	Floor	1.69 ± 1.68
2	Tandem leg	Floor	0.42 ± 0.69
2	Single leg	Foam	3.24 ± 2.00
2	Tandem leg	Foam	1.06 ± 1.21
3	Single leg	Floor	1.77 ± 1.60
3	Tandem leg	Floor	0.52 ± 0.82
3	Single leg	Foam	3.17 ± 1.86
3	Tandem leg	Foam	1.02 ± 1.24

BESS, Balance Error Scoring System.

DISCUSSION

These studies found that the modified BESS protocol provides a more reliable measure. Performing minor modifications to the BESS protocol improved the reliability of the measure and reduced the practice effects. This study concurs with previous reports of limited errors performed with the double-leg stance and was modified appropriately.[3,8,16] The double-leg stance accounted for limited variance to the BESS with virtually no errors with the double-leg stance on both firm and foam surfaces. As previous studies have provided limited data regarding the double-leg stance due to lack of variance, and the inability to differentiate between injured and control groups beyond day 1 postinjury, this research demonstrated that the double-leg stance lowers the reliability of the

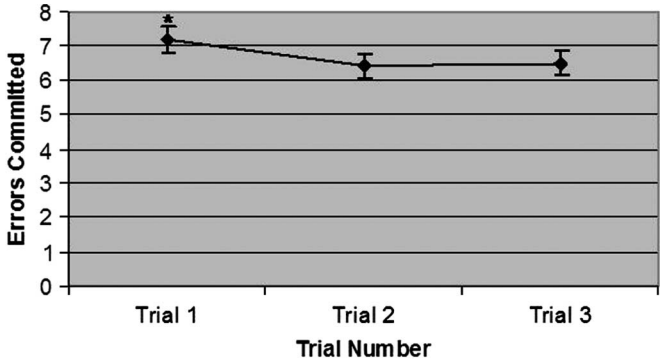

FIGURE 2. Mean number of errors per trial for the revised BESS protocol. BESS, Balance Error Scoring System.

TABLE 4. Intraclass Reliability Coefficients for Trials

Trials	No. of Conditions	Intraclass Reliability Coefficient
1, 2, and 3	12	0.88
2 and 3	8	0.84
2 only	4	0.73

BESS.[6,7,9,11] Study 1 found a significant increase in the reliability of the measure when the double-leg stance was removed. When the double-leg stance for both the firm and floor surfaces was removed, the reliability of the BESS increases from $r = 0.60$ to $r = 0.71$.

Examination of the modified protocol revealed a practice effect between the first and second trials; however, scores stabilized for subsequent trials. Previous studies revealed that a practice effect existed with decreased errors committed with each subsequent session.[7] Given that a practice effect existed for the first trial, we recommend that the first trial should not be scored or used in the analysis and return to participation decisions. It should be administered and treated as a practice trial only.

Administering and scoring 3 trials after the practice trial would, theoretically, increase the intraclass reliability coefficient. We obtained reliability coefficients of 0.84 and 0.88 using the mean of 2 and 3 trials, respectively. Furthermore, analyzing trial 1 with the practice effect might increase the reliability but would not increase validity evidence. Therefore, it is recommended to administer at least 3, preferably 4, trials without using the first trial to obtain good reliability.

CLINICAL RELEVANCE

As standardized concussion assessment tools are suggested within the field of athletic training, having psychometrically sound instruments is a necessity.[17–19] The recommendation for a multifaceted approach including physical examination, self-reported symptoms, and neuropsychological and posturography assessment has created a significant amount of interest in the BESS.

The BESS is a modified static protocol designed to evaluate postural stability in concussed athletes, with better sensitivity and specificity than the original Rhomberg test.[20–22] The standard BESS protocol introduced an objective, time-efficient, easily administered measure of postural ability. This tool is easily inserted into numerous concussion assessment batteries that can be administered virtually anywhere.

Low to moderate intraclass and test–retest reliabilities for the traditional BESS protocol limit the validity of the interpretations of scores ($r = 0.60, 0.67$). Psychometricians would say that the current reliability would not be adequate for a measurement with implications for return to participation decisions. The modified BESS protocol provides greater reliability of the BESS scores. In contrast to the standard BESS protocol, the reliabilities of the modified BESS protocol met the traditionally accepted thresholds for reliability measures. Enhancing the reliability of the BESS will ultimately affect the validity of the interpretations made from the measure to aid in clinical interpretation.

Modifications proposed here maintain the measure as an objective, time-efficient, easily administered tool of postural ability, which can be incorporated into any concussion assessment battery. We suggest administering at least 2, and preferably 3, trials to obtain good reliability. It may appear that 4 trials of 4 conditions might be too time consuming or introduce fatigue. Performing 3 trials within the modified BESS protocol takes approximately 5 minutes to administer and maintains the true purpose of the original BESS protocol.

LIMITATIONS

First, this research was performed using only nonconcussed healthy athletes. Examining the revised protocol with concussed athletes will allow for evaluation of the specificity and sensitivity of the revised BESS protocol. Although it seems that there were no differences between injured and control participants on the double-leg stance past day 1 postinjury, this study should be replicated using injured participants to determine the necessity for the double-leg stance as a practice or control stance and clinical sensitivity of the modified BESS protocol.[3]

Second, the number of trials used was based on the number of trials used in similar protocol measures with the Neurocom Sensory Organization Test. Correlating the modified BESS protocol to the Sensory Organization Test and investigating fatigue and stability of scores after administering 3 or more trials for both baseline and injured data may be helpful to determine the best protocol for use in concussion assessment.

CONCLUSIONS

Concussion assessment has evolved with the influx of concussion-related research and reaffirming that postural stability is an important piece of a multifaceted approach to concussion management. The modified protocol provides greater reliability of the BESS scores. Psychometrically sound instruments support the ability to make and interpret clinical decisions regarding injury and return to participation. Further analysis of the model is being conducted on baseline and concussed subjects to determine if the double-leg stance is truly an important factor in the BESS during the acute phases of assessment after concussive injuries.

REFERENCES

1. Guskiewicz KM, Perrin DH. Research and clinical applications of assessing balance. *J Sport Rehabil*. 1996;5:45–63.
2. Ingersol CD, Armstrong CW. The effects of closed-head injury on postural sway. *Med Sci Sports Exerc*. 1992;24:739–743.
3. Riemann BL, Guskiewicz KM. Effects of mild head injury on postural stability as measured through clinical balance testing. *J Athl Train*. 2000; 35:19–25.
4. Guskiewicz KM, Ross SE, Marshall SW. Postural stability and neuropsychological deficits after concussion in collegiate athletes. *J Athl Train*. 2001;36:263–273.
5. McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on concussion in sport, 3rd international conference on concussion in sport, held in Zurich, November 2008. *Clin J Sport Med*. 2009;19:185–200.
6. Riemann BL, Guskiewicz KM, Shields EW. Relationship between clinical and forceplate measures of postural stability. *J Sport Rehabil*. 1999;8:71–82.
7. Valovich-McLeod TC, Perrin DH, Guskiewicz KM, et al. Serial administration of clinical concussion assessments and learning effects in healthy young adults. *Clin J Sport Med*. 2004;14:287–295.

8. 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*. 2003;38:51–56.
9. Susco TM, Valovich-McLeod TC, Gansneder BM, et al. Balance recovers within 20 minutes after exertion as measured by the balance error scoring system. *J Athl Train*. 2004;39:241–246.
10. Wilkins JC, Valovich McLeod TC, Perrin DH, et al. Performance on the Balance Error Scoring System decreases after fatigue. *J Athl Train*. 2004; 39:156–161.
11. McCrea M, Guskiewicz KM, Marshall SW, et al. Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study. *JAMA*. 2003;290:2556–2563.
12. Valovich TC. *The Use of the Standardized Assessment of Concussion and Balance Error Scoring System and Learning Effects in Youth Sports Participants* [dissertation]. Charlottesville: University of Virginia; 2002.
13. Brennan RL. *Elements of Generalizability Theory (rev. ed.)*. Iowa City, IA: American College Testing; 1992.
14. Crick J, Brennan R.GENOVA: *A Generalized Analysis of Variance System* (FORTRAN IV) [computer software and manual]. Dorchester, MA: Computer Facilities, University of Massachusetts at Boston; 1982.
15. Baumgartner TA, Jackson AS. *Measurement for Evaluation*. Boston, MA: McGraw-Hill; 2003:96–119.
16. Guskiewicz KM, Riemann BL, Perrin DH, et al. Alternative approaches to the assessment of mild head injury in athletes. *Med Sci Sports Exerc*. 1997;29(Suppl 7):S213–S221.
17. Aubry M, Cantu R, Dvorak J, et al. Summary and agreement statement of the 1st international symposium on concussion in sport, Vienna 2001. *Clin J Sport Med*. 2002;12:6–11.
18. Guskiewicz K, Bruce S, Cantu R, et al. National Athletic Trainers' Association position statement: management of sport-related concussion. *J Athl Train*. 2004;39:280–297.
19. McCrory P, Johnston K, Meeuwisse W, et al. Summary and agreement statement of the 2nd international conference on concussion in sport, Prague 2004. *Clin J Sport Med*. 2005;15:48–55.
20. Jansen ER, Larsen R, Oleson M. Quantitative Rhomberg's test. *Acta Neurol Scand*. 1982;66:93–99.
21. Riemann BL, Guskiewicz KM. Assessment of mild head injury using measures of balance and cognition: a case study. *J Sport Rehabil*. 1997;6: 283–289.
22. Ekdahl C, Jarnlo G, Anderson S. Standing balance in healthy subjects. *Scand J Rehabil Med*. 1989;21:187–195.