ABSTRACT

Background: Balance is a complex construct, affected by multiple components such as strength and co-ordination. However, whilst assessing an athlete's dynamic balance is an important part of clinical examination, there is no gold standard measure. The multiple single-leg hop-stabilization test is a functional test which may offer a method of evaluating the dynamic attributes of balance, but it needs to show adequate intra-tester reliability.

Purpose: The purpose of this study was to assess the intra-rater reliability of a dynamic balance test, the multiple single-leg hop-stabilization test on the dominant and non-dominant legs.

Design: Intra-rater reliability study

Methods: Fifteen active participants were tested twice with a 10-minute break between tests. The outcome measure was the multiple single-leg hop-stabilization test score, based on a clinically assessed numerical scoring system. Results were analysed using an Intraclass Correlations Coefficient (ICC$_{2,1}$) and Bland-Altman plots. Regression analyses explored relationships between test scores, leg dominance, age and training (an alpha level of $p = 0.05$ was selected).

Results: ICCs for intra-rater reliability were 0.85 for the dominant and non-dominant legs (confidence intervals = 0.62-0.95 and 0.61-0.95 respectively). Bland-Altman plots showed scores within two standard deviations. A significant correlation was observed between the dominant and non-dominant leg on balance scores ($R^2 = 0.49$, $p<0.05$), and better balance was associated with younger participants in their non-dominant leg ($R^2 = 0.28$, $p<0.05$) and their dominant leg ($R^2 = 0.39$, $p<0.05$), and a higher number of hours spent training for the non-dominant leg $R^2 = 0.37$, $p<0.05$).

Conclusions: The multiple single-leg hop-stabilisation test demonstrated strong intra-tester reliability with active participants. Younger participants who trained more, have better balance scores. This test may be a useful measure for evaluating the dynamic attributes of balance.

Level of Evidence: 3

Key words: Assessment, balance, reliability, hop testing
INTRODUCTION

Normal balance requires the interaction between multisensory organ systems (proprioceptive, visual and vestibular) and the brain and spinal cord, which ultimately control the multi-joint musculoskeletal system. These systems can be affected by factors such as nutrition, age, injury and disease. At an optimal level they work to maintain the center of gravity within a defined base of support, as well as the task specific orientation of body parts.

Within sports medicine, assessing an athlete's balance is an important part of a clinical examination. It is within this domain that an emphasis is placed upon proprioceptive / balance exercises as both a tool for injury prevention and as a rehabilitation strategy. However, the physical demands of sport are extremely diverse, and balance and postural control appear to be influenced by other performance attributes. For example, strength training programs lead to significant improvements in both static (Romberg) and dynamic (Star Excursion Balance Test) measures of balance.

Despite the implementation of balance training for both injury prevention and rehabilitation, no gold standard outcome measure exists with which to quantify balance within the athletic population. While it is acknowledged that balance can be measured statically or dynamically, the population being examined should direct the nature of the test selected. Furthermore it should not be assumed that static balance ability is positively correlated with dynamic balance performance. Therefore it appears appropriate to use a dynamic measure of balance when examining the athletic population, as all sports require a "dynamic" attribute of balance in some way.

The purpose of looking at athletic balance stems from the results of a series of single case studies evaluating the use of clinically targeted compression in athletes, whereby compression was delivered to the pelvic girdle via a customised orthosis in the form of shorts. Questionnaire responses from the participating athletes suggested that this type of external pelvic compression may have had a positive effect upon balance. In order to investigate whether this is the case, the intention was to incorporate a functional measure of athletic balance in future clinical trials. On the basis of the current literature and discussion with clinical colleagues, it is anticipated that a functional single leg test may be an appropriate measure of dynamic balance.

Previous researchers have found that knee instability is positively correlated with one-legged tests, and that a single leg hopping test can demonstrate good test re-test reliability. The multiple single-leg hop-stabilization test (MSLHST) is a single leg dynamic measure involving forwards, and diagonal movements in a unipedal stance, that incorporates periods of statically maintaining this stance. Athletes are scored on both a balance and landing scale, according to the errors that they commit in each period of the test; these scores are summed to give the total error score. It has been argued that this type of functional test is important because it challenges athletes in a way which reflects the forces and directions of movement that are integral to sport.

Although this test has been reported to have very good inter-tester reliability (ICC values 0.70-0.92), intra-rater reliability was shown to be lacking. Closer inspection of the intra-rater reliability reveals that this lack of reliability only refers to the balance scores which significantly differed between tests; no significant difference was observed with the landing scores. Further, this study assessed three test sessions, each 48 hours apart; a different scenario to the current intra-rater reliability study in which the testing was completed in one session.

A further consideration for any balance study involving athletes with a lower limb injury is the influence of lower limb dominance. In football, a player's dominant (preferred kicking leg) has been shown to be significantly stronger than their non-dominant leg in terms of hip adductor strength, and hip flexor strength, but not in all muscle groups. It has been suggested that any rehabilitation of injury needs to take leg dominance into consideration. As a strength deficit may potentially contribute to poor balance, it is important that a study considers the role of limb dominance, and examines how this may influence the reliability of the balance measure used.

The purpose of this study was to assess the intra-rater reliability of a dynamic balance test, the MSLHST on the dominant and non-dominant legs. A secondary purpose was to explore whether...
relationships exist between the MSLHST scores and leg dominance, age, and time spent engaging in exercise (training).

METHODS

Design
An intra-rater reliability study was undertaken. All of the testing was undertaken by a single investigator, using portable equipment; the test was scored in “real time” while the balance measure was being performed.

Participants
A convenience sample of volunteers was recruited from Plymouth University staff and students, and from local sports clubs. To maximise recruitment the study was conducted at the University (Human Movement Laboratory) to accommodate the staff and student participants. Ethical approval was gained from a local University Ethics Commitee (Plymouth University).

Eligibility Criteria
To be included, participants had to be over the age of 18, and able to give informed consent, be self-declared as healthy, and have sustained no lower limb musculoskeletal injuries in the prior three months. Participants were excluded if they were pregnant, had a current illness / unresolved condition, or had any neurological, musculoskeletal or cardiorespiratory impairment.

Sample Size
Reliability coefficients greater than 0.7 are deemed to be acceptable for most clinical trials. A power calculation indicated that 15 people were needed to be recruited in order to demonstrate an ICC of >0.7 (power = 0.88; \( \alpha = 0.05 \)). This is in keeping with the work of Fleiss\(^2\) and their discussion of the numbers required for a reliability study involving quantitative measures.

Participant Characteristics
Participant demographics (age, gender, height, weight), their leg dominance (as defined by which side they would kick a ball), and the average number of hours spent training / performing sports in a week were recorded.

Measurement of the MSLHST
Testing was undertaken in standard sports attire (shorts, t shirt and athletic shoes) and conducted in the same undisturbed environment, in order to minimize external influences and allow for standardization. Standardized written instructions were given to all participants prior to testing; this included photographs of stances. Participants also received verbal instructions from the researcher while viewing the MSLHST set up, and before completing their practice attempts.

The distances between each of the boxes (Table 1) were standardized according to the participants’ height. Diagonal distances represented 45% of the participants’ height (wearing athletic shoes), and Pythagoras Theorem used to calculate the distances in the frontal plane, for the adjacent boxes. The mat was labelled according to the height related distances prior to testing to ensure that during testing, there was minimal delay in setting up the mat. This was achieved using hook and loop combinations of numbered Velcro\(^\circ\) squares.

One practice attempt on each leg was undertaken for familiarization of the procedure while avoiding fatigue. Both the dominant leg (as defined as the leg that people would prefer to kick a ball with) and the non-dominant leg were tested in a randomized order (randomization was undertaken using the Microsoft

<table>
<thead>
<tr>
<th>Table 1. Hop distances according to height, Reimann and Manske(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height in Centimetres (cm)</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>150-159.9</td>
</tr>
<tr>
<td>160-169.9</td>
</tr>
<tr>
<td>170-179.9</td>
</tr>
<tr>
<td>180-189.9</td>
</tr>
<tr>
<td>190-199.9</td>
</tr>
<tr>
<td>200-209.9</td>
</tr>
</tbody>
</table>
The MSL-HST scoring was defined as:

**Balance score.** Up to three error marks were given for participants committing the following in any balance period:

- Touching the floor with the non-weight bearing limb;
- Removing hands from iliac crests;
- Non-weight bearing limb touching the weight bearing limb;
- Non-weight bearing limb moving into excessive flexion, extension or abduction (this was defined as movement beyond the predetermined stance (>30 degrees of movement); displayed to the participants in a photographic format).

**Landing score.** Up to 10 error marks were given for participants committing the following in any landing period:

- Removing hands from iliac crests;
- Foot not covering the numbered square;
- Stumbling on landing;
- Landing foot not facing forwards with 10 degrees of inversion or eversion.

Therefore potential test scores could range from 0-130 (0-100 for the landing component, and 0-30 for the balance element).

**Statistical Analyses**

Statistical analyses were performed using SPSS 20 for Windows (IBM). Two-way random absolute agreement intra-class correlation ($ICC_{2,1}$) and 95% confidence intervals were used to assess the intra-rater reliability.\(^{24}\)

Bland Altman plots were presented to show a visual representation of intra-rater reliability. Using more than one measure of reliability has been advised as no one measure is suitable for all reliability studies.\(^{25}\) ICCs give a relative view of reliability, therefore it has been advised not to draw conclusions before using methods of examining the absolute reliability.\(^{26}\)

A paired t-test was used to ascertain if there was a significant difference between the balance ability of the dominant and non-dominant leg ($p = <0.05$). Regression analyses were undertaken to explore possible relationships between balance ability on the dominant leg.
The strength of the correlation coefficients were interpreted as: 0 = zero, 0.1-0.3 = weak, 0.4-0.6 = moderate, 0.7-0.9 = strong and 1 = perfect.27

The time spent training each week was further explored using t tests to determine the possibility of predicting test performance according to the amount of training undertaken (< or > five hours per week). Such a relationship has been observed in previous work, showing that lifelong football trained men demonstrated significantly superior balance to age matched untrained men.28

RESULTS

Fifteen participants (males = 8), aged 22-57 participated in the study. The demographics of the tested population are presented in Table 2.

Table 2. Demographic data

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Weight (kg)</th>
<th>Height (cm)</th>
<th>Gender</th>
<th>Dominant Leg</th>
<th>Average Weekly Training Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>32.8</td>
<td>71.4</td>
<td>174.2</td>
<td>Female = 7</td>
<td>Left = 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Male = 8</td>
<td>Right = 13</td>
<td>5.5</td>
</tr>
<tr>
<td>SD</td>
<td>9.2</td>
<td>9.5</td>
<td>7.5</td>
<td></td>
<td>4.3</td>
</tr>
<tr>
<td>Range</td>
<td>22-57</td>
<td>53.8-88</td>
<td>162.5-184.5</td>
<td></td>
<td>0.3-14</td>
</tr>
</tbody>
</table>

Yrs = years; kg = kilograms; cm = centimetres; SD = standard deviation

Table 3. Intra-rater reliability results. ICC$_{(2,1)}$

<table>
<thead>
<tr>
<th>Leg</th>
<th>95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraclass Correlation Coefficient</td>
<td>Lower Bounds</td>
</tr>
<tr>
<td>Dominant Leg</td>
<td>0.85</td>
</tr>
<tr>
<td>Non-Dominant Leg</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Table 4. Intra-rater reliability results for the non-dominant leg balance and landing scores. ICC$_{(2,1)}$

<table>
<thead>
<tr>
<th>Leg</th>
<th>95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraclass Correlation Coefficient</td>
<td>Lower Bounds</td>
</tr>
<tr>
<td>Landing Score</td>
<td>0.78</td>
</tr>
<tr>
<td>Balance Score</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Landing score = balance errors made in the one second period immediately after landing
Balance score = balance errors made in the period prior to undertaking each jump and the period one to five seconds after landing and stabilizing the position.

Table 5. Intra-rater reliability results for the dominant leg balance and landing scores. ICC$_{(2,1)}$

<table>
<thead>
<tr>
<th>Leg</th>
<th>95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraclass Correlation Coefficient</td>
<td>Lower Bounds</td>
</tr>
<tr>
<td>Landing Score</td>
<td>0.72</td>
</tr>
<tr>
<td>Balance Score</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Landing score = balance errors made in the one second period immediately after landing
Balance score = balance errors made in the period prior to undertaking each jump and the period one to five seconds after landing and stabilizing the position.

The time spent training each week was further explored using t tests to determine the possibility of predicting test performance according to the amount of training undertaken (< or > five hours per week). Such a relationship has been observed in previous work, showing that lifelong football trained men demonstrated significantly superior balance to age matched untrained men.28
Paired t-tests revealed no significant differences between performance of the dominant and non-dominant legs in the first or second performance of the test \((p = >0.05)\), therefore the scores for the dominant and non-dominant legs were averaged across the two tests (Figure 4).

There was a significant positive and strong relationship\(^{29}\) between the scores obtained on the dominant and non-dominant legs; higher scores on one leg were associated with higher scores on the other leg \((R^2 = 0.49, p<0.05)\) (Figure 5).

There was a significant positive and moderate relationship\(^{29}\) between the scores obtained on both the dominant / non-dominant legs and the age of the participant. Higher scores (indicating more errors) were associated with advancing age. The relationship was stronger on the dominant leg (non-dominant leg \(R^2 = 0.28, p<0.05\), Figure 6; dominant leg \(R^2 = 0.39, p<0.05\), Figure 7).
Greater number of training hours per week were associated with lower scores on the MSLHST. This relationship, which was of moderate strength, was significant for the non-dominant leg only ($R^2 = 0.37$, $p < 0.05$).

Further analysis using t-tests showed a significant difference ($p = <0.05$) in overall scores between those training more and those training less than five hours per week. This was seen for the average scores for both dominant and non-dominant legs.

**DISCUSSION**

ICC values can be interpreted as follows; 0.75 and above indicates excellent reliability, 0.4-0.75 is fair to good reliability and <0.4 is seen as poor reliability. The ICC results for both the dominant and non-dominant leg both demonstrate a mean value of 0.85. Whereas this may be considered as demonstrating excellent intra-rater reliability, examination of the 95% CI urges more caution. The intervals ranging from 0.62-0.95 for the dominant leg, and, 0.61-0.95 for the non-dominant leg, should be interpreted as showing that the MSLHST demonstrates good to excellent intra-rater reliability in a healthy, exercising population.

The varying degrees of reliability shown in Tables 4 and 5 allows a comparison with previous findings on the differences in the landing and balance score reliability. The current findings show that ICCs range from 0.72-0.88; indicating good to excellent reliability. The finding that reliability is greater with the balance scores than landing is in contrast to prior work. While this may reflect the difference in the prescribed scores given for landing and balance errors, for the purpose of this work the focus upon intra-rater reliability is with the overall MSLHST score which is derived by totalling the balance and landing scores.

While ICCs were examined to provide a quantitative assessment of reliability in terms of consistency of agreement; Bland Altman plots were examined as a qualitative method of assessing reliability and determining degree of absolute agreement. Inspection of these plots (Figures 2 and 3) show that the MSLHST intra-rater scores all lay within the two standard deviation limits. Considering these findings together with those of previous research, it appears that the MSLHST could be a reliable functional outcome measure, and may be considered for inclusion in future clinical trials in a similar population.

Thorborg et al suggested that one may expect to see a difference in balance ability between the dominant/ non-dominant legs. However, paired t-tests used to examine the current data demonstrated that
there was no significant difference between the dominant and non-dominant limbs in this sample \( (p > 0.05) \). Furthermore a significant strong, positive correlation was observed between the MSLHST scores of the dominant and non-dominant leg. Those making less errors completing the test on their dominant leg, tend to perform similarly on their non-dominant leg. This finding has also been observed in the sedentary population,\(^{31}\) although future work is warranted to explore this in athletes.

A moderate and significant positive relationship was demonstrated between balance scores and age; higher error scores (indicative of worsening balance) occurred with increasing age when both the dominant and non-dominant legs were assessed. A deterioration of balance with age has been reported previously.\(^{32}\) Changes include an increased amplitude and speed of postural sway, reduced dynamic balance and greater instability when sensory inputs controlling balance are perturbed or reduced.\(^{33}\) Many of these studies compared balance ability in younger (<30 years) and older (>60 years) age groups.\(^{32,33}\) It is of note that this measure of dynamic balance appeared able to detect variations in performance with age even within the relatively narrow age band of the current sample (22-57 years).

People who trained for greater time periods each week had lower scores on the MSLHST (indicating better balance ability). This was only significant on the non-dominant leg. Interestingly, the task used to define the dominant leg was kicking a ball in which the opposite non-dominant leg is balancing, supporting the body weight. The moderate relationship seen between the hours spent training and better performance on the non-dominant leg balance scores might be because this leg is used more frequently for balancing activities; especially during asymmetric activities like football that involve phasic movements of the dominant leg.

Predicting performance scores through other variables can be useful in forecasting future performance outcomes. Led by the findings of earlier research\(^{29}\) the number of training hours undertaken each week was explored as a predictor of subjects MSLHST scores; a significant difference \( (p < 0.05) \) was shown between participants when grouped in terms of the time spent engaged in exercise activities each week. More specifically the results show that it is possible to predict how well a participant will do on the MSLHST by looking at the number of hours that they spend training each week; more than five hours of training per week is a strong indicator that a participant will have a lower error score (indicative of better balance). This is supported by literature in other populations where engagement in sport and physical activities has been shown to be associated with better balance and postural control.\(^{34}\)

**CONCLUSION**

The results of the current study demonstrate that the MSLHST demonstrates good to excellent intra-rater reliability in a healthy, active population. Furthermore simple regression analyses suggest that predictions may be made as to participants’ MSLHST error scores, based on known factors such as their age and training hours. The latter showing a significant difference \( (<0.05) \) in performance between those training more and less than five hours per week. However further work is required to confirm these findings. Similar to the findings of previous work,\(^{18}\) it appears that this test could be an appropriate functional measure of athletic balance for use in future studies.

**REFERENCES**


