

Original Research

# Reliability and Measurement Error of the Qualitative Analysis of Single Leg Loading (QASLS) Tool For Unilateral Tasks

Gemma N Parry<sup>1a</sup>, Lee C Herrington<sup>2</sup>, Allan G Munro<sup>1</sup><sup>1</sup> School of Health and Society, The University of Salford, <sup>2</sup> School Health and Society, The University of Salford

Keywords: kappa paradox, movement quality, QASLS, single leg squat, single leg land

<https://doi.org/10.26603/001c.88007>

---

## International Journal of Sports Physical Therapy

Vol. 18, Issue 5, 2023

---

### Background

Movement quality assessment is popular within clinical and sporting practice, due to the contribution diminished or suboptimal movement quality is believed to have on musculoskeletal (MSK) injury risk. Various movement quality assessments exist, many are limited to bilateral or jumping movements evaluation. Qualitative analysis of single leg loading (QASLS) is a new clinical assessment tool for unilateral tasks that utilizes a dichotomous scoring system of ten questions relating to the segmental body regions of the trunk, lower and upper limb.

### Purpose

To determine the intra and inter-rater, within- and between-session reliability of the QASLS tool during two unilateral movement tasks, and provide insight to measurement error and smallest detectable difference (SDD).

### Study Design

Reliability Study

### Methods

Fifteen healthy females (mean age 19 years SD2; height 167 cm, +/- 6; weight 56 kg, +/- 6) completed two unilateral tasks, single leg squat (SLS) and single leg landing (SLL), within session data collection occurred on the same day, with between session data collection occurring seven days later. Tasks were scored with the QASLS tool via video playback. Intra-Class correlation coefficients ( $ICC_{k,3}$ ) were used to measure within and between session reliability, and Kappa coefficients and percentage of exact agreement (PEA%) were used to determine intra and inter-rater reliability. Standard error of measurement (SEM) and the SDD for the compound score of each limb was calculated.

### Results

Within session reliability of QASLS scores was good ( $ICC = 0.82-0.86$ ) for SLS and moderate ( $ICC = 0.67-0.87$ ) for SLL. Between session reliability was moderate ( $ICC = 0.69-0.87$ ) for SLS and excellent ( $ICC = 0.92-0.93$ ) for SLL. SEM was less than 1 point, and SDD for compound score ranging from 1.0-2.5 points. Intra-rater agreement of compound QASLS score was near perfect ( $k = 0.85-1.00$ ; PEA% 90-100%) and agreement of individual components was substantial- near perfect ( $k = 0.13-0.74$ ; PEA% 78-100%). Inter-rater agreement for compound QASLS scores ranged from non-substantial ( $k = 0.13-0.74$ ; PEA% 43.3-90%) for SLS and non-slight ( $k = 0.03-0.17$ ; PEA% 43.3-60%) for SLL.

---

<sup>a</sup> Corresponding Author:

Gemma N Parry

School of Health and Society, University of Salford, Manchester, M6 6PU, United Kingdom

0161 295 6222

[G.Parry1@salford.ac.uk](mailto:G.Parry1@salford.ac.uk)

## Conclusions

The QASLS movement analysis tool can be used to analyze movement quality during two unilateral loading tasks with moderate to excellent within and between session reliability. PEA% was acceptable for inter-rater agreement, however rater education training is recommended to develop more acceptable levels of reliability.

## Level of Evidence

3

## INTRODUCTION

Musculoskeletal (MSK) profiling tools, particularly of the lower limb, are widely used to highlight injury risk, and influence the composition of rehabilitation and conditioning programming for return to training (RTT).<sup>1</sup> Suboptimal movement quality, or movement that is impaired, inefficient, asymmetrical, functionally compensated or diminished<sup>2</sup> is believed to have an impact on injury risk,<sup>2-4</sup> as such, the assessment of movement quality by practitioners is popular in clinical practise. Movement quality is considered to be a modifiable factor that can influence injury risk, as research has continued to show associations between movement variability and musculoskeletal (MSK) injury,<sup>5</sup> therefore tools that capture and monitor changes in movement quality are of growing interest to practitioners.

Quantitative and qualitative human movement analysis is widely utilized in sport and clinical practice. Laboratory-based three-dimensional (3D) analysis is purveyed as the “gold standard”,<sup>6,7</sup> however, in the non-research environment it is expensive, time consuming, and often unfeasible to set-up. The real-world need to capture data on large numbers of participants frequently, has led to several qualitative visual rating criteria emerging as a cheaper, more accessible means of human movement analysis.<sup>7</sup>

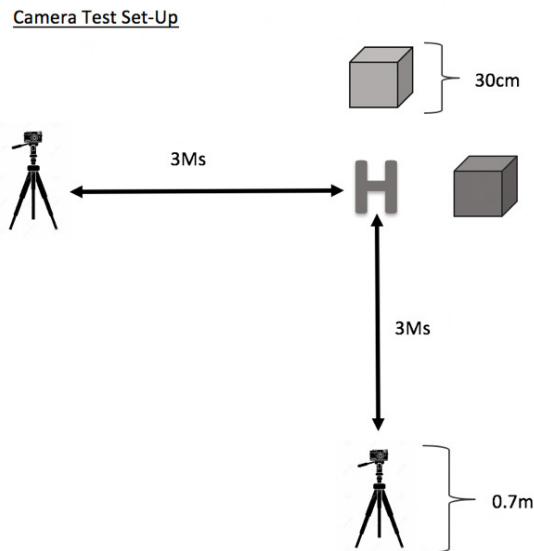
Several authors<sup>3,8-11</sup> cultivated and explored the use of lower limb qualitative scales during functional movements, to provide clinicians with simply applied means of identifying movement quality issues within the MSK system. The development of the Landing Error Scoring System (LESS) provided practitioners with a reliable and valid tool<sup>9,12</sup> with minimal set-up time and efficient post-test evaluation through the assessment of a jump landing technique. Unfortunately, analysis of trunk position is limited, with no evaluation of the upper limb and evaluation of bilateral jumping movements only. Torso and upper limb positioning has been shown to influence the lower limb during landing<sup>13-15</sup> potentially impacting lower limb loading, patterning, movement quality, and subsequent injury risk. While these studies have acknowledged the contributions and impact the torso and upper limb may have on the biomechanics on the lower limb, protocols measuring and capturing torso and upper limb movement within whole movement patterning are lacking. Additional investigation into developing a methodology via qualitative means is therefore warranted.

The Qualitative Analysis of Single Leg Loading (QASLS) is a relatively new clinical assessment tool that incorporates biomechanical analysis of movement patterns of the lower limb, upper limb, and torso during single-leg loading

tasks in addition to providing a compound score.<sup>16</sup> This allows for comparison between limbs but is also arguably more replicative of the unilateral hopping, landing, and change of direction patterns observed in sport. Unilateral limb evaluation is important because it remains the most common mechanism of the majority of lower limb overuse and traumatic injuries.<sup>11</sup> Furthermore, the effective evaluation of unilateral movement quality provides valuable markers for identifying both sporting and non-sporting individuals at risk of injury.

Research into QASLS use is limited, with only one study to date reporting on intra-rater reliability.<sup>17</sup> However, the study was limited by sample size of participants and raters and no insight into absolute measurement error was presented. Measurement error values are an integral element of understanding the value of a tool, task, or intervention, as they inform a clinician if any notable changes have occurred and whether they are representative of a truly observed change and not attributed to systematic error, chance or an intervention. While ICCs allude to the reliability, they remain insensitive to sample variety.<sup>18</sup> It is therefore recommended that a standard error of measurement (SEM) and the smallest detectable difference (SDD) also be presented to accurately identify and establish parameters to classify changes in performance.<sup>19</sup> The SEM informs clinicians of the measurement error of a test, is presented in the same units as the measurements and therefore allows scrutiny to other SEM presented within the literature. The SDD provides a base value which should be surpassed to distinguish real change from random error.

Currently, no investigation has documented measurement error values or within and between session values of the QASLS system. If the measurement error, reliability and validity of the qualitative method can be established, practitioners will be able to use the QASLS system with certainty. This will assist with informing observation around individual and group performances, movement variability and associated injury risk, to support the development of better profiling practises. To determine the intra and inter-rater, within- and between-session reliability of the QASLS tool during two unilateral movement tasks, and provide insight to measurement error and smallest detectable difference (SDD). A secondary purpose was to report on the associated measurement error and SDD. It was hypothesised that QASLS scores would demonstrate good to excellent reliability for all tasks, however it was expected that inter-rater reliability would demonstrate greater variability depending on rater experience.



**Figure 1. Test-space set-up for unilateral loading tasks: H = tape markings on floor, M = meters**

## MATERIALS AND METHODS

### PARTICIPANTS

Fifteen vocationally trained elite pre-professional<sup>20</sup> female dancers,<sup>21</sup> volunteered for this study. The colleges medical team approved participation by all participants who were uninjured and had no history of surgical intervention within the prior six months. Informed written consent was provided by each participant. This study was approved by the University Research and Ethics Committee.

### PROCEDURES

Participants attended testing within their performance facility on three separate occasions during a three-week testing period. Within-session data collection occurred on the same day with Session 2 occurring one hour after Session 1, and between-session data collection occurring one week later. All testing sessions were conducted at the same time of day to account for circadian rhythm changes that may affect performance tasks. Participants performed single-leg squat (SLS) and a single- leg landing (SLL) on both the right and left legs. The order of the movement task was randomly selected (by participants selecting face down cards with the tasks written) in Session 1 and repeated in Session 2.

### MOVEMENT ASSESSMENT TASKS

Each movement trial was recorded from frontal and sagittal planes with video cameras (Panasonic Lumix DMC-FZ200) at a sampling rate of 100 Hz, positioned three meters from the testing/landing zone, with each camera lens set upon tripods set to a height of 0.7m. (Figure 1).

### SINGLE-LEG SQUAT (SLS) (FIGURE 2A)

Participants were asked to stand on one limb (self-selected) facing the frontal plane, they were verbally instructed to squat as low as possible as if sitting back and down on a chair and return to the start position. Participants were then asked to repeat on the opposite limb. No further instructions were provided so as not to influence the individual's movement strategy.

### SINGLE-LEG LAND (SLL) (FIGURE 2B)

Participants stood on a 30cm high box, they were asked to step forward and land onto the contralateral limb holding the landing for at least two seconds. No further instructions were provided so as not to influence the individual's movement strategy.

### QASLS PROCEDURE

QASLS is a visual rating tool that provides segmental scoring of an observed unilateral loaded movement pattern on a 10-point scale. Adopting a dichotomous scoring strategy of six body segments (Arm, Trunk, Pelvis, Hip, Knee and Ankle) the tool utilizes a region criteria where appropriate strategy scores a zero and suboptimal strategy scores a one (Figure 3). A higher QASLS score indicates a greater number of suboptimal strategies used to complete a task, and a lower QASLS score indicates fewer component strategies required to complete the tasks. Within the QASLS framework, operational definitions are provided in conjunction to the movement strategies observed at each segmental level, along with instruction relating to compound dichotomous scoring. The QASLS system is advocated to be used so that the compound score, irregardless of if from a singular or multiple effort, is comprised of the total number of strategies required by an individual to complete the task irregardless of frequency. Namely, if three or five repetitions of a unilateral task are completed, even if a sub-optimal strategy is observed once or five times the practitioner awards one mark, resulting in the cumulation of a "sub-optimal" trial. Five repetitions were evaluated based on previous reported procedures within the only article to evaluate reliability of the tool,<sup>17</sup> to designate a compound overall QASLS score.

Videos were analyzed using QASLS scoring sheet (Figure 3), the scoring performance was derived for each participant from both the frontal and sagittal plane views, with each video viewed then marked and scored. Three raters (LH, AM and BO) independently scored participants across the five trials via QASLS scoring sheet (Figure 3) having viewed both the frontal and sagittal place videos for each participant. The three raters were provided with written instructions on how to assess the movement tasks via QASLS, could review the videos as many times as required to obtain a score and were blinded to the other raters scores, to avoid potential bias.

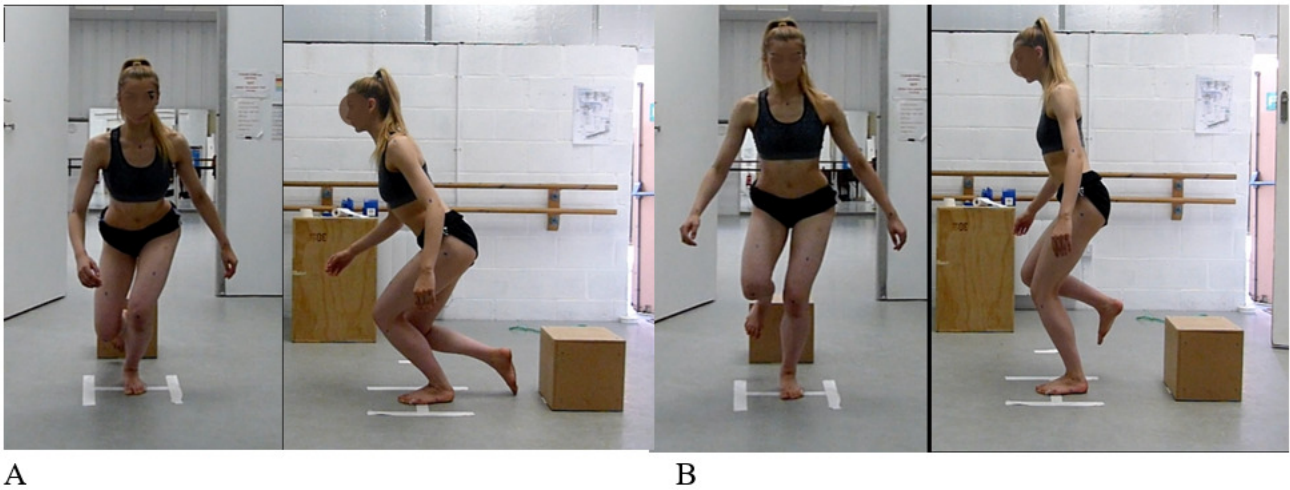


Figure 2. Example of Single-Leg Squat (SLS) (A) and Single-Leg Landing (SLL) (B) movement tasks

Qualitative analysis of single leg loading			
Date:	Patient:		
Condition:	Left	Right	Bilateral
QASLS	Task: Single leg squat Single leg step down Single leg hop for dist	Left	Right
Arm strategy	Excessive arm movement to balance		
Trunk alignment	Leaning in any direction		
Pelvic plane	Loss of horizontal plane		
	Excessive tilt or rotation		
Thigh motion	WB thigh moves into hip adduction		
	NWB thigh not held in neutral		
Knee position	Patella pointing towards 2 <sup>nd</sup> toe (noticeable valgus)		
	Patella pointing past inside of foot (significant valgus)		
Steady stance	Touches down with NWB foot		
	Stance leg wobbles noticeably		
	Total		

Figure 3. QASLS Scoring Sheet, displaying the 10-point visual rating system. If a rater observes an appropriate strategy at the segmental region the participant scores zero, if a rater observes a suboptimal strategy the participant scores one. A higher QASLS score indicates a greater number of segmental strategies used to complete a unilateral loading task, and a lower QASLS score indicates a lesser number of required segmental strategies to complete the unilateral loading tasks

## STATISTICAL ANALYSIS

SPSS for windows (version 25) (SPSS Inc, Chicago, IL.) was used to determine within and between session reliability agreement. Within and between session reliability agreement<sup>22</sup> of the QASLS rating criteria was determined using intra-class correlations ( $ICC_{k,3}$ ) for each limb and movement assessment task, with 95% confidence intervals (%CI). A custom-made spreadsheet (Microsoft Excel Version 16.16.22) calculated standard error of measurement (SEM) and smallest detectable change (SDD) values. Within and between-session reliability of composite scores were calculated using a mean rating ( $ICC_{k,3}$ ) 2-way mixed-effects absolute agreement model.  $ICC_{k,3}$  values were interpreted as > 0.90 excellent, 0.75-0.9 good, 0.50-0.75 moderate, and < 0.50 as poor.<sup>18</sup> Statistical significance was set at  $p < 0.05$ .

Due to the dichotomous nature of the QASLS system,<sup>22</sup>,<sup>23</sup> intra and inter-rater compound scores and individual components of the QASLS tool were determined via the percentage of exact agreement (PEA%) and kappa co-efficient. Cohens Scales<sup>24</sup> were selected to interpret kappa values where 0.81-1.00 is almost perfect agreement, 0.61-0.80 substantial, 0.41-0.61 moderate, 0.21-0.40 fair and 0.01-0.20 none to slight. Acceptable PEA% has been described<sup>25</sup> as between 75-90%, however, this remains specific to each study. As there are no current universally accepted interpretations, and in the absence of literature supporting clear interpretation of PEA,  $\geq 66\%$  has been chosen as a reflection of majority agreement (55-75% as defined in other papers.<sup>24,26</sup> SEM and SDD were calculated to represent and establish the smallest worthwhile change and identify random error scores between test sessions. With formulas taken from previously reported methods.<sup>16,19,27-29</sup>

## RESULTS

Fifteen participants originally volunteered for this study, due to corrupted video data, one participant was excluded resulting in an analysis of 14 participants (age  $19 \pm 2$  height  $167 \pm 6$  cm body mass  $56 \pm 6$  kg).

### WITHIN AND BETWEEN SESSION RELIABILITY OF QASLS (TABLE 1)

No significant differences were noted between limbs ( $p = 0.20$ ) or testing sessions for both tasks. Within and between-session reliability for both tasks were moderate to excellent ( $ICC = 0.67-0.93$ ). Within-session reliability of the QASLS composite score (0-10) for SLS was good for both limbs (Right  $ICC = 0.82$ , 95%CI = .36-.96; Left  $ICC = 0.86$ , 95%CI = .49-.97). SEM for within-day reliability was 0.82 and 0.72 points the SDD 2.28 and 2.00 points on a ten-point scale for the right and left limbs respectively. Similar results were observed in the right SLL task ( $ICC = 0.87$ , 95% CI .42-.97, SEM 0.45, SDD 1.26) however, left limb performance was moderate ( $ICC = 0.67$ , 95%CI .25-.92, SEM 0.89, SDD 2.45).

Between-session reliability of the composite QASLS score for the SLL was slightly less than the within-session

scores (Right  $ICC = 0.72$  95%CI .15-.93, Left  $ICC = 0.69$  95%CI .07-.92) graded as moderate. SEM for SLS between-session reliability was 0.96 and 0.99, the SDD was 2.65 and 2.75 for the right and left limbs respectively. The SLL task demonstrated greater between-session reliability than the SLS task ( $ICC = 0.92-0.93$ ) with SEM of 0.41 and SDD of 1.14 for the right limb and 0.47 and 1.52 for the left limb. SEMs for both within-session and between-session were less than 1 with the SDD ranging from 1.0-2.5 points. This suggests an error measurement of 1 across testing time frames and that a change of 1-3 points would be necessary to demonstrate a minimal detectable change.

### INTRA-RATER RELIABILITY OF QASLS

Intra-rater reliability was "perfect to excellent" agreement ( $k=0.85-1.0$ ) for both movement tasks. Except for right SLL ( $k=0.85$ , PEA = 90%) where items 7 and 8 in the QASLS criteria were disagreed on for participants 1 & 5 respectively. Therefore, individual components of the QASLS tool were further analyzed with details found in Tables 2 and 3.

### INTER-RATER RELIABILITY OF QASLS

Table 4 presents the inter-rater reliability for compound QASLS scores, rater reliability for SLS ranged from non-to substantial ( $k=0.13-0.74$ ) and for SLS non-slight for SLL ( $k = 0.03-0.17$ ). Single leg squat demonstrated the biggest discrepancy between PEA%. Rater 2 (R2) demonstrated the greatest difference between Rater 1 (R1) and Rater 3 (R3) (43%-90% respectively). R2 and R3 demonstrated the highest levels of PEA% (53.3%-90%) with each other, agreement with R1 was lower for both R2 (43-47%) and R3 (53-60%).

Inter-rater reliability for individual and categorical components ranged from non-substantial ( $k = .000-0.80$ ) (table 5). Kappa values were unable to be established for all raters and participants scores, due to the lack of variance in 1 or both rater scores. Despite high values of PEA% (such as 100%) low kappa scores were still noted. During the SLS raters demonstrated the best agreement for pelvic, knee and touchdown components (items 3,4,7,8 and 9 on the criteria), however this was different for SLL where raters demonstrated the best agreement for upper limb, trunk and ankle components (items 1,2 and 10 on the criteria).

## DISCUSSION

The purpose of this study was to determine the intra- and inter-rater and within- and between-session reliability of the QASLS tool during two unilateral movement tasks, the SLS and SLL. A secondary purpose was to report on the associated measurement error. Overall compound QASLS scores suggest moderate to excellent reliability ( $ICC = 0.82-0.87$  and  $ICC = 0.69-0.93$  for within and between session, respectively), indicating the QASLS tool is sufficiently reliable for movement analysis of the unilateral movement tasks of squatting and landing. Results highlighted that there was a measurement error of 1 between testing time-frames and that a change in 1-3 points is required to de-

**Table 1. Within and between session reliability of qualitative criteria (QASLS) for Single Limb (SL) tasks. Values are presented as mean  $\pm$ SD**

Within-Session							
Task/Limb	QASLS Score	SD	ICC	95%CI	SEM	SDD	CV%
SL Squat Right	4.78	1.79	0.82	.359-.956	0.82	1.28	37
SL Squat Left	4.78	1.86	0.86	.491-.966	0.72	1.00	40
SL Landing Right	4.33	1.50	0.87	.423-.970	0.45	1.26	25
SL Landing Left	4.78	1.30	0.67	.025-.917	0.89	2.45	27
Between-Session							
SL Squat Right	4.78	1.79	0.72	.146-.929	0.96	1.65	40
SL Squat Left	4.78	1.86	0.69	.068-.922	0.99	1.75	35
SL Landing Right	4.38	1.46	0.93	.716-.983	0.41	1.14	34
SL Landing Left	4.78	1.46	0.92	.393-.989	0.47	1.52	40

**Table 2. Intra-Rater Reliability of Qualitative Rating Criteria (QASLS)**

Testing Occasion	Kappa coefficient (95%CI)		Percentage of Exact agreement (PEA%)	
	R SLS	L SLS	R SLL	L SLL
1	1.0	1.0	100%	100%
2	0.85 (0.73-0.98)	1.0	90%	100%

L= Left limb, R= Right limb, SLL = Single-Leg Landing, SLS = Single-Leg Squat

**Table 3. Intra-Rater reliability scoring of Right SLL by individual QASLS component**

QASLS Items	QASLS Component	PEA%	Kappa Value (95%CI)
1	<b>Arm Strategy</b> (Excessive arm movement to balance)	100	1.0
2	<b>Trunk Alignment</b> (Leaning in any direction)	100	1.0
3	<b>Pelvic Plane</b> (Loss of horizontal plane)	100	1.0
4	<b>Pelvic Plane</b> (Excessive tilt or rotation)	100	1.0
5	<b>Thigh motion</b> (WB thigh moves into hip adduction)	100	1.0
6	<b>Thigh motion</b> (NWB thigh not held in neutral)	100	1.0
7	<b>Knee position - noticeable valgus</b> (Patella pointing towards 2 <sup>nd</sup> toe)	78	0.78 (0.69-0.92)
8	<b>Knee position - significant valgus</b> (Patella pointing past inside of foot)	78	0.75 (0.64-0.90)
9	<b>Steady Stance</b> (Touch down with NWB foot)	100	1.0
10	<b>Steady Stance</b> (Stance leg wobbles noticeably)	100	1.0

termine a change in performance. This is believed to be the first study to provide within and between session reliability specifically for the QASLS tool, therefore there is no prior research to compare results to. Other qualitative movement screens that use dichotomous scales similar to QASLS such

as the Functional Movement Screen (FMS™) and LESS have reported similar test-retest reliability values.<sup>27</sup> Shultz et al. established that compound FMS scoring was relatively good (ICC = 0.6) for elite female athletes when tested seven days apart. The reliability values within this study are consis-

**Table 4. Inter-Rater Reliability of QASLS**

Raters	Kappa Coefficient (95%CI)		Percentage of Exact Agreement (PEA%)	
	SLS	SLL	SLS	SLL
1 Vs 2	.125 (-0.18-0.43)	.030 (-0.30-0.63)	43.3	46.7
1 Vs 3	.182 (-0.16-0.52)	.171 (-0.18-0.52)	53.3	60
2 Vs 3	.737 (0.51-0.97)	.129 (-0.17-0.43)	90	53.3

SLL = Single-Leg Landing, SLS = Single-Leg Squat

**Table 5. Inter-Rater PEA% and Kappa Scoring QASLS participants during SLS and SLL**

Participant	Rater	No of Agreements	Total Tasks	PEA%	Discrepancy between PEA	Kappa Coefficient (95%CI)	Agreement
<b>Single Leg Squat (SLS)</b>							
1	1 Vs 2	6	10	60		.310 (0.18-0.44)	Fair
1	1 Vs 3	6	10	60	20	.200 (0.07-0.33)	Non-slight
1	2 Vs 3	8	10	80		.600 (0.47-0.73)	Moderate
2	1 Vs 2	3	10	30		.310 (0.18-0.44)	Fair
2	1 Vs 3	4	10	40	60	.000*	N/A*
2	2 Vs 3	9	10	90		.800 (0.67-0.93)	Substantial
3	1 Vs 2	9	10	90		.286 (0.16-0.42)	Fair
3	1 Vs 3	7	10	70	30	.400 (0.27-0.53)	Fair
3	2 Vs 3	6	10	60		.800 (0.67-0.93)	Substantial
<b>Single Leg Land (SLL)</b>							
1	1 Vs 2	4	10	40		.200 (-.36-.76)	Non-slight
1	1 Vs 3	5	10	50	10	.000*	N/A*
1	2 Vs 3	5	10	50		.087 (-.50-.68)	Non-slight
2	1 Vs 2	6	10	60		.310 (-0.07-0.69)	Fair
2	1 Vs 3	8	10	80	20	.524 (-0.05-1.0)	Moderate
2	2 Vs 3	6	10	60		.310 (-0.07-0.69)	Fair
3	1 Vs 2	4	10	40		.200 (-.41-0.81)	Non-slight
3	1 Vs 3	5	10	50	10	.000*	N/A*
3	2 Vs 3	5	10	50		.000*	N/A*

\*Kappa values unable to be calculated due to lack of variance between 1 or both raters

tent with those reported in the above literature, 95% CI remain large and are potentially due to the variability within human movement. Despite ICCs being commonly reported in reliability studies, within qualitative research many interpretations of the ICC exist. Therefore clarification of excellent or good reliability is elusive with studies classifying broad values (from 0.40->0.80) as excellent or fair to good.<sup>30</sup>

Intra-rater reliability was found to be excellent (PEA% 90-100%,  $k = 0.85-1.0$ ) and is in agreement with other work<sup>17</sup> (although limited to SLS) that has analyzed rater-reliability. There are believed to be no comparable papers currently available concerning the reliability of the QASLS tool and a SLL task.

Overall inter-rater compound QASLS scoring was non-substantial for SLS ( $k = 0.03-0.17$ ), which is lower than previously reported reliability,<sup>17</sup> however PEA% ranged from 43-90%. Results were comparative to other qualitative measures that have analysed SLS.<sup>2</sup> Chmielewski et al. showed PEA of 32-48% during SLS via segmental approach and weighted kappa values of 0.00-0.53.<sup>31</sup> Schultz et al. (2013) described inter-rater agreement via Krippendorff's  $\alpha$  ( $k_a$ ) as poor ( $k_a = .38$ ) when using the FMS on female athletes.

Inter-rater reliability of each QASLS component was fair to almost perfect ( $k = 0.40-1.0$ ). Regarding individual component analysis, best scores appeared to be between R2 and R3 during the SLS with 100% agreement in 8/10 categories. The three raters demonstrated differences of agreement in components 6, 7 and 8 (NWB thigh movement and knee valgus). Previous findings<sup>17</sup> have also concluded rater disagreement of the valgus components during the scoring of SLS in university participants. The raters in this research, as with the raters in this current study received no formal training and were reliant on the operational differences presented within the tool.

The operational differences presented in components 7 and 8 of the QASLS tool are very similar in their description, which might not be concise enough for raters to deduce the difference between the terms "noticeable" and "significant". It is unclear if the reliability results observed in this study are attributed to the level of rater training or vagueness of the operational definition of knee valgus. This might also provide an explanation for why these differences were not reflected in SLL results where the greater complexity of the task suggests valgus is easier to spot within the movement pattern.

Inter-rater reliability was unable to be calculated for some categorical components due to the lack of variance between raters and observations of movement errors, and is described as the kappa paradox.<sup>24</sup> When conceiving this study, due to minimal research regarding the QASLS tool, important decisions regarding the interpretation of the variable generated by the QASLS tool were considered, as this would dictate the statistical approach. Unlike quantitative variables seen in 2D or 3D movement data that follow interval or ratio principles that can be parametrically analyzed, a case could be made for QASLS being classified as ordinal (due to the dichotomous element of the segmental evaluation where the outcome falls into two cate-

gories of yes or no) and interval (compound scores that run on a scale of 0-10 where the gaps are proportional), thus, how best to establish tool performance relating to reliability and agreement was open to debate. Previous visual rating methods that also use dichotomous scoring, treat data as interval.<sup>9-11</sup> The QASLS tool has been designed as a clinical instrument to provide a score that guides practitioners in evaluation of single-leg loading patterns of the whole system, the decision was therefore made to evaluate data as an interval variable.

## STUDY APPLICATIONS AND LIMITATIONS

A strength of this study was the presence of both the kappa and PEA% analysis methods, yet, neither method is without fault. PEA is a precise, interpretable and easily determined statistic but does not account for chance rater guesses.<sup>24</sup> The kappa value eliminates any chance rater choices, but is limited in sensitivity in data prevalence that clusters very high or very low, or in homogenous populations where estimate agreement appears exclusively lowered.<sup>22</sup>

Described as the "base rate problem,"<sup>32</sup> and usually seen in a moderate to high PEA and a low kappa score, the paradox has been shown to occur in very simple cases with only two evaluators and two outcomes (similar to this paper's design), at equal points of the sensitivity and specificity of the raters, or if the prevalence of one of the raters assigns one specific outcome more frequently<sup>33</sup> as observed between R1 and R2, and R1 and R3.

Data indicated that at individual participant level, movement variability was high with different movement patterns deployed within the same movement pattern, but as an overall cohort movement patterns were consistent and therefore variability was low. It is unsurprising that this data set has high levels of homogeneity that is likely unavoidable in the analysis of a sub-elite population. Analysis of movement quality remains a key aspect of profiling and programming within the sporting environment, it is therefore likely that future research will continue to be focused within this population. It is prudent to acknowledge the limitations of this non-heterogeneous sample and the likely impact that would have on a kappa score, and establishing a truly heterogeneous elite sporting population would be difficult to achieve. The argument is therefore made that the limitation is within the statistic rather than the direct relevance of the population. Future research into other sporting populations such as injured or adolescents where a cohort could be relatively heterogeneous in their construct would be warranted.

A final limitation of the study is the level of rater training provided in using the QASLS tool. The findings of the kappa results are potentially suggestive of a redesign of the test instrument or retraining of the raters.<sup>24</sup> Given the robustness of the intra-rater and between and within-session results, the requirement for full instrument redesign appears unlikely. Raters were provided with the same standardized instructions on how to administer the tool along with the basic component operational definitions embedded within the tool. It is possible that each rater interpreted



each section in a specific way which ultimately impacted agreement.

While training around the use and interpretations of other movement visual rating criteria is standardized by other authors, currently there are no training programs available for the QASLS tool. It was therefore decided that understanding the current interpretations, limitations, and strengths of the QASLS tool as it is currently used without training within clinical practice, was more pertinent for this study, to better guide any future recommendations around QASLS training content.

Rater training is an important component to qualitative analysis<sup>34</sup> but rarely appears to be delivered in a standardized way.<sup>3,9,10,31,32</sup> Providing raters with greater instruction around operational differences and providing potential examples of each observable segmental strategy (e.g. trunk dominant, hip avoidant, knee dominant) may assist raters clinically in standardizing their scoring methods. This is particularly evident around components 7 and 8 of the QASLS tool where identifying minor deviations in knee movement appeared more difficult. This is also supported by the better levels of reliability and agreement observed during SLL, where the larger deviations seen within that movement pattern are more discernible. The QASLS tool has demonstrated satisfactory within and between and intra-rater reliability for its use by practitioners. Results demonstrate that the current operational definitions within the tool are adequate for intra-rater use, further work on rater-education to include standardized examples, may maintain more consistent and objective analysis to improve agreement ratings before more widespread use.

## CONCLUSION

The QASLS tool demonstrated moderate to excellent within- and between-session reliability, and excellent intra-rater reliability, and could be used as a movement quality tool to evaluate unilateral squatting and landing tasks by a single rater. PEA% was acceptable for inter-rater agreement, but results should be interpreted with caution. It would be beneficial to explore the operational definitions used within the tool, so inter-rater agreement could be elevated to more acceptable levels. A potentially homogenous population was selected, and while not unrepresentative of a healthy, elite sporting population, it is unclear how the QASLS tool may be influenced by more heterogenous samples such as injured populations or adolescent younger age groups. Future additional investigation within additional groups of athletes will provide greater understanding into the application and continuing development of the QASLS and other visual observation tools of movement quality.

## CONFLICTS OF INTEREST

The authors report no conflicts of interest.

Submitted: June 27, 2022 CDT, Accepted: August 31, 2023 CDT  
© The Author(s)



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

## REFERENCES

1. Bahr R, Clarsen B, Ekstrand J. Why we should focus on the burden of injuries and illnesses, not just their incidence. *Br J Sports Med.* 2018;52(16):1018-1021. [doi:10.1136/bjsports-2017-098160](https://doi.org/10.1136/bjsports-2017-098160)
2. Whittaker JL, Booyesen N, De La Motte S, et al. Predicting sport and occupational lower extremity injury risk through movement quality screening: A systematic review. *Br J Sports Med.* 2017;51(7):580-585. [doi:10.1136/bjsports-2016-096760](https://doi.org/10.1136/bjsports-2016-096760)
3. Chmielewski TL, Hodges MJ, Horodyski M, Bishop MD, Conrad BP, Tillman SM. Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: A comparison of 2 rating methods. *J Orthop Sports Phys Ther.* 2007;37(3):122-129. [doi:10.2519/jospt.2007.2457](https://doi.org/10.2519/jospt.2007.2457)
4. Maclachlan L, White SG, Reid D. Observer rating versus three-dimensional motion analysis of lower extremity kinematics during functional screening tests: a systematic review. *Int J Sports Phys Ther.* 2015;10(4):482-492. <http://www.ncbi.nlm.nih.gov/pubmed/26346642><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4527195>
5. Baida SR, Gore SJ, Franklyn-Miller AD, Moran KA. Does the amount of lower extremity movement variability differ between injured and uninjured populations? A systematic review. *Scand J Med Sci Sports.* 2018;28(4):1320-1338. [doi:10.1111/sms.13036](https://doi.org/10.1111/sms.13036)
6. Bohn MB, Petersen AK, Nielsen DB, Sørensen H, Lind M. Three-dimensional kinematic and kinetic analysis of knee rotational stability in ACL-deficient patients during walking, running and pivoting. *J Exp Orthop.* 2016;3(1):27-35. [doi:10.1186/s40634-016-0062-4](https://doi.org/10.1186/s40634-016-0062-4)
7. Schurr SA, Marshall AN, Resch JE, Saliba SA. Two-dimensional video analysis is comparable to 3D motion capture in lower extremity movement assessment. *Int J Sports Phys Ther.* 2017;12(2):163-172. <http://www.ncbi.nlm.nih.gov/pubmed/28515970><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5380858>
8. Ekegren CL, Miller WC, Celebrini RG, Eng JJ, MacIntyre DL. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. *J Orthop Sports Phys Ther.* 2009;39(9):665-674. [doi:10.2519/jospt.2009.3004](https://doi.org/10.2519/jospt.2009.3004)
9. Padua DA, Marshall SW, Boling MC, Thigpen CA, Garrett WE Jr, Beutler AI. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The jump-ACL Study. *Am J Sports Med.* 2009;37(10):1996-2002. [doi:10.1177/0363546509343200](https://doi.org/10.1177/0363546509343200)
10. Crossley KM, Zhang WJ, Schache AG, Bryant A, Cowan SM. Performance on the Single-Leg Squat Task Indicates Hip Abductor Muscle Function. *Am J Sports Med.* 2011;39(4):866-873. [doi:10.1177/0363546510395456](https://doi.org/10.1177/0363546510395456)
11. Whatman C, Hume P, Hing W. Kinematics during lower extremity functional screening tests in young athletes – Are they reliable and valid? *Phys Ther Sport.* 2013;14(2):87-93. [doi:10.1016/j.ptsp.2012.06.001](https://doi.org/10.1016/j.ptsp.2012.06.001)
12. Onate J, Cortes N, Welch C, Van Lunen B. Expert versus novice interrater reliability and criterion validity of the landing error scoring system. *J Sport Rehabil.* 2010;19(1):41-56. [doi:10.1123/jsr.19.1.41](https://doi.org/10.1123/jsr.19.1.41)
13. Williams SJ, Barron TR, Ciepley AJ, Ebben WP. Kinetic analysis of the role of upper extremity segmental inertia on vertical jump performance. *J Exerc Physiol Online.* 2017;20(2):28-34.
14. De Blaiser C, Roosen P, Willems T, Danneels L, Bossche LV, De Ridder R. Is core stability a risk factor for lower extremity injuries in an athletic population? A systematic review. *Phys Ther Sport.* 2018;30:48-56. [doi:10.1016/j.ptsp.2017.08.076](https://doi.org/10.1016/j.ptsp.2017.08.076)
15. Dingenen B, Blandford L, Comerford M, Staes F, Mottram S. The assessment of movement health in clinical practice: A multidimensional perspective. *Phys Ther Sport.* 2018;32:282-292. [doi:10.1016/j.ptsp.2018.04.008](https://doi.org/10.1016/j.ptsp.2018.04.008)
16. Herrington L, Munro A. A preliminary investigation to establish the criterion validity of a qualitative scoring system of limb alignment during single-leg squat and landing. *J Exerc Sports Orthop.* 2014;1(2):1-6.
17. Almangoush A, Herrington L, Jones R. A preliminary reliability study of a qualitative scoring system of limb alignment during single leg squat. *Phys Ther Rehabil.* 2014;1(1):2. [doi:10.7243/2055-2386-1-2](https://doi.org/10.7243/2055-2386-1-2)

18. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
19. Munro A, Herrington L, Carolan M. Reliability of 2-dimensional video assessment of frontal-plane dynamic knee valgus during common athletic screening tasks. *J Sport Rehabil*. 2012;21(1):7-11. doi:10.1123/jsr.21.1.7
20. Coutinho P, Mesquita I, Fonseca AM. Talent development in sport: A critical review of pathways to expert performance. *Int J Sports Sci Coach*. 2016;11(2):279-293. doi:10.1177/1747954116637499
21. Munro A, Herrington L, Comfort P. The relationship between 2-dimensional knee-valgus angles during single-leg squat, single-leg-land, and drop-jump screening tests. *J Sport Rehabil*. 2017;26(1):72-77. doi:10.1123/jsr.2015-0102
22. Hernaez R. Reliability and agreement studies: A guide for clinical investigators. *Gut*. 2015;64(7):1018-1027. doi:10.1136/gutjnl-2014-308619
23. de Vet HCW, Terwee CB. The minimal detectable change should not replace the minimal important difference. *J Clin Epidemiol*. 2010;63(7):804-805. doi:10.1016/j.jclinepi.2009.12.015
24. McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med*. 2012;22(3):276-282. doi:10.11613/bm.2012.031
25. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability - practical assessment, research and evaluation. *Pract Assessment Res Eval*. 2004;9(4):1-11. http://pareonline.net/getvn.asp?v=9&n=4
26. Starling LT, Lambert MI. Monitoring rugby players for fitness and fatigue: what do coaches want? *Int J Sports Physiol Perform*. 2018;13(6):777-782. doi:10.1123/ijsp.2017-0416
27. Herrington L, Alenezi F, Alzhrani M, Alrayani H, Jones R. The reliability and criterion validity of 2D video assessment of single leg squat and hop landing. *J Electromyogr Kinesiol*. 2017;34:80-85. doi:10.1016/j.jelekin.2017.04.004
28. Kropmans TJB, Dijkstra PU, Stegenga B, Stewart R, De Bont LGM. Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint. *J Dent Res*. 1999;78(3):784-789. doi:10.1177/00220345990780031101
29. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Evidence-Based Practice*. FA Davis; 2015.
30. McCunn R, Aus der Fünten K, Govus A, Julian R, Schimpchen J, Meyer T. The Intra- and inter-rater reliability of the soccer injury movement screen (SIMS). *Int J Sports Phys Ther*. 2017;12(1):53-66. http://www.ncbi.nlm.nih.gov/pubmed/28217416%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5294947
31. Shultz R, Anderson SC, Matheson GO, Marcello B, Besier T. Test-retest and interrater reliability of the functional movement screen. *J Athl Train*. 2013;48(3):331-336. doi:10.4085/1062-6050-48.2.11
32. Morris R, MacNeela P, Scott A, et al. Ambiguities and conflicting results: The limitations of the kappa statistic in establishing the interrater reliability of the Irish nursing minimum data set for mental health: A discussion paper. *Int J Nurs Stud*. 2008;45(4):645-647. doi:10.1016/j.ijnurstu.2007.07.005
33. Zec S, Soriani N, Comoretto R, Baldi I. High agreement and high prevalence: the paradox of Cohen's kappa. *Open Nurs J*. 2017;11(1):211-218. doi:10.2174/1874434601711010211
34. Minick KI, Kiesel KB, Burton L, Taylor A, Plisky P, Butler RJ. Interrater reliability of the functional movement screen. *J Strength Cond Res*. 2010;24(2):479-486. doi:10.1519/jsc.0b013e3181c09c04