

Article

Not peer-reviewed version

Establishing the Reliability of a Functional Performance Test Battery that Incorporates the QASLS Tool in Pre-Elite Female Field Hockey Players

[Rosalyn Cooke](#)*, [Lee Herrington](#), James Martin, [Alison B. Rushton](#), [Nicola R. Heneghan](#), [Andrew Soundy](#)

Posted Date: 10 February 2026

doi: 10.20944/preprints202602.0803.v1

Keywords: reliability; functional performance tests; female athletes



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

Establishing the Reliability of a Functional Performance Test Battery that Incorporates the QASLS Tool in Pre-Elite Female Field Hockey Players

Rosalyn Cooke ^{1,*}, Lee Herrington ², James Martin ³, Alison B Rushton ⁴, Nicola R Heneghan ¹ and Andy Soundy ¹

¹ School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Birmingham

² The School of Health and Society, Mary Seacole Building, Frederick Road Campus, Broad St, Frederick Road Campus, Salford

³ Institute of Applied Health Research, Public Health Building, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

⁴ School of Physical Therapy, Faculty of Health Sciences, Western University, London, Canada

* Correspondence: rcc004@student.bham.ac.uk

Abstract

Pre-elite female field hockey players experience a high incidence of lower-extremity injury, highlighting the need for practical and reliable screening tools. A dual assessment approach that combines functional performance tests (FPT) with observational movement quality scoring, such as the Qualitative Assessment of Single Leg Loading (QASLS), may offer valuable insight; however, its reliability in this population requires further investigation. This study aimed to (1) evaluate the test-retest re-liability of an FPT battery incorporating QASLS in pre-elite female field hockey players, (2) identify the most reliable performance metric for QASLS application, (3) determine intra and inter-rater reliability of QASLS scoring using still images, and (4) assess the practicability of dual assessment in pre-elite team sport settings. Fifteen U18 female field hockey players completed a FPT battery—Anterior Reach, Single Hop for Distance, Side Hop, and Drop Vertical Jump Landing—on two occasions, 28 days apart. Three experienced raters independently scored QASLS for each test. All FPT measures demonstrated good to excellent test-retest reliability ($ICC_{2,1} = 0.63\text{--}0.90$). Intra-rater agreement for composite QASLS scores was good to excellent ($ICC_{2,1} = 0.79\text{--}0.90$), with high agreement across individual components. Inter-rater agreement varied by test, ranging from moderate to near-perfect. These findings indicate that combining FPT with QASLS offers a sufficiently reliable dual assessment method to inform targeted injury prevention strategies in pre-elite female athletes.

Keywords: reliability; functional performance tests; female athletes

1. Introduction

The Field Hockey athletes adopt a sports-specific posture that requires a combination of flexion through the trunk and lower limb with both upper limbs involved in the manipulation of the stick used to handle the ball travelling along the ground [1,2]. This posture is maintained whilst running, cutting and decelerating during play. The impact of this posture on lower limb biomechanics has found that field hockey athletes adopt greater trunk and hip flexion angles with large lateral trunk flexion angles during unanticipated sidestepping [3]. These trunk positions have impact at the knee, potentially leading to injury due to the increase in knee internal rotation and valgus angles and

moments which can be further increased when UL position is constrained [4]. Injury rates within female field hockey have been reported for mixed-age and collegiate cohorts, with the most recent data covering the 2018/2019 NCAA season [2,5]. Consistently across these reports, lower limb injuries account for 51% of injuries, with ankle sprain being the most commonly reported, alongside concussion [5]. O'Neil [6] longitudinal analysis of injury highlighted that collegiate field hockey players have a high risk of recurrent injury within 61 days of their index injury and that this is a window for injury prevention. This is particularly relevant for a pre-elite population, as they represent the players that will feed into elite senior international teams in time.

Prevention of an index injury is a priority at this stage and will contribute to availability for training and competition as well as improved musculoskeletal health throughout their athletic career and beyond [7,8]. There is a clear need to optimise injury prevention strategies for this population and undertaking athlete profiling to assess physical and movement qualities with Functional Performance tests (FPT) is a component of this [9].

Functional Performance Tests (FPT) is a term used to describe a variety of different movement tasks which seek to replicate the demands of sport-specific movements [10]. These are often closed skill motor tasks which have a defined measured outcome such as distance hopped, height jumped or time to complete.

Poor movement quality or asymmetry between lower limbs can predispose an athlete to risk of lower extremity injury (LEI). Within field hockey there is limited reports of FPT testing despite the multiplanar demands of the sport and high burden of LEI.(REF) The Y-balance test has been used with community and collegiate field hockey athletes previously and shown to indicate risk of LEI [10–12].

For a comprehensive evaluation of an individual's movement ability, there is a need to select tasks which use different planes of movement, and there is an assessment of the movement quality [13]. Emerging evidence from within anterior cruciate ligament (ACL) rehabilitation populations has highlighted that FPT can under or overestimate an individual's ability to return to sport when we only consider the performance-based metrics (distance, height) in single planes of motion and when using the uninjured limb as a comparator (Limb Symmetry Index LSI) [13–15]. Traditional hop test batteries evaluate multiple horizontal hop tasks, [16] which could be considered redundant due to the strong correlations between different forward hop variations (Single hop for distance and triple hop $\rho=0.96$) [13]. Movement quality can be assessed through a variety of kinematic methods, such as 3-Dimensional (3D) motion capture. Although this is considered the gold standard it is often not practicable within pre-elite training environments due to the financial and time costs associated with it. Observational movement quality tools such as the Qualitative Assessment of Single Leg Loading (QASLS) have shown sufficient criterion validity against 3D motion capture (QASLS $k=0.9$ for hop landing, $k=0.97$ single leg squat) [17–19]. The QASLS tool has previously been applied to unilateral tasks such as single-leg squat and single-leg land [17,20,21]. Adapted QASLS criteria have also been applied to hop tests within a rehabilitation population and found to identify those at risk of secondary ACL injury [19]. We define the application of observational movement quality tools such as QASLS to FPT as 'Dual Assessment'. The use of this term avoids confusion between descriptors such as quantitative and qualitative or quantity and quality when describing the variables being assessed within a single FPT [15,19,22–24].

Dual assessment has been utilised and explored within rehabilitation populations with a focus on prediction of secondary ACL injury [15,19,22–24]. There is currently a lack of studies evaluating dual assessment in uninjured populations or within specific sports [13].

While some studies have considered the biomechanical demands, injury risk and rates in field hockey this has primarily been evaluated in elite [2,11,25] or community populations [26], leaving pre-elite female players an under-investigated group despite their high risk of LEI [7]. The Dual Assessment approach may identify specific performance metrics or movement quality issues that need to be addressed. In order to achieve this, there is a need to establish the reliability of a dual assessment approach when applied to an FPT battery in this population.

1.1. Aims

- To report the test-retest reliability of a FPT battery that incorporates QASLS within a female pre-elite field hockey cohort
- To identify which performance metric of the included FPT is the most reliable for application of the QASLS tool
- To report the intra and inter-rater reliability of the QASLS tool when applied to a still image
- To report the practicability of Dual Assessment when completing FPT with pre-elite female team sport athletes.

2. Materials and Methods

This test retest, intra and inter-rater reliability study was designed and is reported in line with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [27] to improve the quality of evidence available in relation to measurement properties of this novel application of QASLS with a FPT battery.

2.1. Participants

All participants were over sixteen years of age and part of the under eighteen training squad for a single national governing body (NGB). Participants were not currently injured or completing rehabilitation and able to complete single leg hopping and landing activities in multiple planes of motion. Participants were excluded if they had competed at senior international competition level or regularly trained (>1 training session per month) in the senior training group. Fifteen pre-elite female field hockey players competing at junior international level were invited and consented to participate in the study.

2.2. Sampling and Sample Size

A convenience sampling technique was adopted. A priori power analysis was conducted to determine reliability of FPT battery when using QASLS [28]. The expected intraclass correlation coefficient (ICC) was set at 0.70, with an alpha level of 0.05, with three raters and a statistical power of 0.80. This indicated that a minimum of 169 observations would be required. Fifteen participants completed both profiling sessions and were included in the analysis. The test battery included 4 FPT with 20 expected observations recorded per subject. This generated 300 observations per session exceeding the minimum requirement.

2.3. Ethical Approval

Participant information and consent were provided and collected utilising the REDCAP data collection system hosted at the University of Birmingham [29]. This study was approved by the University of Birmingham Ethics Committee (ERN_21-1414).

2.4. Study Setting

Participants completed the FPT battery on two separate occasions in an indoor sports hall. The two profiling sessions were completed 28 days apart and took place prior to any other hockey training or physical activity. Participants were allocated an identification number and randomly allocated to a test station for one of the FPT and then proceeded to complete all FPT in a block order. All testing was completed in training footwear and normal hockey training apparel, including shorts/skort or leggings and t-shirt. Subjects' performance in all FPT was filmed using a GoPro Hero 4 camera at 60fps to enable analysis of movement performance using QASLS.

2.5. Data Collection

Functional Performance Tests

Four FPT were selected to be included in the test battery. Selection of the FPT was based on certainty of evidence for measurement properties, practicability and the requirement to assess movement quantity and quality in frontal and sagittal planes.

Standardised test protocols were developed based on previously reported methods and are outlined in Table 1. All subjects were allowed to perform a self-directed warm-up that they would normally complete before gym-based training. As this was not a variable of interest, it was not standardised or recorded. All subjects completed three practice trials on each leg before AR, DVJ and SHFD and a maximum of 10 seconds of SH on each leg before measured trials. Subjects self-selected which limb they tested first in each of the FPT.

Table 1. Functional Performance Test Protocols.

Test	Equipment	Testing Process and Scoring	Indication for repeat test
Anterior Reach (AR) [11,30]	Tape Measure	3 practice trials on each leg	Kicking push box Not returning to starting position under control Touching down during reach Foot on top of stance plate
	Tape to mark floor	Subject selects leg to test first	
	1x camera	3 measured trials on one leg and then repeated on other leg	
	1x tripod	Distance pushed measured to nearest 0.5cm	
	Y Balance Testing Kit	Scoring completed at time of testing	
Single Leg Drop Vertical Jump (DVJL) [14,31,32]	Tape Measure	3 practice trials each side with minimum of 1-minute rest	Loss of balance – steps out of landing Extra hop on landing Touching down with either contralateral leg or with hand
	Tape to mark floor	Subject selects leg to test first	
	1 x camera	Completes 3 measured trials on first leg and then repeats on other side	
	1 x tripod	Landing to be held for at least 2 seconds on completion of DVJ	
	30cm box	Scoring completed post testing using video footage and My Jump Lab App to calculate:	
Single Hop for Distance (SHFD) [16]	Landing area marked 30cm in front of box with a strip of tape	<ul style="list-style-type: none"> • Jump height • Contact Time • Flight Time • Reactive strength index (RSI) 	Loss of balance – steps out of landing Extra hop on landing Touching down with either contralateral leg or with hand
	Tape Measure	3 practice trials each side with minimum of 1-minute rest	
	Tape to mark floor	Subject selects leg to test first	
	1 x camera	Completes 3 measured trials on first leg and then repeats on other side	
	1 x tripod	Landing to be held for at least 2 seconds	
Side Hop (SH) [33,34]	Tape measure set to 250cm and secured to the floor with tape	Measurement taken from heel of landing foot	Loss of balance during the test Forgets to keep hands on hips
	Tape Measure	Distance measured to the nearest cm	
	Tape to mark floor	Scoring completed at time of testing	
	2 lines of 1metre 40cm apart	Up to 10secs practice each side with minimum of 1 minute rest before completing testing	
	1 x camera	Completes 1 set on each leg	
1 x tripod	Self-selected rest between each side minimum of 1min	Loss of balance during the test Forgets to keep hands on hips	
Timer	Subject selects leg to test first		
	Scoring completed post testing using video footage to calculate:		
	Total Hops,	Adjusted Score (total hops – total errors) % Error (total errors/total hops x 100)	
	Total Errors,		

2.7. Movement Quality Assessment: Qualitative Assessment of Single Leg Loading (QASLS)

Visual assessment of functional movement performance utilising video has been established as a reliable and practicable alternative to 3D motion capture [17,20,21]. The QASLS tool has been developed and validated for assessing the performance of single-leg movement tasks by evaluating the body segment movement strategies. This was initially developed for a single-leg squat [17] but has been utilised in other single-leg activities [20].

The QASLS tool score is calculated by the assignment of a score of 1 for the presence of suboptimal movement strategies at each segment as described on the score sheet (Figure 1). In the absence of these suboptimal movement patterns a score of 0 is given with a potential range of scores from 0-10 with a higher score indicating greater suboptimal movement in the task.

Within this study, the QASLS tool was applied to a still image captured in the frontal plane of the terminal landing or reach at the point of deepest knee flexion for the AR, DVJ and SHFD in line with previously reported methods [23,24,35,36]. For the SH due to it being a repeated movement task the QASLS was assessed at 5-second intervals throughout the duration of the test, which lasted 30 seconds. Additional descriptors were added to the QASLS tool when evaluating the performance of the FPT where the arms are fixed (AR, LH) and in the SH, where the movement demands of the task produce a trunk lean (Figure 1). All still images scored by the raters using QASLS were predetermined by the lead researcher (RC).

Intra and Inter-rater reliability of the QASLS tool was calculated based on the performance of 5 randomly selected participants from within the sample by 3 raters. This provided 150 observations per rater, 450 total observations for inter-rater evaluation and 300 observations for intra-rater evaluation [3]. As analysis and scoring takes approximately 30 seconds per observation this equates to a minimum of 75 minutes of data processing.

DUAL ASSESSMENT WITH THE QUALITATIVE ANALYSIS OF SINGLE LEG LOADING TOOL (QASLS)

Ros Cooke - Doctoral Researcher University of Birmingham

Supervisors: Dr Andy Soundy, Dr Nicola Heneghan, Dr James Martin, Dr Alison Rushton, Dr Lee Herrington

How do we score QASLS?

- A score of 1 is given if any of the movement faults in the description are seen.
- A score of 0 is given if no movement faults are observed
- Each limb is scored separately and given a total score out of 10
- The higher the score the greater the number of sub-optimal movement strategies

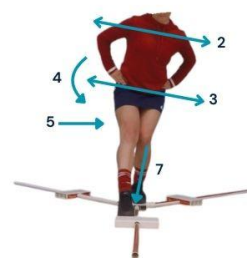
The QASLS Tool – extended criteria for Dual Assessment of FPT

QASLS Component	Description	Score Left	Score Right
ARM STRATEGY	1) Excessive arm movement to balance or unable to maintain hands on hips		
TRUNK ALIGNMENT	2) Leaning in any direction so that the shoulder girdle and pelvic girdle are not aligned		
PELVIC PLANE	3) Loss of horizontal plane		
	4) Excessive tilt or rotation		
THIGH MOTION	5) Weight bearing thigh moves into hip adduction		
	6) Non-weight bearing thigh not held in neutral		
KNEE POSITION	7) Patella pointing towards 2nd toes (noticeable valgus)		
	8) Patella pointing past inside of foot (significant valgus)		
STEADY STANCE	9) Touches down with NWB foot		
	10) Stance leg wobbles noticeably		
	Total		

Example: QASLS & Anterior Reach

QASLS Components Labelled

Left Leg Score: 5/10



Example: QASLS & Single Leg Drop Vertical Jump

QASLS Components Labelled

Right Leg Score: 0/10



Example: QASLS & Side Hop

QASLS Components Labelled

Right Leg Score: 4/10



Example: QASLS & Single Hop for Distance

QASLS Components Labelled

Right Leg Score: 2/10



Figure 1. QASLS Tool Scoring Sheet with examples of how to score each FPT.2.8. Rater Characteristics.

Each testing station was allocated a scorer to oversee and record the participant's performance. All scorers were provided with written instructions and verbal guidance on how the FPT was to be conducted and scored. This was standardised due to the variation in experience and consistency of scorers (4 physiotherapists and 2 field hockey technical coaches). All scorers were blinded to the

performance of participants from each of the profiling sessions. The lead researcher (RC) was responsible for the scoring of the SH and DVJ using the methods outlined in Table 1. This was conducted after the completion of both testing sessions and QASLS assessments to maintain blinding to performance in the FPT.

The QASLS assessment evaluated performances across all trials of the FPT by the lead researcher (RC, Rater 1) and two additional highly experienced physiotherapists with over 15 years of clinical practice and research experience, including the use of QASLS in athletic populations (Rater 2,3). Raters were provided with instructions for QASLS scoring, images to demonstrate criteria (Figure 1) and QASLS scoring sheet within an Excel database. The order of participants was randomised for each FPT to minimise familiarisation. All raters were blinded to the performance of the participants and completed the scoring independently in a single session.

2.9. Data Processing

For the FPT test battery (AR, SHFD, SH, SLDVJ) all measured trials were included for analysis with the fifteen participants who completed both data collection sessions (Supplementary materials 2). QASLS assessments were completed for five athletes selected at random from the fifteen included. Screenshots were captured at point of deepest knee flexion using frame by frame review of video footage for each measured trial of the AR, SHFD and DVJ. For the SH screenshots were captured every 5 seconds for the 30-second testing period for QASLS assessment.

2.10. Statistical Analysis

In this study, statistical analyses were performed using STATA 18 (StataCorp 2023). The code used for these analyses is provided in the supplementary materials 3 to ensure reproducibility and transparency. To assess test-retest reliability, Intraclass correlation coefficients (ICC^{2,1}) were calculated for the mean (mean of three measured trials), best (highest scoring trial), and worst (lowest scoring trial) scores for each limb in sessions one and two for AR, SLDVJ and SHFD. Evaluation of the SH required the total number of hops (total) the number of errors (error) and score (total number of hops – errors) to be calculated for each limb in sessions one and two to calculate ICC^{2,1}. Interpretation of these measures was based on previous methods, [20,37] where Excellent ≥ 0.90 , Good 0.75-0.90, Moderate 0.50-0.75, Poor < 0.50 . Measurement error was also calculated by reporting the standard error of measurement ($SEM = s\sqrt{1-R}$) and smallest detectable change ($SDC = 1.96 \times \sqrt{2 \times SEM}$).

Inter-rater reliability of the QASLS composite score was assessed using intra-class coefficients (2,k) [37]. ICC were used as the data was considered to be continuous. To investigate agreement between raters across each component of the QASLS, percentage exact agreement (PEA) was calculated using the formula $PEA = \left(\frac{A}{N}\right) \times 100$ where A represents the number of agreements between the raters and N is the total number of ratings. Interpretation of what is acceptable PEA% is difficult due to a lack of interpretation guidance but has been previously used when evaluating rater agreement for components of QASLS [20].

In line with previous analysis of intra and inter-rater reliability of the QASLS tool the following interpretations of PEA were applied; 50-60% moderate, 61-75% substantial, 76-100% near perfect [20,21]. Intra-rater reliability of the QASLS composite score was analysed using intra-class coefficients (ICC^{2,k}) and PEA were calculated for the individual components of the QASLS score and interpreted as previously described [20].

3. Results

Test-retest reliability included fifteen subjects (age 16.7 ± 0.73 , height $167.9\text{cm} \pm 8.08$, weight $61.6\text{kg} \pm 7.26$), across two sessions, twenty-eight days apart. 300 observations were completed in each testing session, and all observations were included within the analysis. 450 observations were used

to assess inter-rater reliability, and 300 for intra-reliability analysis of the QASLS tool. This exceeded the number of required observations agreed a priori (n=169).

3.1. Test-Retest Reliability of FPT Battery

ICC^{2,1} and 95% confidence intervals were calculated for the mean, best and worst trials for each participant for each limb for the AR, DVJL and SHFD. The results are reported in Table 2. SEM and SDD were calculated for each of the FPT within and between session and between limbs. and reported in Table 3.

AR mean distance reach for both sides achieved moderate to excellent reliability (Left ICC^{2,1} 0.85 95%CI,0.54-0.95, Right ICC^{2,1} 0.88 95%CI, 0.64-0.96). The best and worst scores achieved poor to good reliability ratings although the best reach score ICC^{2,1} for both limbs achieved ICC over 0.70 with a smaller 95% CI compared to the worst reach score. SEM between sessions was found to be 2.15-2.36cm left (SDD 8.43-9.24cm) and 2.12-3.66cm for the right (SDD 8.31-14.34cm).

The DVJ jump height reliability showed differences between sides with left mean, best and worst scores demonstrating good to excellent reliability (Mean 0.90, Best 0.79, Worst 0.80) whilst the right showed poor to moderate (Mean 0.65, Best 0.39, Worst 0.48). These differences were less apparent in the SEM with the left having a greater range from 1.43-1.76cm between sessions (SDD = 5.61-6.90cm) than right 1.51-1.54cm (SDD 5.90-6.05). SHFD demonstrated moderate to good reliability across all measures on both sides (Mean 0.73-0.81, Best 0.65-0.72, Worst 0.52-0.67).

SEM for both sides was similar (Left 5.64-8.93cm, right 5.06-9.23cm), although SDD was 22.11-35.0cm and 19.85-36.17cm for left and right limbs respectively.

For SH test retest reliability was assessed for each of the score components as originally described by Gustavasson [33]. The total number of hops showed mean moderate reliability on both sides (ICC^{2,1} 0.63-0.64) whilst the number of errors showed a poor to moderate rating and score poor to good. As score is calculated by subtracting the number of errors from the total, this is not a surprising finding. SEM between sessions was found to be 3.72 and 4.16 for left and right sides, respectively, with a SDD of 10.32 for left and 11.52 for right. This suggests that a change of 10-12 hops would be necessary to indicate a change in performance.

Table 2. Test-Retest reliability of the FPT Battery.

FPT	Left								
	Mean ICC	Mean CI 95%	Class	Best ICC	Best CI 95%	Class	Worst ICC	Worst CI 95%	Class
AR (distance cm)	0.85	(0.54-0.95)	Mod-Exc	0.73	(0.35-0.90)	Poor-Exc	0.67	(0.26-0.88)	Poor-Good
DVJL (height cm)	0.9	(0.71-0.97)	Good-Exc	0.79	(0.48-0.92)	Poor-Exc	0.8	(0.50-0.93)	Mod-Exc
SHFD (distance cm)	0.73	(0.16-0.91)	Poor-Exc	0.65	(0.21-0.87)	Poor-Good	0.52	(0.01-0.81)	Poor-Good
	Total ICC	Total CI 95%	Class	Errors ICC	Errors CI 95%	Class	Score ICC	Score CI 95%	Class
SH (count)	0.64	(0.19-0.86)	Poor-Good	0.4	(-0.06-0.74)	Poor-Mod	0.53	(0.07-0.81)	Poor-Good
	Right								
	Mean ICC	Mean CI 95%	Class	Best ICC	Best CI 95%	Class	Worst ICC	Worst CI 95%	Class
AR (distance cm)	0.88	(0.64-0.96)	Mod-Exc	0.71	(0.35-0.89)	Poor-Good	0.7	(0.29-0.89)	Poor-Good
DVJL (height cm)	0.65	(-0.10-0.88)	Poor-Good	0.39	(-0.15-0.74)	Poor-Mod	0.48	(-0.05-0.79)	Poor-Good

SHFD (distance cm)	0.85	(0.54-0.95)	Mod-Exc	0.72	(0.36-0.90)	Poor-Exc	0.67	(0.26-0.88)	Poor-Good
	Total ICC	Total CI 95%	Class	Errors ICC	Errors CI 95%	Class	Score ICC	Score CI 95%	Class
SH (count)	0.63	(0.21-0.86)	Poor-Good	0.33	(-0.12- 0.71)	Poor-Mod	0.43	(-0.04-0.75)	Poor-Good

FPT: Functional Performance Test. AR: Anterior Reach. DVJL: Single Leg Drop Vertical Jump Land. SHFD: Single Hop for Distance. SH: Side Hop Mean: Mean of 3 trials, Best: Best scoring trial, Worst: Worst scoring trial, Total: total number of hops, Errors: number of errors, Score: Score = total number of hops – errors. ICC: Intraclass correlation coefficients (ICC^{2,1}), CI 95%: 95% Confidence Interval for ICC. Class: Classification of ICC. Exc: Excellent ≥ 0.90 , Good: Good 0.75-0.90, Mod: Moderate 0.50-0.75, Poor: poor < 0.50 .

Table 3. Summary of Standard Error of Measurement (SEM) and Smallest Detectable Difference (SDD) for individual FPT over each measured trial and between sessions.

	Trial	Testing Session 1				Testing Session 2			
		Left		Right		Left		Right	
		SEM	SDD	SEM	SDD	SEM	SDD	SEM	SDD
AR	1v2	2.75	10.77	1.87	7.32	2.13	8.34	3.46	13.57
	1v3	1.89	7.41	2.15	8.41	2.07	8.11	4.28	16.78
	2v3	2.44	9.55	2.35	9.22	2.26	8.85	3.23	12.66
	Mean	2.36	9.24	2.12	8.31	2.15	8.43	3.66	14.34
DVJ	1v2	2.16	8.47	1.64	6.43	1.43	5.62	1.28	5.01
	1v3	1.38	5.42	1.61	6.29	1.38	5.40	1.43	5.59
	2v3	1.74	6.81	1.27	4.99	1.48	5.79	1.92	7.53
	Mean	1.76	6.90	1.51	5.90	1.43	5.61	1.54	6.05
SHFD	1v2	13.49	52.87	11.19	43.87	4.82	18.88	4.18	16.39
	1v3	6.76	26.50	9.36	36.70	7.11	27.85	5.13	20.10
	2v3	6.54	25.63	7.13	27.94	5.00	19.59	5.89	23.07
	Mean	8.93	35.00	9.23	36.17	5.64	22.11	5.06	19.85
SH		3.72	10.32	4.16	11.52				

3.2. Intra-Rater Reliability of QASLS

Intra-rater reliability of the composite score of the QASLS tool was calculated for each of the FPT in the test battery and found to be good to excellent for AR (ICC^{2,1} 0.90) and DVJL (ICC^{2,1} 0.90) and moderate to good for SH (ICC^{2,1} 0.79) and moderate to excellent for SHFD (ICC^{2,1} 0.83). Table 4 reports these results.

Table 4. Intra-Rater reliability of Composite Score of QASLS by FPT.

Functional Performance Test	ICC ^{2,1} (95% CI)	ICC Classification
Anterior Reach	0.90 (0.81-0.95)	Good-Excellent
Drop Vertical Jump Land	0.90 (0.79 -0.95)	Good-Excellent
Side Hop	0.79 (0.66 - 0.87)	Moderate -Good
Single Hop for Distance	0.83 (0.62-0.92)	Moderate -Excellent

PEA for the individual components of the QASLS achieved agreement of excellent to perfect (86.6-100%) for the AR, DVJL and SHFD (PEA: 83.3-100%). SH PEA was substantial to perfect (66.7-100%). This is reported in supplementary materials 4.

3.3. Inter-Rater Reliability of QASLS

Table 5 presents the inter-rater reliability of the QASLS composite scores for each of the FPT included within the test battery. Inter rater reliability for AR ranged from poor ($ICC^{2,k} 0.38$) to good ($ICC^{2,k} 0.82$) with Rater 2 demonstrating the greatest difference to Raters 1 and 3. This was also observed in the SHFD where reliability was poor between Rater 1 & 2 ($ICC^{2,k} 0.48$ 95%CI -0.23-0.80) and Rater 2 & 3 ($ICC^{2,k} 0.46$ 95%CI -0.14-0.81) but good between Rater 1 & 3 ($ICC^{2,k} 0.77$ 95%CI 0.52-0.89). Intra-rater reliability for both the DVJL and SH was good between Raters 1&2 and 1&3. Raters 2&3 achieved moderate reliability (DVJL $ICC^{2,k} 0.70$ 95%CI-0.16-0.90, SH $ICC^{2,k} 0.54$ 95%CI -0.14-0.81)

Table 5. Inter-Rater Reliability of QASLS Composite Score by FPT.

QASLS Composite Score	Rater 1v2	Classification	Rater 1v3	Classification	Rater 2v3	Classification
Anterior Reach	0.38 (-0.23 -0.71)	Poor-Moderate	0.82 (0.61-0.91)	Moderate-Excellent	0.38 (-0.23- 0.71)	Poor-Moderate
Drop Vertical Jump Land	0.85 (0.53-0.93)	Moderate-Excellent	0.76 (0.48-0.88)	Poor-Good	0.70 (-0.16-0.90)	Poor-Good
Side Hop	0.75 (0.31- 0.88)	Poor-Good	0.78 (0.54-0.89)	Moderate-Good	0.54 (-0.15 -0.79)	Poor-Good
Single Hop for Distance	0.48 (-0.23 -0.80)	Poor-Good	0.77 (0.52-0.89)	Moderate-Good	0.46 (-0.14 -0.81)	Poor-Good

Inter-rater agreement of the individual components of the QASLS presented in supplementary materials 4 demonstrated moderate to perfect agreement across all the FPT (PEA% 60.0-100). QASLS component 4; presence of tilt or rotation in the pelvic plane showed the lowest levels of agreement (PEA% 60.0-78.3) within the AR, SH and SHFD tests. Arm strategy (QASLS Component 1) in SHFD only showed substantial agreement (PEA% 70.0-73.3) between all raters. SH had two other components, which only achieved substantial agreement amongst the three raters. These were loss of horizontal pelvic plane (QASLS Component 3 PEA% 66.7) and position of NWB thigh (QASLS Component 6 PEA% 68.3).

4. Discussion

The purpose of this study was to report the test-retest, intra and inter-rater reliability of an FPT battery which incorporates the QASLS tool in a dual assessment approach within a cohort of female pre-elite field hockey players. This is believed to be the first study reporting a dual assessment approach within an uninjured athlete population who are at high risk of LEI. The methods within our study are novel and have applied expanded QASLS criteria to a still frontal plane image across four movement tasks performed in sagittal (AR, SHFD) and frontal planes (DVJ, SH). As this is a new methodology and application of the QASLS tool it is important to establish knowledge of the measurement properties of this approach [19,20,38].

4.1. Which Score to Use in FPT?

Mean scores of performance metrics within FPT are commonly used to assess performance and calculate limb symmetry index (LSI) [39]. These methods improve reliability as the measurement error is reduced, but may lead to an overestimation of an individual's ability and produce misleading results [23,40]. Figure 2 demonstrates hypothetical data where an athlete's performance in the SHFD, although different between limbs or between sessions, achieved the same mean score and falsely

indicates symmetry when applying an LSI calculation.

THE CHALLENGE OF USING MEAN SCORE WITH FPT

Between Limbs



	Trial 1	Trial 2	Trial 3	Mean	Limb Symmetry Index (Mean Left/Mean Right x100)
Left	150	150	150	150	100%
Right	122	158	170	150	100%

In both these examples the performance metrics suggest symmetry in SHFD performance between limbs and between sessions.

The variability in performance exceeds the standard error of measurement for this FPT

Between Sessions



	Trial 1	Trial 2	Trial 3	Mean	Limb Symmetry Index
Session 1	150	150	150	150	100%
Session 2	160	180	110	150	100%

Figure 2. The challenge of using mean score with FPT.

The interpretation of these hypothetical hop tests would suggest that by solely considering the mean score, there is symmetry between limbs or between sessions. A critique of the scores would indicate a lack of consistency in performance between limbs and between sessions. This also does not consider the movement quality associated with each of these testing sessions, which may also provide context to the scores achieved [13,19]. This temporality of performance in FPT for uninjured populations is not well reported within the literature, with a need for longitudinal studies with repeated measures using FPT with known measurement properties [9].

The use of mean score does not facilitate a dual assessment approach, and therefore, as part of this study, there was a need to understand the test-retest reliability of a participant's best and worst performance within the individual tests of the FPT battery. Within the FPT literature, there is great variability in the reporting of metrics, although best and worst scores are not reported [41,42].

Best and worst scores for both AR and SHFD demonstrated moderate test-retest reliability across both limbs, whereas both these tests have previously reported excellent test-retest reliability (AR $ICC^{2,1}$ 0.84-0.93, SHFD $ICC^{2,1}$ 0.92-0.97) [30,43]. These differences may arise due to previous studies using mean scores and the cohorts being mixed-sex, elite and collegiate-level athletes. DVJL showed differences between sides and measures with the best score having poor to moderate reliability ($ICC^{2,1}$ 0.39-0.79) and worst score having moderate to good reliability ($ICC^{2,1}$ 0.48-0.80). Jump height was used as the measure in line with previous literature as this has shown to be valid against traditional force platform assessment and other measures such as restive strength index (RSI) are influenced by age group [44].

For SH best and worst performances are not recorded as metrics within the test as it is a single effort. This study found total score demonstrated moderate reliability ($ICC^{2,1}$ Left 0.64 95%CI 0.19-0.86, Right 0.63 95%CI 0.21-0.86). These results report lower $ICC^{2,1}$ values than previously found with a collegiate cohort (n=14) where test-retest reliability was good to excellent ($ICC^{2,1}$ Left =0.96 Right =0.84) [43]. The differences in our results could be in part due to the age and sex difference in cohorts

and the time interval between testing sessions (1 week versus 1 month). As athlete profiling is not routinely completed on a weekly basis it is important to indicate the stability of measurement over longer time periods which are representative of the sports environment. Error count and score showed poor reliability with wide confidence intervals. This may reflect the high variability in the repeated performance of this task over the 30-second testing period and the variability within human movement particularly at the pre-elite stage of athletic development [20]. These findings supported the approach to perform the QASLS assessment at 5-second intervals across the testing period. This is a prudent approach and allows for variation in the number of hops performed by participants which is not consistent within and between individuals [13,19]. SEM and SDD are important considerations when utilising a FPT and inform the practitioner if they have observed a true change in performance.

These measurement properties are commonly under or poorly reported [38,42]. The findings of our study report SEM values for AR and SHFD that are higher than previously reported [43] although the populations are not comparable as these studies have utilised male collegiate-level athletes.

For DVJL, SEM is not readily reported in the literature, despite multiple validity and reliability reports for the My Jump app [44,45]. SEM for SH (SEM 3.72-4.16 hops SDD 10.32-11.52 hops) is in line with previous reported research within collegiate and pre-elite Scandinavian populations (SEM 0.5-5.4 hops) [34,43].

4.2. Is QASLS a Reliable Tool?

Intra-rater reliability was assessed for the composite score of the QASLS tool applied to each of the FPT and found to be excellent for AR and DVJL and good for SH and SHFD. The requirements of the SH test for repeated efforts in a dual task over 30 seconds could give rise to greater variability in performance and account for the difference in reliability. This is supported when considering the PEA of the individual components of the QASLS score, which was 66.7% for the pelvic plane component.

Inter-rater reliability of the composite QASLS score varied by FPT and by rater combination. DVJL and SH demonstrated moderate to good reliability (DVJ ICC^{2,k} 0.70-0.85, SH ICC^{2,k} 0.54-0.75) across all rater combinations. AR demonstrated poor to good inter-rater reliability (ICC^{2,k} 0.38-0.82).

The AR is comparable to a single-leg squat task and previously, studies have found poor inter-rater reliability when applying QASLS to this task [20]. As good reliability was seen between Raters 1 & 3 (ICC^{2,k} 0.82 95%CI 0.61-0.91) it may suggest that rater training and examples may be beneficial to improve the consistency of application of QASLS in different FPT.

This is also seen in the inter-rater reliability scores for the SHFD test, which varied from poor to good (ICC^{2,k} 0.46-0.77) across raters, with Rater 2 again showing greatest difference in scoring. Considering the components of the QASLS score through PEA enables us to understand the level of agreement between raters and provide context to ICC measurements. An example of this is AR, where PEA for QASLS component 4 (pelvic plane excessive tilt or rotation) where all rater combinations only achieved substantial agreement (%PEA 63.3-76.7) and this lack of agreement would support the observed relatively poorer ICC scores.

PEA is an indication of measurement error when the data is considered to be ordinal or nominal and is therefore the most appropriate for use with the QASLS score where the components are considered to be ordinal [20].

As identified within AR test, QASLS component 4 accounted for the greatest difference between raters, and this was also seen within SH and SHFD. Previous studies have not found this component to be a source of rater disagreement [20,21]. Analysis of the pelvis using gold standard 3D motion capture methods has also found greatest variation when modelling pelvic motion within transverse plane [17] it is therefore unsurprising that this is a source of variability in observational methods. Arm positioning (QASLS component 1) during the SHFD lacked agreement amongst the raters (%PEA 70-73.3) and may result from the vagueness of the descriptor 'excessive' for this component. It is the only FPT in the battery where arm position is not fixed, and a variety of positions have been observed. In line with the findings from Parry, [20] agreement between raters about NWB thigh

position during the SH was lower (%PEA 68.3). Due to trunk position and frontal plane demands of this task, the position of the NWB thigh varies as subjects try and stabilise through adducting the NWB limb, which is sub-optimal [17,20,21]. Examples and expanding the detail within the QASLS criteria may help to address these issues with both components. QASLS component 10 (steady stance) achieved 100% agreement by all raters across all FPT. This level of agreement arises from the lack of ability to assess this component when a still image is used and is therefore redundant with the current methods. The use of a still image may also account for the difference in findings to previous work, where QASLS components relating to knee valgus were the source of greatest disagreement between raters. The use of a still image may be more appropriate if phases of landing or movement tasks are the focus of the movement quality assessment.

4.3. Strengths and Limitations of Study

Our study aimed to establish reliability metrics for the new concept of Dual Assessment and to enhance the robustness of the methods and analysis this study was developed and reported using GRRAS guidance. The study also acknowledges the findings from previous published work reporting the practicability and measurement properties of FPT and has attempted to address the limitations in the level of evidence [34–36]. The QASLS tool, although frequently used within clinical settings has not been subject of many studies in comparison to other observational tools such as the LESS, FMS or TJ [46]. Comparison to other studies is limited as the QASLS tool has not been applied to the included FPT, with the adapted criteria and to a still frontal plane image. The reliability results presented here are sufficient according to CONsensus-based Standards for the selection of health Measurement Instruments (COSMIN) criteria for good measurement properties to support use of these methods to enable dual assessment of FPT [47,48].

A common limitation from this study and within the available published literature relates to the detail within the descriptors of each of the QASLS components. This is relevant to this study as some changes had been made due to a change in the method of application of the QASLS tool (still image scoring versus video footage).

Although these were indicated to the raters involved, their previous experience with the tool may have led to them not considering the additional descriptions. Rater training does not follow a formalised process or calibration, and consideration of this may improve rater agreement and reliability. Rater populations also need to be considered, and within this study, only involved physiotherapists.

Although physiotherapists are trained to evaluate, assess and quantify movement through their undergraduate training, this is often concerning normal movement tasks and not sporting actions. It may therefore be wrong to assume that physiotherapists are the only profession within the sports science and medicine team that can deliver analysis in this area. Pelvic motion assessed with the QASLS tool showed the greatest disagreement between raters during AR, SH and SHFD. Addressing this using a marker on the anterior superior iliac spine and pelvic crest may improve visualisation of the pelvis and agreement between raters. Disagreement between raters may also reflect that only a frontal plane still image was used for analysis.

This study did not standardise the warm-up completed by participants due to their status as trained athletes who regularly complete self-led warm-up activities. While this reflects real-world practice, it may be considered a limitation. Participant completion of familiarisation trials was standardised to ensure all subjects had had sufficient exposure to the movement task. The necessity and impact of a standardised warm-up on FPT performance should be considered and justified based on the characteristics of the target population (training age, injury status), the FPT selected (single versus multiple tests) and familiarisation of participants with the movement task [49]. Learning effects have been reported for some FPT, with score stabilisation occurring after one to four trials, depending on the test [50]. This study included three familiarisation trials for AR, SLDVJ, and SHFD, consistent with previous methods [11,14,16,30–32,50]. Given the complexity of the SH, a longer

familiarisation period may be beneficial, though this may be constrained by the test's physical demands [50].

The sample size included within this study is a limitation, although it is in line with previous research [20] and reflects the nature of purposeful sampling in a pre-elite athletic population, of which there are limited numbers. The use of 5 subjects provided a sufficient number of observations for inter- and intra-rater reliability analysis and did not place an excessive time demand on raters to complete the QASLS tool. Using a small sample may not truly reflect the variation of subjects' performance and is a limitation of this study.

5. Conclusions

The test-retest reliability of the FPT battery varied by the individual FPT and by the performance metric included for analysis (mean of three trials, best trial, worst trial). The best effort from measured trials for the AR, DVJL and SHFD demonstrated sufficient reliability as assessed by COSMIN (ICC<0.70) and can be used for dual assessment in pre-elite female athletes. The SH demonstrated moderate reliability for total hops (ICC² 0.63-0.64) across both sessions, and therefore, assessing QASLS at 5-second intervals will provide a comprehensive assessment independent of the number of hops performed. The QASLS tool demonstrated good to excellent intra-rater reliability across the FPT battery, and this is supported by PEA%, which was substantial to near perfect (66.7-100%). Inter-rater reliability of the composite QASLS score varied by FPT and rater combination in line with previous research. PEA% was found to be moderate to perfect (60-100%) across all tests, with assessment of the pelvic plane being the component with the least agreement. The findings of our study support the methods of utilising the QASLS tool with a still image at a defined point within a movement task (point of deepest knee flexion on landing or 5-second time intervals for duration of the test).

This supports a consistent approach to dual assessment regardless of the FPT involved and may be beneficial in the profiling of athletes at risk of LEI.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, <https://doi.org/10.25500/edata.bham.00001255>. 1.: Functional Performance Testing Procedures. 2.: Raw Data of participants performance in FPT and QASLS Scores. 3.: Stata Syntax for statistical analysis. 4.: Intra and Inter-Rater Reliability of Individual QASLS Components.

Author Contributions. RC is a PhD student and AS (lead supervisor), JM, AR, NR and LH are supervisors. RC: conceptualization, methodology, Investigation, writing-original draft, visualization, project administration. LH: conceptualization, methodology, investigation, writing-reviewing & editing, supervision. JM: conceptualization, formal analysis, writing reviewing& editing. AR – conceptualization, methodology, writing reviewing & editing. NH - conceptualization, methodology, writing reviewing & editing. AS: conceptualization, writing- review & editing, supervision.

Funding: This article received no external funding.

Institutional Review Board Statement: The study was in accordance with the principles outlined in the Declaration of Helsinki. The study received ethical approval from the University of Birmingham (ERN_21-1414) and was conducted with the participants' informed and written consent collected utilising the REDCAP data collection system hosted at the University of Birmingham.

Consent for publication: Not applicable.

Availability of data and materials: All data generated or analysed for this study are included in this published article and within the supplementary materials which are linked to UBIRA eData Repository hosted by the University of Birmingham <https://doi.org/10.25500/edata.bham.00001255>.

Acknowledgments: The authors would like to acknowledge the support and engagement by players and staff at England Hockey who cooperated and participated in the data collection. The authors would also like to acknowledge the contribution of Dr Gemma Parry and Dr Liz Arnold to the inter-rater reliability data collection.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AR – anterior reach direction of the Y-balance test
 COSMIN – Consensus-based Standards for the selection of health Measurement INstruments
 DVJL – drop vertical jump land test
 FPT – Functional Performance Test
 GRRAS – Guidelines for Reporting Reliability and Agreement Studies
 ICC – Intraclass Correlation Coefficients
 NGB – National Governing Body
 PEA –Percentage Exact Agreement
 QASLS – Qualitative Assessment Single Leg Loading
 SH – Side Hop Test
 SHFD – Single Hop for Distance Test

References

- Braun HJ, Shultz R, Malone M, Leatherwood WE, Silder A, Dragoo JL. Differences in ACL biomechanical risk factors between field hockey and lacrosse female athletes. *Knee Surg Sports Traumatol Arthrosc.* 2015 Apr;23(4):1065–70.
- Murtaugh K. Injury patterns among female field hockey players. *Med Sci Sports Exerc.* 2001 Feb;33(2):201–7.
- Smith M, Weir G, Donnelly CJ, Alderson J. Field hockey sport-specific postures during unanticipated sidestepping: Implications for anterior cruciate ligament injury prevention. *J Sports Sci.* 2020 Nov;38(22):2603–10.
- Chaudhari AM, Hearn BK, Andriacchi TP. Sport-dependent variations in arm position during single-limb landing influence knee loading: implications for anterior cruciate ligament injury: Implications for anterior cruciate ligament injury. *Am J Sports Med.* 2005 June;33(6):824–30.
- Nedimyer AK, Boltz AJ, Robison HJ, Collins CL, Morris SN, Chandran A. Epidemiology of injuries in National Collegiate Athletic Association women’s field hockey: 2014–2015 through 2018–2019. *J Athl Train.* 2021 July 1;56(7):636–42.
- O’Neil B, Mentele P, Smith X, Pohlgeers K, Rimer E, Stamatis A. Longitudinal analysis of subsequent musculoskeletal injuries and predictive value of index injuries in collegiate women’s field hockey. *Int J Exerc Sci.* 2025 Jan 1;18(5):182–92.
- Smyth EA, Newman P, Waddington G, Weissensteiner JR, Drew MK. Injury prevention strategies specific to pre-elite athletes competing in Olympic and professional sports – A systematic review. *J Sci Med Sport.* 2019 Aug 1;22(8):887–901.
- Toohey LA, Drew MK, Cook JL, Finch CF, Gaida JE. Is subsequent lower limb injury associated with previous injury? A systematic review and meta-analysis. *Br J Sports Med.* 2017 Nov 17;51(23):1670–8.
- Verhagen E, van Dyk N, Clark N, Shrier I. Do not throw the baby out with the bathwater; screening can identify meaningful risk factors for sports injuries. *Br J Sports Med.* 2018 Oct;52(19):1223–4.
- Clark N. Functional performance testing following knee ligament injury. *Phys Ther Sport.* 2001 Jan 1;2:91–105.
- Hoch MC, Welsch LA, Hartley EM, Powden CJ, Hoch JM. Y-Balance Test Performance Following a Competitive Field Hockey Season: A Pretest-Posttest Study. *Journal of sport rehabilitation PB - [Internet].* 2017 Jan 1;26(5 PG-SP-EP-AN-). Available from: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med14&NEWS=N&AN=28530486>

12. Stokes MJ, Witchalls J, Waddington G, Adams R. Can musculoskeletal screening test findings guide interventions for injury prevention and return from injury in field hockey? *Phys Ther Sport*. 2020 Nov;46:204–13.
13. Girdwood MA, Crossley KM, Rio EK, Patterson BE, Haberfield MJ, Couch JL, et al. Hop to it! A systematic review and longitudinal meta-analysis of hop performance after ACL reconstruction. *Sports Med*. 2025 Jan;55(1):101–13.
14. Kotsifaki A, Van Rossom S, Whiteley R, Korakakis V, Bahr R, Sideris V, et al. Single leg vertical jump performance identifies knee function deficits at return to sport after ACL reconstruction in male athletes. *Br J Sports Med*. 2022 Feb 8;56(9):490–8.
15. Wren TAL, Mueske NM, Brophy CH, Pace JL, Katznel MJ, Edison BR, et al. Hop distance symmetry does not indicate normal landing biomechanics in adolescent athletes with recent anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther*. 2018 Aug;48(8):622–9.
16. Noyes F, Barber SD, Mangine R. Abnormal lower limb symmetry determined by function hop tests after anterior cruciate ligament rupture. *Am J Sports Med*. 1991 Sept 1;19:513–8.
17. 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. Available from: <http://dx.doi.org/10.15226/2374-6904/1/2/00113>
18. Straub RK, Powers CM. Biomechanical predictors of primary ACL injury: A scoping review of prospective studies. *Gait Posture*. 2025 Feb;116:22–9.
19. Butler L, Martinez A, Entessari M, Cardenas G, Wright M, Sugimoto D. Qualitative and quantitative return-to-sport test battery and second anterior cruciate ligament injury risk factors. *BMJ Open Sport Exerc Med*. 2024 Aug 3;10(3):e002000.
20. Parry GN, Herrington LC, Munro AG. Reliability and Measurement Error of the Qualitative Analysis of Single Leg Loading (QASLS) Tool For Unilateral Tasks. *Int J Sports Phys Ther*. 2023 Oct 1;18(5):1136–46.
21. Almangoush A, Herrington L, Jones R. A preliminary reliability study of a qualitative scoring system of limb alignment during single leg squat. Available from: <https://www.researchgate.net/publication/262735520>
22. van Melick N, Pronk Y, Nijhuis-van der Sanden M, Rutten S, van Tienen T, Hoogeboom T. Meeting movement quantity or quality return to sport criteria is associated with reduced second ACL injury rate. *J Orthop Res*. 2022 Jan;40(1):117–28.
23. Weber M, Müller M, Mathieu-Kälin M, Caminada S, Häberli M, Baur H. Evaluation of hop test movement quality to enhance return to sport testing. A cross-sectional study. *Front Sports Act Living*. 2024 Mar 4;6:1305817.
24. Welling W, Benjaminse A, Seil R, Lemmink K, Gokeler A. Altered movement during single leg hop test after ACL reconstruction: implications to incorporate 2-D video movement analysis for hop tests. *Knee Surg Sports Traumatol Arthrosc*. 2018 Oct;26(10):3012–9.
25. Lehr ME, Cheek W, Dacko S, Stramara T, Miller C, Antensteiner I, et al. Movement patterns and neuromusculoskeletal impairments observed in a female Olympic Field Hockey team: An observational cohort study. *J Bodyw Mov Ther*. 2021 Apr;26:128–33.
26. Clark NC, Clacher LH. Lower-limb motor-performance asymmetries in English community-level female field hockey players: Implications for knee and ankle injury prevention. *Phys Ther Sport*. 2020 May;43:43–51.
27. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud*. 2011 June;48(6):661–71.
28. Monti CB, Ambrogi F, Sardanelli F. Sample size calculation for data reliability and diagnostic performance: a go-to review. *Eur Radiol Exp*. 2024 July 5;8(1):79.
29. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*. 2019 July;95(103208):103208.
30. Munro AG, Herrington LC. Between-session reliability of the star excursion balance test. *Physical Therapy in Sport* PB - CY -. 2010 Jan 1;11(4):128-132 AN-DO-.

31. Nilstad A, Andersen TE, Kristianslund E, Bahr R, Myklebust G, Steffen K, et al. Physiotherapists Can Identify Female Football Players With High Knee Valgus Angles During Vertical Drop Jumps Using Real-Time Observational Screening. *J Orthop Sports Phys Ther.* 2014 May 1;44(5):358-365 AN-.
32. Dingenen B, Malfait B, Vanrenterghem J, Verschueren SMP, Staes FF. The reliability and validity of the measurement of lateral trunk motion in two-dimensional video analysis during unipodal functional screening tests in elite female athletes. *Physical therapy in sport : official journal of the Association of Chartered Physiotherapists in Sports Medicine PB -.* 2014 May 1;15(2):117-23 AN-.
33. Gustavsson A, Neeter C, Thomeé P, Silbernagel KG, Augustsson J, Thomeé R, et al. A test battery for evaluating hop performance in patients with an ACL injury and patients who have undergone ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc.* 2006 Aug 1;14(8):778-88.
34. Fältström A, Hägglund M, Hedevik H, Lindblom H, Kvist J. The side hop test: Validity, reliability, and quality aspects in relation to sex, age and anterior cruciate ligament reconstruction, in soccer players. *Phys Ther Sport.* 2023 July;62:39-45.
35. Mathieu-Kälin M, Müller M, Weber M, Caminada S, Häberli M, Baur H. Content validity, interpretability, and internal consistency of the 'Quality First' assessment to evaluate movement quality in hop tests following ACL rehabilitation. A cross-sectional study. *Front Sports Act Living.* 2023 June 16;5:1180957.
36. Epstein E, Huse C, Link M, Greenberg E. Reliability of a qualitative movement assessment tool during a single-leg triple hop landing. *Phys Ther Sport.* 2024 May;67:104-9.
37. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016 June;15(2):155-63.
38. Lidwine B, Mokkink M, Bouter J, Donald L. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measure.
39. Gokeler A, Welling W, Benjaminse A, Lemmink K, Seil R, Zaffagnini S. A critical analysis of limb symmetry indices of hop tests in athletes after anterior cruciate ligament reconstruction: A case control study. *Orthopaedics & traumatology, surgery & research : OTSR PB -.* 2017 Jan 1;103(6):947-951 AN-.
40. West TJ, Bruder AM, Crossley KM, Culvenor AG. Unilateral tests of lower-limb function as prognostic indicators of future knee-related outcomes following anterior cruciate ligament injury: a systematic review and meta-analysis of 13 150 adolescents and adults. *Br J Sports Med.* 2023 July;57(13):855-63.
41. Hegedus EJ, McDonough S, Bleakley C, Cook CE, Baxter GD. Clinician-friendly lower extremity physical performance measures in athletes: a systematic review of measurement properties and correlation with injury, part 1. The tests for knee function including the hop tests. *Br J Sports Med.* 2015 Apr 2;49(10):642-8.
42. Cooke R, Rushton A, Martin J, Herrington L, Heneghan NR. Practicability of lower extremity functional performance tests and their measurement properties in elite athletes: protocol for a systematic review. *BMJ Open.* 2020 Dec 22;10(12):e042975.
43. Kockum B, Heijne AILM. Hop performance and leg muscle power in athletes: Reliability of a test battery. *Phys Ther Sport.* 2015 Aug 1;16(3):222-7.
44. Haynes T, Bishop C, Antrobus M, Brazier J. The validity and reliability of the My Jump 2 app for measuring the reactive strength index and drop jump performance. *J Sports Med Phys Fitness.* 2019 Feb;59(2):253-8.
45. Gallardo-Fuentes F, Gallardo-Fuentes J, Ramírez-Campillo R, Balsalobre-Fernández C, Martínez C, Caniuqueo A, et al. Intersession and intrasession reliability and validity of the My Jump app for measuring different jump actions in trained male and female athletes. *J Strength Cond Res.* 2016 July;30(7):2049-56.
46. Fox AS, Bonacci J, McLean SG, Spittle M, Saunders N. A Systematic Evaluation of Field-Based Screening Methods for the Assessment of Anterior Cruciate Ligament (ACL) Injury Risk. *Sports Med.* 2016 May 1;46(5):715-35.
47. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Practical guides to biostatistics and epidemiology: Measurement in medicine: A practical guide. Cambridge, England: Cambridge University Press; 2011. 347 p.
48. Mokkink LB, Eekhout I, Boers M, van der Vleuten CPM, de Vet HCW. Studies on Reliability and Measurement Error of Measurements in Medicine - From Design to Statistics Explained for Medical Researchers. *Patient Relat Outcome Meas.* 2023 July 7;14:193-212.

49. Read P, Mc Auliffe S, Wilson MG, Myer GD. Better reporting standards are needed to enhance the quality of hop testing in the setting of ACL return to sport decisions: a narrative review. *Br J Sports Med.* 2021 Jan;55(1):23–9.
50. Munro A, Herrington L. BETWEEN-SESSION RELIABILITY OF FOUR HOP TESTS AND THE AGILITY T-TEST. *J Strength Cond Res.* 2011 Apr 19;25(5):1470–7.

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.