ABSTRACT

**Study Design:** Clinical Measurement, Correlation, Reliability

**Objectives:** To assess the relationship between the Single Leg Balance (SLB), modified Balance Error Scoring System (mBESS), and modified Star Excursion Balance (mSEBT) tests and secondarily to assess inter-rater and test-retest reliability of these tests.

**Background:** Ankle sprains often result in chronic instability and dysfunction. Several clinical tests assess postural deficits as a potential cause of this dysfunction; however, limited information exists pertaining to the relationship that these tests have with one another.

**Methods:** Two independent examiners measured the performance of 34 healthy participants completing the SLB Test, mBESS test, and mSEBT at two different time periods. The relationship between tests was assessed using the Pearson Correlation and Fisher’s Exact Tests. Inter-rater and test-retest reliability were assessed using the intraclass correlation coefficient (ICC) and Kappa statistics.

**Results:** A significant correlation ($r = -0.35$) was observed between the mSEBT and the mBESS. Fisher’s Exact Test showed a significant association between the SLB Test and mBESS ($P = .048$), but no association between the SLB and mSEBT ($P = 1.000$). Inter-rater reliability was excellent for the mSEBT and fair for the mBESS (ICCs of .91 and .61 respectively). Excellent agreement was observed between raters for the SLB test ($\kappa = 1.00$). Test-retest reliability was excellent for the mSEBT (ICC = 0.98) and fair for the mBESS (ICC = 0.74). There was poor test-retest agreement for the SLB test ($\kappa = .211$).

**Conclusion:** There was a significant relationship observed between the SLB Test, mBESS test, and mSEBT; however, strength of association measures showed limited overlap between these tests. This suggests that these tests are interrelated but may not assess equal components of postural stability.

**Key Words:** reliability, postural stability, star excursion balance test, single leg balance, balance error scoring system

---

1 Resident, USM-BU Sports Medicine Physical Therapy Doctoral Program, United States Military Academy, West Point, NY.
2 Slippery Rock University Student, Slippery Rock, PA.
3 Director of Orthopaedic Research, John A. Feagin Sports Medicine Fellowship, Keller Army Hospital, West Point, NY.
4 Director, USM-BU Sports Medicine Physical Therapy Doctoral Program, United States Military Academy, West Point, NY.

This research was conducted at the United States Military Academy, Arvin Cadet Physical Therapy Clinic, Physical Therapy Department, West Point, NY. No funding support was received for this research study. At the time of the research, Dr. Clark was completing a post-professional sports physical therapy residency under the mentorship of Dr. Gerber as well as completing his DScPT in Physical Therapy from Baylor University, Waco, TX. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the US Air Force, US Army, the Department of Defense, or the United States Government.

**CORRESPONDENCE**

Richard C. Clark, 10th Medical Operations Squadron, Department of Physical Therapy, USAFA, CO 80840. E-mail: airforcept@gmail.com
INTRODUCTION
Ankle sprains are the most common traumatic injury in sports, with injury rates among athletes between 12-33%.1, 17, 22, 24, 25 Athletes who participate in sports that require sudden stopping, planting, and cutting such as football, soccer, and basketball are especially vulnerable to these injuries.2, 4, 10, 18, 19, 26 Full recovery after an ankle sprain is often time consuming and incomplete. Because ankle sprains account for 16-23% of all sport injury time lost, and athletes commonly report lingering dysfunction, it is important to determine reliable methods to identify individuals at risk to ankle sprains and to develop preventative initiatives to reduce their risk.31, 17

One risk factor associated with an increased incidence of ankle sprains is impaired postural stability.3, 6, 8, 9, 29-31, 34 Researchers commonly report that there is an established relationship between postural stability deficits and chronic ankle instability.4-6, 30, 32 To identify those who have an increased risk for ankle sprains (both first time and recurrent), devices such as stabilometry have been effectively used to objectively assess postural stability.18 The disadvantages with stabilometry are that it is expensive, time consuming, and not readily available to most clinicians. In contrast, several clinical measures of postural stability such as the Single Leg Balance (SLB) test, Balance Error Scoring System (BESS), and the modified Star Excursion Balance Test (mSEBT) are inexpensive, quick to administer, and typically accessible in clinical and field settings. Poor performance on these tests may also prove to be an indicator that a person has an increased risk of a future ankle sprain.

The development, utilization, and understanding of these measures of postural stability are in the beginning stages. Of the commonly used clinical and field measures of postural stability, the mSEBT appears to be the “gold standard”. However, while the mSEBT has shown consistently strong inter-rater and intra-rater reliability, it is more time consuming than the SLB and BESS.13-16, 20, 23 The latter assessments are relatively new tests for predicting lower extremity injury and have demonstrated varying degrees of reliability in assessing impaired postural stability.7, 21, 27, 28, 33

The primary purpose of this study was to assess the relationship between the SLB, a modified version of the BESS (mBESS), and mSEBT tests, utilizing the mSEBT as the criterion measure for identifying postural deficits. A secondary purpose of this study was to assess inter-rater and test-retest reliability of these 3 tests. If the SLB and mBESS tests demonstrate strong relationships with the mSEBT and are reliable to perform, clinicians may opt for a simpler, less resource intensive measure to assess postural stability.

METHODS
Subjects
Thirty-four volunteers from the United States Military Academy, West Point, New York enrolled in the study. They were included if they were generally healthy and between the ages of 17 and 25. Participants were excluded from the study if they reported any history of lower extremity or head injury (concussion) within the previous 12 weeks, a history of ankle sprain within the past twelve months, any inner ear disturbance at the time of testing, or a history of prior ankle surgery. Participants were also excluded if they were currently in a boxing course or boxing club due to an increased risk of concussion and potential loss to follow-up testing (Figure 1). This study was reviewed and approved by the Institutional Review Board at Keller Army Hospital (West Point, NY) with secondary review by the U.S. Army Clinical Investigation Regulatory Office (Ft. Sam Houston, TX). All participants read and signed an informed consent and HIPPA addendum form prior to enrollment and the rights of all participants were protected throughout the study.

Procedures
Participants were randomly assigned a sequence in which they were to perform the SLB Test, mBESS test, and mSEBT (Figure 1). Each participant was given a brief orientation and demonstration of the testing procedures. Testing was performed in a semi-private area with the participant barefoot, wearing loose fitting shorts and a t-shirt. The subject's dominant leg (one he or she kicked with) was always tested first. Participants were allowed to practice each test prior to performance to minimize learning effect, and they received a minimum of a one minute rest period between each test. One practice attempt was allowed for the SLB and mBESS tests,
and a minimum of one and maximum of 6 practices for the mSEBT as per the protocol of Hertel et al.15

To examine inter-rater reliability, two examiners performed assessments independently at the same time during the balance tests. To examine test-retest reliability, the subjects were scheduled for re-testing approximately 15-28 days (mean = 23 days) after the initial test to minimize a potential learning effect33. Participants were again randomly assigned a sequence of balance testing for the second assessment.

**Postural Stability Testing**

The SLB test was performed in accordance with the Trojan protocol.28 Participants stood on one foot with the contralateral knee slightly bent and not touching the weight bearing leg, hips level to the ground, and eyes open and fixed on a spot marked on the wall. Once participants had obtained their balance, they closed their eyes for 10 seconds. Investigators noted if the participant's legs touched each other, the ipsilateral foot moved on the floor, the contralateral foot touched down, the eyes opened, or the arms moved from their start position. If the participant had a positive test (failed to remain balanced) during the first trial, a second trial was performed, with the results of the second trial counting for analysis.

Docherty et al7 noted in their BESS study on athletes that there was no statistical difference during the test between subjects with and without functional ankle instability in the double leg stance positions and the tandem stance on a firm surface. Due to these results, a modified BESS (mBESS) test was derived by this facility for this study in attempts of creating a clinical test that would be appropriate for testing athletes with impaired postural stability. The mBESS test was performed by having participants stand unsupported with their eyes closed under 6 conditions: 3 different stances (single-limb dominant, single-limb non-dominant, and tandem with

---

**Figure 1.** Flow of participants through each stage of the study.
the intersection of 3 lines (Figure 2) with the most distal aspect of their stance limb great toe at the 10 cm line. Participants maintained single-leg stance while reaching with the contralateral leg to touch as far as possible along the chosen line. All participants were given the same instructions: to maintain foot-flat stance during the test maneuver. If the foot moved at any time during the test, the test was stopped, the foot repositioned, and the test repeated. They touched the furthest point possible on the line with the most distal part of their foot as lightly as possible so that the reach leg did not provide any significant support in the maintenance of upright posture. The examiners marked the point touched along the line and then manually measured the grid. The participant repeated this three times on each line, and completed all three lines for a total of 9 reaches. The distance reached was averaged and recorded for each direction and also summed to produce the participant’s total mSEBT score. In addition, data was normalized as recommended by Gribble et al.12 by dividing the average reach distance by the participant’s leg length, and then multiplying the value by 100.

Statistical Analysis
Descriptive statistics for categorical and continuous variables were calculated to summarize the data. Tests for outliers and assumptions for utilization of the statistical tests were performed. Separate one-way ANOVAs were performed to assess for significant differences in the total mBESS and mSEBT scores. The independent variable evaluated was whether subjects passed or failed the SLB test. Plots of the means were performed to demonstrate the relationships between each of these tests, and the SLB. Pearson Correlation studies were performed between continuous data of the mSEBT and mBESS. To allow for correlational analysis between the SLB, the data for the mSEBT and mBESS was collapsed to nominal, dichotomous data. This was done by determining if an individual would be considered “positive” or “negative” based on the performance on the test. A “positive” for these tests was operationally defined as poor postural stability. For the mSEBT, a participant was considered “positive” if he or she demonstrated a dominant to non-dominant difference of ≥4 cm.23 For the mBESS, a participant was considered “positive” if the number of balance errors during the test was greater than one standard deviation above the mean.

Figure 2. Modified Star Excursion Balance Test with reach directions labeled in reference to the right stance foot.
Similarly, subjects who failed the SLB also had significantly lower mSEBT total scores, which also represent poorer postural stability \( (P = .025) \) compared to those who passed the SLB test (Table 1). A plot of the means of the mBESS and mSEBT by outcome on the SLB is presented in Figure 3. A significant \( (P < .05) \) correlation \( (r = -0.35) \) was observed between the mSEBT and the mBESS (Figure 4). Subjects with poorer postural stability scores on the mSEBT also seemed to have poorer postural stability scores on the mBESS. Strength of association measures suggest that the total mBESS score accounts for approximately 12% of the variability in the mSEBT score \( (r^2 = 0.12) \). When scores from the mBESS and mSEBT were collapsed, Fisher’s Exact Test revealed a significant association between the SLB Test and the mBESS \( (P = .048) \), but no association was observed between the SLB and mSEBT \( (P = 1.00) \).

**Reliability**

Inter-rater reliability was excellent for the mSEBT and fair for the mBESS (ICCs of 0.91 and 0.61 respectively).

**Table 1. ANOVA Table - Associations between the SLB and mBEss Tests and the SLB and mSEBT Tests.**

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mBEss</td>
<td>Between Groups</td>
<td>423.534</td>
<td>1</td>
<td>423.534</td>
<td>8.030</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>1500.966</td>
<td>32</td>
<td>46.905</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1924.500</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mSEBT</td>
<td>Between Groups</td>
<td>256.308</td>
<td>1</td>
<td>256.308</td>
<td>5.665</td>
</tr>
<tr>
<td></td>
<td>Within Groups</td>
<td>1473.766</td>
<td>32</td>
<td>46.055</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1730.074</td>
<td>33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**RESULTS**

Thirty-four participants (20 males, 14 females; mean ages of 20.9 ± 1.92 and 20.43 ± 2.21 years respectively) enrolled and completed the study. Twelve of the 34 participants failed to maintain balance on the SLB during one of the 2 testing periods. Of those twelve, 4 failed only during the first testing period, 5 failed only during the second testing period, and 3 failed during both testing periods. The mean number of total errors on the mBESS was 16.5 ± 7.6. Most errors occurred while testing the non-dominant leg on the foam surface \( (4.5 ± 2.0 \text{ errors}) \) while the fewest errors occurred while testing tandem on firm surface \( (0.8 ± 0.9 \text{ errors}) \). Five participants demonstrated a dominant to non-dominant difference of >4 cm on the mSEBT. The duration for each test (from explanation to completion) was approximately 78 seconds for the SLB, 273 seconds for the mBESS, and 311 seconds for the mSEBT. The authors of this paper are confident that no learning effect was observed between trials, as subject performance did not improve during the second trials.

**Associations between tests**

Subjects who failed the SLB test made significantly more errors when performing the mBESS \( (P = .005) \), suggesting that they had poorer postural stability when compared to those who passed the SLB.
a significant association between the 2 static tests, but not between the SLB and the mSEBT. Additionally, while there was a significant correlation between the mBESS and mSEBT, the relationship was not as strong as we would have anticipated \((r = -0.35)\). This suggests that only 12% of the variability in scores of one test can be explained by the results of the other test. Practically, this means that while these tests are interrelated it appears they do not measure the same component of postural stability. The SLB test is a simple, static pass/fail test that may be strongly influenced by external factors that provided only nominal data. As the subject is only provided a maximum of two opportunities to test, little room for error is available for a “bad test.” The mBESS is also a static test, but provides a greater amount of continuous data and tests six postural positions. If a participant has difficulty in a single position, the poor performance may be less apparent in their overall score if they perform well on the remaining tests. The mSEBT is a dynamic, functional test that averages the performance of three trials in each direction, diminishing the potential effect of a single outlying (poor) performance.

The SLB test is a clinically alluring measure of postural stability as it is the most efficient, least

**DISCUSSION**

The primary finding of this study was that while there was a significant relationship between several clinical postural stability tests, the strength of association was limited. This suggests that these tests may be interrelated but may not assess similar components of postural stability. Secondarily, of these three clinical measures of postural stability, the combination of inter-rater and test-retest reliability is highest for the mSEBT.

Because the SLB, mBESS, and mSEBT are all measures of postural stability, the authors of this study hypothesized there would be a strong relationship between the tests. However, the results do not support this hypothesis. Despite the similarities between the three tests, there is one fundamental difference between these tests. The SLB and the mBESS are static tests, whereas the mSEBT is a dynamic, functional test. This could explain why the authors found a significant association between the 2 static tests, but not between the SLB and the mSEBT. Additionally, while there was a significant correlation between the mBESS and mSEBT, the relationship was not as strong as we would have anticipated \((r = -0.35)\). This suggests that only 12% of the variability in scores of one test can be explained by the results of the other test. Practically, this means that while these tests are interrelated it appears they do not measure the same component of postural stability. The SLB test is a simple, static pass/fail test that may be strongly influenced by external factors that provided only nominal data. As the subject is only provided a maximum of two opportunities to test, little room for error is available for a “bad test.” The mBESS is also a static test, but provides a greater amount of continuous data and tests six postural positions. If a participant has difficulty in a single position, the poor performance may be less apparent in their overall score if they perform well on the remaining tests. The mSEBT is a dynamic functional test that averages the performance of three trials in each direction, diminishing the potential effect of a single outlying (poor) performance.

The SLB test is a clinically alluring measure of postural stability as it is the most efficient, least

**Figure 4.** A significant \((p < .05)\) but weak correlation was observed between the modified Star Excursion Balance Test and modified Balance Error Scoring System.
expensive, and easiest test to administer. Indeed, this is why the SLB test is perhaps the most widely used measure for postural stability in clinical settings and why it has been suggested that the SLB test could be used in a preparticipation setting to identify athletes with an increased risk of ankle sprains. Similar to the Trojan et al study, we found excellent inter-rater reliability with the SLB test, however, Trojan did not assess test/retest reliability. The authors of the current study were surprised that the test/retest reliability was extremely poor on this healthy sample. While 22 of 34 participants (65%) maintained balance on the SLB test during both testing periods, 9 of 34 (26%) failed the SLB test during a single testing period. These results may have been due to the lack of extrinsic controls that were utilized in this study. While increasing the extrinsic controls may have increased the test/retest reliability, it is the belief of the authors that it would have compromised the external validity of the test. Based on these poor test/retest results, using the SLB test as a simple objective pass/fail measure of postural stability or to identify athletes with an increase risk of ankle sprains does not appear warranted. The SLB test may still have value to the clinician as a subjective measure, but it appears to have limited value as an objective pass/fail measure.

Compared to the SLB, the mBESS test takes longer to administer and requires some equipment (i.e., foam surface pad). A potential advantage, however, is that the mBESS provides more detailed information than simply a pass/fail score that the SLB provides. The mBESS demonstrated fair interrater reliability (ICC = .61) and test/retest reliability (ICC = .74) suggesting that the test appears to be a more reliable objective test when compared to the SLB, but in this study showed lower intratester reliability than those reported in previous studies on the original BESS. However, for an athletic population, this version of the mBESS needs further refinement and testing to improve its test/retest reliability, but it appears to have the potential to be a useful objective test in identifying athletes with impaired static postural stability. Further research continues to be important to determine the clinical applicability of the SLB and mBESS tests.

Compared to the SLB and mBESS tests, the mSEBT takes the most time to administer and requires the most equipment. It was the only measure in this study, however, to demonstrate both excellent inter-rater and test/retest reliability. These findings are consistent with others who have studied reliability using the mSEBT. In addition, the mSEBT has been shown extensively to be sensitive in identifying postural stability deficits. While it would be appealing to use a more efficient and easier test, ultimately, the results of the current study support that in the clinical setting where stabilometry is not accessible, the mSEBT should continue to be considered as the best available criterion measure to assess postural stability. However, as discussed above, it should be noted that the mBESS and mSEBT may assess different components of postural stability (i.e., static versus dynamic postural stability). Further investigation is needed to determine how these measures may complement each other in assessing static and dynamic aspects of postural stability as well as determining the potential role of these assessments in injury screening and prevention initiatives.

Some limitations associated with this study should be noted. This study was performed on a relatively small sample size, only using healthy individuals. Stronger relationships between the postural stability tests may have been observed if both injured and uninjured participants were tested. It is possible that a ceiling effect was observed in this study as these tests simply may not have been sensitive enough to discriminate postural stability differences in healthy individuals. Perhaps the results would have been different for example in those with ankle pathology. Therefore, future research looking at a larger sample that includes both injured and uninjured individuals would be beneficial. Additionally, no controls were instituted for a variety of external factors that could affect performance between testing periods. Several factors such as fatigue, time of day, activities prior to the testing, amount of sleep the night prior, concentration, and a multitude of other factors could have affected the ability of participants to perform on these postural stability tests. While it is possible that controlling such factors could have improved test/retest reliability, these factors realistically are not controlled for in most clinical settings. Given the variety of external factors that could have negatively affected participants’ performance over time, the excellent test/retest reliability demonstrated by the
mSEBT provides further evidence of its reliability as a clinical postural stability test.

**CONCLUSION**

There was a statistically significant relationship observed between the mBESS and mSEBT, but the strength of relationship was weak. Additionally, there was a significant association observed between the SLB and the mBESS. This suggests that these tests are interrelated but that they may not assess equal components of postural stability. While there was perfect (κ = 1.00) interrater reliability with the SLB test, this test demonstrated poor test/retest reliability significantly impacting its clinical utility. The mBESS demonstrated fair interrater reliability and fair test/retest reliability. The mSEBT demonstrated excellent interrater and test/retest reliability. Further research to determine the best applications and utility for each of these clinical postural tests is warranted.

**REFERENCES**


