Test-retest Reliability of the NeuroCom® VSR™ Sport in Division I Collegiate Female Soccer Players

Crystal Davisson  
Utah State University

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

Part of the Medicine and Health Sciences Commons

Recommended Citation
Davisson, Crystal, "Test-retest Reliability of the NeuroCom® VSR™ Sport in Division I Collegiate Female Soccer Players" (2014). All Graduate Plan B and other Reports. 419.  
https://digitalcommons.usu.edu/gradreports/419

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.
Test-retest Reliability of the NeuroCom® VSR™ Sport in Division I Collegiate Female Soccer Players

By
Crystal Davisson, ATC

A plan B proposal submitted in partial fulfillment of the requirements of the degree

Of
MASTER OF SCIENCE

In
HEALTH AND HUMAN MOVEMENT

Approved:

_________________________    _______________________
Dr. Dennis Dolny      Dr. Breanna E. Studenka
Major Professor      Committee Member

__________________________
Trek Lyons, MD
Committee Member

Utah State University
Logan, UT
2014
Abstract

The NeuroCom® VSR Sport stability evaluation test (SET) protocol was designed to identify postural control deficits in athletes who have sustained a concussion. The purpose of this study was to determine the test-retest reliability of the NeuroCom® VSR Sport SET protocol in non-concussed Division I collegiate female soccer players. A secondary purpose of this study was to determine whether participants displayed a learning effect when tested using the SET protocol across three days. A repeated measures design was used, testing participants on days 1, 7, and 8. Nineteen participants (20.05 ± 1.26 years old) from Utah State University’s women’s soccer team completed data collection. Participants were not excluded based on history of lower extremity injury, concussion, or current illness if it did not affect participation in normal soccer activities. The VSR Sport force plate and NeuroCom® Balance Manager® software was used to complete the SET. The SET protocol instructions were followed, with the modification of all single-leg stances performed on the right leg. Test-retest reliability was calculated using intraclass correlation coefficients (ICC) and typical error. An ICC of > .75 was considered clinically acceptable. Learning effects were analyzed using repeated measures analysis of variance and pairwise comparisons. Composite score ICC values ranged between .56 - .66, indicating moderate clinical reliability. ICC values by condition ranged from .16 - .71. Small to moderate typical error values were found. A significant learning effect was found between days 1 and 8 for both the tandem foam stance and the composite scores. The results indicate that the SET protocol has similar reliability to the reported reliability of the balance error scoring system (BESS).

Keywords: concussion, balance, postural control, sway velocity, NeuroCom, stability evaluation test, VSR sport, balance error scoring system, soccer
Introduction

Sport-related concussions have received an increasing amount of attention over the last few years, both in the news and in research. The American Medical Society for Sports Medicine defines concussion as a traumatically induced transient disturbance of brain function (Harmon et al., 2013). The Centers for Disease Control and Prevention (CDC) estimates sports-related concussion incidents to be between 1.6 and 3.8 million annually in the USA—approximately 5-9% of all sports-related injuries (Harmon, et al., 2013). Some studies show that the incidences of concussion are increasing (Guskiewicz, Weaver, Padua, & Garrett, 2000; Hootman, Dick, & Agel, 2007). Some experts attribute this increase in concussion incidence to improvements in the detection of concussion, as well as a possible increase in concussion rates over time (Daneshvar, Nowinski, McKee, & Cantu, 2011; Hootman et al., 2007). According to the National Athletic Trainer’s Association (NATA), certified athletic trainers care for an average of seven concussive injuries per year (Guskiewicz et al., 2004). Concussions are particularly common in football, ice hockey, lacrosse, women’s soccer, and basketball (Broglio et al., 2007; Daneshvar et al., 2011; Gessel, Fields, Collins, Dick, & Comstock, 2007; Guskiewicz et al., 2000; Marar, McIlvain, Fields, & Comstock, 2012; Harmon et al., 2013). With the number of concussions on the rise, sports medicine team members need to have a concussion evaluation and management protocol that accurately identifies and safely manages athletes with concussions.

Concussion is one of the most complex sports injuries to diagnose and manage (McCrory et al., 2013). Both the NATA and the international conference on concussion in sport recommend a multifaceted approach to evaluating and managing concussions (Guskiewicz et al., 2004; McCrory et al., 2013). The assessment of concussion should include a neurological
examination of mental status, cognitive function, balance, and self-reported signs and symptoms (Guskiewicz et al., 2004; McCrory et al., 2013). Often, when determining whether an athlete has concussive symptoms, medical professionals rely on self-reported symptoms such as headache, sleep disturbances, difficulty concentrating, emotional changes, ringing in the ears, nausea, sensitivity to light and sound, etc. Athletes are often reluctant to report these symptoms to coaches or medical personnel out of fear of losing playing time or losing their spot on the team (Riemann & Guskiewicz, 2000). Additionally, signs and symptoms may resolve soon after injury, regardless of the injury severity (Guskiewicz et al., 2000). The transient nature of these self-reported symptoms can make it difficult for clinicians to determine the severity of a concussion, and in turn, when the athlete can safely return to participation in athletic activities.

Once an athlete has sustained a concussion, he/she is at an increased risk for additional concussions (Guskiewicz et al., 2004). This risk is higher in younger athletes and for those who return to play before complete recovery from an initial concussion (Guskiewicz et al., 2004). Warnings about returning to play too soon, and the dangers of multiple concussions, are becoming stronger as more research on concussions is conducted. Recent research on former NFL players with histories of concussion indicates a possible link between multiple concussions and Alzheimer-like changes in the brain (Harmon et al., 2013). Athletes who either return to play too soon or who experience multiple concussions can experience learning deficits or chronic neurocognitive impairment (Harmon et al., 2013; Guskiewicz et al., 2004). Some concussion experts also warn of a condition called second-impact syndrome, a condition where athletes who return to play before recovering from a concussion sustains a second head injury (Bey & Ostick, 2008; Guskiewicz, Ross, & Marshall, 2001). This second brain injury can result in diffuse cerebral swelling, brain herniation, and death (Bey & Ostick, 2008). Because of the risks of
returning to play too soon, the NATA suggests clinicians use objective assessment tools to
measure injury severity and monitor the athlete’s recovery, rather than relying solely on
subjective measures, such as self-reported symptoms (Guskiewicz et al., 2004). In addition to
the subjective symptoms of concussion, multiple studies have shown balance deficits in those
with concussions. According to Guskiewicz (2011), balance problems are the 6th most commonly
reported concussive symptom, and occur in 30% of concussed individuals. Balance assessment is
an objective measure that can be used to evaluate concussive symptoms (Guskiewicz, 2011;
Guskiewicz et al., 2004). This paper will focus on the use of balance assessment as a reliable,
objective measure of concussion.

Guskiewicz (2011) defines balance as the process of maintaining the center of gravity
within the body’s base of support. The ability to maintain balance is called postural control
(Pollock, Durward, Rowe, & Paul, 2000). The human vestibular system helps to maintain
balance, in conjunction with visual and somatosensory inputs (Guskiewicz, 2011; Hytonen,
Pyykko, Aalto, & Starck, 1993). The vestibular system works together with the eyes and
somatosensory system to perform collective actions called postural reflexes. The central nervous
system maintains the body’s upright posture by relying on sensory organization and on muscle
coordination. Sensory organization is what determines the timing, direction, and amplitude of
corrective postural actions, or postural reflexes. These reflexes rely on information from the
visual, vestibular, and somatosensory inputs. Concussions are thought to interfere with the
communication between the visual, vestibular, and somatosensory systems, thus causing balance
deficits. Concussion symptoms such as dizziness, vertigo, lightheadedness, tinnitus, and blurred
vision also indicate that these systems are affected by concussion. Guskiewicz (2011) suggests
that of the three systems that regulate balance, the vestibular system is the most commonly
affected system following concussion. Because of the correlation between concussion and postural control, most experts suggest utilizing postural control (balance) testing as an objective part of the concussion evaluation (Broglio & Macciocchi, 2007; Guskiewicz, Perrin, & Gansneder, 1996; Guskiewicz et al., 2004; Guskiewicz et al., 2001; McCrory et al., 2013).

There are several different ways to measure postural control, the two most common of which are the balance error scoring system (BESS) and using a force plate. Force plates measure center of pressure, which can then be used to measure postural control as a function of sway velocity (the angular motion of the body’s center of pressure, measured in deg/sec). Some force platforms can be expensive or difficult to use, so the BESS was created to measure postural control without using a force plate (Loughran, Tennant, Kishore, & Swan, 2005). Administering the BESS involves having the athlete stand in three different stances (a narrow double-leg, single-leg, and tandem), on two different surfaces (firm and on a foam pad), and recording the number of errors the participant makes (Broglio, Zhu, Sopiarz, & Park, 2009; Harmon et al., 2013; Valovich, Perrin, & Gansneder, 2003). The participant is instructed to stand in the designated stance for 20 seconds with his/her eyes closed, hands on his/her iliac crests, and to stand as motionless as possible. Examples of errors include the participant opening his/her eyes, lifting their hands off his/her hips, or stepping, stumbling, or falling out of the test stance (Harmon et al., 2013). Clinicians assessing postural control in a concussed athlete interpret an increase in postural sway and/or an increase in BESS errors as an indication of decreased postural control secondary to concussion (Guskiewicz, Perrin, & Gansneder, 1996). Most studies on the BESS have reported that its sensitivity is greatest immediately after injury, and that most athletes’ BESS scores returned to baseline within 3-7 days post-injury (Harmon et al., 2013).
Some researchers question the test-retest reliability of the BESS. Test-retest reliability is a measure of the consistency of the BESS when the same individual performs the test multiple times under the same conditions. Correlations above .75 indicate an excellent reliability, with some researchers contending that ideal test-retest reliability for clinical applicability is > 0.80 (Broglio et al., 2009; Fleiss, 1986). Correlations lower than 0.75 are often attributed to a practice, or learning, effect with multiple administrations of the BESS. Studies have indicated that the test-retest reliability of the BESS is below an acceptable level: Hunt, Ferrara, Bornstein, and Baumgartner (2009) reported an ICC of .60; Broglio et al. (2009) reported reliabilities ranging from .60-.63; and McLeod, Barr, McCrea, and Guskiewicz (2006) reported an ICC of .70 for the BESS.

Because the BESS relies on a subjective assessment of balance, some researchers suggest using a more sensitive measure of postural control when possible, such as that attained on a force platform, rather than relying on individuals counting errors. Guskiewicz (2011) suggests that measuring sway on a force platform can provide valuable information for determining return to play. The gold standard for balance testing is the Sensory Organization Test (NeuroCom International, Inc, Clackamas, Oregon, USA; Harmon, et al., 2013). It uses a force plate system to measure balance under several conditions. Research conducted using the Sensory Organization Test (SOT) has detected balance and sensory interaction deficits up to three days following concussion (Harmon, et al., 2013). Broglio, Macciocchi, and Ferrara (2007) found that the SOT identified balance deficits in 61.9% of individuals with concussion. While the SOT is a useful tool for evaluating concussive symptoms, it is a large, expensive system, and clinical availability is limited (Harmon, et al., 2013). Another downside to the SOT is that the system is
not portable, and therefore not practical for sideline testing or for testing athletes while on the road.

Recently NeuroCom® has developed a device called the VSR™ Sport, designed for use as a portable means of concussion testing. The VSR™ Sport system uses a portable force plate and computer system designed for testing balance. The stability evaluation test (SET) protocol of the VSR™ Sport follows the BESS protocol, but instead of an observer measuring errors, the force platform measures and records changes in sway velocity. The program displays results in a graphical form, with center of pressure (COP) traces for each condition as well as reporting a weighted average composite score of sway velocity in deg/sec (“VSR™ Sport: Play a Smarter Game,” 2011). NeuroCom® states that the SET is specifically designed for the athletic population to assist clinicians make safer return to play decisions following concussion.

To date, few studies have been conducted testing the reliability of the VSR™ Sport. Hanline and Olsen (2013) tested the VSR™ Sport on a pediatric sample (n = 54) with no known history of lower extremity injury or concussion. Participants were tested on days 1, 30, and 35 to assess the test-retest reliability of the SET protocol. The researchers found strong test-retest reliability (r = .86 - .92) for the SET in a pediatric sample. As reported from a collegiate student research showcase, Saba and Wrisley (2013) conducted a study testing the reliability of the SET and the BESS in athletes (unreported number) with a mean age of 21 ± 1.4 years. Each athlete was tested twice, seven days apart. Reliability was calculated using the Intraclass Correlation Coefficient (ICC). Interrater reliability (the number of raters was unreported) for the BESS was 0.76, and intrarater reliability was 0.28. Test-retest reliability for the SET was moderate at 0.69. While the author states that the reliability for the SET is better than that of the BESS, the value
of 0.69 still falls below the 0.75 value suggested by Fleiss (1986). Further research needs to be conducted to determine the test-retest reliability of the SET protocol in different populations.

**Purpose**

The purpose of this study was to determine the test-retest reliability of the NeuroCom® VSR™ Sport SET protocol in Division I collegiate female soccer players. Knowing the reliability of SET can help clinicians determine its place in the concussion assessment protocol. If the SET has good reliability, the values obtained from a post-injury test can be compared to the athlete’s baseline scores; an analysis comparing the pre- and post-injury scores may be useful in determining whether an athlete’s recovery is sufficient to return to play; further studies would need to be completed to determine what amount of variation between pre- and post-injury scores indicates a postural control deficit. Having SET values can provide an additional data point for clinicians to use in their concussion protocol to determine safe return to play. As previously stated, the potentially devastating results of an athlete returning to play before proper healing has occurred is the chief concern of those caring for the athlete. Based on the studies by Hanline and Olsen (2013), and by Saba and Wrisley (2013), the reliability of the SET protocol was expected to be moderate to good, approaching the clinically significant value of .80 for both individual stances and for the composite score.

A secondary purpose of the study was to determine if the participants display a learning effect—for individual stances or composite scores—when tested using the SET protocol across days 1, 7, and 8. A learning effect was not anticipated, based on the studies by Valovich et al. (2003) and by Broglio et al. (2009).
Design

A repeated measures design was used to determine the test-retest reliability of the NeuroCom® VSR™ Sport Stability Evaluation Test (SET) protocol. As three different trials were needed to calculate reliability, participants were tested on days 1, 7, and 8. McCrea et al (2003) reported that by day seven following concussion, only 9% of athletes had balance impairments. Thus the range between days one and eight is a conservative measure of the duration of a typical concussion. For the purposes of this study, day one represented the immediate, post-injury measure, while days seven and eight represented the latest days an athlete would be tested following a concussion.

Subjects

Twenty two participants were recruited from the Utah State University women’s soccer team. Three of the participants did not have complete data for all three testing days, two due to scheduling conflicts, and one due to errors in data collection. The total number of participants whose results were analyzed was 19. Participants’ ages ranged from 18-22 years old (20.05 ± 1.26 years). Due to the limited number of athletes on the women’s soccer team, participants were not excluded based on history of lower extremity injury, concussion, or current illness if it did not affect participation in normal soccer activities. Participants were screened prior to the study to determine any history of a lower extremity musculoskeletal injury within the last year, any concussion history, and any illness that could affect inner ear function and/or balance during data collection. Table 1 lists the participants’ self-reported injury and illness histories. Only one of the
participants had sustained a concussion within 6 months prior to the study. Reported illnesses included three participants with colds or runny noses, one just recovering from stomach flu, and one recovering from bronchitis. Both athletes recovering from the flu and bronchitis were cleared for participation in practices and games at the time of data collection. All participants were required to give written informed consent before participating in the study.

Instrumentation

Data was collected using the VSR™ Sport force plate and the NeuroCom® Balance Manager® software (NeuroCom International, Inc., Clakamas, OR). The Stability Evaluation Test (SET) protocol was selected within this software. The NeuroCom® Balance Manager® software measures postural control as a function of sway velocity, reported in deg/sec. The calculation used to determine sway velocity in the SET protocol is not available to the public. According to the interpretation suggested by NeuroCom, a larger score represents a potential balance deficit. Users have the option to compare SET scores to NeuroCom’s normative data range (for ages 14-25) to assist in identifying potential balance deficits. NeuroCom has not released the source(s) of this normative data.
### Table 1

**Incidence of Participants’ Self-reported Injury & Illness History**

<table>
<thead>
<tr>
<th>Participant</th>
<th>Concussion</th>
<th>LE Injury</th>
<th>LE Surgery</th>
<th>Current Illness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Y</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Y</td>
</tr>
</tbody>
</table>

*Note.* Concussion numbers reported include all concussions participant has sustained. Only participant 1 had sustained a concussion within 6 months prior to data collection. LE = Lower Extremity.
Procedures

Each participant completed the full Stability Evaluation Test (SET) three times, over a period of 8 days. The VSR™ Sport SET protocol consists of six conditions/stances. These include: double leg firm, single leg firm, tandem (right foot behind) firm, double leg foam, single leg foam, and tandem (right foot behind) foam. Each condition required the participant to place her hands on her iliac crests, and to close her eyes. The SET protocol called for the single leg stance to be performed on the participant’s non-dominant leg, however to keep consistent with concussion testing protocol at Utah State University all single leg stances were performed on the right leg. Participants were given instructions on how to perform the stance for each condition prior to testing. The researchers demonstrated each stance prior to beginning the trial condition. Participants were reminded to maintain the stance to the best of their ability for the duration of the testing period. They were instructed that if they moved out of the testing stance, they were to return to the stance as quickly as they could. The researcher started the trial once the participant was in the required position and had their eyes closed. The SET protocol collects data for each stance for 20 seconds, however the trial concludes early if the participant steps off the platform. The researchers then selected the option to mark the trial as a fall and continue on to the next stance. Trials marked as falls were still included in the SET output, with a sway velocity given for the duration of the stance. These trials were also used in determining the composite score, however NeuroCom does not explain if/how these shorter trials are weighted in determining the composite score.
Reliability was calculated for the three testing days using intraclasse correlation coefficients (ICCs) and typical error (as a percent of the coefficient of variation). A confidence interval of 95% was chosen a priori. An ICC less than .40 indicated poor reliability, between .40 and .75 indicated fair to good reliability, and an ICC greater than .75 was considered excellent reliability (Fleiss, 1986). For the purposes of this paper, ICC values in the fair to good range presented by Fleiss were labeled “moderate” reliabilities. The ICC is a ratio of the variance of the between subject differences to the variability of the sample as a whole (Lin, Seol, Nussbaum, & Madigan, 2008). Consequently, little variability among the participants (low levels of between subject differences) can lead to lower ICC values (Santos, Delisle, Lariviere, Plamondon, & Imbeau, 2007; Weir, 2005). This is of particular concern when testing a homogeneous population (Santos et al., 2007; Weir, 2005). Typical error can be used to distinguish whether low ICC values are due to within subject variability or are due to low sample variability (Henriksen, Lund, Moe-Nilssen, Bliddal, & Danneskoid-Samsoe, 2003). Smith and Hopkins (2011) suggest interpreting typical error values according to the following scale: 0.3, 0.9, and 1.6 as small, moderate, and large magnitudes, respectively (reported in the current study as 30%, 90%, and 160% percent typical error). A high typical error indicates a high level of error and implies that the test values are non-reproducible, while a low typical error value may indicate that the test itself may have good reliability, but that the test results were affected by low sample variability (Lin et al., 2008).

ICCs and typical error calculations were completed using the template provided by Hopkins (Hopkins, 2011). Explanations of Hopkins’ calculations are found in his article on...
reliability (Hopkins, 2000). Additionally, repeated measures analysis of variance was used to analyze the changes in sway velocity between the three days to identify any learning effects. Parametric data sets were analyzed with ANOVA tests and non-parametric data were analyzed using Friedman’s test. When a main effect was indicated, pairwise comparisons were used for parametric data, and Wilcoxon signed ranks test was used for non-parametric data. Bonferroni adjustments were applied to all post-hoc tests. Analysis of variance and post hoc tests were run using SPSS version 21 (SPSS IBM, New York, U.S.A.).

Results

All data sets were evaluated to determine normality and to check for outliers. Data sets with non-parametric distributions were log-transformed before correlation tests were run (Hopkins, 2011). No significant outliers were present. Mean sway velocity scores are reported in Table 2. Most of the conditions showed slightly higher sway velocity scores on day one than on day eight. The greatest difference is noted in the tandem foam stance, where the scores for days one and seven were 4.0, and went down to 3.2 for day eight. Table 3 displays the typical error for each condition and for the composite score. The mean typical error for individual stances ranged from 12.7% – 36.0%, with the lowest error in the double leg foam stance, and the highest in the tandem foam stance. The composite score mean typical error was 12.8%, indicating a small magnitude of typical error.
### Table 2

**Mean Sway Velocity Scores (with Standard Deviations in Parentheses)**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Day 1</th>
<th>Day 7</th>
<th>Day 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double Leg Firm</td>
<td>0.7 (0.2)</td>
<td>0.6 (0.1)</td>
<td>0.7 (0.2)</td>
</tr>
<tr>
<td>Single Leg Firm</td>
<td>1.8 (0.5)</td>
<td>1.6 (0.3)</td>
<td>1.6 (0.5)</td>
</tr>
<tr>
<td>Tandem Firm</td>
<td>1.4 (0.5)</td>
<td>1.4 (0.4)</td>
<td>1.3 (0.5)</td>
</tr>
<tr>
<td>Double Leg Foam</td>
<td>1.9 (0.3)</td>
<td>1.7 (0.3)</td>
<td>1.8 (0.3)</td>
</tr>
<tr>
<td>Single Leg Foam</td>
<td>3.2 (0.9)</td>
<td>3.2 (0.9)</td>
<td>3.0 (0.7)</td>
</tr>
<tr>
<td>Tandem Foam</td>
<td>4.0 (2.1)</td>
<td>4.0 (2.3)</td>
<td>3.2 (1.5)</td>
</tr>
<tr>
<td>Composite</td>
<td>2.2 (0.5)</td>
<td>2.1 (0.4)</td>
<td>1.9 (0.3)</td>
</tr>
</tbody>
</table>

*Note. N = 19. Sway velocity scores are reported in deg/sec.*

### Table 3

**Typical Error as a Coefficient of Variation**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean Typical Error</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double Leg Firm</td>
<td>18%**</td>
<td>14.7-23.5%</td>
</tr>
<tr>
<td>Single Leg Firm</td>
<td>17.6%**</td>
<td>14.4-23.0%</td>
</tr>
<tr>
<td>Tandem Firm</td>
<td>22.4%**</td>
<td>18.2-29.3%</td>
</tr>
<tr>
<td>Double Leg Foam</td>
<td>12.7%*</td>
<td>10.4-16.5%</td>
</tr>
<tr>
<td>Single Leg Foam</td>
<td>23.8%**</td>
<td>19.3-31.2%</td>
</tr>
<tr>
<td>Tandem Foam</td>
<td>36.0%**</td>
<td>28.9-47.9%</td>
</tr>
<tr>
<td>Composite</td>
<td>12.8%*</td>
<td>10.5-16.6%</td>
</tr>
</tbody>
</table>

*Note. Smith & Hopkins (2011) suggest doubling the typical error before determining the magnitude. Doubled error values < 30% represent a *small magnitude and those between 30 – 90% ** represent a moderate magnitude.*
Intraclass correlation coefficients (ICC) are reported in Table 4. Composite score ICC values ranged between .56-.66 (moderate clinical reliability). Intraclass correlations by condition ranged from .16-.71, with the highest ICC being for the tandem foam stance between day 1 and day 8, and the lowest ICC being for the single leg foam stance between days 7 and 8.

Table 4

Test-retest reliability (ICC) of the SET Mean Sway Velocity for each condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Timeframe 1 (Day 1 – Day 7)</th>
<th>Timeframe 2 (Day 7- Day 8)</th>
<th>Timeframe 3 (Day 1 – Day 8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double Leg Firm</td>
<td>.58*</td>
<td>.31</td>
<td>.46*</td>
</tr>
<tr>
<td>Single Leg Firm</td>
<td>.49*</td>
<td>.54*</td>
<td>.39</td>
</tr>
<tr>
<td>Tandem Firm</td>
<td>.61*</td>
<td>.63*</td>
<td>.66*</td>
</tr>
<tr>
<td>Double Leg Foam</td>
<td>.49*</td>
<td>.54*</td>
<td>.54*</td>
</tr>
<tr>
<td>Single Leg Foam</td>
<td>.41*</td>
<td>.16</td>
<td>.24</td>
</tr>
<tr>
<td>Tandem Foam</td>
<td>.41*</td>
<td>.55*</td>
<td>.71*</td>
</tr>
<tr>
<td>Composite</td>
<td>.56*</td>
<td>.66*</td>
<td>.65*</td>
</tr>
</tbody>
</table>

* Moderate clinical reliability between .40 - .75 (Fleiss, 1986).

Table 5 displays Friedman and ANOVA test results. The double leg foam results were parametric, so ANOVA was run to check for learning effects. All other data was non-parametric, so the Friedman test was used. Significant differences were identified for the double leg foam stance, tandem foam stance, and for the composite scores.
Table 5

*Evaluation of learning-effects: Friedman and ANOVA Tests*

<table>
<thead>
<tr>
<th>Condition</th>
<th>df</th>
<th>$\chi^2/F$ value</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friedman Tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Double Leg Firm</td>
<td>2</td>
<td>3.364</td>
<td>.186</td>
</tr>
<tr>
<td>Single Leg Firm</td>
<td>2</td>
<td>2.638</td>
<td>.267</td>
</tr>
<tr>
<td>Tandem Firm</td>
<td>2</td>
<td>5.631</td>
<td>.060</td>
</tr>
<tr>
<td>Single Leg Foam</td>
<td>2</td>
<td>2.493</td>
<td>.287</td>
</tr>
<tr>
<td>Tandem Foam</td>
<td>2</td>
<td>6.811</td>
<td>.033*</td>
</tr>
<tr>
<td>Composite Score</td>
<td>2</td>
<td>7.258</td>
<td>.027*</td>
</tr>
<tr>
<td>ANOVA Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Double Leg Foam</td>
<td>2</td>
<td>3.639</td>
<td>.036*</td>
</tr>
</tbody>
</table>

*Note. N = 19. * $p < .05.*

Tables 6 and 7 display the post-hoc analyses of data that identified significant differences between days. Both the tandem and composite scores showed that the significant difference was present between days one and eight. Post-hoc analysis for the double leg foam stance did not identify a statistically significant difference between any of the test days. Figure 1 presents test-retest composite SET scores plotted by day. A positive correlation is noted between days: participants who had a higher composite score on day one tended to have higher scores on days seven and eight as well.
Table 6

*Post-hoc Tests: Wilcoxon Signed Ranks Test with Bonferroni Adjustment*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Day 1: Day 7</th>
<th></th>
<th>Day 7: Day 8</th>
<th></th>
<th>Day 1: Day 8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Z score</td>
<td>p</td>
<td>Z score</td>
<td>p</td>
<td>Z score</td>
<td>p</td>
</tr>
<tr>
<td>Tandem Foam</td>
<td>-0.644</td>
<td>.519</td>
<td>-1.942</td>
<td>.052</td>
<td>-2.38</td>
<td>.017*</td>
</tr>
<tr>
<td>Composite Scores</td>
<td>-1.071</td>
<td>.28</td>
<td>-1.614</td>
<td>.11</td>
<td>-2.489</td>
<td>.013*</td>
</tr>
</tbody>
</table>


Table 7

*Post-hoc Tests: Pairwise Comparison with Bonferroni Adjustment*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Day 1: Day 7</th>
<th></th>
<th>Day 7: Day 8</th>
<th></th>
<th>Day 1: Day 8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean difference</td>
<td>p</td>
<td>Mean difference</td>
<td>p</td>
<td>Mean difference</td>
<td>p</td>
</tr>
<tr>
<td>Double Leg Foam</td>
<td>0.189</td>
<td>.059</td>
<td>0.079</td>
<td>.742</td>
<td>0.111</td>
<td>.415</td>
</tr>
</tbody>
</table>


Figure 1. Composite Mean Sway Velocity Scores Plotted By Day

*Figure 1. Sway velocity tables created using Hopkins’ spreadsheet template (Hopkins, 2011).*
Discussion

The purpose of this study was to determine the test-retest reliability of the NeuroCom® VSR™ Sport SET protocol in Division I collegiate female soccer players using measures of intraclass correlation coefficients (ICCs) and typical error. The results of this study indicated poor to moderate ICC values with small to moderate typical error.

The results of the current study indicate lower test-retest reliability for the SET protocol than those reported by Hanline and Olsen (2013). Hanline and Olsen reported good test-retest reliability for composite scores, with ICCs ranging from .86 to .92, while the current study had a moderate reliability with composite ICCs ranging from .56 - .66. Hanline and Olsen’s reported ICCs for individual stances indicated moderate to good reliability between all timeframes, except for timeframe 3 (between test days 1 and 35) of the double leg foam stance (ICC of .47). Their study tested a pediatric population with no known history of injury or condition within the 6 months prior to testing that could have affected the participants’ balance at the time of testing. The current study did not exclude participants with a history of lower extremity injury, concussion, or illness if the participant was cleared to participate in normal soccer activities. Had the current study contained only healthy participants with no injury history within the past six months and no current illnesses, ICC values may have approximated those reported by Hanline and Olsen.

To further investigate the meaning of the current study’s ICC values, refer to the mean typical error values in Table 3. Typical error values were doubled before interpreting the magnitude as per the suggestion of Smith and Hopkins (2011) and Hopkins (2011). The composite and double leg foam stances had a small magnitude of typical error, while all other
stances had a moderate typical error (Smith & Hopkins, 2011). These small to moderate typical error values suggest that the low ICC values can be partially attributed to low sample variability. The sample tested in the current paper was a homogenous population; all participants were athletic, Division I female soccer players within a narrow age-range (18-22 years; 20.05 ± 1.26 years). Santos et al. (2008) also reported low ICC values when studying a small population of college students (n = 12, age = 26.9 ± 4.7 years). In contrast, the study conducted by Hanline and Olsen (2013) using the SET protocol was comprised of a pediatric population whose ages ranged between 5-15 years, and who participated in various sports; they reported ICCs of .86-.92 for composite sway velocity scores. Hytonen et al. (1993) reports that children and elderly persons have the greatest amount of sway when compared to other populations. Based on this information from Hytonen et al., it is possible that the greater amount of sway in these populations can result in greater ICC values than those found in other populations with lower levels of between subject variability. Thus it is possible that the use of a pediatric population may have resulted in greater amounts of sway, partially contributing to the strength of the ICC values reported by Hanline and Olsen. Lin et al. (2008) conducted a study comparing the sway velocity during an eyes-closed, double leg stance, between an older population and a college-aged population. They reported a greater range in sway velocity for the older population, as well as the older population having a stronger reliability than the younger population (ICCs of .91-.92 and .77-.79, respectively). Based on these studies by Hytonen et al. and Lin et al., it is possible that the narrow age range of the population used in the current study led to a lower variability in sway velocity scores, contributing to the lower ICC values compared to those reported by Hanline and Olsen (2013).
According to the current study, the composite SET scores had a small magnitude of typical error, indicating that the composite scores may be more reproducible than the individual conditions. Neurocom® suggests that the composite score is the most important score when determining whether a participant has a postural control deficit related to concussion. This recommendation aligns with the clinical application of the BESS, where clinicians typically use the total score to assess postural control deficits.

A second difference between the present and Hanline and Olsen’s (2013) study were the instructions given to the participants. Hanline and Olsen instructed participants to try to stay on the force plate during testing and informed them of acceptable errors participants could make if they felt like they were going to fall off the platform or lose their balance. Acceptable errors included opening eyes, taking hands off hips, and touching their non-weight-bearing foot to the ground (for single leg conditions). In the current study, participants were instructed according to the prompts in the SET protocol, which did not include instruction on acceptable errors or the instruction to attempt to remain on the force plate. Had participants in the current study been informed that it was acceptable for them to open their eyes briefly in order to regain balance or to remain on the force plate, participants may have attained better, and more consistent, sway velocity scores. In Hanline and Olsen’s study, if a trial ended early due to the participant stepping off the platform, the participant was given a second trial. For the current study, any trial that ended before the allotted time due to the participant stepping off the platform was marked as a fall, and no further testing was done for that stance. The majority of participants remained on the platform for the duration of the test; however 8 of the 19 participants stepped off the platform prematurely during at least one condition. Two of the eight participants whose trials ended prematurely consistently stepped off the platform as soon as they started losing their balance. A
total of 342 trials were performed, with only 20 of those trials ending prematurely. This equates to 5.85% of trials being marked as a fall. Had participants been given the same instructions and been given a second trial as in Hanline and Olsen’s study, within-subject variation would likely have decreased, leading to higher test-retest reliability. The current study’s aim was to evaluate the SET protocol based on how the instructions indicate its use. The researchers chose to have each participant perform only one trial for each condition as the protocol does not instruct clinicians to perform more than one trial.

The SET protocol is based on the balance error scoring system (BESS), the most commonly used clinical test used by sports medicine professionals to assess post-concussion balance impairment. Some researchers and clinicians question the interrater and intrarater of the BESS. Interrater reliability is the degree of agreement between two or more individuals scoring the BESS. A low interrater reliability is of concern when the clinician who administers a post-injury BESS is not the clinician who administered the athlete’s baseline test. Intrarater reliability is a function of how consistently a single individual scores the BESS over multiple repetitions of the test. Finnoff, Peterson, Hollman, and Smith (2009) conducted a study evaluating the intrarater and interrater reliability of the BESS in which three scorers viewed and rated 30 BESS tests two different times. Reliability was calculated for each stance, using intraclass correlation coefficients (ICCs). The researchers reported that the intrarater reliability ranged from 0.50 to 0.88, with the single-leg foam stance the lowest. Interrater reliability ranged from 0.44 to 0.83, with the lowest correlation the firm tandem stance position. When ICCs were calculated for all the stances combined, intrarater reliability was 0.74 and interrater reliability was 0.57. Most clinicians use the total BESS score when evaluating BESS performance; based on the ideal reliability of 0.80, the total BESS scores in the Finnoff et al. (2009) study were not reliable.
Another concern in using the BESS to evaluate injured athletes is determining how much variation from baseline scores indicates decreased postural control due to concussion rather than differences due to the individual scoring the test. Compared to baseline measure, currently, there is no accepted method for determining how much increase in BESS errors indicates a balance deficit, so institutions and practitioners are forced to determine a number on their own (McLeod et al., 2006). Utah State University’s concussion protocol states that an athlete’s BESS score must be ± 3 of their baseline score (Finnoff & Mildenberger, 2003). McLeod et al. (2006) also suggested that an increase in 3 or more BESS errors indicated balance impairment within a 70% confidence interval (CI). While the results of their study indicated that an increase of 6 errors fell within the 90% CI, McLeod et al. state that this value may be too conservative for sports-related concussion because of the subtle and transient nature of concussion symptoms.

Finnoff et al. (2009) also investigated what change in BESS scores indicated a balance deficit, by calculating the minimum detectable change (MDC) for intrarater and interrater BESS scores. Minimum detectable change was defined as “the number of points required to represent a change in the subject’s postural stability rather than a reflection of inter- or intrarater scoring variability” (p. 53). In this definition, postural stability is synonymous with the term postural control in the current paper. Based on Finnoff et al.’s findings, a subject would have to score 7.3 points greater than his baseline score to determine that the change in score was not due to intrarater variation. The interrater MDC was 9.4, which the authors attribute to greater interrater scoring variability. These high MDC values question whether the BESS accurately assesses postural control deficits in concussed athletes.
The typical BESS score immediately after injury is an average of 5.7 points above baseline, and this difference decreases to 2.7 one day after injury (Guskiewicz, 2011). Both of these values are significantly lower than the MDCs found in the Finnoff et al. (2009) study. Mulligan, Boland, and Payette (2012) reported that athletes who do not have a concussion may still score significantly below their baseline BESS score when tested in a fatigued state, such as after a game. Due to confounding factors such as inter- and intra-rater reliability and fatigue, and the lack of a standard method for interpreting BESS scores to determine balance deficits, it can be difficult for clinicians to determine what variation in BESS scores indicates a postural control deficit due to concussion. The SET protocol measures balance deficits as a function of sway velocity, rather than errors; it is possible that, with additional studies, the more precise measure of postural control as a function of sway velocity may yield a range of values that can be used to better determine what variation in sway velocity may indicate a postural control deficit due to concussion.

While the current study resulted in only moderate test-retest reliabilities (.56-.66) for the composite scores in the SET protocol, reported test-retest reliabilities for the BESS (.57-.74) also fall below the acceptable level of .75 to .80 (Broglio et al., 2009; Finnoff et al., 2009; Hunt et al., 2009; McLeod et al., 2006). Thus the researchers conclude that the SET protocol can be considered to have a similar level of reliability as the BESS, the test that is most often used by clinicians to determine post-concussion postural control deficits.

Studies indicate that the reliability of the BESS, and thus the reliability of the SET, may improve when more trials are performed. Multiple performances of balance tests however may yield a practice effect. This is of concern when testing athletes daily to evaluate an athlete’s improvement in concussion-related balance deficits. A secondary purpose of the current study
was to determine whether participants displayed a learning effect—for individual stances or for composite scores—when tested using the SET protocol across days 1, 7, and 8. Broglio et al. (2009) reported that individuals displayed a learning effect when tested up to five times, but suggest that performing 2-3 trials of the BESS yields a reliable and clinically acceptable measure of balance performance without a significant learning effect. Hunt et al.’s (2009) study investigated the ICCs for up to seven trials of a modified BESS protocol in which the double leg conditions were omitted. They reported the ICC for the modified 4-condition protocol increased from .60 to .71. Reliability increased with multiple trials, with a range of .83 with two trials to .88 with 3 trials. Hunt et al. reported that a significant practice effect occurred between trials one and two, and thus recommend a practice trial be administered before scoring the BESS. When three trials were performed and the first trial was omitted, Hunt et al. reported an ICC of .84. Based on these findings, Hunt et al. recommended performing 3 trials of their modified 4-condition BESS, without using the first trial. They suggest, however, that further research be conducted using the 4-condition protocol on concussed athletes to determine if the double leg conditions are necessary for a practice or control stance to maintain the clinical sensitivity of the BESS.

ANOVA and Friedman tests were used to identify learning effects in the current study. These tests revealed significant differences in the double leg foam stance, tandem foam stance, and composite scores. Although post-hoc tests did not reveal where the significant difference was for the double leg foam stance (see Table 7), the mean difference between day 1 and 7 was greater (with the p-value approaching the significant p-value of .05), than the differences between days 7 and 8, and days 1 and 8. Post-hoc tests for the tandem foam and composite scores, revealed a significant difference between test days 1 and 8 for both the tandem and
composite stances. These significant main effects indicate that a practice effect was present for the SET protocol in the current study. When evaluating the mean sway velocity scores, the tandem foam stance has both a larger sway velocity and a greater standard deviation (1.5-2.3) than any other stance. The tandem foam stance also displays the highest typical error value. It is possible that the learning effect noted with the composite scores may be disproportionately influenced by the tandem scores compared to any of the other stances. As a result, clinicians may want to compare scores on individual stances, as well as the composite score, in determining whether an athlete has regained normal pre-concussion balance.

The current study has several limitations. Previously mentioned is the lack of a “healthy” population. The researchers contend, however, that if an athlete is healthy enough to participate in normal soccer activities, in which the athlete has a potential risk of sustaining a concussion, the athlete is healthy enough to complete the balance protocol of the SET. Additionally, when athletes are baseline tested, or are tested following a concussion, they are rarely without a history of an injury within the six months prior to the test. The researchers consider the population used in the current study to be a “normal” collegiate athlete population, despite the participants being what other researchers may consider “unhealthy”.

Two other limitations to the current study were the location where the SET protocol was performed, and the number of participants used in the study. All baseline and post-concussion testing at Utah State University occurs in a quiet, narrow room attached to the athletic training room. This room has countertops on both sides, which allows participants to “catch” themselves with their hands on the countertop. Although it is impossible to know for certain, had the test been conducted in a wider room, more participants may have stepped off the force platform prematurely, rather than simply catching themselves and continuing on with the trial. Ideally, the
current study would also have included more participants, however as participation was voluntary, the test was limited to 19 participants’ data.

Further studies involving the VSR™ Sport SET protocol should consider baseline testing athletes using 2-3 trials consecutively, as suggested Broglio et al. (2009) and Hunt et al (2009), to determine whether this increases the reliability of the SET protocol. Another consideration for further research is to eliminate the double leg stance conditions from the SET protocol to see if the composite scores are more reliable, as reported by Hunt et al.’s study on the BESS.

Conclusions

This study suggests that the NeuroCom® VSR™ Sport Stability Evaluation Test protocol has moderate test-retest reliability in Division I collegiate female soccer players. Further research should evaluate the test-retest reliability of the SET protocol in other Division I collegiate athletes. Additional consideration should be made for testing a normal collegiate athlete population, rather than limiting research to include only athletes with no history of concussion or recent lower extremity injuries. This additional research can help determine how strong the test-retest reliability of the SET is, which can help determine the SET’s usefulness in determining an athlete’s post-concussion recovery. Once test-retest reliability is established, further studies need to be completed to determine what amount of variation between pre- and post-injury scores indicates a postural control deficit. Further research should also examine the specificity and sensitivity of the SET in the diagnosis of concussions.

This study, along with other studies involving the SET protocol, indicates a moderate to high reliability (Hanline & Olsen, 2013; Saba & Wrisley, 2013). As the SET protocol has a
similar reliability to the BESS test, when available, the SET protocol could be used in place of the BESS test as part of a concussion testing protocol. While the current findings on the test-retest reliability of the SET do not justify using the VSR™ Sport instead of the BESS, the SET does provide a more objective measure of postural control by measuring sway velocity instead of balance errors.
References


McCrea, M., Guskiewicz, K. M., Marshall, S. W., Barr, W., Randolph, C., Cantu, R. C., …Kelly,


