Test–Retest Reliability and Practice Effects of the Stability Evaluation Test

Richelle M. Williams, Matthew A. Corvo, Kenneth C. Lam, Travis A. Williams, Lesley K. Gilmer, and Tamara C. Valovich McLeod

Context: Postural control plays an essential role in concussion evaluation. The Stability Evaluation Test (SET) aims to objectively analyze postural control by measuring sway velocity on the NeuroCom VSR portable force platform (Natus, San Carlos, CA). Objective: To assess the test–retest reliability and practice effects of the SET protocol. Design: Cohort. Setting: Research laboratory. Participants: 50 healthy adults (20 men, 30 women; age 25.30 ± 3.60 y, height 166.60 ± 12.80 cm, mass 68.80 ± 13.90 kg). Interventions: All participants completed 4 trials of the SET. Each trial consisted of six 20-s balance tests with eyes closed under the following conditions: double-leg firm (DFi), single-leg firm (SFi), tandem firm (TFi), double-leg foam (DFo), single-leg foam (SFo), and tandem foam (TFo). Trials were separated by 5-min seated rest periods. Main Outcome Measures: The dependent variable was sway velocity (°/s), with lower values indicating better balance. Sway velocity was recorded for each of the 6 conditions, as well as a composite score for each trial. Test–retest reliability was analyzed across 4 trials with intraclass correlation coefficients (ICCs). Practice effects were analyzed with repeated-measures analysis of variance, followed by Tukey post hoc comparisons for any significant main effects (P < .05). Results: Sway-velocity reliability values were good to excellent: DFi (ICC = .88; 95%CI: .81,.92), SFi (ICC = .75; 95%CI: .61,.85), TFi (ICC = .84; 95%CI: .75,.90), DFo (ICC = .83; 95%CI: .74,.90), SFo (ICC = .82; 95%CI: .72,.89), TFO (ICC = .81; 95%CI: .69,.88), and composite score (ICC = .93; 95%CI: .88,.95). Significant practice effects (P < .05) were noted on the SFi, DFo, SFo, and TFO conditions and composite scores. Conclusions: Our results suggest that the SET has good to excellent reliability for the assessment of postural control in healthy adults. Due to the practice effects noted, a familiarization session is recommended (ie, all 6 conditions) before data are recorded. Future studies should evaluate injured patients to determine meaningful change scores during various injuries.

Keywords: balance, concussion, postural control

An estimated 1.1 to 1.9 million sport- and recreation-related concussions occur each year in the United States in children under 18 years of age.1 After a sport-related concussion, many functional domains are affected, including somatic, cognitive, emotional, vestibular, physical signs, behavioral changes, and sleep disturbances. In particular, postural control is impaired after a concussive injury and can be affected up to 14 days postinjury.2,3

The current guidelines for the assessment and management of a sport-related concussion recommend a multifaceted approach that includes an evaluation of symptoms, cognition, and balance as the most sensitive and recommended for use in clinical practice.3 Objective postural assessment should be incorporated into injury evaluation to assess the motor domain of neurological function. Postural-control assessments have used both advanced force-plate technology and more clinically feasible balance assessments such as the Balance Error Scoring System (BESS).2 Instrumented balance assessment is typically expensive due to the cost of computerized dynamic-posturography equipment, which makes them difficult to acquire for many clinicians. As a result, several companies have developed portable computerized force-platform systems including the BioSway (Biodex, Shirley, NY) and the VSR Sport (Natus, San Carlos, CA). Specifically, the VSR Sport force plate functions as a highly mobile device and allows clinicians to use it in a variety of settings (ie, clinic, on field, during physicals). Furthermore, the VSR Sport enables the Stability Evaluation Test (SET) protocol to be implemented. The SET aims to objectively analyze balance and postural control through measurements of sway velocity and appears to demonstrate clinical benefits; however, little is known about its reliability. As the VSR portable platform is commercially available for use, it is imperative to measure validity and reliability of the testing protocol. Therefore,
the purpose of this study was to assess the test–retest reliability and susceptibility to practice effects of the SET protocol.

**Methods**

**Subjects**

Fifty adults (20 men, 30 women; age $25.26 \pm 3.56$ y, height $166.6 \pm 12.8$ cm, mass $68.8 \pm 13.9$ kg) participated in one 45-minute test session. Participants were at least 18 years of age or older and free of any lower-extremity injury that would limit their ability to balance on 1 foot. Participants were not excluded with previous concussion diagnosis; however, they were excluded if they had any current lower-extremity injury that would limit balance abilities. Subject demographics are presented in Table 1.

**Instrumentation**

The SET is a balance assessment that includes 6 trials performed on a $18 \times 30$-in VSR force plate (Natus, San Carlos, CA). The tasks include balancing under varying stance conditions: double-leg (DFi), single-leg (SFi), and tandem stances (TFi) on the force plate (firm surface) and double-leg (DFo), single-leg (SFo), and tandem stances (TFo) on a foam pad placed on the force plate (foam surface). Each condition was 20 seconds long and required participants to balance with their eyes closed and hands on their hips. Completion of all 6 conditions constituted 1 trial. The SET produces an objective output of postural sway, measured as sway velocity, and expressed in °/s as the balance tasks are performed.

**Procedures**

On arrival at the research laboratory, participants signed an informed consent and completed a brief medical-history form. After completion of demographic forms, they were instructed on the balance tasks. Participants practiced the balance protocol to familiarize themselves with each of the conditions. The initial familiarization period was standardized across participants. Participants were given 5 to 10 seconds to attempt each stance and to understand how to place legs, hands on hips, and get a feel for the foam pad used. Once participants indicated they understood the stance, the next stance was demonstrated and familiarized until participants verbally indicated that they understood the protocol and felt comfortable. The participants completed 1 complete SET trial (6 conditions, ~5 min) 4 different times, with a 5-minute sitting break between trials. Each participant was given standardized instructions for each trial.

**Statistical Analysis**

Intraclass correlation coefficients (ICC$_{2,1}$) were used to evaluate the test–retest reliability of the SET conditions across the 4 trials. A repeated-measures ANOVA was used to evaluate for practice effects across the trials. Post hoc analyses using Tukey pairwise comparisons were used to further evaluate differences in conditions with significant practice effects ($P < .05$).

**Results**

**Intraclass Correlation Coefficients**

The ICCs and 95% confidence intervals are presented in Table 2. The SET demonstrated good to excellent reliability (ICC > .75) for all conditions. The SET composite score had excellent reliability (ICC = .93, 95%CI .88, .95).

**Practice Effects**

Means and SDs for all trials are displayed in Table 2. The repeated-measures ANOVA revealed no significant practice effects for the SET (Figure 1) on the DFi condition ($F_{3,144} = 1.46, P = .23$) and TFi condition ($F_{3,144} = 2.21, P = .09$). A significant difference was noted for the SFi ($F_{3,144} = 8.15, P < .01$), DFo ($F_{3,144} = 23.19, P < .01$), SFo ($F_{3,144} = 12.45, P < .01$), TFo ($F_{3,144} = 6.40, P < .01$), and composite ($F_{3,129.47} = 30.33, P < .01$). The post hoc analysis showed that for SFi, trials 2 (1.90 ± 0.52), 3 (1.93 ± 0.81), and 4 (1.81 ± 0.61) were significantly lower than trial 1 (2.26 ± 0.69). DFo demonstrated that trials 2 (2.28 ± 0.63) and 3 (2.20 ± 0.51) were significantly lower than trial 1 (2.81 ± 0.87), and trial 4 (2.12 ± 0.47) was significantly lower than trials 1 (2.81 ± 0.87) and 2 (2.28 ± 0.63). Condition SFo revealed that trials 2 (3.75 ± 0.93) and 3 (3.73 ± 1.10) were significantly lower than trial 1 (4.33 ± 1.11), and trial 4 (3.50 ± 0.82) was significantly lower than trials 1 (4.33 ± 1.11) and 2 (3.75 ± 0.93). The TFo post hoc analysis revealed that trials 3 (3.82 ± 1.71) and 4 (3.86 ± 1.60) were significantly lower than trials 1 (4.62 ± 1.52) and 2 (4.37 ± 1.50). Finally, the composite score post hoc analysis revealed that trials 2 (2.42 ± 0.53), 3 (2.33 ± 0.53), and 4 (2.26 ± 0.49) were lower than trial 1 (2.73 ± 0.59).
Discussion

The results of this study suggest that the SET demonstrated good to excellent reliability on all conditions and composite scores and may be used as a measurement of sway velocity. The ICC ranges from .75 to .88 and demonstrates that the SET is able to dependably test postural stability over time. These findings are consistent with a study by Lafond et al., using computerized stabilometry on a static balance platform, which demonstrated the

---

Table 2  Scores on the Stability Evaluation Test (Mean ± SD) Across the 4 Trials and the Intraclass Correlation Coefficients (ICCs) for the 4 Trials of the Balance Tasks

<table>
<thead>
<tr>
<th>Trial</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>95% ICC (confidence interval [lower bound, upper bound])</th>
<th>Overall reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFi</td>
<td>0.70 ± 0.19</td>
<td>0.84 ± 0.79</td>
<td>0.84 ± 0.76</td>
<td>0.82 ± 0.59</td>
<td>.88 (.81, .92)</td>
<td>Excellent</td>
</tr>
<tr>
<td>SFi</td>
<td>2.26 ± 0.69</td>
<td>1.90 ± 0.52*</td>
<td>1.93 ± 0.81*</td>
<td>1.81 ± 0.61*</td>
<td>.75 (.61, .85)</td>
<td>Good to excellent</td>
</tr>
<tr>
<td>TFi</td>
<td>1.59 ± 0.62</td>
<td>1.45 ± 0.72</td>
<td>1.43 ± 0.65</td>
<td>1.38 ± 0.63</td>
<td>.84 (.75, .90)</td>
<td>Good to excellent</td>
</tr>
<tr>
<td>DFo</td>
<td>2.81 ± 0.87</td>
<td>2.28 ± 0.63*</td>
<td>2.20 ± 0.51*</td>
<td>2.12 ± 0.47*</td>
<td>.83 (.74, .90)</td>
<td>Moderate to excellent</td>
</tr>
<tr>
<td>SFo</td>
<td>4.33 ± 1.11</td>
<td>3.75 ± 0.93*</td>
<td>3.73 ± 1.10*</td>
<td>3.50 ± 0.82*</td>
<td>.82 (.72, .89)</td>
<td>Moderate to excellent</td>
</tr>
<tr>
<td>TFo</td>
<td>4.62 ± 1.52</td>
<td>4.37 ± 1.50</td>
<td>3.82 ± 1.71*</td>
<td>3.86 ± 1.60*</td>
<td>.80 (.69, .88)</td>
<td>Moderate to excellent</td>
</tr>
<tr>
<td>Composite</td>
<td>2.74 ± 0.59</td>
<td>2.42 ± 0.53*</td>
<td>2.33 ± 0.53*</td>
<td>2.26 ± 0.49*</td>
<td>.93 (.88, .95)</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

Abbreviations: DFi, double-leg firm; SFi, single-leg firm; TFi, tandem firm; DFo, double-leg foam; SFo, single-leg foam; TFo, tandem foam.

*Indicates significant differences from trial 1 ($P < .01$).
mean velocity of center of pressure to range from .83 to .94. Similarly, data from a systematic review of studies using the BESS suggest that the reliability of the BESS ranges from poor to moderate to good. Furthermore, scoring of the BESS is largely subjective and can play a factor in reliability studies. Therefore, to enhance interrater and intrarater reliability among BESS scoring, more-objective balance protocols are needed with an added portability component.

While the reliability was found to be good to excellent, some practice effects with repeated testing were noted. The findings from this study further highlight the need for practice trials to reduce practice effects, with performance after the first trial demonstrating the greatest change in performance. Furthermore, the SET conditions DFO, SFO, and composite scores demonstrated a performance change with better fourth-trial performance than during the second trial. These findings are consistent with the literature demonstrating practice effects after use of the BESS protocol. Our data suggest that the greatest practice effect was displayed during the single-leg condition, as well as while performing the task on the foam pad resulting in greater sway velocity from the first trial to the remainder of the trials.

Many factors can influence the presence and duration of practice effects. The short testing intervals between trials have been shown to increase the severity of practice effects. Therefore, the 5-minute resting time allotted during this study may not have been long enough to mitigate practice effects. Furthermore, there is reason to speculate that the foam surface and single-leg conditions are most prone to practice effects, as these are novel tasks. Consistent with the literature, the results from the current study demonstrated the largest practice effects during those conditions. Valovich et al found that the foam surface and single-leg conditions had the worst performance with the greatest improvement between trials. Similarly, Nordahl et al found that the foam surfaces along with the shortest time interval between sessions demonstrated the greatest learning effects. Therefore, due to the practice effects found during this study and results from previous studies, a full unscored familiarization trial with the stances is recommended before testing begins. Implementing a familiarizing session can decrease the possibility of practice effects.

The current study only examined the reliability and practice effects of sway velocity in healthy participants. This study is limited in looking at within-day reliability, and therefore future studies should attempt to replicate findings with a long-term testing protocol to equate to a baseline-to-postconcussion test paradigm. Further studies will include concussive patients and incorporate baseline and postinjury sessions to determine meaningful change scores for various injuries.

Postural stability is affected after concussion and musculoskeletal injuries. Therefore, it is important to have a reliable postural-stability assessment available as an assessment tool. Force-plate technology allows for objective assessment of postural stability. The SET demonstrates good reliability, which allows clinicians to objectively evaluate postural stability with little error. However, practice effects do occur after multiple trials, so our study indicates that a practice session should be performed before recording testing trials.

References