Development of a Clinical Static and Dynamic Standing Balance Measurement Tool Appropriate for Use in Adolescents

Background and Purpose. There is a need in sports medicine for a static and dynamic standing balance measure to quantify balance ability in adolescents. The purposes of this study were to determine the test-retest reliability of timed static (eyes open) and dynamic (eyes open and eyes closed) unipedal balance measurements and to examine factors associated with balance. Subjects. Adolescents (n=123) were randomly selected from 10 Calgary high schools. Methods. This study used a repeated-measures design. One rater measured unipedal standing balance, including timed eyes-closed static (ECS), eyes-open dynamic (EOD), and eyes-closed dynamic (ECD) balance at baseline and 1 week later. Dynamic balance was measured on a foam surface. Reliability was examined using both intraclass correlation coefficients (ICCs) and Bland and Altman statistical techniques. Multiple linear regressions were used to examine other potentially influencing factors. Results. Based on ICCs, test-retest reliability was adequate for ECS, EOD, and ECD balance (ICC=.69, .59, and .46, respectively). The results of Bland and Altman methods, however, suggest that caution is required in interpreting reliability based on ICCs alone. Although both ECS balance and ECD balance appear to demonstrate adequate test-retest reliability by ICC, Bland and Altman methods of agreement demonstrate sufficient reliability for ECD balance only. Thirty percent of the subjects reached the 180-second maximum on EOD balance, suggesting that this test is not appropriate for use in this population. Balance ability (ECS and ECD) was better in adolescents with no past history of lower-extremity injury. Discussion and Conclusion. Timed ECD balance is an appropriate and reliable clinical measurement for use in adolescents and is influenced by previous injury. [Emery CA, Cassidy JD, Klassen TP, et al. Development of a clinical static and dynamic standing balance measurement tool appropriate for use in adolescents. Phys Ther. 2005;85:502–514.]

Key Words: Adolescent, Balance, Measurement, Proprioception, Reliability.

Carolyn A Emery, J David Cassidy, Terry P Klassen, Rhonda J Rosychuk, Brian H Rowe
Balance training is a key component of rehabilitation following sports injury. It also is quickly gaining recognition as a vital component of injury prevention programs for many athletes, including adolescents. Currently, there is no “gold standard” for the measurement of standing balance in the young active population. Accurate measurement of standing balance is essential in assessing the effectiveness of balance training.

Balance can be defined as the ability to maintain the body’s center of gravity over its base of support with minimal sway or maximal steadiness. There is some evidence to suggest that decreased static unipedal balance is a risk factor for ankle sprain reinjury in soccer. In sports, an athlete is usually visually attentive to the game, and the activity is dynamic in nature at the time of injury. Some authors, therefore, agree that impaired dynamic unipedal balance may be more critical than static balance in sports.

There is some evidence that static unipedal balance ability does improve following balance training using a wobble board. Most of these studies, however, exclusively examined improvement following ankle injury. Other studies have demonstrated that balance training is effective in preventing sport-specific injury. Measurements of balance often are not examined in these prevention studies. Consequently, the effect of these training programs on balance remains unclear.

Factors that may influence balance ability, and thus any measurement of balance, must be considered in examining balance as an outcome measurement in rehabilitation or as a risk factor for injury in sports. Factors to be considered include: leg dominance, fatigue or learning effects, age, sex, height, weight, foot size, physical activity level and specificity, and previous lower-extremity injury.

Numerous techniques have been described to measure standing balance, with varying levels of challenge in different populations. Laboratory balance measures (e.g., stabilometry, accelerometry, motion analysis) use equipment that is costly, highly technical, and often not portable. The test-retest reliability of the measurements obtained with this equipment is extremely variable. Some measurement tools have been developed for use in the clinical setting, but many of these tools were developed for use in elderly people and people with neurological impairments.

All authors provided concept/idea/research design. Dr Emery and Dr Rowe provided writing. Dr Emery provided data collection, and Dr Emery, Dr Cassidy, and Dr Rosychuk provided data analysis. Dr Emery provided project management. Dr Cassidy, Dr Klassen, Dr Rosychuk, and Dr Rowe provided consultation (including review of manuscript before submission). The authors acknowledge the Calgary Board of Education high school principals and physical education teachers. Without their support, this research would not have been possible. They are especially grateful to the many high school students who consented to participate in this study. They acknowledge the financial support of the Department of Pediatrics, University of Alberta.

Dr Emery was supported by the Canadian Institutes of Health Research (CIHR) in partnership with the Physiotherapy Foundation of Canada, the Alberta Heritage Foundation for Medical Research (AHFMR), the CIHR Institute of Musculoskeletal Health and Arthritis, Bone and Joint Health Training Program, and the Walter Johns Graduate Scholarship Fund, University of Alberta. Dr Cassidy was supported by AHFMR as a Health Scholar. Dr Rosychuk was supported by the AHFMR as Population Health Investigator, and Dr Rowe was supported by the CIHR as a Canada Research Chair.

Ethics approval was granted by both the University of Alberta Health Research Ethics Board and Calgary Board of Education.

The authors declare no known conflicts of interest. Specifically, they are not paid consultants nor shareholders in Fitter International Inc and have no vested interest in their products (Airex Balance Pads) or sales.
tools, however, are arguably not challenging enough for adolescents without neurological impairments.

Adequate test-retest reliability for timed static unipedal balance has been reported in both children and adults.\textsuperscript{19,60} To date, attempts to establish adequate reliability of a dynamic balance test using a tilt board or a hop-stabilization test have not been successful.\textsuperscript{17,60,63} The use of a foam or narrow support surface for the measurement of balance relies on an observer scoring system that includes observing sway (ie, minimal, moderate, and large), movement strategy (ie, control of balance primarily initiated at the ankle, hip, or trunk), and time.\textsuperscript{43,64–66} Variable test-retest reliability (intraclass correlation coefficient [ICC] or \( r = .05–.83 \)) has been reported.\textsuperscript{43,64–66} The use of foam to alter proprioceptive feedback from the support surface and create a more dynamic task may be an appropriate tool for the measurement of timed dynamic unipedal balance in adolescents without neurological impairments.

The goals of our study were: (1) to determine the test-retest reliability of data obtained with a timed static and dynamic unipedal balance test in adolescents without neurological impairments, (2) to investigate limits of static and dynamic balance in these adolescents, and (3) to determine the influence of age, sex, leg dominance, body mass index, foot size (length and width), previous injury, sport participation level, sport participation specificity, and visual feedback on static and dynamic balance ability.

**Method**

**Subjects**

The sample was recruited from 15 Calgary Board of Education high schools. We randomly selected the order in which schools were approached to participate using computer generation of random numbers.\textsuperscript{67} Four adolescent subjects without neurological impairments (2 male, 2 female) from physical education (PE) program rosters in each grade from 10 to 12 were randomly approached for recruitment. In Alberta, participation in PE is mandatory in grade 10 and elective in grades 11 and 12. If a subject declined participation or dropped out at the time of the baseline assessment, another student (from the same school, grade, and sex) was recruited. Three additional students were tested at baseline because 3 students indicated that they would be absent from school for the 7-day follow-up assessment. School and subject recruitment are summarized in Figure 1.

Subjects were included if they were between the ages of 14 and 19 years and participated in PE class. Subjects were excluded from the study if they reported a history of a musculoskeletal injury (ie, requiring medical attention and time loss from sporting activity of 1 or more days) in the 6 weeks prior to recruitment, a history of a serious musculoskeletal pathology (eg, fracture, rheumatologic disease, systemic disease, surgery) in the 6 months prior to recruitment, or a serious ongoing medical condition or disability. A description of study subjects and dropout subjects is shown in Table 1.

**Procedure**

Both the subjects and their parent or guardian completed a written informed consent form. Each subject completed a baseline questionnaire regarding previous history of injury and sports participation. Next, the primary examiner (CAE) measured height (in meters), weight (in kilograms), and foot length and width (in...
Balance Pad is a high-density (50-kg/m³) closed-cell performed barefoot on an Airex Balance Pad.* An Airex barefoot on a gym floor surface. The dynamic tests were closed dynamic


* Alcan Airex AG, Speciality Foams, CH-5643 Sins, Switzerland. † StataCorp LP, 4905 Lakeway Dr, College Station, TX 77845.

Table 1.
Comparison of Study and Dropout Subjects

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Subjects Who Completed Study (n=111)</th>
<th>Dropout Subjects (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (95% CI)* or Proportion (95% CI) or Frequency (%)†</td>
<td>Mean (95% CI) or Proportion (95% CI) or Frequency (%)†</td>
</tr>
<tr>
<td>Age (y)</td>
<td>16.59 [16.4, 16.78]*</td>
<td>16.5 [16.05, 16.95]*</td>
</tr>
<tr>
<td>Sex</td>
<td>Male: 56 [50.45%]†</td>
<td>Male: 5 [41.67%]†</td>
</tr>
<tr>
<td></td>
<td>Female: 55 [49.55%]†</td>
<td>Female: 7 [58.33%]†</td>
</tr>
<tr>
<td>Grade</td>
<td>Grade 10: 39 [35.14%]†</td>
<td>Grade 10: 3 [25%]†</td>
</tr>
<tr>
<td></td>
<td>Grade 11: 36 [32.43%]†</td>
<td>Grade 11: 5 [41.67%]†</td>
</tr>
<tr>
<td></td>
<td>Grade 12: 36 [32.43%]†</td>
<td>Grade 12: 4 [33.33%]†</td>
</tr>
<tr>
<td>Previous injury (lower</td>
<td>15/111 = 13.51%</td>
<td>3/12 = 25%</td>
</tr>
<tr>
<td>extremity)</td>
<td>(7.77, 21.31)*</td>
<td>(5.49, 57.19)*</td>
</tr>
<tr>
<td>Previous injury (all)</td>
<td>25/111 = 22.52%</td>
<td>4/12 = 33.33%</td>
</tr>
<tr>
<td></td>
<td>(15.14, 31.43)*</td>
<td>(9.92, 65.11)*</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.70 [1.68, 1.72]</td>
<td>1.69 [1.64, 1.76]</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>68.05 [65.41, 70.68]</td>
<td>69.17 [55.8, 82.53]</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>23.42 [22.57, 24.28]</td>
<td>23.95 [19.58, 28.33]</td>
</tr>
<tr>
<td>Foot width (cm)</td>
<td>9.64 [9.51, 9.77]</td>
<td>9.64 [9.22, 10.05]</td>
</tr>
<tr>
<td>previous 6 wk (hr/wk)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* CI = confidence interval.

centimeters). Then, each subject completed 1 timed static and 2 timed dynamic unipedal balance tests (eyes-closed static=ECS, eyes-open dynamic=EOD, and eyes-closed dynamic=ECD). The static tests were performed barefoot on a gym floor surface. The dynamic tests were performed barefoot on an Airex Balance Pad.* An Airex Balance Pad is a high-density (50-kg/m³), closed-cell foam pad (47 × 39 × 6 cm, 0.7 kg). The order of leg examination (left and right) for each subject was randomly selected for each test. We also randomized the order of testing of all 3 protocols (ECS, EOD, ECD) with blocks of 6. A 30-second rest was provided between protocols. The timed measurements were completed using a stopwatch. For all 3 balance tests, each subject completed 3 trials on each leg. A 15-second rest was allowed between trials. In addition, a 15-second practice session on the foam pad was allowed prior to the start of the dynamic tests so that subjects could gain some familiarity with this support surface.

The dominant leg was identified by asking subjects to kick a ball hard prior to testing. For all trials, the subjects placed their hands on their hips and time started upon elevation of the opposite foot from the floor. Subjects participated in our study. Data were transformed logarithmically if the assumptions of normality and equal variance were not met for statistical tests. We report geometric means, which typically approximate the median and are the best measure of the central tendency when data are skewed, to allow for calculation of a 95% confidence interval (95% CI). The use of an arithmetic mean to describe central tendency is less appropriate if data are skewed. Geometric means and 95% CIs are back-transformed values from the logarithmically transformed data.

For test-retest reliability, all analyses are based on one examiner’s measurements at baseline and follow-up. The main outcome measurements included maximum time achieved over 3 trials for each of 2 legs on each of 3 tests (ECS, EOD, and ECD). One of the weaknesses of the ICC in determining reliability is that, as the between-subject variability of a measurement increases, the estimated ICC also increases. Greater between-subject variability clearly does not indicate increased reliability of that measurement. Analysis using ICCs also fails to examine whether the variability of the measurement (and, as a result, the estimated reliability) is indepen-

---

* Alcan Airex AG, Speciality Foams, CH-5643 Sins, Switzerland.
† StataCorp LP, 4905 Lakeway Dr, College Station, TX 77845.
dent of the magnitude of the measurement. In addition, use of ICCs fails to use the units of measurement in question. It is thus extremely difficult to make decisions regarding clinical relevance of measurement differences. As such, results based on Bland and Altman’s methods of agreement were examined in this study. We first plotted the individual subject differences between test sessions against the individual mean scores for both test sessions. Uniform scatter of points around the mean difference indicates no association between the differences and the magnitude of the measurement. Our data were not uniformly scattered, and we log-transformed them. The final results were then back-transformed and presented as geometric means. We report the ratio of the follow-up to the initial measurement, which is the geometric mean ratio with its 95% limit of agreement. Intraclass correlation coefficients also are reported. The ICC (3,1)‡ and 95% CI, using the method described by Shrout and Fleiss, were calculated to assess test-retest reliability with multiple scores from the same rater. Given the underlying assumptions of repeated-measures ANOVA, log-transformed data were used to estimate ICC. Test-retest reliability was examined for each of 3 unipedal balance stances (ECS, EOD, and ECD).

We used t tests and ANOVA (repeated measures and 1-way) to examine the influence of leg dominance, sex, age, potential learning and fatigue, and sport specificity on all 3 balance tests. We used linear regression to investigate the influence of several other factors—leg dominance, order of testing (blocks 1–6), age (in years), sex, body mass index (in kilograms per square meter), foot length (in centimeters), foot width (in centimeters), previous lower-extremity injury within 1 year, sports participation level (estimated hours per week in previous 6-week

### Table 2.
Geometric Means (95% Confidence Interval [CI]) for Comparison of Study and Dropout Subjects

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Subjects Who Completed Study (n=111) Mean (95% CI)</th>
<th>Dropout Subjects (n=12) Mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECS balance (s)</td>
<td>25.57 (21.91–29.85)</td>
<td>24.17 (17.5–33.38)</td>
</tr>
<tr>
<td>EOD balance (s)</td>
<td>54.59 (46.92–63.54)</td>
<td>52.7 (34.93–79.52)</td>
</tr>
<tr>
<td>ECD balance (s)</td>
<td>5.38 (5.02–5.77)</td>
<td>4.75 (3.83–5.91)</td>
</tr>
</tbody>
</table>

‡ ICC (3,1) is based on Shrout and Fleiss model 3 (based on repeated-measure analysis of variance [ANOVA] where tested raters are the only raters of interest), form 1 (single measurement is the unit of analysis rather than a mean). Calculations are based on a repeated-measures ANOVA with one fixed effect for one rater.
period), and sports participation specificity—on ECS and ECD balance.

Those subjects who reached the ceiling maximum of 180 seconds on any of the 3 balance tests (ECS, EOD, or ECD) were excluded from the analyses involving those particular balance tests because the difference between measurements would be based on censored time and would bias any estimates based on those values.

Results

There were no important differences between study subjects and dropouts (Tabs. 1 and 2), suggesting that our study results are not subject to selection bias as a result of dropouts. On the ECS test, 4 subjects achieved the maximum time (180 seconds); therefore, 107 subjects remained in the analysis of reliability for ECS balance. On the EOD test, 33 subjects were excluded from the analysis for the same reason. Consequently, 78 subjects remained in the analysis of reliability for EOD balance. All 111 subjects who completed the study were included in the analysis for ECD balance.

No important differences were found between maximum balance times for left and right legs or between sexes. The geometric means (and 95% CIs) as well as medians (and ranges) for all 3 balance tests are summarized in Table 3. Box plots (Figs. 2 and 3) demonstrate the success of log-transformation in meeting the assumptions of normality for all balance measurements (based on a maximum of 6 trials).

The Bland and Altman plot for ECS balance is presented in Figure 4. In this plot, the funnel shape suggests that the differences are clearly greater with measurements of a greater magnitude. Log-transformation was somewhat successful in producing differences unrelated to the magnitude of the measurement (Fig. 5). The funnel shape has been eliminated, but the scatter resembles an oval shape rather than a completely even scatter. This can be seen visually with a more even scatter of points about the mean. The geometric mean ratio is 0.95 (Tab. 4, 95% limits of agreement=0.28–3.2). This means that the 1-week follow-up measurement yielded an ECS maximum on average 5% less than baseline. The limits of agreement indicate that 95% of the time the follow-up measurement should be between 0.28 and 3.2 times that at baseline.

The Bland and Altman plot for EOD balance is presented in Figure 6. Log-transformation was highly successful in producing EOD differences unrelated to the mean (Fig. 7). The geometric mean ratio was 0.88 (Tab. 4, 95% limits of agreement=0.25–3.2). This means that the 1-week follow-up measurement yielded an EOD maximum on average 12% less than baseline. The limits of agreement indicate that 95% of the time the follow-up measurement should be between 0.25 and 3.2 times that at baseline.

The Bland and Altman plot for ECD balance is presented in Figure 8. As with the EOD test, log-transformation was highly successful in producing ECD differences unrelated to the mean (Fig. 9). The geometric mean ratio was 1.05 (Tab. 4, 95% limits of agreement=0.48–2.29). This means that the 1-week follow-up measurement yielded an ECD maximum on average 5% greater than baseline. The limits of agreement indicate that 95% of the time the final measurement should be between 0.48 and 2.29 times that at baseline.

For all 3 balance tests (ECS, EOD, and ECD), ICCs (3,1) and 95% CIs were calculated on log-transformed maximums for baseline and follow-up assessments. The results of all analyses examining reliability are summarized in Table 4. Based on the 95% limits of agreement,
examples of baseline measurements for ECS, EOD, and ECD balance and expected measurements 1 week later are given in Table 5.

We found no evidence that learning or fatigue based on order of testing the 3 protocols influenced our findings or that there were any differences among trials 1, 2, and 3 for ECS or EOD balance. There was an apparent learning effect over 3 trials for ECD balance, however (repeated-measures ANOVA: $F = 4.69; df = 2,244; P = .01$).

The final regression model examining factors influencing ECS balance was: $\log_{10}(ECS) = 3.309 - 0.514 \times \text{Injury}$, where $\log_{10}(ECS)$ denotes the log-transformed ECS balance maximum at baseline and Injury indicates previous lower-extremity injury within 1 year ($0 = \text{no injury}$ and $1 = \text{injury}$). The coefficient associated with previous injury indicates that static balance in adolescents with a 1-year history of lower-extremity injury was less than that of those without a previous history of lower-extremity injury ($-0.514, 95\% CI = -0.899$ to $-0.13$). Based on multiple linear regression analysis, no associations were found between EOD balance and other potentially influencing factors.

The final regression model examining factors influencing ECD balance was: $\log_{10}(ECD) = 1.699 - 0.193 \times \text{Injury}$, where $\log_{10}(ECD)$ denotes log-transformed ECD balance maximum at baseline. The coefficient associated with previous injury was less than 0, indicating that dynamic balance at baseline in adolescents with a 1-year history of lower-extremity injury was less than that of those without a previous history of lower-extremity injury ($-0.193, 95\% CI = -0.376$ to $-0.01$). Log-transformed ECS and ECD balance from these regressions were back-transformed to estimate the predicted ECS balance and ECD balance in adolescents with a 1-year history of lower-extremity injury compared with those with no history of lower-extremity injury (Tab. 6).

We also examined the effect of sport specificity on our results. Subjects were grouped by their number one sport for estimated hours spent per week beyond PE class in the past year. Based on 23 different sports reported, no differences were found between groups for log-transformed ECS or ECD balance.
Discussion
This study is the first of its kind to examine timed dynamic balance measurements in adolescents without neurological impairments, using an Airex Balance Pad for a support surface. The Airex Balance Pad is readily available, easy to clean, easy to transport, and resilient to deformation. As such, it is extremely useful in a sport setting or school setting. We determined that timed ECD unipedal balance is an appropriate and adequately reliable clinical measure of standing balance in adolescents without neurological impairments.

Other studies have examined timed ECS balance in subjects without neurological impairments. Hahn et al. found that 1.8% of their sample (competitive athletes aged 14–24 years) reached the maximum time of 180 seconds, which is similar to the findings in our study (1.6%). In our study, the geometric mean for ECS
balance based on maximum time attained over 6 trials (3 on each leg) was 25.43 seconds (Tab. 3, 95% CI = 22.06–29.31). This finding is in accordance with other research, which demonstrated a mean of 29 seconds based on maximum time achieved over 2 trials.33 Ekdahl et al26 demonstrated a mean time of 44 seconds in people 20 to 29 years of age, based on maximum times achieved over 3 trials. In a study by Bohannon et al,25 the maximum time set was 30 seconds. Only 25% of subjects who were between the ages of 20 and 29 years failed to achieve the 30-second maximum. In our study, 55% of the subjects did not achieve 30 seconds at the baseline assessment. The differences among studies may be related to age and the potential inappropriate use of an arithmetic mean if data were skewed.

Table 5. Examples of Expected 1-Week Follow-up for All 3 Balance Measures (in Seconds) Based on 95% Limits of Agreement

<table>
<thead>
<tr>
<th>Balance Test</th>
<th>95% Limits of Agreement (Bland and Altman*70,72)</th>
<th>Baseline Measurement (s)</th>
<th>Expected Follow-up Measurementa (s)</th>
<th>Potential Interval Length for Follow-up Measurement (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyes-closed static</td>
<td>0.28–3.2</td>
<td>3.8</td>
<td>1.06–12.2</td>
<td>11.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.4</td>
<td>4.31–49.28</td>
<td>44.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.4</td>
<td>7.39–84.5</td>
<td>77.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>46.4</td>
<td>12.99–148.5</td>
<td>135.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>166.2</td>
<td>46.53–180</td>
<td>133.47</td>
</tr>
<tr>
<td>Eyes-open dynamic</td>
<td>0.25–3.2</td>
<td>8.6</td>
<td>2.15–27.5</td>
<td>25.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35.7</td>
<td>8.93–114.2</td>
<td>105.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>58.4</td>
<td>14.6–180</td>
<td>165.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>94.7</td>
<td>23.68–180</td>
<td>156.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>174.1</td>
<td>43.53–180</td>
<td>136.47</td>
</tr>
<tr>
<td>Eyes-closed dynamic</td>
<td>0.48–2.28</td>
<td>2.4</td>
<td>1.15–5.5</td>
<td>4.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>1.92–9.1</td>
<td>7.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.9</td>
<td>2.35–11.2</td>
<td>8.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.4</td>
<td>3.07–14.6</td>
<td>11.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19.6</td>
<td>9.41–44.7</td>
<td>35.29</td>
</tr>
</tbody>
</table>

*Calculated by multiplying the baseline measurement by the limits of agreement.

Table 6. Predicted Eyes-Closed Static (ECS) and Eyes-Closed Dynamic (ECD) Balance Times (in Seconds) by Injury Historya

<table>
<thead>
<tr>
<th>Balance Test</th>
<th>Subjects Who Reported No Previous Lower-Extremity Injury</th>
<th>Subjects Who Reported Previous Lower-Extremity Injury</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric Mean</td>
<td>(95% CI)</td>
<td>Geometric Mean (95% CI)</td>
</tr>
<tr>
<td>ECS balance</td>
<td>27.35 (23.58–31.73)</td>
<td>16.36 (10.52–25.43)</td>
</tr>
<tr>
<td>ECD balance</td>
<td>5.47 (5.1–5.87)</td>
<td>4.51 (3.66–5.56)</td>
</tr>
</tbody>
</table>

*All values are back-transformed from log ECS and ECD. CI = confidence interval.

retest reliability of a timed ECS balance test has been demonstrated previously using the ICC alone. Atwater et al*60 examined children 4 to 6 years of age (ICC=.59–.77), Balogun et al*18 examined young adults without neurological impairments (ICC=.96), and Bohannon et al*61 examined adults following stroke (ICC=.44–.75). Differences among these studies included age and disability. The present study demonstrated adequate test-retest reliability for ECS balance, based on ICC alone, consistent with these other studies (ICC=.69, 95% CI=0.57–0.78). Based on the wide limits of agreement, however, caution is needed in interpretation of ECS balance for people in clinical practice (Tab. 4). For example, an adolescent without neurological impairments may demonstrate “average” balance on the ECS balance test one week (ECS=26 seconds), and his or her expected performance on the ECS balance test the following week might be anywhere between 7 and 85 seconds. As such, the use of ECS balance as an outcome measure of change on an individual level is limited, although it may be appropriate for measuring change at a group level.

Riemann et al*43 demonstrated no difference (F=1.08; df=1,10; P=.358) in repeated testing of dynamic balance on a foam surface using an error scoring system measured 1 day apart. In our study, dynamic balance measurements (EOD and ECD) appeared to have moderate and poor reliability based on the ICCs alone (Tab. 4, EOD: ICC=.59, 95% CI=0.43–0.71; ECD: ICC=.46, 95% CI=0.31–0.59). However, reliability based on limits of agreement demonstrated sufficiently narrow limits of agreement for ECD balance to be considered in clinical practice. For example, an adolescent without neurolog-
ical impairments may demonstrate “average” balance on the ECD balance test one week (ECD=5 seconds) and his or her expected performance on the ECS balance test the following week, with no intervention, could be anywhere between 2 and 11 seconds. The reliability of these measurements appears to be greater for relatively low balance ability and decreases as balance ability improves (Tab. 5). Using the limits of agreement presented in Table 5, a clinician may determine the expected change in ECD balance over a 1-week interval. Improvement would be assessed based on change beyond the upper limit of agreement. In the case of ECD balance, improvement would be a follow-up performance of greater than 2.28 times baseline ability.

One of the weaknesses of the ICC in determining reliability is that, as the between-subject variability of a measurement increases, the estimated ICC also increases.68 Greater between-subject variability clearly does not indicate increased reliability of that measurement.68 In our study, between-subject variability of the ECD measurement was very small (Tab. 3, range = 2.38–19.63, with only 3 subjects exceeding 12 seconds). This may have contributed to a poor ICC for this test.

In the EOD test, 18% and 24% of subjects achieved the maximum of 180 seconds in each of 2 test sessions. This ceiling of 180 seconds limits our ability to examine changes in EOD balance over time. As such, EOD balance is not the most suitable clinical examination for dynamic balance in this population, unless the maximum time is extended and reliability is further examined.

There are other limitations that may have contributed to the moderate test-retest reliability found in our study. One week between assessments allows time for potential practicing of balance. Testing was performed on the same weekday and time of day for both assessments; however, it is possible that physical activities extraneous to our study may have affected balance ability at one session and not another. Adolescents also may be influenced by boredom during the testing session, peer pressure, or limited attention span, which may influence the reliability of these measurements.

Diminishing visual feedback with eyes closed consistently results in decreased postural stability in comparison with eyes-open conditions.2,25–27,33,62,69,76–78 The results of our study are consistent with these findings. Based on geometric means, reported balance ability in eyes-open conditions exceeds that of eye-closed conditions. Other studies4,25,27,33,62,76,78 also have failed to demonstrate a difference in balance ability between the dominant and nondominant legs in subjects without neurological impairments. In addition, we found no association between balance and age in our study. This finding is consistent with the results of the study by Hahn et al.33 Most of the studies demonstrating that ECS balance decreases with age examined this relationship over a wider age range (ie, adolescent to elderly, child to adolescent).25,26,62,79 Consistent with other research,26,29,33,78,80,81 our results also failed to demonstrate any relationship between balance ability and sex in adolescents and adults. Some studies26,79,82,85 have demonstrated a relationship between sex and balance ability in elderly people and in young children, with female subjects demonstrating better balance ability than male subjects.

In theory, factors that lower the center of gravity (ie, decreased height) and increase the base of support (ie, foot size) will increase postural stability.13 Odemark and Sandstedt31 found both height and weight to be predictors of increased postural sway in boys and girls aged 5 to 15 years. Habib and Westcott92 found that increased foot length was associated with greater balance ability in children. The age of their subjects differed considerably from that of the subjects in our study. Consistent with our study, however, Peeters et al79 and Ekdahl et al26 demonstrated that height and weight had no direct influence on balance.

Our study failed to demonstrate an association between either hours of sport participation or sport specificity and balance ability. Ekdahl et al26 also failed to demonstrate an association between postural stability and leisure activities. Hahn et al33 demonstrated that timed unipedal balance was not associated with type of sport, but was positively associated with hours per week of basketball and number of years of basketball and was negatively associated with hours of swimming.

We found no learning or fatigue effects over repeated trials on the same day for ECS and EOD balance. For ECD balance, however, there was evidence of a learning effect over 3 trials. Further examination of a potential learning effect beyond 3 trials, by increasing the number of repetitions, is recommended for future study. This may be related to the increased difficulty associated with ECD balance. Other studies22,24,26,84 have demonstrated learning effects over more than 3 repeated trials using laboratory measurements. In our study, there was no evidence of fatigue based on 6 possible orders of ECS, EOD, and ECD tests.

Previous lower-extremity injury (1-year history) appeared to decrease both ECS and ECD balance in our study. Previous injury had no effect on EOD balance, suggesting that vision may have compensated for the effects of previous injury. The difference seems to be more significant based on ECS balance than ECD bal-
ance. This is likely related to the low intersubject variability in ECD balance measurement. Differences in balance measurements between previously injured and uninjured athletes also are consistently reported in the literature.\textsuperscript{2, 14, 19, 34, 37, 35, 36, 83–87}

One of the major strengths of our study is the random recruitment of schools and subjects, which increases the generalizability of the study results. Given that grade 11 and 12 PE is elective in Alberta, however, the generalizability is potentially limited to a more active and healthy population of adolescents. The geometric means provided in our study will give the clinician a relative comparison with a random sample of adolescents without neurological impairments. Other strengths of our study include the high rate of consent to participate (>98%). In addition, the dropout rate was extremely low (<10%). Both of these factors limit selection bias. This study also confirms the need to consider alternate and more appropriate statistical methods, in addition to the commonly used ICC, in the assessment of reliability of outcome measurements in sports medicine.

Future research may include examination of the test-retest reliability of ECS and ECD balance at a shorter interval than 7 days. In addition, EOD balance should be examined without a set ceiling maximum of 180 seconds. Concurrent validity of these timed balance measurements also may be examined by comparison with laboratory stabilometry techniques. Previous lower-extremity injury should always be considered in future research examining timed balance measurements in adolescents.

**Conclusion**

Timed ECS and ECD balance (using an Airex Balance Pad for base of support), with a 180-second maximum for each test, are appropriate clinical balance measures for use in adolescents. Reliability for both ECS and ECD balance appear to be adequate based on ICCs. Based on wide limits of agreement using Bland and Altman methods, however, caution is needed when examining ECS balance at an individual level in clinical practice. Bland and Altman methods\textsuperscript{70, 72} methods of agreement demonstrate sufficient reliability for ECD balance only. Test-retest reliability was moderate for EOD test; however, more than 24% of subjects achieved the maximum of 180 seconds on this test. Consequently, EOD balance using this ceiling maximum is inappropriate for use as a clinical balance measurement in this population. In future research examining balance, it is critical to consider previous lower-extremity injury as a key factor influencing balance.

**References**


22 Holliday PJ, Fernie GR. Changes in the measurement of postural sway resulting from repeated testing. Aggressologie. 1979;20:225–228.


67 Stata Statistical Software. Release 5.0 [computer program]. College Station, Tex: Stata Corp; 1997.


