Intrarater and Interrater Reliability of the Single-Leg Squat Test

Michael D. Kennedy, PhD; Lisa Burrows; and Eric Parent, PhD, PT • University of Alberta

Functional testing is widely used to assess recovery of normal joint function during rehabilitation or as a screening tool to detect impairments during a preparticipation evaluation. Functional testing can indirectly reveal deficiencies in muscular strength and power. Many lower extremity functional testing protocols include unilateral or bilateral squats. The single-leg squat (SLS) is often preferred, because it replicates a common athletic position, and it requires control of the trunk over a single leg thereby challenging all of the lower extremity joints.

Three-dimensional (3D) kinematic analysis is considered the gold standard for assessment of lower extremity function. Because it is expensive and laboratory-based, its use for preparticipation screening of athletes is impractical. The SLS has been widely used for clinical assessment of lower extremity joint function, even though criteria to define an adequately performed squat are limited. The SLS has merit for preseason screening. It is simple to administer and may yield important information about an athlete’s lower extremity function.

Criterion-referenced assessment of the SLS by athletic trainers (0–3 rating: hip flexion < 65 degrees, hip adduction < 10 degrees, and knee valgus < 10 degrees) has been reported to yield poor interrater Kappa values for occurrence of knee valgus (0.28) and hip adduction (0.16). An acceptable level of interrater reliability has been reported for experienced physical therapists who rated a videotaped SLS performance as either good or poor. An acceptable level of intrarater reliability has been reported for the number of repetitions of a single-limb task performed in 30 seconds by patients with patello-femoral pain syndrome (antero-medial lunge ICC = 0.82; step down ICC = 0.94). Discriminate validity of SLS assessment has been established for 3D motion analysis of knee movement in subjects with and without patella-femoral pain syndrome, as well as acceptable test-retest reliability for frontal plane femoral angle (ICC = 0.74–0.83). Intrarater and interrater reliability of SLS performance assessment through repetition counts and subjective ratings have not been established. The purpose of this study was to determine the intrarater and interrater reliability of a standardized single-leg squat screening test that was used to assess a large cohort of varsity athletes.

Key Points

- Single-leg squat (SLS) reliability is adequate for evaluation of acceptable repetitions completed.
- Interrater reliability is poor for identification of a specific limiting factor.
- Interrater reliability improves when choice is either “perfect” or “with any limitation.”
- More standardized criteria are needed to reduce subjectivity assessment of SLS performance.
Procedures

Forty-two varsity athletes were randomly selected from a pool of 281 athletes, representing 11 varsity teams that included the following sports: men’s and women’s basketball, men’s and women’s hockey, men’s and women’s wrestling, men’s and women’s swimming, women’s volleyball, and women’s field hockey. Inclusion criteria included completion of a preparticipation medical questionnaire, active varsity athlete status, and participation in preseason fitness testing. The project was conducted in accordance with the research ethics board guidelines of the institution, and participants provided informed consent. Three board-certified athletic therapists (ATs) with 3–5 years of professional experience and a minimum of 5 years of university education administered the test. Training included an afternoon of pilot work that involved each therapist assessing SLS performance multiple times and collective feedback that included instructions to give to athletes during administration of the test.

Single-Leg Squat Test Administration

The SLS test procedure is depicted in Figure 1. Each participant was provided with a verbal cue at the beginning of the SLS test to “think about sitting on a chair and extend your non-weight bearing leg at the knee with your foot just off the floor.” Prior to assessment, one practice trial was performed on each leg, and the AT asked the participant if any clarification of instructions was needed. A metal bar was positioned at the level of the knee joint line in a standing position so that the participant’s buttocks touched it at full descent. The athlete was instructed to perform five repetitions on each leg, starting with the right leg. The AT stopped the athlete when loss of balance made performance of a subsequent squat impossible. The test was also stopped if the athlete did not achieve adequate squat depth (i.e., failure to touch the buttocks to the metal bar in > 3 seconds). All properly performed SLS repetitions were filmed in the frontal plane by a digital video camcorder (Panasonic, PV-S320) that was located 9 m from the center of the squat station and 1 m above the floor.

Reliability Evaluation

Four investigators reviewed the frontal plane video recordings of the test, including the AT who administered the test. All four of the investigators had a similar level of knowledge in the area of athlete health and performance. The investigators counted the number of repetitions adequately completed and selected the most significant factor that limited performance among the following: trunk, hip, knee, lower leg, and other (Table 1). Perfect performance was defined as the absence of any obvious abnormality in trunk, hip, knee, lower leg, or other lower extremity movement. The limiting factors were selected by a panel of clinicians that included the athletic therapy staff, a physical

Figure 1  Start position of single-leg squat test (A) and completion of descent into position of hip and knee flexion (B). Successful completion of a repetition required the buttocks to touch metal bar that was positioned at knee joint line when participant was standing. All videotapes were obtained in the frontal plane at a distance of 9 meters (C).
therapist, and a sport medicine physician. The raters were provided with a brief description of each limiting factor. The criteria for identification of abnormal movement were based on common interpretations made by clinicians. To ensure that the raters understood the criteria, all raters attended a training session that included a review of five video clips. Each rater had a printed outline of rating criteria available to reference during the SLS performance assessment. The identification of a limiting factor was separated from the assessment of the number of repetitions adequately completed. The criterion for determination of the number of consecutive repetitions completed was loss of postural control that prevented performance of the subsequent squat repetition.

Intrarater reliability was evaluated on the basis of two separate assessments that were performed on successive days by the main rater, and interrater reliability was evaluated on the basis of one assessment performed by each of the four raters. The intraclass correlation coefficient (ICC) and standard error of measurement (SEM) were calculated for the intrarater and interrater counts of adequate repetitions. The intrarater and interrater consistency in identification of limiting factors was determined through calculation of Kappa coefficients. Interrater reliability for dichotomous classification of perfect versus imperfect performance was also evaluated by calculation of the Kappa coefficient.

Findings

Participants were able to complete an average of four consecutive acceptable repetitions on both the right and left leg. The knee (RL) and lower leg (LL) were the most commonly identified limiting factors (Table 2). Table 3 summarizes the reliability values for both intrarater and interrater reliability. The intrarater reliability

<table>
<thead>
<tr>
<th>Limiting Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trunk</td>
<td>Twisting, lateral shift, lateral leaning the torso.</td>
</tr>
<tr>
<td>Hip</td>
<td>Non-weightbearing hip drops below weightbearing hip. The weightbearing hip moves laterally beyond knee and ankle plumb line.</td>
</tr>
<tr>
<td>Knee</td>
<td>The knee moves medially beyond the medial malleolus.</td>
</tr>
<tr>
<td>Lower leg</td>
<td>Shaking of the lower leg is observed above the ankle.</td>
</tr>
<tr>
<td>Other*</td>
<td>Overall pattern of movement deemed abnormal</td>
</tr>
</tbody>
</table>

* Overall pattern was not functionally appropriate, but not primarily identified by one of the other anatomic areas.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Repetitions</th>
<th>Descriptive Statistic (n = 42)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Right (95% CI)</td>
<td>Left (95% CI)</td>
</tr>
<tr>
<td></td>
<td>4.2 ± 1.5 (1.2 – 7.2)</td>
<td>4.3 ± 4.5 (1.3 – 7.2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Limiting Factor</th>
<th>None</th>
<th>Trunk</th>
<th>Hip</th>
<th>Knee</th>
<th>Lower Leg</th>
<th>Other</th>
<th>Any limiting factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Repetitions</td>
<td>21.4% (11.7 - 35.9)</td>
<td>23.8% (13.5 - 38.5)</td>
<td>11.9% (5.2 - 25.0)</td>
<td>26.2% (15.3 - 41.1)</td>
<td>7.1% (2.4 - 19.9)</td>
<td>9.5% (3.8 - 22.0)</td>
<td>78.6% (64.1 - 88.3)</td>
</tr>
<tr>
<td>Limiting Factor</td>
<td>19.0% (10.0 - 33.3)</td>
<td>21.4% (11.7 - 35.9)</td>
<td>9.5% (3.8 - 22.0)</td>
<td>16.7% (8.3 - 30.6)</td>
<td>23.8% (13.5 - 38.5)</td>
<td>9.5% (3.8 - 22.0)</td>
<td>81.0% (66.8 - 90.1)</td>
</tr>
</tbody>
</table>
for dichotomous classification of SLS performance as either “perfect” or “with any limitation” was better than that for identification of specific limiting factors (RL Kappa = 0.62; LL Kappa = 0.40).

**Discussion**

The reliability for counts of acceptable SLS repetitions was adequate, but suboptimal. Intrarater reliability was only marginally better than interrater reliability for repetition count. The results suggest a need to further delineate criteria for judging the number of acceptable repetitions. The reliability for identification of the primary limiting factor was poor. Many athletes displayed more than one limiting factor, and raters differed in their selections of the primary limiting factor. DiMattia et al.² suggested that more than one muscle or joint may be responsible for a change in mechanics, which may preclude identification of a “primary” limiting factor. Previous research that has involved dichotomous classification of a similar squat performance as either good or poor reported similar results, which was reported to provide acceptable reliability.⁶ Acceptable reliability for health research has been defined as a level of agreement greater than 0.8, with minimal variance.⁹ Thus, moderate reliability may be achieved by discriminating between a perfect performance and an impaired performance. Improved standardization of test administration procedures and more clearly delineated criteria for determination of movement abnormalities will be necessary to improve consistency in identification of specific limiting factors.

**Clinical Relevance**

The results of this study illustrate the subjective nature of functional assessments that are routinely performed in the clinical setting. Despite our finding of suboptimal reliability, criteria for preparticipation screening of a large number of athletes must be simple enough to permit a rapid assessment of test performance. Improved criteria for assessment of SLS performance could include range of motion (e.g., no greater than five degrees of valgus during descent) and additional qualitative indicators of neuromuscular control (e.g., absence of visible shaking of the knee during descent).³ Training sessions for evaluators should include videotaped examples of “perfect” and “deficient” performances of the SLS test. Our results suggest that the reliability of the test will be greatest if the evaluator makes a dichotomous classification of performance as either perfect or deficient. Athletes who demonstrate deficient performances could be evaluated further, whereas athletes with perfect performances would not need a follow-up analysis. Such an approach to rating SLS test performance increases the feasibility of screening a large number of athletes as a component of the preparticipation physical examination.

The value of the SLS test for quantification of injury risk has not been established, but deficient SLS performance has been documented in patients with patello-femoral pain syndrome,¹⁰ and deficits in core and hip strength have been shown to adversely affect SLS performance.¹¹,¹² Future research should prospectively assess SLS test performance in a cohort of athletes prior to the start of the season and subsequently relate the test results to injury occurrence to determine the prognostic value of screening.

**Conclusion**

The subjective nature of rating SLS test performance limits its reliability for identification of the primary

---

**Table 3. Summary of Intrarater and Interrater Reliability Values for the Number of SLS Repetitions and Limiting Factor**

<table>
<thead>
<tr>
<th>Repetitions</th>
<th>Limiting Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Right leg</strong></td>
<td><strong>Left Leg</strong></td>
</tr>
<tr>
<td>** ICC (95%CI) **</td>
<td>** ICC (95%CI) **</td>
</tr>
<tr>
<td>Intrarater</td>
<td>0.85 (0.74-0.92)</td>
</tr>
<tr>
<td>Intrarater</td>
<td>0.80 (0.71-0.88)</td>
</tr>
<tr>
<td>Intrarater</td>
<td>0.31 (0.08)</td>
</tr>
</tbody>
</table>

ASE = Asymptotic Standard Error
that contribute to deficient performance. Acceptable reliability was found for dichotomous classification of SLS performance as perfect or deficient, and good to excellent reliability was found for counts of the number of acceptable SLS repetitions. The SLS may have value as a lower extremity screening test, but refinements of test administration procedures and criteria for classification of abnormal movement patterns are needed to improve consistency in identification of specific factors that contribute to deficient performance.

References

Michael Kennedy is an assistant professor and Director of the Student and Athletic Health Research Laboratory, faculty of Physical Education & Recreation at the University of Alberta. Edmonton, Alberta, Canada.

Lisa Burrows is an MSc student and assistant athletic therapist in the Department of Athletics, University of Alberta.

Eric Parent is an assistant professor with the Faculty of Rehabilitation Medicine at the University of Alberta and clinical scientist at Alberta Health Services, Glenrose Rehabilitation Hospital in Edmonton.

Develop rehabilitation programs for various tendinopathies

Lower Extremity Tendinopathies
Ebonie Scase and Jill Cook, PhD
©2010 • Online course • ISBN 978-0-7360-7722-4
$69.00 ($77.95 CDN, £58.95 UK, €70.70 EURO, $122.10 AUS, $152.60 NZ)

For more information or a free course preview, visit www.HKEducationCenter.com.
(800) 747-4457 US • (800) 465-7301 CDN
(44) (0) 113-255-5665 UK
(08) 8372-0999 AUS • 0800 222 062 NZ
(217) 351-5076 International

Audiences: Certified athletic trainers, sport physical therapists, sports medicine professionals, and certified strength and conditioning specialists.

Lower Extremity Tendinopathies is a comprehensive resource that helps clinicians better understand, evaluate, and treat these common injuries. The interactive course with companion e-book—written by leading tendinopathy researcher Jill Cook in conjunction with clinician Ebonie Scase—systematically reviews the pathology of lower limb tendinopathies and teaches you how to treat these challenging and often-recurring injuries.

This evidence-based course and text provides insight into the complex and ongoing noninflammatory, pathological alterations that occur in the tendon matrix and presents a three-stage tendon pathology continuum that will enhance your clinical reasoning process and treatment regimens. After learning the essential background knowledge, you will move on to assessment techniques by looking at intrinsic factors as well as load-related training error to examine how they relate to tendon injuries. The course also includes video of several functional assessments and discusses the advantages and disadvantages of ultrasound and MRI imaging for assessing tendinopathies.