Form Matters—Technical Cues in the Single Leg Heel Raise to Failure Test Significantly Change the Outcome: A Study of Convergent Validity in Australian Football Players

Brady Green 1,2,*, Molly Coventry 3, Tania Pizzari 2,4, Ebonie K. Rio 4,5 and Myles C. Murphy 1,3©

1 School of Health Sciences, The University of Notre Dame Australia, Fremantle, WA 6160, Australia; myles.murphy@nd.edu.au or m.murphy@ecu.edu.au
2 School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, VIC 3086, Australia; t.pizzari@latrobe.edu.au
3 Nutrition and Health Innovation Research Institute, School of Medical and Health Sciences, Edith Cowan University, Joondalup, WA 6027, Australia; m.coventry@ecu.edu.au
4 La Trobe Sport and Exercise Medicine Research Centre, Melbourne, VIC 3086, Australia; e.rio@latrobe.edu.au or ebonier@australianballet.com.au
5 The Australian Ballet, Melbourne, VIC 3006, Australia
* Correspondence: brady.green@nd.edu.au or b.green@latrobe.edu.au

Abstract: Practitioners routinely use the single leg heel raise (SLHR) to quantify calf function in healthy and injured populations. Despite this, approaches vary and the impact of cueing on SLHR performance and results interpretation in athletes is unknown. The primary aim of this study was to quantify the level of agreement of the cued versus non-cued SLHR tests. The secondary aim was to explore test outcomes and the potential impact of intrinsic factors. Cued and non-cued SLHR tests were conducted in fifty-one Australian football players (23 women, 28 men). Metronome pacing (60 bpm) and five key cues were included in the cued condition. The level of agreement (Bland–Altman) between tests was measured for capacity (repetitions to failure) and asymmetry. Data from 100 legs were included. The non-cued and cued SLHR tests demonstrated poor agreement in both capacity and asymmetry. More repetitions to failure were performed in the non-cued SLHR \( [\text{Mean (SD)} = 33.9 (10.3) \text{ vs. } 21.9 (5.3), p < 0.001] \) and men had greater capacity \( (36.8 (10.4) \text{ vs. } 30.3 (9.2), p < 0.001) \). During the cued SLHR, older players \( (\text{age} \geq 30 \text{ years}) \) had reduced calf muscle function. Cueing the SLHR test significantly changes the result—outcomes are not comparable or interchangeable with the commonly used non-cued SLHR. These findings can guide practitioners quantifying calf capacity.

Keywords: heel raise; calf; soleus; gastrocnemius; strength; endurance; injury prevention; muscle strain

1. Introduction

Calf muscle capacity is critical to lower limb function across the spectrum of locomotive and functional activities [1–3]. Practitioners commonly use the single leg heel raise (SLHR) test to quantify calf muscle strength–endurance as a measure of capacity in healthy [4] and clinical populations [5], as well as athletes [6,7].

Methods to perform the SLHR test vary [8,9]. The range of loading rate (beats per minute), ankle dorsiflexion (over the edge of a step vs. 10° incline vs. flat ground) and plantar flexion (partial vs. maximum height) requirements reported in research may have contributed to the inconsistency seen clinically [10]. Specific instructions and cues have been a key consideration in contemporary research and clinical settings to help standardise practice [11]. Performance and technique cues during the SLHR test may alter the number of repetitions to failure performed and potentially provide a more robust
indication of an individual’s capacity when screening athletes [12,13]. Although the long-held clinical hypothesis that cueing is warranted is yet to be substantiated in Australian football players [12]. Whether cueing the SLHR test in this population significantly impacts test outcomes and the practical interpretation of the results between conditions is unknown.

Standardising SLHR performance using cues may increase the strength of the comparisons that can be made between different people and samples [8], or after injury. Conversely, consistency and optimal performance appear unlikely clinically when these efforts are not made [14], resulting from ‘uncontrolled’ or non-cued testing. Sub-optimal use of the calf muscles, or test completion at rates that are not consistent with strength measurement [15], are potential negative consequences of the non-cued SLHR. For example, without metronome pacing, more rapid loading rates may bias elastic tissues and recoil rather than measuring the repeated force-generating capacity (i.e., strength–endurance) of the contractile elements within the calf muscles [15–17].

Clinicians value understanding how test procedures can impact (and potentially improve) their practice and patient outcomes [18]. This is especially apparent for practitioners aiming to use these data to help prevent or manage injuries involving the calf muscle–tendon unit, that could be associated with reduced strength-endurance, such as calf muscle strain injuries [12,19] and Achilles Tendinopathy [20,21]. Calf muscle–tendon unit injuries are problematic in athletes [22,23], and, particularly in Australian football, the consistent prevalence and susceptibility of players to recurrent calf muscle strain injuries represents a growing injury burden [24,25]. For these reasons there is high clinical value in improving practice in the implementation of SLHR testing and the interpretation of the results, based on the approach used.

Despite the potential practical value, studies investigating SLHR outcomes under non-cued versus cued conditions in athletic populations are lacking [9]. An evaluation has not been undertaken in male and female Australian football players either. The primary aim of this study was to quantify the level of agreement of the cued versus non-cued SLHR tests in terms of capacity (i.e., repetitions to failure) and asymmetry (i.e., between-leg differences). The secondary aim was to measure and describe the SLHR test outcomes, including an exploration of the potential impact of intrinsic factors, to help guide future prevention research.

2. Materials and Methods

Human Research Ethics Committee approval was obtained for this study from The University of Notre Dame Australia (Approval: 2022-149F). Fifty-seven participants from the Western Australian Football League men’s and women’s leagues were recruited for SLHR testing during the 2023 preseason. Excluded participants (1) opted out after the first testing session due to soreness (n = 2) or (2) provided incomplete baseline data (n = 4). One leg was excluded for two participants due to an injury that precluded the leg from testing. All participants provided informed consent via an online Qualtrics survey (Qualtrics, Provo, UT, USA. https://www.qualtrics.com (accessed November 2022–March 2023).

Two commonly researched [10,26,27] and clinically utilised SLHR methods were used. Tests were conducted at two separate preseason training sessions spaced ≥6 days apart (i.e., to ensure the first testing session would not impact the second). One of three authors (either BG, MC or MCM) assessed the non-cued SLHR test. One author (BG) performed all cued SLHR tests. To avoid contamination of the cued SLHR, the non-cued SLHR was conducted in the first training session for almost all (92%) players. Reliable methods were used in the cued SLHR, which involved standardised performance and technique cues and associated failure criteria (Supplementary Table S1) [13,26,27].

Baseline characteristics were self-reported using Qualtrics prior to SLHR testing. The following data were recorded: chronological age (years), height (cm), weight (kg), ethnicity (as per Australian Bureau of Statistics recommendations), gender (man, woman, non-binary, prefer not to say), leg dominance (left, right, ambidextrous), competition level (league, reserves), injury history (yes/no for different injury regions), playing age
(years playing Australian Football) and calf resistance training history (completing calf-focused resistance training: yes/no). International Olympic Committee (IOC) standards were used to categorise injury history data by region and type [28]. Chronological age data were handled both as continuous and categorical. Two age categories were used in separate analyses, which were based on current evidence and validated approaches [29]: (1) Age > cohort median, as older athletes are susceptible to recurrent calf muscle strain injuries [30] and (2) age ≥ 30 years, as individuals over 30 years of age have a higher incidence of calf muscle strain and Achilles tendon injuries [31,32].

In the non-cued SLHR, players performed repetitions until volitional failure was reached (i.e., volitional failure was the only reason for test cessation) [10]. During the cued SLHR, the test ended if volitional or technical failure was reached. Technical failure was based on observation and specifically occurred when cues could not be followed (Supplementary Table S1): (1) metronome pace could not be maintained (60 bpm/30 repetitions per minute), (2) knee flexion occurred, (3) a hip propulsive strategy was used, (4) vertical displacement was lost (i.e., movement forwards rather than upwards/trunk lean) and (5) there were marked deviations from plantar flexion height and alignment [7,10,26]. Testers provided generic verbal encouragement in both tests to help ensure the test was maximal. During the cued SLHR only, all players received a verbal reminder if a deviation in technique occurred, but if the error remained the test was ceased (Supplementary Table S1) and these repetitions were not counted [7,10,26]. The number of repetitions for each leg was recorded for both tests. Data were imported into and analysed in SPSS (IBM SPSS Statistics Windows, Version 29.0. Armonk, NY, USA: IBM Corp).

Frequency distributions were generated to describe categorical variables. Continuous data were evaluated for normality (raw visualisation, Shapiro–Wilk test) and were explored using descriptive statistics (mean, median, standard deviation (SD), interquartile range, range). Bland–Altman plots were generated to evaluate the level of agreement between tests for both repetitions to failure and asymmetry. Limits of agreement were calculated (mean difference ± 1.96x*standard deviation) [33,34] and simple linear regressions were used to identify the risk of proportional bias.

The SLHR test scores were non-normally distributed. Performance was measured using the Wilcoxon Signed-Rank test to compare repetitions to failure between the non-cued vs. cued SLHR tests. Parametric (independent samples t-test: cued SLHR) and non-parametric (Mann–Whitney U: non-cued SLHR) tests were used to explore the potential differences in test performance based on categorical intrinsic characteristics (e.g., age (>cohort median (yes/no); ≥30 years (yes/no)), leg dominance (dominant vs. non-dominant leg), gender (women vs. men), ethnicity (Indigenous vs. non-Indigenous) and exposure to calf-focused resistance training (yes/no)). Simple linear regressions were performed for continuous variables to explore potential correlations with performance (e.g., chronological age, playing age, height, weight, body mass index (BMI)). Results were considered significant if \( p < 0.05 \).

3. Results

3.1. Player Characteristics

Fifty-one players [women: 23 (45%); men: 28 (55%)] were included. Players were predominantly of Australian ethnicity (88%), with 6% of participants being of ‘Aboriginal or Torres Strait Islander’ (i.e., Indigenous Australian) ethnicity. Players were predominantly league-level (e.g., ‘senior’ WAFL/W level) (65%). Almost half of the players (45%) had a previous lower limb injury, and the majority (87%) of previous lower limb injuries were muscle–tendon injuries. The complete demographics of the cohort are described in Supplementary Tables S2 and S3.

3.2. Calf Muscle Strength–Endurance

Matched non-cued and cued SLHR results from 100 legs were included. Players were tested before (18%), during (25%) and after (35%) field-based training, or during
a designated lower body gym session on a separate day to field-based training (22%). Cessation of the cued SLHR was due to volitional (43%) or technical (57%) failure. The most common reasons for technical failure were reduced plantar flexion height = 49.1%; knee flexion = 24.6%; hip strategy = 10.5%; trunk rocking back and forth = 8.8%; and foot/ankle alignment during plantar flexion = 7%.

Repetitions to failure between tests were moderately correlated ($r = 0.49$, $p < 0.001$). Players recorded on average 12 fewer repetitions during the cued SLHR [mean (SD) = −12 (8.9) repetitions: $Z = −8.53$, $p < 0.001$] (Figure 1, Table 1). Only four limbs (4%) performed the same or more repetitions during the cued SLHR (equal: 1%; more: 3%). Men had a greater non-cued SLHR capacity compared to women [mean (SD) = +6.5 (18.6) repetitions: $U = 755$, $p < 0.001$] (Table 1). Indigenous players had a reduced cued SLHR capacity compared to non-Indigenous Australian players [mean (SD) = −3.4 (7.7) repetitions: $t(15.5) = −3.8$, $p = 0.002$].

![Figure 1. Scatter, box and whisker plots for repetitions to failure during the non-cued ('uncontrolled') and standardised ('cued') single leg heel raise tests.](image)

Table 1. Calf strength–endurance during the non-cued and cued single leg heel raise tests.

<table>
<thead>
<tr>
<th></th>
<th>Non-Cued SLHR</th>
<th>Cued SLHR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Median (IQR)</td>
</tr>
<tr>
<td>All legs</td>
<td>33.9 (10.3) *</td>
<td>32 (27–40)</td>
</tr>
<tr>
<td>Leg dominance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominant</td>
<td>34.1 (9.9)</td>
<td>32 (27.8–40.3)</td>
</tr>
<tr>
<td>Non-dominant</td>
<td>33.6 (10.8)</td>
<td>31 (25–40.5)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>30.3 (9.2)</td>
<td>29 (24–33.5)</td>
</tr>
<tr>
<td>Men</td>
<td>36.8 (10.4) *</td>
<td>35 (29–45)</td>
</tr>
<tr>
<td>RT history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completing Calf RT</td>
<td>35.7 (10.4) *</td>
<td>33 (28–41)</td>
</tr>
<tr>
<td>Not completing Calf RT</td>
<td>30.9 (9.6) *</td>
<td>29.5 (25–36.3)</td>
</tr>
</tbody>
</table>

Legend: SLHR = single leg heel raise; SD = standard deviation; IQR = interquartile range; RT = resistance training. * significant difference between test conditions (i.e., uncontrolled vs. standardised SLHR); † significant difference within test condition.

Chronological age was not associated with SLHR performance when handled as a continuous variable (non-cued: $r = −0.14$, $p = 0.16$; cued: $r = −0.12$, $p = 0.22$). A subgroup analysis revealed that players ≥30 years old (n = 8 legs) were weaker during the cued [mean (SD) = 17 (2.2) vs. 22.1 (5.3) repetitions: $t (98/4.7) = −4.2$, $p = 0.01$] but not during the
non-cued SLHR (mean (SD) = 31.5 (7.1) vs. 33.9 (10.4) repetitions: U = 169, p = 0.69). Players who reported completing calf muscle resistance training performed a greater number of SLHR repetitions to failure during both tests compared to players not undertaking this form of loading (non-cued: (SD) = +4.8 (19) repetitions, U = 873.5, p = 0.03; cued: +3.3 (10) repetitions, t (90.6) = 3.3, p = 0.001). Leg dominance and all other intrinsic player characteristics were not associated SLHR repetitions to failure under both conditions, including injury history (p > 0.05 for all).

Between-leg asymmetry during the cued [mean (SD) = 2.7 (2.1) repetitions, median: 2, range: 0–9] and non-cued [mean (SD) = 2.9 (2.9) repetitions, median: 2, range: 0–12] SLHR tests were similar (Z = −0.46, p = 0.65). Individual asymmetries during both tests were not correlated (r = 0.06, p = 0.67). Intrinsic player characteristics were not associated with asymmetry during either SLHR test (p > 0.05 for all).

3.3. Agreement between Tests

Clinically wide limits of agreement were identified when comparing the cued and non-cued SLHR repetitions to failure (mean difference ± 1.96x*SD = −5.0 to 29.5 repetitions) (Figure 2A) and asymmetry (mean difference ± 1.96x*SD = −6.9 to 7.6 repetitions) (Figure 2B). Linear regression also revealed a risk of proportional bias, with a moderate–strong correlation (r = 0.63, p < 0.001) between mean SLHR repetitions to failure and the difference in repetitions to failure performed between tests (Figure 2A). When comparing asymmetry, a proportional bias (r = 0.29, p = 0.15) or a bias towards greater asymmetry during either test (mean difference: 0.33 (3.7) repetitions, 95%CI −0.7–1.4) (Figure 2B) was not identified. Interpreted together, these findings demonstrate poor agreement between the SLHR tests and indicate that the scales of measurements within the non-cued and cued SLHRare not clinically equivalent.
Figure 2. Bland–Altman plots of the differences in repetitions to failure (A) and asymmetry (B) measured during the non-cued and cued single leg heel raise tests.
4. Discussion

Three key clinically relevant outcomes were demonstrated from this study: (1) the cued and non-cued SLHR tests do not measure the same construct (i.e., calf muscle strength–endurance) interchangeably—the test results are not directly comparable; (2) intrinsic factors (i.e., sex, chronological age, calf-focused resistance training exposure, ethnicity) may impact the repeated force-generating capacity of the calf muscles and warrant investigation in a larger cohort; and (3) the strength data generated from this research now exist to guide practitioners.

The repetitions to failure obtained from the cued and non-cued versions of the SLHR tests are not comparable clinically [33,34]. Our results support a long-standing hypothesis that these test conditions should not be considered interchangeable field-based or clinical measures of calf muscle function in practical settings [8]—form matters. By proxy, this may extend into SLHR interventions and explain why some SLHR interventions have a minimal effect on SLHR repetitions to failure [35] and even clinical improvement. Our results revealed that the scales of measurement within these tests are not equivalent and that Australian football players are biased towards recording a greater number of repetitions during the non-cued SLHR.

From our results and current evidence, we hypothesise that the cued SLHR test provides a more accurate representation of the repeated force-generating capacity of the contractile elements within the calf muscle–tendon unit [7,8]. Task constraints (e.g., controlling the loading rate at 60 bpm and the provision of technical cues) likely make the cued SLHR test preferential for muscular performance and increase the time under tension (i.e., cumulative load) encountered [15].

Non-cued conditions often involve the use of faster loading rates in order to utilise the stretch–shortening cycle and reduce the metabolic cost to the working muscles—incurring a greater contribution from the elastic elements of the muscle–tendon unit [16,17]. It is also possible that non-cued testing alters calf muscle work. For example, uncontrolled mechanics may result in an altered range of motion, a loss of vertical displacement (peak and cumulative), suboptimal calf muscle recruitment and the utilisation of other muscles when executing the task. From these perspectives, cueing the SLHR test may provide clinicians with a robust method to determine the repeated force-generating capacity (i.e., strength–endurance) of the calf muscles. In addition to the role of cues in standardising SLHR performance, clinicians should consider the potential impact of intrinsic factors on calf muscle strength–endurance.

We identified intrinsic factors that may be associated with reduced calf muscle strength–endurance and warrant future evaluation. Non-modifiable (female athletes, age ≥ 30 years, Indigenous ethnicity) and modifiable (previous calf muscle resistance training exposure) factors impacted the number of repetitions to failure in the SLHR of Australian football players in the current study. While the low number of players with Indigenous ethnicity should be acknowledged, the exploratory nature of the second study aim and the need for further research is highlighted. When considering the susceptibility of women, Indigenous and older athletes to injuries involving the calf muscles (e.g., calf muscle strain [24,36–38] or tendon injury [32]), combined with the findings of the current study, future exploration of the potential preventative role of strengthening in these athletes may be supported as well.

Intrinsic factors can negatively impact calf muscle ‘performance fatigability’ and strength [7]. The association between older age and a previous calf muscle–tendon unit injury with reduced strength qualities, or the structural (e.g., cross-sectional area) and functional (e.g., total displacement) correlates of them, are established [7,26,39,40]. Our results support the potential impact of ageing on SLHR strength–endurance. The fewer cued repetitions to failure in players aged >30 years is consistent with previous data, highlighting the reduction in SLHR capacity with each decade of life [26]. These data could reflect the onset of more pronounced age-related changes affecting calf muscle structure or function.
Our findings may suggest that the SLHR test has the capacity to detect differences in the calf function of men and women athletes participating in the same sport. Previous research has not always demonstrated these differences according to sex, including maximum voluntary isometric contraction [4]. Nonetheless, while reduced calf muscle strength–endurance is hypothesised to increase the injury risk clinically, due to mechanisms such as reduced load tolerance and performance fatigability [12,41], whether different strength qualities mitigate the impact of non-modifiable intrinsic factors (e.g., age, injury history) on injury susceptibility warrants prospective evaluation in women’s and men’s league athletes.

This study provides calf muscle strength–endurance data, assessed using the SLHR test for both men’s and women’s Australian football codes. Calf strength–endurance is an important capacity to quantify clinically [8,26], as well as from a performance perspective [6]. Despite being a clinical measure of strength, it is not always closely correlated to maximum isometric calf strength [4]. Strength–endurance represents a discrete capacity [42,43] that may be especially important to athletes that require the calf muscles to possess ‘performance fatigability’ (i.e., fatigue resistance) and repeatedly generate force, such as during running [6,41]. Our cued SLHR data are also comparable to recent findings in New Zealand rugby athletes using a standardised protocol [mean (SD) = 22 (5) vs. 20 (5) [7] repetitions]. We also found that completing specific loading for the calf muscles (in the form of calf-focused resistance training) was associated with better strength–endurance. While evidence for the preventative role of strengthening in the calf muscles is lacking [44], this preliminary finding potentially highlights the practical impact of devoted loading on improving calf muscle function in sporting populations.

Strengths and Limitations

Our study is the first evaluation of the role of clinically cueing the SLHR test in any Australian football code. Our study is also the first quantification of SLHR strength–endurance in Western Australian Football League men’s and women’s players. The order of the tests were not randomised to reduce the risk of bias associated with performing the cued SLHR prior to the non-cued SLHR. The risk of confounding was also minimised by separating the testing sessions by at least 6 days, reducing the likelihood that performance would be affected by fatigue from the first testing session [4]. When conducting the cued SLHR test, all players were cued to focus on the same technical elements—contributing to standardised task completion and participant intent. In addition to controlling the loading rate, a metronome (at 60 bpm) standardised the time under tension, though it is not known what the optimal pace is, despite its common use in research [10]. As a result of our findings, clinicians are provided with data that may act as a minimum acceptable standard. It is unavoidable that some athletes may have experienced fatigue due to completing training either prior to or concurrent with the testing session, or as a result of other training activities completed in the preceding days. Due to availability, it was also unavoidable that four players performed the cued SLHR test first.

5. Conclusions

The non-cued SLHR test commonly used in clinical practice does not provide comparable capacity or asymmetry data to the cued version of the SLHR test. Strength testing methods that provide the most value for calf injury prevention is unknown and warrants further research.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/biomed4020007/s1, Table S1: clinical procedure for non-cued and cued single leg heel raise tests; Table S2: intrinsic player characteristics; Table S3: player demographic information.


35. Murphy, M.C.; Travers, M.; Chivers, P.; Debenham, J.; Docking, S.I.; Rio, E.K.; Gibson, W.; Ardern, C. Can we really say getting stronger makes your tendon feel better? No current evidence of a relationship between change in Achilles tendinopathy pain or disability and changes in Triceps Surae structure or function when completing rehabilitation: A systematic review. *J. Sci. Med. Sport.* 2023, 26, 253–260. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.