The Reliability of the Lateral Step Test

Annabelle King, Mark Hancock, and Joanne Munn

Context: Functional strength measures correlate more closely with functional performance than non-functional strength measures. Objectives: To determine the reliability of the lateral step test as a measure of maximal strength. Setting: Research laboratory. Participants: Twenty four healthy, pain free subjects. Intervention: Two protocols (A and B) were evaluated. The protocols were identical except protocol B involved a three second pause. Participants performed a one repetition maximum (1RM) for each protocol on two occasions separated by one week. Main Outcome Measures: Step height (nearest cm) representing 1RM. Results: Both protocols demonstrated excellent reliability, protocol A: ICC = 0.94 (95% CI, 0.87 to 0.97), SEM = 1.47 cm. Protocol B: ICC = 0.94 (95% CI, 0.85 to 0.97). Percent close agreement within 2 cm was 83.3% for protocol A and 79.1% for protocol B. Conclusion: Both protocols demonstrated excellent inter-tester reliability as measures of functional lower limb strength. Key Words: strength, reliability, 1RM

Strength impairments affect broad populations including elite athletes and older persons. Reliable strength tests are essential to accurately measure baseline strength and determine the effectiveness of any intervention. In addition to having high reliability, the ideal strength test is related to function: sensitive to strength changes and cost, time, and space efficient.

Nonfunctional strength tests commonly used in a clinical setting include manual muscle testing, hand held dynamometry, and isokinetic dynamometry. Manual muscle testing can be easily performed in a clinical setting; however, previous studies have found manual muscle testing to have poor reliability and be insensitive to strength changes over time.1-5 In contrast, hand held dynamometry and isokinetic testing of the lower limb are reliable in many populations with well designed studies reporting intraclass correlation coefficients (ICCs), ranging from 0.82 to 0.996-9 and 0.71 to 0.99, respectively.10-23 Despite high reliability, the use of isokinetic dynamometry is somewhat limited due to the size and cost requirements of the equipment. Also, a significant limitation associated with hand held and isokinetic dynamometry is that they tend to be a poor reflection of functional ability.24-28 This is because muscle strength is posture, angle, and velocity specific,29 and movement demands during functional tasks require not only the generation of muscle tension, but also the initiation of appropriate motor synergies and postural
adjustments. Consequently, the use of functional strength tests has been advocated as they demonstrate higher correlation with functional performance.\textsuperscript{18,25,28,30,31} Functional tests of the lower extremity are generally closed kinetic chain movements, which emulate forces and neuro-motor demands experienced during weight bearing activity. Because of the specific nature of strength,\textsuperscript{29} it is reasonable that functional strength tests are consistently more predictive of function than nonfunctional strength tests. There are several functional strength tests for the lower limb; however, the reliability of many of these tests has not been investigated. Commonly used functional strength tests include hopping for maximum distance (as opposed to timed hopping tests) and sit to stand tests. The hop test for maximum distance has excellent reliability (ICC 0.87 to 0.99)\textsuperscript{18,32-34} when the average of at least two trials is used. The reliability of this test has only been investigated in participants following anterior cruciate ligament (ACL) injury or reconstruction or in relatively young healthy participants. The reliability and appropriateness of the maximal hop test in the broader population is unclear.

The sit-to-stand test is a functional measure used in clinical practice and research. There are many variations of how the test is performed.\textsuperscript{30,35-38} Generally, these tests have excellent reliability. Jones et al\textsuperscript{37} reported ICCs ranging between 0.77 and 0.95; however, the protocol tested the number of chair stands completed in 30 seconds. Functional tests (such as this type of sit-to-stand test), which involve many repetitions, do not reflect maximum strength. Rather, such tests may be used to assess other aspects of muscle performance such as endurance.\textsuperscript{1} Other sit-to-stand protocols measure the time to perform a defined small number of repetitions such as timed 5-repetition test ($r = 0.95$, $P = 0.0001$)\textsuperscript{35,38} and the timed 1-repetition test\textsuperscript{36} (which demonstrates poor reliability ICC = 0.25). These tests, however, are a reflection of the speed of muscle force development (power) rather than maximal strength.\textsuperscript{39}

The lateral step exercise is used to both measure lower extremity strength and as a strengthening exercise.\textsuperscript{21,46} Biomechanical and electromyographic (EMG) studies have shown that the lateral step primarily involves coordinated activation of the knee and hip extensors, the hip adductors, and the ankle plantar flexors.\textsuperscript{40} The lateral step has the benefits of being functional, quick, simple, and inexpensive. To date, the lateral step test’s reliability as a measure of maximal lower limb strength has not been investigated. In the only study into the lateral step test’s reliability as a measurement tool, high reliability (ICC\textsubscript{(2,1)} 0.90 to 0.96) was reported for the number of repetitions completed on a preset step height in 15 seconds and for the time taken to complete 50 lateral step repetitions.\textsuperscript{41} These tests involved a high number of repetitions and as such are more a measure of endurance than maximal strength. The high reliability previously found can not be generalized to the use of the lateral step as a maximal strength test. The aims of this study were to design a lateral step protocol to assess maximal lower limb strength and to test the reliability of this protocol. Secondary aims were to establish if reliability of the lateral step was influenced by the inclusion of a three-second pause during the protocol or by the level of clinical experience of the tester.
Method

Participants

Twenty-four adult participants (6 male and 18 female) volunteered for this study. Participants had a mean age of 37.8 years (SD ± 19.9) years, mean height of 170.0 centimeters (SD ± 12.9), and mean weight of 68.4 kilograms (SD ± 12.6). The general health of all participants was initially screened using the Modified Physical Activity Readiness Questionnaire (PARQ). If participants did not satisfy all criteria of the PARQ (for example age greater than 69 years), their general practitioner was consulted to make a decision based on the patients’ overall health status. No participants were excluded by the GP. Participants were also excluded from this study if they currently reported pain in either lower extremity or were involved in a lower limb strengthening program at the time. IRB approval was granted in the spirit of the Helsinki declaration.

Testers

Four testers were used in this study. Testers 1 and 2 were third year physical therapy students. Testers 3 and 4 were qualified physical therapists with 12 and 13 years experience, respectively. Prior to data collection, each tester was provided with written guidelines detailing both protocols (see Appendix). All testers met on a single occasion for 30 minutes to practice the protocols.

Procedure/Testing Sequence

An intertester repeated measures study design was used to assess the interrater reliability of two slightly different lateral step protocols (A and B). The lateral step test used here determined the maximum step height a person can negotiate only once (ie, 1 repetition maximum, RM, measured in cm). Each participant was tested with protocol A on one leg and protocol B on the other leg. The selection of the tester from the pool of 4 testers was counterbalanced so each of the four testers assessed six participants for their first test (protocols A and B) and a different 6 participants for their second test (protocols A and B). The second test occurred seven days after the initial test. During the second testing session, each participant repeated the same tests in the same order as the first testing session. Both the participant and the second tester were blind to the results from the first testing occasion. At the commencement of each session, the participant was familiarized with the testing procedure and performed a warm up of 10 lateral step repetitions for each leg on a 7 cm step. During this time, the participant received verbal feedback with demonstration if required to ensure their technique was correct and to attempt to minimize compensation strategies that constituted a failed attempt, such as lateral deviation of the trunk and trunk flexion with associated posterior movement of the pelvis.

Two different protocols (A and B) were tested as during pilot testing it appeared that the inclusion of a short pause at the point of maximal resistance may reduce compensation strategies and result in a more reliable test. The outcome for both protocols was the 1RM step height (cm).
Protocol A (No Pause). Participants stood on their test leg at the top of a step facing a wall with the great toe positioned 30 cm back from the wall. The index fingertip on the side of the test leg was allowed to rest on the wall for balance, at a comfortable height for the participants. If poor balance became a limiting factor, the index fingertip from the other hand was also permitted on the wall, and the participants were retested in the same way (this was only needed for two of the 24 participants). The lateral step test began with the participants directly lowering the non-test leg (maintained in knee extension and ankle dorsiflexion) by flexing their weight bearing limb at the hip, knee, and ankle until the heel of the non-test leg made light contact with a 1 cm thick sponge placed on the floor (Figure 1). Only slight depression of the sponge was allowed to ensure participants did not weight bear on the floor. If participants did weight bear through the heel of the non-test leg (as evident by total depression of the sponge), the attempt was considered a failure. The tester gave standardized encouragement to ensure maximal effort by the participants. Once the participants’ heels made contact with the sponge they

Figure 1 — Lateral step test technique.
extended their test leg to return to the start position, thus completing the repetition. One second rest was allowed before the next repetition. If the participants were able to successfully complete two repetitions, the test was stopped and the step height increased. If the participants were unable to complete a single repetition, the step height was reduced. Repetitions were completed at a smooth speed that was comfortable for participants. They were instructed to look straight ahead and keep their trunk upright.

Protocol B (3 Second Pause). The lateral step was performed as above; however, once contact had been established between the sponge and the heel of the non-test leg, the participant was required to pause for three seconds without weight bearing through the heel. The three second pause was timed using a stop watch and counted out loud by the tester.

For both protocols the step height was adjusted until the 1 RM was established (to the nearest 1 cm). Participants were able to discount attempts they felt were submaximal. Feedback from participants regarding the ease or difficulty of the previous trial guided the magnitude of step height change. A rest period of two minutes (or longer if the participants still reported fatigue) was used between each attempt for both protocol A and B. Once the 1RM for the first protocol tested had been determined, testing of the other leg commenced. The step height achieved in the initial testing session was not disclosed to participants, and they were asked not to vary their usual physical regimen between testing occasions.

Statistical Analysis

Interrater reliability was determined for both protocols using intraclass correlation coefficients (ICC(2,1)) with 95% confidence intervals. Standard error of the mean (SEM) was calculated using the formula $SD \times \sqrt{1 - ICC}$. Percent close agreement (PCA) within 2 cm was determined for both protocols. This represents the percent of cases where the two tester’s scores for 1RM were different by 2 cm or less.

Table 1 One Repetition Maximum Lateral Step Test Heights (cm) for Protocol A and B

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Testing Occasion 1</th>
<th>Testing Occasion 2</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>No pause (A) (cm)</td>
<td>27</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>Pause (B) (cm)</td>
<td>25</td>
<td>26</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 Reliability Statistics for Lateral Step-Up Protocols A and B

<table>
<thead>
<tr>
<th>No pause protocol (A)</th>
<th>Pause protocol (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra class correlation coefficient (95% confidence interval)</td>
<td>0.94 (0.87 to 0.97)</td>
</tr>
<tr>
<td>Standard error (cm)</td>
<td>1.5</td>
</tr>
<tr>
<td>Percent agreement within 2 cm</td>
<td>83.3</td>
</tr>
</tbody>
</table>
Results

All 24 participants completed all testing. Mean values (± SD) for 1 RM lateral step height is given in Table 1 for protocol A and B. Both lateral step protocols demonstrated excellent interrater reliability with ICC\(^{(2,1)}\) values of 0.94 for both protocols (95% CI, 0.87 to 0.97 and 0.85 to 0.97 for protocol A and B, respectively; Table 2). There was no significant difference between the reliability of protocol A and protocol B (Table 2). The SEM was 1.5 cm for protocol A and 1.5 cm for protocol B. The percent close agreement within 2 cm for all testers was 83.3% and 79.2% for protocol A and protocol B, respectively (Table 2). Combining data from both test protocols, excellent reliability was found regardless of the experience of the tester. ICCs were 0.83 and 0.90 for experienced physical therapists and physical therapy students, respectively.

Comments

The results of this study demonstrate that the lateral step test has excellent intertester reliability when used as a 1RM test in healthy adults (18 to 81 yrs). The reliability is not significantly influenced by modifying the protocol to use a 3 second pause time.

The lateral step test has not previously been examined as a measure of maximal lower limb strength, so the reliability shown here cannot be directly compared with other studies. The high reliability found here, however, is comparable to other commonly used measures of maximal strength, including hand held dynamometry and isokinetic dynamometry. In the lower limb high intraclass correlation coefficients have been reported for hand held dynamometry (0.82 to 0.99)\(^{6-9}\) and isokinetic testing (0.71 to 0.99).\(^{10-23}\) The reliability of the lateral step test performed here is also comparable to other functional strength tests such as the single leg hop test for maximum distance (ICC 0.89 to 0.97).\(^{18,32-34}\) Importantly, unlike the single leg hop test for maximum distance, the reliability of the lateral step test has demonstrated reliability in a wide age range (18 to 81 years).

The SEM can be used to determine the range in which a participant’s “true score” could be expected to fall when the amount of error associated with the measurement is considered. The results of this study indicate the error associated with lateral step measures obtained using protocol A is 1.47 cm and protocol B is 1.51 cm. The magnitude of change required to be 95% confident genuine change has occurred is calculated by multiplying the SEM by 1.96. Thus, a change in step height of approximately 3 cm (whether protocol A or B is being used as an outcome measure) would be required to confidently attribute the observed change to a difference in strength rather than measurement error.

The results of this study indicate that the lateral step test is not significantly influenced by a learning effect when 10 warm up repetitions are performed prior to testing. The mean step height for the first and second testing occasions was identical for protocol A and differed by less than 1 cm for protocol B, which highlights the stability of the measure (Table 1). In contrast, the single leg hop for maximum distance has been found to be associated with a learning effect with significant improvements across repeated trials being reported.\(^{18,32,34,45}\)
There are several benefits offered by the lateral step test that make it clinically useful as a measure of lower limb strength. First, regardless of the tester’s clinical experience, the lateral step test was found to be reliable. As reliability was very similar between protocol A and B, data was combined to calculate ICCs for experienced and inexperienced testers. ICCs for both sets of testers were similar with values for experienced physical therapists being 0.83 and for physical therapy students being 0.90, indicating clinical experience does not influence the reliability of the test. This high reliability was achieved with minimal tester training, despite testers having no previous experience using the test. Prior to data collection, testers received written instructions detailing the protocols and attended a single 30-minute training session to ensure consistent testing criteria. These findings support the generalization of the results to other physical therapists implementing the test.

Second, the demands of the lateral step are highly specific to important functional activities such as stair ascent or descent and standing up or sitting down and are therefore highly likely to reflect functional ability.

A further benefit of the lateral step test as a strength measure is that the lateral step motion is widely used as a strengthening exercise during rehabilitation of the lower limb following ACL injury, stroke, and hip fracture. The most appropriate strength measures reflect the manner in which strength is trained according to the specificity of training principle whereby strength gains at the angles, velocities, and postures in which training occurred are usually greatest. Thus, the lateral step test would be suitable to measure strength changes in rehabilitation programs where the lateral step is used as part of the training program.

In this study, the demands of the lateral step test to determine 1 RM was increased or decreased by changing step height rather than weight as is typically done in RM testing. Adjusting height is commonly used in strength training programs using step or sit to stand exercises to progress the intensity of the program. Although we cannot be certain how much the increase in physical demand associated with increasing step height is related to demands on strength, participants regularly expressed that strength was the limiting factor during testing, although this information was not formally recorded. Furthermore, delayed onset muscle soreness reported by some of the subjects implied that the test required very high levels of muscle contraction. Also in support of the relationship between increase in step height and strength demands, a previous study on patients with poliomyelitis found that maximum step height was a predictor of knee extensor strength. As the physical demands of the lateral step test at any given height are related to the individual’s height, it would not be appropriate to compare step heights between individuals as is the case with other functional strength measures such as sit to stand or hop tests, without first normalizing the data.

The lateral step test is simple and inexpensive to administer. In this study steps of 1, 2, 7, 11, 22, and 46 cm heights were combined to allow step height changes of 1 cm increments to be measured for all patients whose age, strength, and physical ability varied considerably. The step heights achieved by participants ranged from 10 to 40 cm. No specialized or expensive equipment was required. The portability of the test allows it to be used in both the clinic and the home environment. This is especially beneficial for physiotherapists treating older people, for whom a greater proportion of rehabilitation is occurring at home.
It is important to note that participants were excluded from this study if they had pain in either lower extremity, and all participants had sufficient strength to weight bear unilaterally. The reliability of the lateral step needs to be investigated in participants with lower limb pain, pathology, or disease.

The results of this study demonstrate that the lateral step test for measuring maximal lower extremity strength has excellent reliability (ICC = 0.94) in healthy subjects. Adding a 3-second pause to the protocol does not change the reliability. The test required minimal tester training, regardless of clinical experience. As well as demonstrating excellent reliability, the lateral step up test is functional, inexpensive, and easy to administer.

References


**Appendix**

**Testing Procedure**

1. Participants need to be given a demonstration by the tester along with a verbal explanation of correct technique, as well as compensatory movements which will constitute a failed attempt. The participant should then perform 10 repetitions at a height of 7 cm, during which time they are able to ask questions and receive feedback regarding their technique from the tester.

2. The first leg will be tested with 2 minutes rest between attempts, longer if the patient reports to be still feeling fatigued.
3. The 2nd leg will be tested using the opposite protocol to that used for the first leg tested.

4. One week later the exact protocol and order of tests will be repeated by a different tester. Results of the first session will not be available until completion of the 2nd testing session.

**No Hold Protocol**

For all testing, participants should wear shorts and remove their shoes.

1. The participant starts standing on top of the step, with the first toe of the leg to be tested positioned 30 cm back from a wall that they are facing, with the medial edge of the foot being tested in line with the edge of the step. The participant may place one fingertip on the wall directly in front of them, on the side of their tested leg, for balance. The fingertip can be at a height which is comfortable for the participant.

2. The nontested leg is to remain in knee extension and ankle dorsiflexion throughout the test.

3. The nontested leg is to be lowered straight down from the step until their heel just makes contact with a sponge placed on the floor. The sponge should be positioned so the long side runs parallel with the foot on the step, and such that the end of the sponge is in line with the heel on the step. The tester may give verbal/visual guidance as to the distance participants have to go until they reach the sponge. The participant is to touch the sponge with his or her heel, but not to weight bear on the floor. Once contact has been made as judged by tester observation they will indicate that the participant can straighten the tested leg to complete the repetition.

   During testing, stand at the side of the participant on the nontesting leg. Watch for compensatory movements. It is normal to use some lateral pelvic tilt, and this will be allowed during the test as long as the participant does not deviate the trunk to either side. The participant is to maintain head and shoulders upright through the entire movement. Watch for the tendency of the participant to flex the trunk and move the hips posteriorly. This is not acceptable and should be considered a failed attempt.

4. The tested leg is then straightened to lift the nontested side back up onto the step. The participant may not push off in any way from the nontested leg. It is critical that the knee remains extended and the ankle dorsiflexed at this time. A short 1-second rest is allowed but not essential before the next repetition begins.

5. The movement is to be completed at a smooth slow speed that is comfortable for the participant.

**Hold Protocol**

1. Protocol as above, however when participants touch the sponge with their heel, they are required to pause with their heel touching the sponge but not weight bearing on the floor.
2. The pause position is to be maintained for 3 seconds. *The 3 seconds begins when the participant is in the pause position and is still.* This is to be judged by the tester. The tester is then to start a stop watch and count the 3 seconds out loud. “One . . . two . . . go.”

3. On “go” the participant will straighten the tested leg and complete the repetition. Again the participant should receive verbal encouragement to ensure a maximal effort.