

Intra- and Inter-Rater Reliability of Features of Movement Commonly Associated with Movement
Competency

by

Dilara Sen

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Kinesiology and Health Sciences

Waterloo, Ontario, Canada, 2025

© Dilara Sen 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Movement screening tools, such as the Functional Movement Screen™ and the Movement Competency Screen, are commonly used to assess fundamental movement patterns. These movement patterns, such as squatting and lunging, serve as the foundation for human movement, essential for daily living and athletic performance. By assessing an individual's ability to control their joints during fundamental movements, screens help identify potential areas for improving performance and reducing injury risk. However, existing screens are limited. They do not adequately assess movement competency, the ability to move competently, without limitation or impairment, under different task demands (e.g., varying loads or speeds).

The Physical Literacy Screen™ (PLS) was developed to more comprehensively assess movement competency. Unlike previous screens, the PLS considers how different task demands (e.g., varying load or speed) influence how movers control key movement features, to inform exercise, education, and coaching recommendations based on a more thorough understanding of an individual's movement abilities. Because PLS feature scoring relies entirely on visual observation, the reliability of judging these features must be established before the framework can be applied with confidence. Therefore, the purpose of this study is to determine the intra-rater and inter-rater reliability of visually assessing key movement features commonly associated with movement competency.

This study examined the reliability of raters visually judging three key movement features, knee control, back control, and shoulder control, using standardized video recordings of athletes performing PLS-derived tasks. Twenty-two kinesiology students independently assessed pre-recorded videos of varsity volleyball athletes and re-scored the same videos at least 14 days later. Intra-rater reliability was measured using Cohen's Kappa, and inter-rater reliability using Fleiss' Kappa.

Results showed moderate intra-rater reliability for Knee ($\kappa = 0.42$) and Back ($\kappa = 0.47$) features and fair reliability for Shoulder ($\kappa = 0.28$). Inter-rater was poor across all features, with Fleiss' Kappa values below zero.

The study provides initial reliability evidence for the visual judgment of key movement features using PLS-derived tasks. These findings highlight the importance of refining feature definitions, improving rater training and calibration procedures, and clarifying scoring criteria to enhance consistency. Strengthening these elements may support more reliable use of visual movement assessments in sport, education, and health contexts.

Acknowledgements

This thesis was completed with the guidance and support of many individuals, whose contributions are gratefully acknowledged.

I would like to begin by thanking my supervisor, Dr. Steve Fischer, for his guidance, patience, and consistent support throughout my graduate studies. Your perspective and mentorship provided steady direction across all stages of this project and were invaluable in navigating both the research process and its challenges.

I am also thankful to my committee members, Dr. Tyson Beach and Dr. Jordan Canon, for their time, expertise, and constructive feedback. Your insights helped refine the scope, clarity, and overall quality of this work.

Thank you to the members of the OBEL Lab for fostering a supportive and collaborative research environment. I am especially grateful to Devon Frayne for his assistance with data collection and for his continued support, expertise, and willingness to help throughout this project.

I would also like to acknowledge the University of Waterloo women's volleyball team and head coach Richard Eddy for their trust and participation. Their openness and commitment made this research possible.

Finally, I would like to thank my friends and family for their unwavering support throughout this journey. Thank you to my mom, dad, sister Emily, and my cousin Zelal for their encouragement and understanding during challenging times. Your support played an essential role in the completion of this thesis.

Table of Contents

Author's Declaration	ii
Abstract	iii
Acknowledgements	iv
List of Figures	vi
List of Tables	vii
List of Abbreviations	viii
Chapter 1	1
Chapter 2	4
Chapter 3	15
Chapter 4	21
Chapter 5	25
Chapter 6	34
References	36
Appendix A Consent Forms	42
Appendix B Physical Literacy Screening Instructions and Standards	47
Appendix C Physical Literacy Screening Criteria	57
Appendix D Scoring Data Structure for Intra-Rater Reliability Analysis	58
Appendix E Rater Completion Time Summary	59

List of Figures

Figure 1. Sample of the Screening and Assessment Sheet for Physical Literacy Screen™, with the Key Feature Score Circled. Retrieved from Patterns & Features, KIN 204 Lecture Content.....	15
Figure 2. Cropped Video Display Showing One Athlete Performing the Lunge as Part of the PLS, from a Frontal and Sagittal View.....	17
Figure 3. Example of Microsoft Forms Quiz Layout for PLS Rating Task, Displaying the One Leg Squat Exercise with Control Checkboxes for Knee and Back.....	18
Figure 4. Intra-Rater Agreement on Knee, Back, and Shoulder Control Features (Mean ± SD Cohen's Kappa).....	22
Figure 5. Frontal and Sagittal Views of Participant Performing a One Leg Hip Bridge with Shifting of Foot's Centre of Pressure that May Have Increased Rater Difficulty in Scoring Knee Feature Control	26
Figure 6. Frontal and Sagittal Views of participant Performing an Inverted Row with a Loose-Fitting Shirt that May Have Increased Rater Difficulty in Scoring Shoulder Feature Control.....	27
Figure 7. Consent Form for Sharing Athlete General Performance Testing Data Within the Research Group at the University of Waterloo.....	43
Figure 8. Consent Form for the Athlete Use of Data for Thesis on Intra- and Inter-Rater Reliability of the Physical Literacy Screen™	44
Figure 9. Consent Form for the Raters Scoring the Reliability of the Physical Literacy Screen™	45
Figure 10. Participant Renumeration and Self-Declaration Form.....	46
Figure 11. Illustration of the presence or absence of RoM and FITNESS criteria for the Physical Literacy Screen™ during the hinge (Back Motion Control).....	57
Figure 12. Illustration of the presence or absence of RoM and FITNESS criteria for the Physical Literacy Screen™ during the lunge (Knee Motion Control).....	57
Figure 13. Illustration of the presence or absence of RoM and FITNESS criteria for the Physical Literacy Screen™ during the push-up (Shoulder Motion Control).....	57

List of Tables

Table 1. Inter-Rater Reliability (Fleiss' Kappa) by Key Feature.....	23
Table 2. Screening Instructions for Physical Literacy Screen™.....	47
Table 3. Physical Literacy Screen™ Standards.....	51
Table 4. Sample of Raw Rater Scores for One Athlete Across Two Assessment Attempts.....	58
Table 5. Average Completion Time per Rater Across Both Scoring Sessions.....	59

List of Abbreviations

FMS	Functional Movement Screen
MCS	Movement Competency Screen
PLS	Physical Literacy Screen
MSDs	Musculoskeletal Disorders
RULA	Rapid Upper Limb Assessment
REBA	Rapid Entire Body Assessment
OWAS	Ovako Working Posture Analysis System
GRRAS	Guidelines for Reporting Reliability and Agreement Studies

Chapter 1 Introduction

1.1 Background

Physical literacy is a construct integrating motivation, confidence, movement competence and movement knowledge. Physical literacy influences how one moves, which in turn can influence quality of life (Whitehead, 2001). Movement competency is a subset of physical literacy and refers to the strategies an individual adopts to move safely and efficiently across various tasks. The demands of a task influence movement competency, where one may move differently when the demands of a task are perceived as low versus elevated (e.g., self-paced vs fast). An individual's competency to leverage their capacity is important to ensure they can meet the demands of a task safely and efficiently (Armstrong et al., 2019). For an individual to optimize use of their capacity, they must build on their movement capacity, which can be done by improving how they control key movement features (e.g. frontal plane knee control and foot-ground center of pressure control). These key features are critical as they regulate stability, alignment, and movement across major joints and muscle groups, allowing for coordinated and efficient movement. Conserving the key features when moving can help increase movement capacity to meet the demands of a task.

Movement screening was popularized by the Functional Movement Screen™ (FMS), developed by Cook et al. (2006), to identify limitations or asymmetries in fundamental movement patterns. However, there remain outstanding concerns regarding its validity in measuring its intended outcomes, as studies have demonstrated problems with the ambiguous grading criteria, the influence of learned behaviour on scores, and the lack of correlation with other movement measures (Bonazza et al., 2017). For instance, the study by Frost et al. (2015d) found that providing individuals with feedback on the grading criteria of the FMS facilitates passing when they previously failed, questioning its utility in assessing the transfer of training. The Movement Competency Screen (MCS) advanced from the FMS by focusing on movement competency during functional tasks rather than identifying movement asymmetries during functional movements (Kritz, 2012). However, the MCS, along with the FMS, continued to have outstanding limitations in comprehensively assessing movement competency due to an inability to consider how movement might change under varying task demands by only assessing bodyweight exercises at a self-selected pace (Frost et al., 2015c). As a result, the emerging Physical Literacy Screen™ (PLS) is an advance from existing movement screening paradigms and was developed to address those limitations.

The PLS includes 15 tasks that challenge movers to maintain control of seven key features under three task demands: strength, speed, and endurance. Raters use the PLS scoring guide to assess seven movement patterns, squat, lunge, hinge, push, and pull, by evaluating key features across task demands to establish movement competency (Performance Redefined, 2019). The PLS addresses limitations of previous movement screening tools by assessing movement patterns under varying task demands rather than performing all at a self-selected pace increasing the likelihood of revealing an individual's movement behaviours (i.e., traits not states).

Common to the use of any screen, it is important to understand the measurement properties. This lack of clarity can lead to inconsistencies in visual assessments, affecting the accuracy of the evaluations. While the PLS requires administration by a trained rater, discrepancies may still arise due to variations in foundational knowledge of movement competency. Visual assessment of human movement is fundamental in sport, rehabilitation, ergonomics, and health settings. Despite its widespread use, visual judgment is influenced by perceptual limitations, viewing angle constraints, rater training, and the structure of scoring systems, leading to differences in how raters score the same movement. Because visual evaluation relies on consistent interpretation of movement cues, determining reliability is essential for meaningful application. Establishing intra-rater and inter-rater reliability determines whether differences reflect true changes or inconsistencies in scoring.

Intra-rater reliability refers to the consistency of scores given by the same rater when assessing the same performance at different time points. Inter-rater reliability is the level of agreement between different raters scoring the same performance. Establishing these measurement properties ensures that observed differences in movement reflect true behavioural change rather than scoring variability. By evaluating rater consistency in visually judging key movement features within PLS-derived tasks, the present study addresses an essential gap in understanding the measurement properties of visually based movement assessments.

1.2 Problem Statement

Although the PLS provides a structured framework for evaluating movement competency, little is known about the reliability of visually judging the key movement features it targets. No published studies have examined how consistently raters can evaluate these features, particularly when assessments are made from video recordings rather than live observation. Establishing the

reliability of these visual judgments is essential for determining whether observed differences in scores reflect true variation in movement behaviour or inconsistencies in rater interpretation.

1.3 Research Questions and Objective

This study examined the reliability of visually judging three key movement features (knee control, back control, and shoulder control) during tasks derived from the PLS. The following research questions guided the investigation:

- **Question 1:** Among undergraduate students trained in PLS scoring principles, what is the intra-rater reliability of visually judging key movement features when the same raters reassess the same performances after a minimum of 14 days?
 - **Hypothesis 1:** The intra-rater reliability of the PLS scores will be strong (Cohen's Kappa = 0.61-0.80), indicating substantial agreement when raters reassess the same performances after a minimum of 14 days.
- **Question 2:** Among undergraduate students trained in PLS scoring principles, what is the inter-rater reliability of visually judging key movement features when the different raters assess the same performances?
 - **Hypothesis 2:** The inter-rater reliability of the PLS scores will be moderate (Fleiss's Kappa = 0.41-60), indicating moderate agreement when different raters assess the same performances.
- **Objective:** To determine the reliability of visually judging knee, back, and shoulder control from standardized video recordings of athletes performing PLS-derived tasks when evaluated by the same raters at two different time points (intra-rater) and by different raters (inter-rater).

To achieve this objective, trained student raters evaluated video recordings of movers performing the PLS. The same raters reassessed the recordings after 14 days to determine intra-rater reliability, and different raters assessed the same recordings to determine inter-rater reliability.

Chapter 2 Literature Review

2.1 Development and Adoption of Physical Literacy

Physical literacy is a longstanding concept with varying definitions and interpretations. Whitehead (2001) is credited with having advanced the construct to the level it is known and recognized today. Her definition suggests that physical literacy is more than the body's ability to complete physical movement; it also considers how one reads the environment and interacts accordingly. It is a widely supported notion that has been adopted across Canada and the rest of the world. The Canadian Sport for Life was one of the first organizations to endorse physical literacy, subsequently influencing many other Canadian organizations endorsing physical activity to promote this concept (Balyi et al., 2005). With its growth in popularity, there has been variability in its accepted definition and elements comprising the concept. Some definitions emphasize physical competence, while others focus more on the cognitive and or behavioural aspects of movement. An essential component supported across all definitions of physical literacy is movement competency, a subset of the physical domain of physical literacy, which is defined as the ability to move in a controlled and safe manner. Jurbala (2015) argues that all humans capable of movement possess some degree of physical literacy; the variability lies in the extent of their literacy, which is unique to each individual due to differences in their movement competency.

Movement competency is closely linked to movement capacity, which refers to the body's ability to handle the load demands required to perform a task safely and effectively. Movement capacity must exceed the demands of a task for it to be performed safely and effectively. The Demand-Capacity-Competency model suggests that movement acts as a fulcrum, balancing the physical demands of a task and an individual's capacity to meet those demands (Armstrong et al., 2019). This particularly applies to athletic contexts, where tasks often involve external loads, varied speeds, and prolonged durations, which can cause individuals to modify their movement patterns. These factors can cause individuals to modify their movement patterns, leading to compensatory movement strategies that may alter key movement features, increasing the risk of injury (Hewett et al., 2005).

2.2 Evidence-Based Training for Movement Features

Strength and conditioning training objectives aim to enhance athletic performance and ensure the long-term health and wellness of athletes. This dual focus is crucial as performance improvements are important but should not come at the expense of an increased risk of injury for the athlete.

Fundamental movement patterns, such as the squat, lunge, hinge, push, and pull, essential for daily activities and sport-specific training, provide insights into an individual's movement competency that may contribute to soft tissue pain, discomfort, and injury (Kritz et al., 2009a, 2009b, 2010). Frost et al.'s study (2015a) investigated the predictive value of general tasks in estimating occupational task performance in firefighters. They found that spine and knee motion magnitudes were often higher in general tasks compared to job-specific ones, yet the strategies used to control those movements were consistent across both contexts. This suggests that the way individuals organize and control movement, rather than the exact magnitude, may represent a stable characteristic of how they move. Therefore, general movement assessments can still provide meaningful information about a person's habitual movement patterns and control strategies, supporting their use in strength and conditioning contexts to identify potential movement deficiencies before they appear in more demanding, sport-specific tasks.

To effectively support athletic performance, strength and conditioning training must also focus on movement traits that are representative of how one moves across most tasks rather than capturing a "snapshot" of a movement state. Movement "states" are temporary movement solutions influenced by immediate conditions, whereas "traits" are habitual movement patterns within a series of movement solutions that more accurately reflect an individual's baseline movement competency. Frost et al. (2015d) emphasize the importance of this distinction, noting that assessments capturing state-related variations may reflect temporary adaptations rather than stable traits. Therefore, in strength and conditioning programming, emphasis should be placed on traits that influence long-term movement behaviours. Good movement behaviour forms the foundation for effective movement and injury prevention, providing alignment of the musculoskeletal system and reducing stress on joints and soft tissues (McGill, 2007). Therefore, an understanding of good movement behaviour is essential in a strength and conditioning environment.

Strength and conditioning programs often focus on foundational movement patterns like the squat, hinge, lunge, push, and pull in programming. Squat and hinge patterns are particularly valuable for assessing lower body strength and mobility; when executed correctly, they engage key muscle

groups and enhance movement efficiency (Kritz et al., 2009a). Screening of the squat involves movement initiation, hinging at the hips, ensuring the knees are aligned over the toes, maintaining a neutral spine and keeping the chest upright throughout (Kritz et al., 2009a). The hinge is also initiated at the hips with a slight knee flexion, a neutral spine and back, hips moving back, and the weight centred over the midfoot (see Figure 9) (Kritz et al., 2009a). The lunge is another fundamental movement pattern, crucial in activities of daily living and sports. There are various forms of the lunge pattern to assess strength, flexibility, and balance, the most common being the forward lunge, as it mimics lower body movements during the gait cycle (Kritz et al., 2009b). During the lunge, the knees should be aligned over the knee, the torso should be upright, and the back knee should descend without touching the ground (see Figure 10) (Kritz et al., 2009b). Upper body movements are also integral to sports and daily living, and upper body strength, scapular control, and grip strength are evaluated. In a push/pull pattern, the core should be stabilized, a centred head position, shoulders down and away from the ears and a controlled protraction/ retraction of the scapulae during the movement (see Figure 11) (Kritz et al., 2010).

Control and range of motion of these specific movement features are common objectives in strength and conditioning training. For instance, knee valgus, where the knee collapses inward during movements like squatting or landing, has been identified as a risk factor for injuries like anterior cruciate ligament tears (Hewett et al., 2005). Strength and conditioning professionals focus on correcting these patterns by prescribing corrective exercises to improve alignment and reduce injury risk. Therefore, identifying key movement features across fundamental movement patterns may better inform training decisions to improve performance and decrease injury risk.

2.3 Current Movement-Based Assessments

2.3.1 Functional Movement Screen (Cook et al., 2006)

Gray Cook and Lee Burton developed the Functional Movement Screen™ (FMS) to assess movement patterns and identify asymmetries that may lead to injury in asymptomatic, healthy populations (Cook et al., 2006). The FMS consists of seven fundamental movement tasks: deep squat, hurdle step, in-line lunge, shoulder mobility, active straight leg raise, trunk stability push-up, and rotary stability. Each task is scored on a scale from 0 to 3, where higher scores indicate better ability to maintain movement with sufficient control and mobility, referred to as movement quality. The selected movements were chosen to comprehensively assess an individual's movement quality

relevant to daily activities, athletic performance and injury prevention. The FMS has been widely adopted in both research and practical applications. Studies have attempted to investigate the reliability and effectiveness of the FMS and generally have found good intra-rater and inter-rater variability, and some studies have found a positive predictive ability for injury risk (McCunn et al., 2016).

Despite its broad implementation, a significant limitation of the FMS is that the tasks included are relatively low demand, which may not be sufficient to expose problematic movement behaviours (Frost et al., 2015b). Additionally, the FMS is often used to interpret changes in movement quality after training or to inform personalized exercises which deviate from the screen's original intent (Frost et al., 2017). Furthermore, the study by Frost et al. (2017) revealed that the FMS scores were inconsistent with pre-training and post-training, even in the control group, indicating potential issues with the test's reliability. This suggests that the FMS may capture a person's movement state, how they move at a particular time, rather than their movement traits, which refers to their habitual movement behaviours. The distinction between state and trait is important because training interventions should ideally impact long-term behavioural patterns rather than temporary performances.

2.3.2 Movement Competency Screen (Kritz, 2012)

Matthew Kritz developed the Movement Competency Screen (MCS) to provide a screen for sports and health professionals to assess whole-body functional movement competency. Due to the recognized importance of movement competency in athletic development, performance, and injury, the study aimed to create and validate a reliable screening tool. The screen was developed based on recommendations obtained from Delphi surveys of strength and conditioning specialists and sports physiotherapists to determine its structure and function. The investigators then examined user perception following its development to ensure its applicability and adoption. The MCS consists of five key movement tasks: squat, lunge & twist, push-up, bend & pull, and single-leg squat. These were selected based on fundamental movement patterns related to daily activities, sports and sport-specific training: the squat, lunge, upper body push, upper body pull, trunk flexion or bend, trunk rotation or twist, and unilateral lower limb or single leg squat. The tasks are evaluated based on criteria for each body segment and movement. The development process of the MCS was rigorous and well-founded in its choice of assessment exercises for the squat, lunge, push-up, and row through

supported evidence and expert opinion. In the study by Pagaduan & Kritz (2022) examining the reliability of the MCS, there was good intra-rater and inter-rater reliability for scoring done by experts in sports and conditioning. In addition to the screen, Kritz (2012) has developed “load levels” that can be used as progressions of the original MCS exercises. The load levels were developed to inform sports and conditioning and physiotherapy training and to progress the athlete’s performance in a safe and informed manner. Each pattern has external load levels, where each pattern may receive a score independent from the other. The MSC is scored with a final composite score ranging between 7-21, where 7-10 indicates poor overall movement competency, 11-16 indicates moderate movement competency, and 17-21 indicates good movement competency.

While the MSC provides a more detailed assessment of movement competency, it has limitations. A notable concern is that although the screen includes “load demands” to progress exercises based on an athlete’s performance, it does not consider the demands of the task. The demands of a task have been shown to influence movement behaviour, where different demands require specific adaptations in movement strategies (Frost et al., 2015c). Additionally, the reliability study by Pagaduan & Kritz (2022) only examined two expert raters, which does not provide a comprehensive assessment of the tool’s reliability across a broad range of evaluators, limiting its application.

2.3.3 Physical Literacy Screen (Performance Redefined, 2019)

The Physical Literacy Screen™ (PLS), developed by Performance Redefined, aims to provide a comprehensive physical literacy assessment by including the evaluation of movement competency and fitness (Table 1). The screen includes components designed based on evidence to evaluate movement patterns, such as squatting and lunging, while focusing on key features like knee alignment, trunk stability and shoulder control (Performance Redefined, 2019). These tasks are performed under varying task demands, such as changes in speed, to challenge different aspects of an individual’s movement abilities. Although the PLS has yet to be validated, it is built on validated methodologies for assessing movement competency.

The PLS consists of 15 tasks divided into three fitness domains: repetition-based, work-based, and tempo-based. Scoring criteria are designed to guide the assessment of movement behaviours under different task demands, expanding on the limitations of the previously mentioned screens. Including varying fitness domains in the PLS ensures that the screen evaluates how

individuals perform fundamental movements under different task demands. This approach incorporates the findings from Frost et al. (2015c), who demonstrated that different task demands, such as increased speed and load, significantly influence movement behaviours.

Within each domain, five movement patterns are evaluated in the PLS: squat, lunge, hinge, push, and pull are examined to assess physical literacy (Performance Redefined, 2019). Dynamic alignment is the control of movement throughout its full range and provides a comprehensive understanding of an individual's movement capabilities. Focusing on the continuous patterns of motion between positions rather than static positions alone, the assessment offers a more accurate reflection of daily activities and sports performance. While the PLS is not designed to predict injury, evaluating these fundamental movement patterns helps identify movement inefficiencies, such as poor muscle coordination or restricted joint mobility, which are documented in literature as affecting joint function and overall movement competency (Hewett et al., 2005).

When evaluating movement patterns, the PLS emphasizes seven key movement features that are important in understanding an individual's movement competency (see Appendix C): frontal plane knee control and foot-ground centre of pressure control, shoulder elevation/depression control and shoulder protraction/retraction control, sagittal plane spine control, frontal plane spine control, and transverse plane spine control (Performance Redefined, 2019). These features were chosen based on their role in maintaining joint stability and supporting efficient movement patterns during dynamic tasks. Verbal instructions on how to maintain ideal form during these movements are provided in Appendix A, ensuring consistency in assessment practices. By examining how individuals control these key movement features under varying loads and speeds, the PLS could help better inform training and injury prevention strategies.

The PLS is similar to the FMS and MCS in that it is a movement-based screen that assesses fundamental movement patterns; however, it places a greater emphasis on movement competency, particularly focusing on the transitions of the movement rather than isolated postures and the demands of the task. The PLS challenges movers to perform specific movements under different task demand requirements, which increases the likelihood of capturing traits rather than states. This emphasis on dynamic movement control under different task demands reinforces the utility of the PLS for real-world applications, as opposed to the other screens.

2.4 Learning from Adjacent Fields Leverage Visual Assessment of Movement

Several visual observation tools have been developed in occupational ergonomics to evaluate posture and movement risk. Commonly used observational tools include the Rapid Upper Limb Assessment (RULA), Rapid Entire Body Assessment (REBA), and Ovako Working Posture Analysis System (OWAS) (Kee & Karwowski, 2007). Like movement screens, these tools rely on raters to visually evaluate body positioning and assign categorical risk scores, making rater reliability an important consideration.

RULA was designed to quickly evaluate a worker's exposure to ergonomic risk factors associated with musculoskeletal disorders (MSDs) of the upper extremities (McAtamney & Corlett, 1993). Its simplicity has supported widespread use, but studies have reported mixed reliability outcomes. Depending on rater training and task complexity, the tool's inter-rater agreement has ranged from moderate to substantial. Its intra-rater reliability tends to be stronger, where reliability improves when assessments are repeated by the same individual (Dockrell et al., 2012). REBA was developed to quickly and simply evaluate whole-body postures and is particularly applied in healthcare and manual handling settings (Hignett & McAtamney, 2000). Research on its reliability has shown variable findings where experienced raters using standardized protocols often achieve substantial agreement, while novices show only fair levels of consistency (Chiasson et al., 2012). These results highlight the importance of expertise when applying REBA in occupational settings. OWAS is also a whole-body posture classification system that categorizes trunk, arm, and leg postures into risk codes for identifying potentially harmful working positions (Karhu et al., 1977). While it is also considered a relatively simple tool, reliability studies have reported moderate inter-rater agreement, with higher consistency observed for gross postures than for more subtle classifications (Kee & Karwowski, 2007).

Beyond ergonomics, multiple studies have evaluated the accuracy and reliability of visual observation relative to instrumented motion analysis. Systematic reviews comparing observer ratings to three-dimensional motion analysis for functional screening tasks (e.g., squats, jumps, and landing tasks) report variable validity and reliability, with better agreement for slower or more constrained tasks and larger errors for complex, high-speed movements (Whatman et al., 2015). Similarly, work on visual posture observation suggests that untrained observers often misestimate joint angles, and that targeted training can reduce, but not eliminate these errors (Covalla, 2003).

Clinical and sport biomechanics research has also highlighted the cognitive demands of qualitative movement analysis. Knudson and Morrison (2002) describe qualitative analysis as a multi-step process requiring selective attention, interpretation spatiotemporal patterns, and integration of biomechanical principles, all of which can influence reliability. Furthermore, it was shown that observers must filter large amounts of kinematic information, and their judgments are shaped by experience, expectations, and prior knowledge. These cognitive demands help explain why expert raters tend to demonstrate higher accuracy and agreement than novices across visual movement assessments, as consistently reported in studies of functional screening and lower-extremity alignment evaluation (Herrington & Munro, 2014).

Collectively, these fields show that visual observation can be informative, but reliability depends on task complexity, viewing conditions, scoring rules, and rater training.

2.5 Challenges to Consider for Movement-Based Screening Tools

Visually assessing patients and athletes is essential for practitioners and trainers to give meaningful feedback and conduct assessments. However, equipment and time are often not always feasible for quantitative biomechanical assessments. The reliability of visually appraising specific movement features, such as those assessed in the PLS, requires further exploration. Studies have examined the reliability of joint angles or positions, but the ability to visually appraise movement features is less understood.

Among movement-based assessments, rater expertise has been established as an important consideration of rater reliability. For example, Whatman et al. (2012) examined physiotherapists with varying levels of experience rating participants using a lower extremity functional test and found excellent intra-rater reliability and fair to good inter-rater reliability. There was also evidence that inter-rater agreement improves with experience, where experienced clinicians tend to have higher levels of agreement and less variation in their ratings (Whatman et al., 2012). Furthermore, the study by Mann et al. (2022) investigating composite scoring of a movement competency screen in pre-elite youth athletes found that the final scores could be reliable; however, expert raters detected more errors than novices. Therefore, this emphasizes the importance of the rater's training and expertise when evaluating a movement screen, and it is important to ensure raters are adequately trained to minimize intra-rater and inter-rater variability.

Another consideration is the impact of the viewing medium on rater variability. Many movement screens are designed as observational tools, assuming the rater has unobstructed visibility and can move around to better observe the tasks. In a systematic review of rater reliability for FMS scores by Moran et al. (2017), it was found that studies with video ratings had more conflicting evidence compared to live ratings, making live scoring more favourable. This becomes especially important when considering the real-world applicability of the screens, where they are likely to be administered in person. However, video assessment is sometimes necessary, especially when evaluating intra-rater reliability, as it allows for consistent reevaluation of the same performances. Additionally, it enables remote scoring, which may be preferred or more feasible for some practitioners or trainers. Therefore, although live scoring may be a preferred method for movement assessments, video assessments may be necessary in certain circumstances, and their limitations should be considered.

2.6 Reliability of Visual Observation of Human Movement

Visual observation is widely applied in ergonomics, clinical gait analysis, rehabilitation, and sport performance to evaluate how individuals move. Research across these fields shows that visually assessing human movement is fundamentally constrained by selective attention, occlusion, limited viewing angles, and variability in how raters interpret movement cues (Knudson & Morrison, 2002). These limitations are observed consistently across posture assessment, gait evaluation, lifting analysis, and task-based movement classification, demonstrating that visual analysis is inherently constrained even under ideal viewing conditions.

Studies comparing visual judgments with three-dimensional motion analysis repeatedly show that raters struggle to identify subtle deviations, joint misalignments, and multi-planar movements, particularly during dynamic or complex tasks. For example, MacLachlan and White (2015) showed that observer ratings of lifting technique differed significantly from motion capture data, consistent with findings in other manual-handling contexts (Karhu et al., 1977). Recent work has expanded this evidence base, with observational studies reporting reduced accuracy when viewing speed increases, when multiple joints must be monitored simultaneously, or when dynamic alignment changes rapidly (Whatman et al., 2015). Research in clinical posture assessment further demonstrates that even trained clinicians misestimate joint orientation from single-plane photographs or video, particularly when spinal curvature or shoulder positioning are involved (Covalla, 2003). Collectively, these

studies demonstrate that visual analysis is best understood as a perceptual estimation process shaped by both movement complexity and observer characteristics.

Another important determinant of reliability to consider is the level of rater training and calibration. Knudson and Morrison (2002) emphasize qualitative movement evaluation as a multi-step analytical process requiring domain knowledge, biomechanical understanding, and systematic attention to movement phases. This is supported by numerous studies showing that expert raters demonstrate higher intra- and inter-rater reliability than novices (Whatman et al., 2012; Mann et al., 2022). Training and calibration are especially important when scoring subtle qualities such as spinal deviations, scapular positioning, or knee alignment during multi-joint tasks. Similar observations exist in ergonomics research using tools such as RULA, REBA, and OWAS, where inter-rater reliability is often only moderate and improves when raters undergo standardized calibration sessions (Kee & Karwowski, 2007; Chiasson et al., 2012).

Reliability is also limited when raters must judge individual movement features from restricted viewpoints. Research in qualitative movement analysis shows that reliable visual judgment typically depends on integrating multiple cues into an overall assessment, rather than evaluating isolated segments (Knudson & Morrison, 2002). Studies comparing visual judgments with biomechanical reference measures demonstrate that novices commonly focus on a single joint or feature and miss compensatory patterns occurring elsewhere in the kinetic chain, whereas experts process movement more holistically (Whatman et al., 2012). This challenge is also reflected in ergonomics research, where observers evaluating posture or segmental alignment show reduced accuracy and reliability when required to attend to isolated anatomical regions rather than broader movement patterns (Kee & Karwowski, 2007; Chiasson et al., 2012). In the context of the PLS, this is particularly relevant because the present study required raters to score only three isolated features (knee, back, and shoulder control) rather than providing holistic movement ratings. Literature across similar fields suggests that this scoring approach increases perceptual load and reduces reliability.

Additionally, newer research in ergonomics and clinical movement analysis has strengthened the understanding of how fixed-camera video affects observer accuracy. Studies using photogrammetry, smartphone-based posture analysis, and clinical video assessments report that deviations outside the primary plane of view are commonly missed, and reliability decreases when raters cannot reposition themselves or inspect the movement from multiple angles (Karhu et al., 1977; MacLachlan & White, 2015). Another more recent ergonomics study by Valentim et al. (2025)

further demonstrated that even small changes in camera placement, such as slight shifts in height or lateral positioning, significantly alter raters' ability to detect spinal curvature, scapular asymmetry, and frontal-plane knee alignment. The authors showed that occlusion created by limb segments or clothing reduced agreement between raters by up to 40%, and that errors were greatest for features requiring multi-planar observation, such as shoulder elevation/depression and thoracic rotation. These findings reinforce that fixed-camera video inherently restricts the visual information available to raters.

Overall, the literature on visual posture and movement assessment demonstrates that observational methods are inherently variable and sensitive to perceptual, cognitive, and methodological constraints. Incorporating these insights provides context and highlights the need to consider viewing angle limitations, rater training, and task complexity when evaluating the reliability of visually judged movement features.

2.7 Key Learnings from the Literature Review

The literature shows that visual assessment of movement is inherently limited by perceptual constraints, rater variability, and restricted viewing conditions. Expert raters consistently demonstrate higher agreement than novices, and video-based assessment can further reduce reliability by limiting access to key movement cues. Although the PLS improves on earlier movement screens by challenging individuals across different task demands, its reliability remains unknown. Given the documented challenges of visual judgement, it is essential to determine whether trained raters can consistently evaluate the key movement features targeted by the PLS so that observed changes in movement reflect true adaptations rather than inconsistencies in scoring.

Chapter 3 Methodology

3.1 Research Design

This study aimed to assess reliability of visually assessing three key movement features of movement competency (knee, back, and shoulder control) using PLS-derived movement tasks. Trained kinesiology student raters evaluated video recordings of volleyball athletes performing the PLS, following the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (Kottner et al., 2011). Reliability was assessed using the aggregated key feature scores, which reflect whether specific movement control criteria are consistently met across the entire test battery (shown in Figure 1). Rather than evaluating features within each individual task domain (strength, endurance, speed), this study focuses on total scores for each key feature (knee, back, shoulder) by summing outcomes across all domains. This aggregation approach offers a simplified and comprehensive measure of movement competency based on key features. Ethics clearance and consent were obtained for part of the study involving the collection of athlete data and the rater component (see Appendix A).

— RECORDING FORM —

PERFORMANCE REDEFINED.

Use the screening protocol to assess control of general patterns with elevated demands

SCREENING CRITERIA (Fitness and Movement)		REPETITIONS (Assess capacity to perform reps - STRENGTH)					WORK (Assess capacity to perform for time - ENDURANCE)					TEMPO (Assess capacity to perform at tempo - SPEED)				
		SQUAT	LUNGE	HINGE	PUSH	PULL	SQUAT	LUNGE	HINGE	PUSH	PULL	SQUAT	LUNGE	HINGE	PUSH	PULL
FITNESS STANDARDS		15 REPS / SIDE					30 SECONDS CONTINUOUS					10 REPS / 15 SECONDS				
FITNESS Fitness Score (E.g. 15 reps per side)																
MOVEMENT STANDARDS		7 KEY FEATURES					7 KEY FEATURES					7 KEY FEATURES				
CONTROL Knee Control (Aligned with hips and feet)																
CONTROL Back Control (Normal low back curve)																
CONTROL Shoulder Control (Down and back)																
ASSESSMENT SCORE (Total number of challenges completed)		FITNESS SCORE (Number of times the fitness and movement standards were met)					PATTERN SCORE (Number of times the fitness and movement standards were met for each of the 5 patterns)					KEY FEATURE SCORE (Number of times movement standards were met for each key feature)				
FITNESS ONLY (/15)		REPS (/5)					SQUAT (/3)					KNEE (/9)				
COMBINED (/15)		WORK (/5)					LUNGE (/3)					BACK (/15)				
		TEMPO (/5)					HINGE (/3)					SHOULDER (/6)				
							PUSH (/3)									
							PULL (/3)									

Figure 1. Sample of the Screening and Assessment Sheet for Physical Literacy Screen™, with the Key Feature Score Circled. Retrieved from Patterns & Features, KIN 204 Lecture Content.

3.2 Participants

Recruited raters: Twenty-two participants (8 male, 14 female) were recruited through a convenience sample. Eligible participants had either (1) completed KIN 204: Movement Assessment and Exercise Prescription within the past two years (since Winter 2024), or (2) administered the Physical Literacy Screen (PLS) within the past two years. All participants provided signed informed consent prior to participating. Participants received a \$25 Amazon gift card upon completing both scoring sessions. Renumeration was documented using the Self-Declaration Form (see Appendix A).

Sample Size Justification: The sample size was informed by previous movement-based assessment studies and agreement statistic guidelines. Previous movement-based assessment studies, such as studies evaluating the intra- and inter-rater reliability of the FMS, have employed a sample size ranging from 3 to 38 raters and 3 to 64 participants, with intra-rater sessions conducted anywhere from 48 hours to 1 week apart (Cuchna et al., 2016). With 22 raters and 32 movers (called participants in the cited previous studies), this study falls within prior sample sizes. Each mover was assessed by approximately three to four raters, providing a strong base for inter-rater analysis.

3.3 Experimental Protocol

Video Recording Procedure: A video bank of previously recorded PLS sessions was accessed, including footage of 32 athletes completing 14 movement tasks (excluding handstand). Each video originally included two movers, captured using frontal and sagittal views from a front camera and two side-view cameras on tripods. However, for the purpose of this study, the footage was cropped to isolate a single mover per video, as shown in Figure 2. The PLS was performed as per standard guidelines (Appendix B).



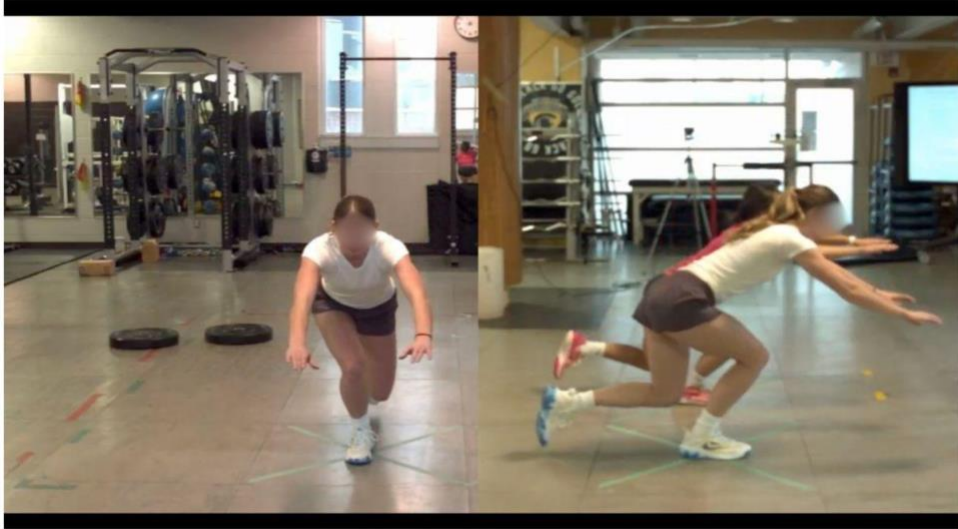
Figure 2. Cropped Video Display Showing One Athlete Performing the Lunge as part of the PLS, from a Frontal and Sagittal View.

Videos were made available as a rating quiz via Microsoft Forms, with each rater receiving a unique quiz link. Raters completed the task independently, adhering to the University's code of conduct to maintain confidentiality and minimize communication between raters. To control for potential biases and order effects, video sequences were randomized, and raters were blinded to their own scores and those of others to avoid recall bias (Cuchna et al., 2016).

Scoring Procedure: Each rater evaluated five movers per session, reviewing the videos based on key features. Raters evaluated movements based on specific PLS scoring criteria by checking or leaving the "control" box blank. Frontal and sagittal views were presented simultaneously for each exercise, and the name of the exercise was specified for each video to help orient the rater. Each key feature check box was marked with a "✓" if successfully controlled on three consecutive repetitions on the same side or with a blank box if control was not achieved (Figure 3). This evaluation was applied across all key features within each of the five movement patterns in the three domains. For the purposes of this study, intra- and inter-rater reliability focus on the scores for each key feature totaled across domains, resulting in a total of three observations for each key feature. Each rater took approximately 30-60 minutes to score for five movers; however, the quiz was made available for 90 minutes to account for buffering and orientation.

1

One Leg Squat (check if Control is achieved)



☐ Knee Control (Aligned with hips and feet)
☐ Back Control (Normal low back curve)

Figure 3. Example of Microsoft Forms Quiz Layout for PLS Rating Task, Displaying the One Leg Squat Exercise with Control Checkboxes for Knee and Back

Intra-Rater Reliability Protocol: Each rater scored the same five movers at two separate time points, with a minimum of 14 days between sessions to minimize recall bias. Videos were re-randomized for the second session, and each rater was blinded to both prior scores and those of others to ensure independent, objective scoring.

Inter-Rater Reliability Protocol: Each of the 22 raters rated five movers with an overlap among raters, to ensure each mover is scored by an average of 3-4 different raters. Specifically, each rater assessed five athletes, and these assessments overlapped with the assessments of 3-4 other raters per athlete, creating redundancy that ensures consistency across raters can be measured.

3.5 Statistical Analysis

Intra-Rater Reliability: Cohen's Kappa coefficient was employed to assess the intra-rater reliability, which is the consistency of the same rater's score at different time points (initial and follow-up). Cohen's Kappa measures the degree of agreement, adjusting for the possibility of

agreement occurring by chance (Cohen, 1960). A Kappa value was computed for each key feature (Knee, Back, Shoulder) across raters. To do this, initial and follow-up scores were paired for each unique combination of athlete, movement pattern, and domain, and Cohen's Kappa was calculated for each rater-feature pair.

Inter-Rater Reliability: Fleiss' Kappa coefficient, an adaptation of Cohen's Kappa, was employed to assess inter-rater reliability, which is the consistency of scores among different raters. Fleiss' Kappa is suitable for nominal data where multiple raters (≥ 2) score the same item and adjusts for agreement expected by chance (Fleiss, 1971). A frequency matrix was constructed for each key feature where rows represented unique movement instances (i.e., combinations of athlete, pattern, and domain), and columns represented the number of raters assigning each score (e.g., control = 1, no control = 0). Fleiss' Kappa was then calculated using these frequency tables.

The formula for Cohen's and Fleiss' Kappa can be computed as follows: $K = \frac{P(o) - P(e)}{1 - P(e)}$, where $P(o)$ denotes the proportion of observed agreement, and $P(e)$ denotes the proportion of chance agreement.

In addition to Kappa coefficients, two descriptive statistics were calculated to support interpretation of agreement. Percent agreement (P_o) was computed as the proportion of matched ratings between sessions (intra-rater) or across raters (inter-rater), calculated as the total number of matching scores divided by the total number of ratings. Category proportions were also calculated to determine the distribution of 0 and 1 scores for each key feature. These proportions indicate whether scores were heavily skewed toward one category, a pattern known to influence Kappa estimates and contribute to suppressed or negative values even when raw agreement is high. Including percent agreement and category prevalence provides essential context for interpreting Kappa values and clarifies whether low coefficients reflect true disagreement or imbalanced scoring distributions.

Statistical analysis was conducted in RStudio (RStudio Inc., Boston, MA, USA), using the *dplyr* and *irr* packages (Hallgren, 2012). Datasets were structured into appropriate formats for each reliability analysis (paired matrices for intra-rater, frequency matrices for inter-rater). Outputs included Cohen's Kappa values for each rater-feature pair (intra-rater reliability), Fleiss' Kappa values for each feature (inter-rater reliability), and the number of scoring instances contributing to each estimate. For intra-rater reliability, outputs were also used to calculate the average Kappa across raters per feature, providing a broader summary of rating consistency. For inter-rater reliability, additional outputs included the number of movement instances rated and the average number of raters

per item. These outputs provide insight into the reliability of visually judging key movement features within the PLS framework at both the individual and group level.

To ensure methodological accuracy, the statistical analysis plan and R code were reviewed in consultation with a statistics specialist from the University of Waterloo Statistics Help Desk (Statistical Consulting and Survey Research Unit, SCSRU). The consultation confirmed that the selected reliability measures (Cohen's and Fleiss' Kappa) and analytical procedures were appropriate for the study design and data structure.

The results were classified using the scale by Landis and Koch (1977), developed for the interpretation of kappa values of observer agreement for categorical and ordinal data: 0.81-1.0: almost perfect reliability; 0.61-0.80: substantial reliability; 0.41-0.60: moderate reliability; 0.21-0.40: fair reliability; 0.01-0.20: slight reliability; <0.00: poor reliability.

Chapter 4 Results

4.1 Category Proportions

Across all ratings in the dataset, there were 1,742 instances scored as 0 (“no control”) and 4,414 instances scored as 1 (“control”). This distribution indicates that raters selected the “control” category approximately 72% of the time, resulting in a skewed prevalence of scores. This skewed distribution reflects the overall high movement competency of the varsity athlete sample and demonstrates how uneven category use can inflate expected chance agreement (P_e) and suppress Kappa values, even when raw agreement is moderate.

4.2 Intra-Rater Reliability

Intra-rater reliability was assessed using Cohen’s Kappa. Each rater assessed the same five athletes twice, with a minimum 14-day interval between assessments. Scores were paired at the level of each key feature (Knee, Back, Shoulder) across all domains (strength, speed, endurance). Cohen’s Kappa was computed for each rater-feature pair where both time points were available.

The average values across all raters were as follows:

- Knee Control: $\kappa = 0.415$ (SD = 0.216, $n = 22$ raters), indicating moderate reliability,
- Back Control: $\kappa = 0.472$ (SD = 0.175, $n = 22$ raters), indicating moderate reliability,
- Shoulder Control: $\kappa = 0.284$ (SD = 0.312, $n = 22$ raters), indicating fair reliability.

These results indicate that raters scored Knee and Back features more consistently than Shoulder features. The Shoulder also demonstrated greater variability across raters compared to the Knee and Back, as reflected in the higher standard deviation (Figure 4).

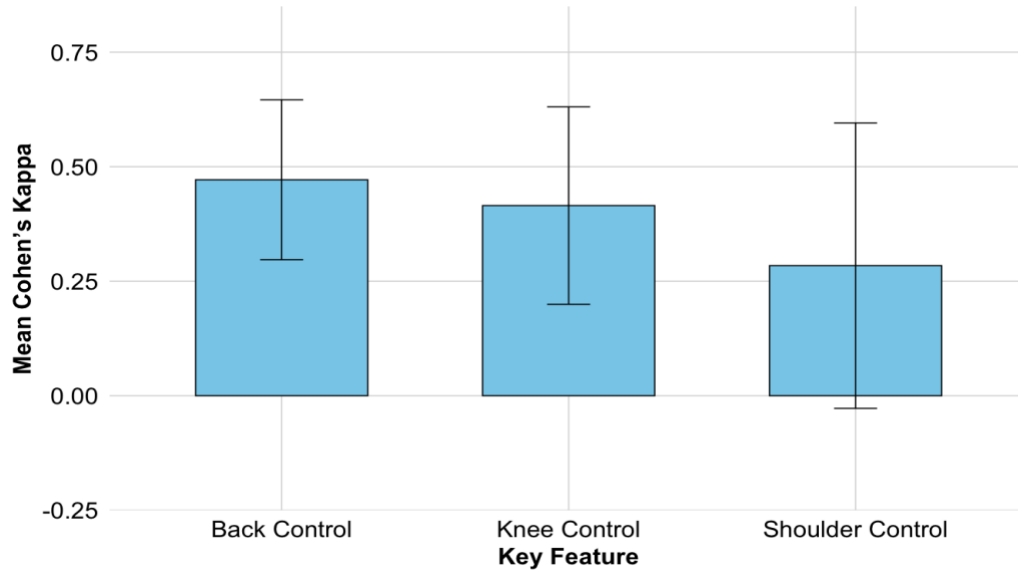


Figure 4. Intra-Rater Agreement on Knee, Back, and Shoulder Control Features (Mean $\pm$ SD Cohen's Kappa)

To provide additional transparency into the raw rating data, a sample of raw rater scores for one athlete across two assessment attempts is included in Appendix D. This sample illustrates how an individual raters' decision differed across sessions and how those differences contributed to the observed Kappa values.

4.2.1 Intra-Rater Percent Agreement

Percent agreement (Po) was calculated to complement Kappa and describe how often raters reproduced their own scores across sessions. The average percent agreement for Knee Control was 66.8% (SD = 37.0%), with a 95% confidence interval of [61.2%, 72.5%]. Shoulder Control showed a similar mean agreement of 63.6% (SD = 39.3%), with a confidence interval of [55.6%, 71.6%]. Back Control had the lowest agreement at 60.2% (SD = 36.9%), with a confidence interval of [55.7%, 64.7%].

These findings suggest that raters generally showed high consistency when re-evaluating the same movement trials after a 14-day interval. Although Shoulder Control exhibited more variation in rater consistency, the overlap in confidence intervals indicates that average intra-rater agreement was broadly similar across all three features.

4.3 Inter-Rater Reliability

Inter-rater reliability was assessed using Fleiss' Kappa. For each key feature, a frequency matrix was constructed where rows represented unique movement instances (athlete $\times$ movement pattern $\times$ domain) and columns captured how many raters assigned each category (control = 1, no control = 0). Only movement instances with at least two rater scores were included.

The calculated values for each feature were:

Table 1. Inter-Rater Reliability (Fleiss' Kappa) by Key Feature

Feature	Mean κ	Movement Instances (N)	Average Rater Count (n)	Reliability Classification
Knee Control	-0.089	288	3.44	Poor
Back Control	-0.096	448	3.44	Poor
Shoulder Control	-0.076	160,	3.43	Poor

Note. Kappa values are interpreted based on the Landis and Koch (1977) scale, where <0.01 = moderate reliability and 0.21 - 0.40 = fair reliability.

All three features demonstrated low agreement, with negative Kappa values suggesting disagreement greater than expected by chance. However, given the highly skewed distribution of scores (high proportion of 1s), negative Kappa values likely reflect prevalence effects, not necessarily disagreement. When one category is overwhelmingly selected, Kappa is known to underestimate agreement even when raters frequently match.

4.3.1 Inter-Rater Percent Agreement

For inter-rater comparisons, the average percent agreement was computed across all rater pairs for each movement feature, alongside measures of variability. Agreement was highest for Knee Control ($M = 66.8\%$, $SD = 37.0\%$, $95\% \text{ CI } [61.2\%, 72.5\%]$), followed by Shoulder Control ($M = 63.6\%$, $SD = 39.3\%$, $95\% \text{ CI } [55.6\%, 71.6\%]$), and Back Control ($M = 60.2\%$, $SD = 36.9\%$, $95\% \text{ CI } [55.7\%, 64.7\%]$). While the mean values reflect moderate levels of agreement, the high standard deviations suggest substantial variability in how consistently raters agreed across different movement instances.

The combination of moderate inter-rater percent agreement and highly uneven category distributions helps to explain the low and negative Fleiss' Kappa values observed for all three

features. Because Kappa is sensitive to category prevalence, heavy use of a single category (e.g., frequent “control” scores) can substantially reduce Kappa even when raters agree a majority of the time. In this context, the percent agreement and category proportions suggest that the low Kappa values are at least partly driven by prevalence effects rather than solely by widespread disagreement between raters.

4.4 Summary of Findings

Intra-rater reliability was generally higher for Knee and Back features compared to Shoulder, suggesting that raters were more consistent when evaluating these features over time. Inter-rater reliability was lower, with negative Fleiss’ Kappa values across all features despite moderate percent agreement ranging from 58.7% to 71.1%. This pattern suggests that while raters matched each other on more than half of all scoring opportunities, the highly skewed distribution of scores, where “control” ratings dominated, elevated expected chance agreement and suppressed Kappa values.

Together, these findings indicate that intra-rater consistency was reasonable, but inter-rater consistency was limited. Percent agreement and category proportions provide important context by showing that low Kappa values likely reflect prevalence effects rather than solely poor rater agreement.

Chapter 5 Discussion

5.1 Summary and Interpretation of Findings

This study examined the reliability of visually judging three key movement features, knee, back, and shoulder control, from PLS-derived video tasks scored by trained kinesiology students. Hypothesis 1 predicted substantial intra-rater agreement ($\kappa = 0.61-0.80$). This hypothesis was not supported as Knee and Back showed only moderate agreement ($\kappa \approx 0.41$ and 0.47), and Shoulder showed fair agreement ($\kappa \approx 0.28$). However, percent agreement (P_o) showed that raters reproduced their own scores in roughly 78-80% of cases across all features, indicating higher raw agreement than the Kappa values alone suggest. This pattern is consistent with literature showing that Kappa may underestimate reliability when category distributions are unbalanced.

Hypothesis 2 predicted moderate inter-rater agreement ($\kappa = 0.41-0.60$). This hypothesis was not supported. Fleiss' Kappa values were negative across all features despite moderate inter-rater percent agreement ($\approx 59-71\%$). The discrepancy between raw agreement and Kappa reflects the prevalence paradox, where a highly skewed distribution of scores inflates expected chance agreement (P_e) and suppresses Kappa values. In this sample, “control” ratings comprised approximately 70% of all inter-rater scores, with the imbalance most pronounced for Shoulder control (78% of 1s). Because Kappa penalizes agreement on the majority category, even moderate levels of matching can produce very low or negative coefficients. Therefore, the negative Kappa values should not be interpreted as evidence of rater disagreement worse than chance, but that Kappa is being distorted by the imbalanced scoring distribution and the subtle, difficult-to-observe features.

A consistent theme was that features with clearer visual cues (e.g., knee control) produced higher reliability than features with subtler or less easily observed criteria (e.g., scapulae alignment). For example, one athlete often shifted her foot's center of pressure during repetitions. Although this behaviour reflects a loss of Knee control by PLS criteria, the variability in rater scoring suggests that not all instances of this behaviour were identified in the same way (Figure 5). This pattern aligns with foundational work showing that raters more reliably detect gross deviations than subtle segmental movements (Knudson & Morrison, 2002).



Figure 5. Frontal and Sagittal Views of Participant Performing a One Leg Hip Bridge with Shifting of Foot's Centre of Pressure that May Have Increased Rater Difficulty in Scoring Knee Feature Control

Similarly, another athlete wore a loose shirt that obscured scapular and trunk alignment from the posterior view (Figure 6), which limited raters' ability to evaluate shoulder positioning directly. This challenge is consistent with research showing that upper-body posture is especially difficult to assess visually because clothing, occlusion, and small angular changes in the thoracic and scapular regions reduce visibility and perceptual clarity (Covalla, 2003). Even trained raters often misestimate scapular elevation, protraction, or trunk rotation when anatomical landmarks are not clearly visible or when only frontal and sagittal perspectives are available.

To address some of this variability, the PLS scoring criteria could be refined to better distinguish between clear control and partial control, similar to recommendations in qualitative movement analysis literature (Knudson, 2013). For features like shoulder and scapular alignment, clearer operational definitions or visual reference points, such as marking key anatomical landmarks (e.g., acromion, scapulae), could help raters identify movement deviations that may otherwise be missed, as suggested by ergonomics and posture-assessment research (Kee & Karwowski, 2007; Chiasson et al., 2012). Similarly, for knee and foot control, the scoring guide could specify the acceptable range of movement at the center of pressure, distinguishing between small, natural weight shifts and loss of stability. These refinements would reduce ambiguity when raters evaluate borderline performances and improve the consistency of visual judgments for features that are inherently more difficult to evaluate.



Figure 6. Frontal and Sagittal Views of Participant Performing an Inverted Row with a Loose-Fitting Shirt that May Have Increased Rater Difficulty in Scoring Shoulder Feature Control

Another possible improvement is the introduction of a graded scoring system, rather than a strict pass/fail approach. The binary nature of the PLS scoring system (“√” or “X”) is intentionally designed as an all-or-nothing approach, where a feature is only marked as controlled if all criteria are satisfied on three consecutive repetitions. While this standardization is clear, it eliminates gradation and may mask nuance, such as partial control or compensatory strategies. Similar concerns have been raised in other visual assessments studies. In ergonomics and clinical movement analyses it was shown that binary scoring tends to magnify disagreement and reduce reliability because borderline performances force raters into categorical decisions that may not reflect observed variability (Covalla, 2003). Therefore, in practice, raters may have been inclined to select “pass” when the movement appeared competent overall, even if one of the required criteria was not fully met. The absence of a scoring range to capture “almost controlled” performances could have contributed to variability in how consistently raters applied the criteria.

A three-point scale (e.g., “fully controlled,” “partially controlled,” “uncontrolled”) could capture nuance and better represent real-world variability in movement quality. Such a system might improve reliability by reducing the need for raters to make overly binary judgments in borderline cases. Research in qualitative movement analysis supports the use of multi-level scoring for features that vary subtly across repetitions because it allows raters to classify small deviations without forcing

an overly binary judgment (Knudson & Morrison, 2002). However, graded systems can introduce subjectivity if intermediate categories are not clearly defined. To prevent this, clear operational definitions and example videos for each level would be essential, as explicit reference standards have been shown to improve consistency in visual assessments (MacLachlan & White, 2015).

Although all raters had received formal instruction in the PLS through coursework, structured calibration sessions could further improve scoring consistency by ensuring that trained raters interpret and apply the criteria in the same way. Calibration differs from basic training in that it emphasizes shared interpretation rather than procedural knowledge. For instance, consensus-based video reviews, where raters collectively evaluate sample performances and discuss ambiguous cases, could help clarify what constitutes “control” versus “no control” across borderline examples. To verify readiness before data collection, a short calibration quiz could also be administered, where rater scores are compared against those of an expert evaluator. This would allow raters to demonstrate sufficient accuracy and consistency before participating in scoring tasks, which has been shown to reduce variability in visual movement evaluations (Whatman et al., 2012). Collectively, such calibration procedures could enhance both the clarity and reliability of the PLS scoring system, improving its practical application across educational, athletic, and occupational contexts. They would also strengthen confidence in the tool’s ability to detect meaningful differences in movement competency while maintaining feasibility as a visual screening method.

5.2 Implications and Generalizability

This study represents the first formal evaluation of the reliability of visually judging key movement features within the PLS framework and highlights its potential and limitations. The results support the use of the PLS for within-rater applications, such as tracking changes over time when the same individual conducts assessments. Percent agreement values close to 80% indicate that raters were generally able to reproduce their own judgments, suggesting that the framework can be applied consistently at the individual rater level. However, the poor inter-rater agreement raises concerns about its use across multiple evaluators in educational settings, research teams, sport organizations, or occupational contexts. This emphasizes the importance of refining criteria for clearer operational definitions and structured rater calibrations. These findings align with previous visual-based ergonomic assessments such as RULA, REBA, and OWAS, where reliability improves with rater training and experience but remains variable when criteria are ambiguous or task complexity is high

(Kee & Karwowski, 2007). Similar trends are well documented across clinical movement analysis, gait assessment, and posture evaluation, where observer agreement depends strongly on both the clarity of visual cues and the amount of shared training among raters (Knudson & Morrison, 2002). As with those tools, the effectiveness of the PLS will depend on how consistently raters can interpret and apply its scoring standards.

In practical terms, the PLS offers more structure than many commonly used movement assessments, often relying on vague descriptors like “good form” or “controlled.” By providing explicit pass/fail criteria, the PLS can potentially reduce ambiguity, but only if raters are prepared to apply those criteria consistently. For coaches, clinicians, rehabilitation practitioners, educators, and occupational professionals, this emphasizes the importance of training and calibration protocols before implementing the PLS at scale. These strategies reflect effective calibration practices in movement assessment, where shared interpretation and structured training significantly improve inter-rater agreement (Whatman et al., 2012).

The generalizability of these results depends on the populations studied. The raters were kinesiology students trained through a university course (KIN 204). While this reflects the level of user likely to employ the PLS in educational or sport environments, it limits generalizability to other professionals such as clinicians, teachers, or occupational assessors who have not taken the course. For instance, the Physical Literacy Screen is commonly used with firefighters trained by Performance Redefined and applied to occupational settings. Similarly, the mover sample consisted exclusively of varsity volleyball athletes. As a highly trained and physically active group, these athletes likely demonstrated greater overall movement competency than the general population. Their high fitness levels and regular exposure to the five foundational movement patterns emphasized in strength and conditioning programs (squat, hinge, lunge, push, and pull) may have contributed to consistently higher-quality movement performances where “control” was more frequently selected. This could reduce variability in observed errors and potentially inflate percent agreement while simultaneously amplifying the prevalence paradox observed in the Kappa statistics, relative to what might be expected in more heterogeneous populations.

Finally, this study relied exclusively on video-based assessments. While videos allowed for standardization and repeated scoring, they restricted the visual information available compared to live evaluations. Raters were limited to one sagittal and one frontal perspective, which prevented them from detecting asymmetries or subtle postural deviations that might have been visible in live

assessments. This is consistent with research showing that limited viewpoints impair detection of subtle deviations in spine, hip, and shoulder alignment (MacLachlan & White, 2015). Clothing, occlusion, and reduced depth cues further complicate visual evaluation, as widely reported in posture and movement observation studies (Covalla, 2003). The original recordings included two athletes performing the tasks simultaneously, and although the footage was cropped to isolate a single mover, the second participant was still occasionally visible in the background or along the frame's edge. While this did not obscure the main athlete's movement, it may have introduced slight visual distractions or influenced rater focus. In applied environments, assessors often move around the athlete to view movements dynamically from multiple angles. These contextual differences underscore the importance of interpreting video-based reliability findings with caution.

5.3 Limitations

This study faced several methodological and technical limitations that likely contributed to variability in results. An unexpected technical issue arose partway through data collection due to institutional changes in cloud storage policy. The University of Waterloo implemented OneDrive storage restrictions midway through the study, where all video files for the scoring quizzes were originally hosted and embedded within Microsoft Forms. As a result, some participants encountered video playback issues, including buffering and failures to load. All videos were manually migrated to an unlisted YouTube playlist to address this, ensuring privacy and accessibility via direct links. This required relinking over 400 videos and redistributing new scoring forms. Consequently, the first eight scorers used the original OneDrive-based forms, while the remaining scorers were transitioned to the YouTube-based version. During the YouTube phase, an additional delay occurred when the account hosting the videos was unexpectedly suspended and flagged for spam-like activity. Although the suspension was successfully appealed and access was regained, this disruption temporarily prevented several participants from completing their scoring tasks. The content and instructions were consistent across platforms; however, differences in streaming quality or user familiarity with the platform may have influenced the rating experience.

The video presentation itself may also have impacted scoring accuracy. Although frontal and sagittal views were provided, these limited perspectives constrained raters' ability to detect subtler features such as scapular control, transverse-plane trunk motion, and asymmetries not visible from fixed positions. These limitations align with past studies showing that visual movement evaluation is

sensitive to the number and orientation of viewing angles, with constrained perspectives producing lower accuracy and reliability (Knudson & Morrison, 2002). Moreover, only one side of the sagittal plane was captured, potentially obscuring left-right asymmetries. Video resolution, playback settings, and viewing behaviours (such as whether raters used full screen or 1080p) likely varied and contributed to additional inconsistency. These factors are known sources of perceptual error in observational assessments and could have contributed to lower agreement, particularly for subtle upper-body features.

Another factor that may have limited rater agreement was the absence of a standardized reference frame for each athlete's neutral posture. Raters were required to judge deviations in spine, shoulder, and knee control without first observing each mover's natural standing alignment. Research in posture and qualitative movement analysis shows that raters rely heavily on individualized baselines to detect subtle changes in curvature, scapular positioning, and weight distribution (Knudson & Morrison, 2002; Covalla, 2003). Without knowing an athlete's typical spine shape, such as their natural lumbar lordosis or thoracic curvature, raters may have disagreed on whether a perceived deviation represented a true loss of control or simply the mover's normal posture. The lack of an initial reference pose likely disproportionately affected features like back and shoulder control, which depend on detecting relatively small changes from baseline. Incorporating a standardized stance trial at the start of scoring sessions could provide raters with a clearer visual reference, potentially improving reliability for these subtler features.

Rater-related factors further limited generalizability. All raters were kinesiology students from a single course (KIN 204). Although they had foundational training in movement evaluation, they were novice judges compared to clinicians, certified strength and conditioning professionals, or ergonomic assessors. Numerous studies have shown that observer expertise significantly influences reliability in gait, posture, and functional movement analysis (Whatman et al., 2012; Chiasson et al., 2012). Participation was voluntary and incentivized with a gift card, so levels of engagement, attentiveness, and motivation likely varied. Guidelines were provided (e.g., stable internet connection, uninterrupted viewing), but adherence could not be verified. The speed at which raters completed their quizzes ranged considerably (see Appendix E for average completion times), and it is unclear whether faster completions reflected efficiency, overconfidence, or reduced attentiveness. Exploratory analysis did not reveal any apparent relationship between completion time and intra-rater reliability, suggesting that differences in pacing alone did not strongly influence consistency.

However, the reliance on video-only assessment prevented the study from capturing how reliability might differ in live settings, where assessors can reposition themselves to obtain optimal angles, a practice known to improve detection of multi-planar deviations (Knudson, 2013).

The choice of reliability metrics introduced another constraint. While Cohen's and Fleiss' Kappa are appropriate for categorical data, they are sensitive to prevalence and bias effects. In this study, scores were predominantly in the "control" category, consistent with the high movement competency expected of the varsity athlete sample. When one category dominates, the expected chance agreement (P_e) becomes artificially inflated, which reduces the numerator of the Kappa statistic ($P_o - P_e$) even when observed agreement (P_o) is high. This phenomenon, known as the Kappa paradox, results in suppressed or negative Kappa values even in datasets with moderate to high raw agreement (Derksen et al., 2024). In the present study, this paradox likely contributed to negative Fleiss' Kappa values, which may misleadingly suggest disagreement worse than chance. The combination of a highly skilled mover sample and a binary scoring system further skewed the distribution of scores, amplifying these effects. Consequently, the reported Kappa coefficients should be interpreted cautiously, recognizing that they may underestimate the true level of agreement due to the distribution of scores rather than inconsistency among raters.

5.4 Recommendations for Future Research

Future research should focus on evaluating the reliability and validity of the PLS under more ecologically valid conditions. Live testing environments, where assessors can dynamically reposition themselves, may produce different reliability outcomes than video-based methods (Knudson & Morrison, 2002). Investigating reliability during live assessments would clarify whether the limitations observed in the present study are inherent to the PLS or are a consequence of fixed-camera video constraints.

Studies should also examine the role of rater background by comparing novice, intermediate, and expert raters across professions such as clinical therapy, strength and conditioning, and occupational ergonomics. Prior research consistently demonstrates that expert raters achieve higher accuracy and agreement than novices in gait, posture, ergonomics, and sport technique evaluation (Whatman et al., 2012; Chiasson et al., 2012). Including a more diverse rater pool would not only enhance the generalizability of findings but also clarify whether professional experience or discipline-specific training leads to more consistent interpretations of movement control. This is particularly

important because the PLS is intended for broad use across educational, athletic, and occupational settings, where rater expertise and interpretive tendencies may vary substantially.

Another priority is to establish the validity of PLS feature scoring by comparing visual ratings to objective biomechanical reference measures, such as 3D motion capture, wearable sensors, or force plates. Doing so would help differentiate between reliability (consistency among raters) and validity (accuracy relative to an objective standard). Studies such as MacLachlan & White (2015) suggest that observers systematically underestimate or misclassify joint positions compared with instrumented data. Validating the PLS against objective standards would therefore help determine whether the tool captures movement control with sufficient precision for applied settings.

Future work should explore alternative scoring models. Because binary scoring is highly susceptible to prevalence effects and magnifies the Kappa paradox, graded scales (e.g., a 3-level or 4-level rating system) may better capture the continuum of movement quality and reduce forced decisions in borderline cases. However, these systems require clear operational definitions and validated exemplar videos to prevent increases in subjectivity.

Additionally, a factor or regression analysis of PLS tasks could identify whether all 15 tasks are necessary or whether a smaller subset of key movements could provide comparable reliability and efficiency for field applications. This may increase feasibility and improve consistency by focusing attention on the tasks most sensitive to movement deviations.

Finally, including mover populations beyond varsity-level athletes will be essential for understanding the reliability of PLS feature scoring in more heterogeneous groups. Including recreational athletes, youth populations, adults with varied physical literacy, and occupational groups would introduce greater variability in movement behaviours and reduce the category imbalance observed in this study. This may mitigate the Kappa paradox and provide a clearer picture of how reliably PLS features can be judged under different movement competencies.

By addressing these research directions, future studies can refine both the methodology and the scoring framework of the PLS, enhancing its utility as a standardized movement screening tool across sport, health, and occupational settings.

Chapter 6 Conclusions

This thesis investigated the intra- and inter-rater reliability of visually judging three key movement features (knee, back, and shoulder control) within the Physical Literacy Screen™ (PLS) using standardized video recordings. The study addressed two key research questions: (1) the consistency of scores when raters reassessed the same performances after a delay (intra-rater reliability), and (2) the extent to which different raters agreed when scoring the same performances (inter-rater reliability).

6.1 Summary of Findings

The results showed that intra-rater reliability was moderate for Knee and Back features and fair for Shoulder control. Percent agreement values were 79.9% for Back, 80.4% for Knee, and 78.3% for Shoulder, supporting the observed Kappa results. This suggests that individual raters could reproduce their judgments across time but not at the level anticipated. Inter-rater reliability was poor across all features, with Fleiss' Kappa values below zero, indicating substantial disagreement between raters. Inter-rater percent agreement values ranged from 58.7% (Back) to 66.5% (Knee) and 71.1% (Shoulder). This discrepancy reflects the prevalence paradox, where high category imbalance inflates expected chance agreement and suppresses Kappa values even when raw agreement is moderate. These findings suggest that while the PLS can provide stable results when used by the same evaluator, its application across multiple raters remains limited.

6.2 Contributions and Implications

This study provides the first formal reliability evidence for PLS feature scoring and contributes to the broader understanding of the challenges associated with visual movement assessment. The results contribute to the broader discussion of visual-based assessments by highlighting the challenges of scoring subtle or obscured features, such as scapular positioning and foot center of pressure, which are especially vulnerable to rater disagreement. The study also reinforces insights from ergonomics, clinical gait analysis, and movement screening literature by showing that subtle or partially occluded features, such as scapular positioning, small center-of-pressure shifts, and minor trunk deviations, are particularly difficult to judge consistently. These challenges underscore the need for clearer operational definitions and structured calibration procedures to support consistent scoring.

6.3 Future Directions

Further research should clarify the criteria for difficult-to-observe features such as shoulder control. Incorporating structured calibration sessions or consensus scoring protocols could also enhance inter-rater agreement. Additionally, future studies should explore the reliability of live administration in real-world contexts, where assessors can reposition themselves and gain access to multiple viewing angles that are not available in fixed-camera video recordings. Additional research comparing PLS feature scores with biomechanical reference standards, such as motion capture or wearable sensors, will help clarify whether observed disagreement reflects true ambiguity in the movement or limitations of visual perception. Expanding reliability testing to more diverse populations with a broader range of movement competency may offer a clearer understanding of how reliably PLS features can be judged across different movement contexts.

In summary, this thesis establishes the first reliability evidence for the visual judgment of three key movement features, knee, back, and shoulder control, when assessed using PLS-derived video tasks. The findings show that these features can be evaluated with reasonable consistency by the same rater over time but that agreement between different raters remains limited. By addressing these limitations, the PLS has the potential to become a more robust tool for assessing movement competency in sport, education, and health contexts.

References

- Armstrong, D. P., Ross, G. B., Graham, R. B., & Fischer, S. L. (2019). Considering movement competency within physical employment standards. *Work (Reading, Mass.)*, 63(4), 603–613.
<https://doi.org/10.3233/WOR-192955>
- Balyi, I., Way, R., Norris, S., Cardinal, C. & Higgs, C. (2005). *Canadian sport for life: Long-term athlete development resource paper*. Vancouver, BC: Canadian Sport Centres.
- Bonazza, N. A., Smuin, D., Onks, C. A., Silvis, M. L., & Dhawan, A. (2017). Reliability, Validity, and Injury Predictive Value of the Functional Movement Screen: A Systematic Review and Meta-analysis. *The American Journal of Sports Medicine*, 45(3), 725–732.
<https://doi.org/10.1177/0363546516641937>
- Chiasson, M. É., Imbeau, D., Aubry, K., & Delisle, A. (2012). Comparing the results of eight methods used to evaluate risk factors associated with musculoskeletal disorders. *International Journal of Industrial Ergonomics*, 42(5), 478–488.
<https://doi.org/10.1016/j.ergon.2012.07.003>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cook, G., Burton, L., & Hoogenboom, B. (2006). Pre-participation screening: the use of fundamental movements as an assessment of function - part 1. *North American Journal of Sports Physical Therapy: NAJSPT*, 1(2), 62–72.
- Covalla, C. (2003). Visual posture observation error and training. *Journal of Manipulative and Physiological Therapeutics*, 26(9), 564–570.

- Cuchna, J. W., Hoch, M. C., & Hoch, J. M. (2016). The interrater and intrarater reliability of the functional movement screen: A systematic review with meta-analysis. *Physical Therapy in Sport*, 19, 57–65. <https://doi.org/10.1016/j.ptsp.2015.12.002>
- Derksen, B. M., Bruinsma, W., Goslings, J. C., & Schep, N. W. L. (2024). The Kappa Paradox Explained. *The Journal of hand surgery*, 49(5), 482–485. <https://doi.org/10.1016/j.jhsa.2024.01.006>
- Dockrell, S., O’Grady, E., Bennett, K., Mullarkey, C., McConnell, R., Ruddy, R., Twomey, S., Flannery, C., & Mulligan, K. (2012). An investigation of the reliability of Rapid Upper Limb Assessment (RULA) as a method of assessment of children’s computing posture. *Applied Ergonomics*, 43(3), 632–636. <https://doi.org/10.1016/j.apergo.2011.09.009>
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. <https://doi.org/10.1037/h0031619>
- Frost, D. M., Beach, T. A. C., McGill, S. M., & Callaghan, J. P. (2015a). The predictive value of general movement tasks in assessing occupational task performance. *Work (Reading, Mass.)*, 52(1), 11–18. <https://doi.org/10.3233/WOR-141902>
- Frost, D. M., Beach, T. A. C., McGill, S. M., & Callaghan, J. P. (2015b). A proposed method to detect kinematic differences between and within individuals. *Journal of Electromyography and Kinesiology*, 25(3), 479–487. <https://doi.org/10.1016/j.jelekin.2015.02.012>
- Frost, D. M., Beach, T. A. C., Callaghan, J. P., & McGill, S. M. (2015c). The Influence of Load and Speed on Individuals’ Movement Behavior. *Journal of Strength and Conditioning Research*, 29(9), 2417–2425. <https://doi.org/10.1519/JSC.0000000000000264>
- Frost, D. M., Beach, T. A. C., Callaghan, J. P., & McGill, S. M. (2015d). FMS Scores Change With Performers’ Knowledge of the Grading Criteria—Are General Whole-Body Movement Screens Capturing “Dysfunction”? *Journal of Strength and Conditioning Research*, 29(11),

- 3037–3044. <https://doi.org/10.1097/JSC.0000000000000211>
- Frost, D. M., Beach, T. A. C., Campbell, T. L., Callaghan, J. P., & McGill, S. M. (2017). Can the Functional Movement Screen™ be used to capture changes in spine and knee motion control following 12 weeks of training? *Physical Therapy in Sport*, 23, 50–57.
<https://doi.org/10.1016/j.ptsp.2016.06.003>
- Hallgren K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34.
<https://doi.org/10.20982/tqmp.08.1.p023>
- Herrington, L.C., & Munro, A.G. (2014). 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* 1(2): 1-6.<http://dx.doi.org/10.15226/2374-6904/1/3/00113>.
- Hewett, T. E., Myer, G. D., Ford, K. R., Heidt, R. S., Jr, Colosimo, A. J., McLean, S. G., van den Bogert, A. J., Paterno, M. V., & Succop, P. (2005). Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *The American Journal of Sports Medicine*, 33(4), 492–501. <https://doi.org/10.1177/0363546504269591>
- Hignett, S., & McAtamney, L. (2000). Rapid Entire Body Assessment (REBA). *Applied Ergonomics*, 31(2), 201–205. [https://doi.org/10.1016/S0003-6870\(99\)00039-3](https://doi.org/10.1016/S0003-6870(99)00039-3)
- Jurbala, P. (2015). What Is Physical Literacy, Really? *Quest*, 67(4), 367–383.
<https://doi.org/10.1080/00336297.2015.1084341>
- Karhu, O., Kansi, P., & Kuorinka, I. (1977). Correcting working postures in industry: A practical method for analysis. *Applied Ergonomics*, 8(4), 199–201. [https://doi.org/10.1016/0003-6870\(77\)90164-8](https://doi.org/10.1016/0003-6870(77)90164-8)
- Kee, D., & Karwowski, W. (2007). A comparison of three observational techniques for assessing

- postural loads in industry. *International Journal of Occupational Safety and Ergonomics*, 13(1), 3–14. <https://doi.org/10.1080/10803548.2007.11076704>
- Knudson, D. V. (2013). *Qualitative diagnosis of human movement: improving performance in sport and exercise*. Human kinetics.
- Knudson, D., & Morrison, C. (2002). *Qualitative analysis of human movement* (2nd ed.). Human Kinetics.
- Kritz, M., Cronin, J., & Hume, P. (2009a). The Bodyweight Squat: A Movement Screen for the Squat Pattern. *Strength and Conditioning Journal*, 31(1), 76–85.
<https://doi.org/10.1519/SSC.0b013e318195eb2f>
- Kritz, M., Cronin, J., & Hume, P. (2009b). Using the Body Weight Forward Lunge to Screen an Athlete's Lunge Pattern. *Strength and Conditioning Journal*, 31(6), 15–24.
<https://doi.org/10.1519/SSC.0b013e3181c1b480>
- Kritz, M., Cronin, J., & Hume, P. (2010). Screening the Upper-Body Push and Pull Patterns Using Body Weight Exercises. *Strength and Conditioning Journal*, 32(3), 72–82.
<https://doi.org/10.1519/SSC.0b013e3181df5cfe>
- Kritz, M. (2012). *Development, reliability and effectiveness of the Movement Competency Screen (MCS)* [Unpublished PhD Thesis]. Auckland University of Technology.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Maclachlan, L., White, S. G., & Reid, D. (2015). Observer rating versus three-dimensional motion analysis of trunk and knee motion during lifting. *International journal of sports physical therapy*, 10(4), 482–492.
- Mann, K. J., O'Dwyer, N., Bruton, M. R., Bird, S. P., & Edwards, S. (2022). Movement Competency Screens Can Be Reliable in Clinical Practice By A Single Rater Using The Composite Score.

- International Journal of Sports Physical Therapy*, 17(4), 593–604.
- McAtamney, L., & Corlett, E. N. (1993). RULA: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24(2), 91–99.
[https://doi.org/10.1016/0003-6870\(93\)90080-S](https://doi.org/10.1016/0003-6870(93)90080-S)
- McCunn, R., Aus der Fünten, K., Fullagar, H. H., McKeown, I., & Meyer, T. (2016). Reliability and Association with Injury of Movement Screens: A Critical Review. *Sports medicine (Auckland, N.Z.)*, 46(6), 763–781. <https://doi.org/10.1007/s40279-015-0453-1>
- McGill, S. (2007). Low back disorders: Evidence-based prevention and rehabilitation (2nd ed.). *Human Kinetics*.
- Moran, R. W., Schneiders, A. G., Mason, J., & Sullivan, S. J. (2017). Do Functional Movement Screen (FMS) composite scores predict subsequent injury? A systematic review with meta-analysis. *British Journal of Sports Medicine*, 51(23), 1661–1669.
<https://doi.org/10.1136/bjsports-2016-096938>
- Pagaduan, J., & Kritz, M. (2022). Interrater and Intrarater Reliability of Movement Competency. *Silliman Journal*, 55(2). Retrieved from
<https://sillimanjournal.su.edu.ph/index.php/sj/article/view/114>
- Performance Redefined. (2019). Assessment and Exercise Design Level 2: Physical Literacy Screen [PDF file]. Retrieved from <https://performanceredefined.ca/wp-content/uploads/2019/10/Physical-Literacy-Manual-Preview.pdf>
- Valentim, D. P., Comper, M. L. C., Sandy Medeiros Rodrigues Cirino, L., da Silva, P. R., Padilha Alonso Gomes, M., Martins da Silva, A., & Padula, R. S. (2025). Observational methods for the analysis of biomechanical exposure in the workplace: a systematic review. *Ergonomics*, 68(10), 1561–1582. <https://doi.org/10.1080/00140139.2024.2427864>

Whatman, C., Hing, W., & Hume, P. (2012). Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. *Physical Therapy in Sport*, 13(2), 87-96

Whatman, C., Hume, P., & Hing, W. (2015). The reliability and validity of visual rating of dynamic alignment during lower extremity functional screening tests: A review of the literature.

Physical Therapy Reviews, 20(3), 210–224.

<https://doi.org/10.1179/1743288X15Y.0000000006>

Whitehead, M. (2001). The Concept of Physical Literacy. *European Journal of Physical Education*,

6(2), 127–138. <https://doi.org/10.1080/1740898010060205>

Appendix A

Consent Forms



Consent Form

Project Title: Exploring the utility of performance testing data to strengthen athlete development.

Researchers: Tyson Beach, Steve Fischer, Devon Frayne, Dilara Sen

Background

- As part of your normal pre-, in-, and post-season training you routinely complete physical testing to inform your athletic development plan.
- Your physical test data are retained by Athletics for use by you (the athlete), your coach, the athletic therapy team, and the strength and condition team at the University of Waterloo.
- Researchers in Department of Kinesiology and Health Sciences (KHS) invite you to allow Athletics to share your physical testing performance data for use in research.

Purpose and Outline of the Research Study

- We are exploring trends in physical testing data between athletes of different sexes, from different sports (as applicable), as a function of age, and as a function of training profile to help inform and enhance athlete development at the University of Waterloo.
- To participate in this study, you do not need to do anything other than consent to allow your physical testing data to be shared with our research team; there will be no additional visits or tests.
- With your consent, we will receive your physical testing data over the course of the year and over the course of your University of Waterloo playing career (i.e., each year that you participate on the team). The data set will include your name, age, year of study, sport, S&C plan, and performance testing results (e.g., movement screen results, counter-movement jump performance, 10-yard dash time, etc.). Please note that performance in some physical tests is recorded by video for post assessment. We will receive the video and it will show your full face until post-processed to de-identify the video. It will remain in an identifiable state for up to two weeks.
- All data will be stored on a secure, WatIAM password protected server hosted by the University of Waterloo.
- Sharing your data is voluntary and you are not under any obligation to do so.
- There will be no impact on your athletic experience if you decide not to share your data and Athletics will not be told who does or does not share their data with KHS.
- You may withdraw your consent at any time by contacting Dr. Steven Fischer (sfischer@uwaterloo.ca). If consent is withdrawn no future data will be shared with the research team and previously shared data will be destroyed where feasible.

**Possible Risks and Benefits**

- You will not receive any direct benefit from sharing your data. However, sharing your data will contribute to research that can help Athletics continuously improve athlete development.
- Although there are security measures in place, there remains a possibility that someone could identify you from the data collected. We take all reasonable precautions including deidentifying your data and storing on a secure password protected server to minimize any risks.

Data Retention Period

- Your data will be retained for a minimum of seven (7) years.

If you have any questions regarding this study or would like additional information to assist you in reaching a decision about participation, please contact Steven Fischer at 519-888-4567, ext. 42368 or by email at steven.fischer@uwaterloo.ca.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB 45663). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca

Figure 7. Consent Form for Sharing Athlete General Performance Testing Data Within the Research Group at the University of Waterloo.



Research Project Consent Form

Project Title: Intra- and Inter-rater Reliability of the Physical Literacy Screening™

Researchers: Dilara Sen, Steve Fischer, Devon Frayne, Tyson Beach

Project Outline:

This research study aims to evaluate the inter- and intra-rater reliability of the Physical Literacy Screen™ in assessing movement competency in varsity volleyball athletes. This involves recording your performance while performing the screen using markerless motion capture and video cameras, then having trained student raters evaluate your movements. The videos will be anonymized using face blurring to protect your identity. The results will help establish the Physical Literacy Screen™ as a reliable tool for physical literacy screening, supporting better performance standards and injury prevention protocols.

Consent:

I have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I understand that I have been asked to allow researchers to access the data that was collected for athletic testing purposes, in order to learn new ways to enhance athlete development.

If you have any questions regarding this study or would like additional information to assist you in reaching a decision about participation, please contact Dilara Sen at dilara.sen@uwaterloo.ca.

By providing your consent, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

I consent to participate in this study.

Name

Signature

Date

Figure 8. Consent Form for the Athlete Use of Data for Thesis on Intra- and Inter-Rater Reliability of the Physical Literacy Screen™.

ONLINE CONSENT FORM

Study Title: Intra- and Inter-rater Reliability of the Physical Literacy Screen™

By providing your consent, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

Conflict of Interest Disclosure:

The Physical Literacy Screen™ was co-founded by Tyson Beach, who is affiliated with Performance Redefined. This affiliation does not influence the conduct or outcomes of this research study.

YES NO I agree to score video recordings of participant movement remotely via Microsoft Forms for this study.

YES NO I agree to participate in two separate scoring sessions for reliability purposes, understanding that I may withdraw at any time without penalty.

YES NO I agree to the use of de-identified data (e.g., my scoring results) in any publications, presentations, or reports arising from this research.

YES NO I agree to allow my scoring data to be used for future research purposes, as described in the information letter.

If you have any questions about this study, please contact Devon Frayne, Email: dfrayne@uwaterloo.ca.

Thank you for your participation!

Name

Signature

Date

Figure 9. Consent Form for the Raters Scoring the Reliability of the Physical Literacy Screen™.

**University of Waterloo
Research Participant's Acknowledgement of
Receipt of Remuneration
and
Self-Declared Income**

Section A: To be completed by Principal Investigator or designate

Principal/Faculty Investigator's Name: Dr. Steve Fischer

Student Investigator(s)'s Name: Dilara Sen

Department: Department of Kinesiology and Health Sciences

Study Title: Intra- and Inter-rater Reliability of the Physical Literacy Screen (PLS)

Section B: To be completed by research participant

In appreciation of my involvement as a research participant in the above study,
I acknowledge that I have received \$ \$25 from the University of Waterloo.

I further acknowledge that:

- this amount received from the University of Waterloo is taxable;
- that it is my responsibility to report the amount received for income tax purposes; and
- the University of Waterloo will not issue a tax receipt for the amount received.

Participant's Name: _____

Participant's Signature: _____

Date: _____

Witness' Name _____

Witness' Signature: _____

Date: _____

Figure 10. Participant Renumeration and Self-Declaration Form

Appendix B

Physical Literacy Screening Instructions and Standards

Table 1. Screening Instructions for Physical Literacy Screen™

Screening for Physical Literacy Screen	
<i>Repetition Based Challenges (Load Demand)</i>	
1. Squat: ONE LEG SQUAT	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Squats to a knee angle of 90° while keeping the heel on the floor REPETITIONS: Performs 15 repetitions with each leg in an alternating fashion KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Stance knee is aligned with the stance hip and foot throughout the motion Note position of knee with respect to line drawn between the stance hip and foot BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature). Note change in relationship between lines drawn through the upper back and hips
2. Lunge: FRONT LUNGE	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Lunges to a front knee angle of 90° while keeping the heel on the floor REPETITIONS: Performs 15 repetitions with each leg in an alternating fashion KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Front knee is aligned with the front hip and foot throughout the motion Note position of knee with respect to line drawn between the front hip and foot BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips
3. Hinge: ONE LEG ROMANIAN DEADLIFT	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Hinges to a hip angle of 90° while keeping the heel on the floor REPETITIONS: Performs 15 repetitions with each leg in an alternating fashion KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Stance knee is aligned with the stance hip and foot throughout the motion Note position of knee with respect to a line drawn between the stance hip and foot BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn

	through the upper back and hips
4. Push: ONE ARM PLANK	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Body held in a push-up position with the hips extended and feet stationary REPETITIONS: Performs 15 repetitions with each arm in an alternating fashion BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips SHOULDER MOTION (Fitness standards must be met to comment on shoulder control, otherwise unknown) CONTROL: The shoulders are held down away from the ears (and scapula do not wing) Note change in relationship between lines drawn through shoulders and ears
5. Pull: ONE ARM ROW	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Body held in a bridge position with the hips extended and feet stationary REPETITIONS: Performs 15 repetitions with each arm in an alternating fashion BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips SHOULDER MOTION (Fitness standards must be met to comment on shoulder control, otherwise unknown) CONTROL: The shoulders are held down away from the ears (and scapula do not wing) Note change in relationship between lines drawn through shoulders and ears
<i>Work Based Challenges (Time Demand)</i>	
1. Squat: JUMP SQUAT	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Descends to a knee angle of 90° while keeping the heels on the floor WORK TIME: Performs continuously for 30s with no pause at the bottom KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Knees are aligned with the hips and feet throughout the motion Note position of knees with respect to lines drawn between the hips and feet BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature). Note change in relationship between lines drawn through the upper back and hips FITNESS (Both criteria must be met to meet the fitness standard) ROM: Lunges to a knee angle of 90° while keeping the heel on the floor WORK TIME: Performs continuously for 30s to left and right side (no pause at bottom or top)
2. Lunge: LATERAL LUNGE	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Lunges to a knee angle of 90° while keeping the heel on the floor WORK TIME: Performs continuously for 30s to left and right side (no pause at bottom or top)

	KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Knee is aligned with the hip and foot throughout the motion Note position of knee with respect to line drawn between the hip and foot BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips
3. Hinge: ONE LEG HIP BRIDGE	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Raises leg to a hip angle of 90° while keeping opposite hip extended WORK TIME: Performs continuously for 30s with left and right leg (no pause at bottom or top) KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Plant knee is aligned with the same side hip and foot throughout the motion Note position of knee with respect to a line drawn between the hip and foot BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips
4. Push: HANDSTAND	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Body held in an inverted position with the hands 1-1.5 feet from the wall WORK TIME: Performs continuously for 30s BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips SHOULDER MOTION (Fitness standards must be met to comment on shoulder control, otherwise unknown) CONTROL: The shoulders are held down away from the ears (and scapula do not wing) Note change in relationship between lines drawn through shoulders and ears
5. Pull: PULL UP HANG	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Body held in a hang position with a prone or neutral grip REPETITIONS: Performs continuously for 30s BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips SHOULDER MOTION (Fitness standards must be met to comment on shoulder control, otherwise unknown) CONTROL: The shoulders are held down away from the ears (and scapula do not wing) Note change in relationship between lines drawn through shoulders and ears

<i>Tempo Based Challenges (Speed Demand)</i>	
1. Squat: AIR SQUAT	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Squats to a knee angle of 90° while keeping the heels on the floor TEMPO: Performs 10 reps in less than 15s 79 KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Knees are aligned with the hips and feet throughout the motion Note position of knees with respect to lines drawn between the hips and feet BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature). Note change in relationship between lines drawn through the upper back and hips
2. Lunge: BACK LUNGE	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Lunges back to a front knee angle of 90° while keeping the heel on the floor TEMPO: Performs 10 repetitions in less than 15s KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Front knee is aligned with the front hip and foot throughout the motion Note position of knee with respect to line drawn between the front hip and foot BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hip
3. Hinge: GOOD MORNING	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Hinges to a hip angle of 90° while keeping the heels on the floor and hands on head TEMPO: Performs 10 repetitions in less than 15s KNEE MOTION (Fitness standards must be met to comment on knee control, otherwise unknown) CONTROL: Knees are aligned with the hips and feet throughout the motion Note position of knees with respect to lines drawn between the hips and feet BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips
4. Push: PUSH UP	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Chest touches floor while hands are shoulder width apart and hips extended TEMPO: Performs 10 repetitions in less than 15s 80 BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips SHOULDER MOTION (Fitness standards must be met to comment on shoulder control, otherwise unknown) CONTROL: The shoulders are held down away from the

	ears (and scapula do not wing) Note change in relationship between lines drawn through shoulders and ears
5. Pull: INVERTED ROW	FITNESS (Both criteria must be met to meet the fitness standard) ROM: Chest touches bar while hands are shoulder width apart and hips extended TEMPO: Performs 10 repetitions in less than 15s BACK MOTION (Fitness standards must be met to comment on back control, otherwise unknown) CONTROL: A neutral low back curvature is maintained (reference standing curvature) Note change in relationship between lines drawn through the upper back and hips SHOULDER MOTION (Fitness standards must be met to comment on shoulder control, otherwise unknown) CONTROL: The shoulders are held down away from the ears (and scapula do not wing) Note change in relationship between lines drawn through shoulders and ears

Table 2. Physical Literacy Screen™ Standards

Physical Literacy Screen Standards	
Task Standards:	Instructions:
<i>Repetition Based Challenges (Load Demand)</i>	
1. Squat: ONE LEG SQUAT • Heel on the floor • Hands off hips • 90° knee angle • Hip extended • Alternate legs 15 reps per leg	• With the shoes on, stand on one leg and place the arms in front or beside the body (not on hips). Keep the heel in contact with the floor at all times. • Squat to a knee flexion angle of 90° and return to standing with the hip fully extended. • Attempt to keep the stance knee in line with the stance hip and foot, and the back in a neutral position (same as standing) throughout the motion. 81 • Perform 15 repetitions on each leg in an alternating manner and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated if the desired depth cannot be achieved on 3 consecutive repetitions (on same leg).
2. Lunge: FRONT LUNGE • Front heel on the floor • Hands off hips • 90° front knee angle	• With the shoes on, stand with the feet hip width apart and place the arms in front or beside the body (not on hips). • Lunge forwards to a front knee flexion angle of 90° and return to standing with the hip fully extended. The front heel must be kept in contact with the floor during the descent phase. • Attempt to keep the front knee in line with the front hip and foot, and the back

• Hip extended • Alternate legs 15 reps per leg	in a neutral position (same as standing) throughout the motion. • Perform 15 repetitions on each leg in an alternating manner and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated if the desired depth cannot be achieved on 3 consecutive repetitions (on same leg).
3. Hinge: ONE LEG ROMANIAN DEADLIFT • Heel on the floor • Hands off hips • 90° hip angle • Hip extended • Alternate legs 15 reps per leg	• Hinge to a hip flexion angle of 90° and return to standing with the hip fully extended. Adopt a slight bend in the stance knee. • Attempt to keep the stance knee in line with the stance hip and foot, and the back in a neutral position (same as standing) throughout the motion. • Perform 15 repetitions on each leg in an alternating manner and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated if the desired depth cannot be achieved on 3 consecutive repetitions (on same leg).
4. Push: ONE ARM PLANK • Hips extended • Hands shoulder width • Feet > shoulder width • Feet stationary • Hand touch shoulder • 15 reps per arm	• With the shoes on, adopt a push-up position with the hands shoulder width apart. The feet must be wider than the shoulders. • Raise one hand, touch the opposite shoulder, and return to the push-up position. • Attempt to keep the shoulders down away from the ears and the back in a neutral position (same as standing) throughout the activity. • Perform 15 repetitions with each arm in an alternating manner and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated if the shoulder cannot be contacted with the hand on 3 consecutive repetitions (on same arm).
5. Pull: ONE ARM ROW • Hips extended • Hands shoulder width • Feet > shoulder width • 90° knee angle • Hand touch shoulder • 15 reps per arm	• With the shoes on, adopt an inverted position beneath the bar with the hands shoulder width apart (bar positioned ~1-2" from fingers on floor). The feet must be wider than the shoulders. • Bend both knees to 90°. Raise one hand, touch the opposite shoulder, and return to the start position. • Attempt to keep the shoulders down away from the ears and the back in a neutral position (same as standing) throughout the activity. • Perform 15 repetitions with each arm in an alternating manner and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated if the shoulder cannot be contacted with the hand on 3 consecutive repetitions (on the same arm).
Work Based Challenges (Time	

<i>Demand)</i>	
1. Squat: JUMP SQUAT • Heels on floor • Hands off hips • Leave the floor • Hips extended • 90° knee angle • 30s continuous	• With the shoes on, stand with the feet hip width apart the arms in front or beside the body (not on hips). • Jump, land and squat to a knee flexion angle of 90°. Perform continuously, fully extending the hips when leaving the ground (height is not important). The arms can be used. • Attempt to keep the knees in line with the hips and feet, and the back in a neutral position (same as standing) throughout the activity. The heels must be kept on the floor while squatting. • Perform continuously for 30s (no pause at bottom) and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated (or repeated) if the desired depth cannot be achieved on 3 consecutive repetitions.
2. Lunge: LATERAL LUNGE • Heel on the floor • Hands off hips • 90° knee angle • Hips extended • Alternate legs • 30s continuous	• With the shoes on, stand with the feet hip width apart and place the arms in front or beside the body (not on hips). • Lunge sideways to a knee flexion angle of 90° and return to standing with the hip fully extended. The front heel must be kept in contact with the floor during the descent phase. • Attempt to keep the front knee in line with the front hip and foot, and the back in a neutral position (same as standing) throughout the activity. • Perform continuously for 30s, alternating to the left and right side. Provide as much feedback as necessary to help the performer desired fitness and movement standards • The test can be terminated (or repeated) if the desired depth cannot be achieved on 3 consecutive repetitions (on same leg).
3. Hinge: ONE LEG HIP BRIDGE • Heel on the floor • Hands off hips • 90° knee angle • 90° hip angle • Alternate legs • 30s continuous	• With the shoes on, lay on the floor with the knees bent, feet hip width apart and the arms beside the body. • Raise both hips off the floor (to knee angle of 90°) and flex one hip to 90°. Return to the bridge position with the hips fully extended and repeat with opposite leg. • Attempt to keep the stance knee in line with the stance hip and foot, and the back in a neutral position (same as standing) throughout the activity. • Perform continuously for 30s, raising each leg in an alternating manner. Provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated (or repeated) if the desired range cannot be achieved on 3 consecutive repetitions (on same leg).
4. Push: HANDSTAND	• Adopt an inverted position with the arms straight and hands shoulder width apart, placed approximately 1-1.5 feet from the wall. Assistance can be provided

• Hands shoulder width • Legs straight • Feet stationary • 1-1.5 feet from wall • Facing away from wall • 30s continuous	to get into this position. • Must face away from the wall with the legs straight and the feet supported by the wall at all times. • Attempt to keep the shoulders down away from the ears and the back in a neutral position (same as standing) throughout the activity. • Perform continuously for 30s and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated (or repeated) if the inverted position cannot be maintained for 30s
5. Pull: PULL UP HANG • Hands shoulder width • Legs straight • Feet stationary • 1-1.5 feet from wall • Facing away from wall • 30s continuous	• Adopt a hang position with the arms straight and hands shoulder width apart in a neutral or prone grip. Assistance can be provided to get into this position. • The legs straight must be kept straight without crossing the feet at all times. • Attempt to keep the shoulders down away from the ears and the back in a neutral position (same as standing) throughout the activity. • Perform continuously for 30s and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be terminated (or repeated) if the hang position cannot be maintained for 30s.
<i>Tempo Based Challenges (Speed Demand)</i>	
1. Squat: AIR SQUAT • Feet hip width • Heels on the floor • Hands off hips • 90° knee angle • Hips extended • 10 reps in 15s	• With the shoes on, stand with the feet hip width apart and the arms in front or beside the body (not on hips). Keep the heels in contact with the floor at all times. • Squat to a knee flexion angle of 90° and return to standing with the hips fully extended. • Attempt to keep the knees in line with the hips and feet, and the back in a neutral position (same as standing) throughout the activity. • Perform 10 repetitions in less than 15s and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be repeated if the tempo is not met or the depth cannot be achieved on 3 consecutive repetitions.
2. Lunge: BACK LUNGE • Front heel on the floor Hip extended • Hands off hips	• With the shoes on, stand with the feet hip width apart and place the arms in front or beside the body (not on hips). • The test can be repeated if the tempo is not met or the depth cannot be achieved on 3 consecutive repetitions. • Lunge backwards to a front knee flexion angle of 90° and return to standing with the hip fully extended. The front heel must be kept in contact with the floor

• 90° front knee angle • Alternate legs • 10 reps in 15s	during the descent phase. • Attempt to keep the front knee in line with the front hip and foot, and the back in a neutral position (same as standing) throughout the motion. • Perform 10 repetitions in less than 15s on and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be repeated if the tempo is not met or the depth cannot be achieved on 3 consecutive repetitions.
3. Hinge: GOOD MORNING • Feet hip width • Heels on the floor • Hands on head • 90° hip angle • Hips extended • 10 reps in 15s	• With the shoes on, stand with the feet hip width apart and place both hands on the back of the head. Keep both heels in contact with the floor at all times. • Hinge to a hip flexion angle of 90° and return to standing with the hip fully extended. Adopt a slight bend in the knees. • Attempt to keep the knees in line with the hips and feet, and the back in a neutral position (same as standing) throughout the activity. • Perform 10 repetitions in less than 15s and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be repeated if the tempo is not met or the depth cannot be achieved on 3 consecutive repetitions.
4. Push: PUSH UP • Hands shoulder width • Feet shoulder width • Hips extended • Chest touches floor • Elbows extended • 10 reps in 15s	• With the shoes on, adopt a push-up position with the hands and feet shoulder width apart. The feet must be held stationary at all times. • Lower the chest to the floor, touch the floor and return to the start position with the elbows fully extended. • Attempt to keep the shoulders down away from the ears and the back in a neutral position (same as standing) throughout the activity. • Perform 10 repetitions in less than 15s and provide as much feedback as necessary to help the performer achieve the desired fitness and movement standards. • The test can be repeated if the tempo is not met or the depth cannot be achieved on 3 consecutive repetitions.
5. Pull: PULL UP HANG • Hands shoulder width • 90° knee angle • Chest touches bar • Hips extended • Elbows extended	• With the shoes on, adopt an inverted position beneath the bar with the hands and feet shoulder width apart (bar positioned ~1- 2" from fingers on floor) and the hips extended. • Bend both knees to 90°. Raise the body until the chest touches the bar and return to start position with elbows fully extended. • Attempt to keep the shoulders down away from the ears and the back in a neutral position (same as standing) throughout the activity. • Perform 10 repetitions in less than 15s and provide as much feedback as necessary to help the performer achieve the desired fitness and movement

• 10 reps in 15s	standards. • The test can be repeated if the tempo is not met or the range cannot be achieved on 3 consecutive repetitions.
--	--

Appendix C

Physical Literacy Screening Criteria

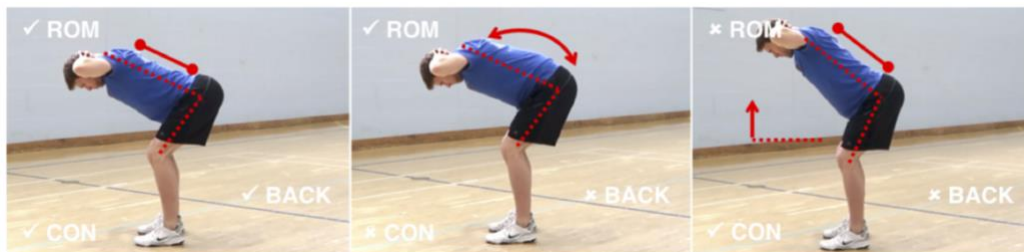


Figure 11. Illustration of the presence or absence of RoM and control criteria for the Physical Literacy Screen™ during the hinge (Back Motion Control).



Figure 12. Illustration of the presence or absence of RoM and control criteria for the Physical Literacy Screen™ during the lunge (Knee Motion Control).



Figure 13. Illustration of the presence or absence of RoM and control criteria for the Physical Literacy Screen™ during the push-up (Shoulder Motion Control).

Appendix D

Scoring Data Structure for Intra-Rater Reliability Analysis

Table 3. Sample of Raw Rater Scores for One Athlete Across Two Assessment Attempts

Rate r	Athlete	Movement Pattern	Domain	Feature Rated	Attempt 1	Attempt 2	Agreement
R01	W_P01	Squat	Reps	Knee Control	0	0	Yes
R01	W_P01	Squat	Reps	Back Control	1	1	Yes
R01	W_P01	Squat	Tempo	Knee Control	1	1	Yes
R01	W_P01	Squat	Tempo	Back Control	0	0	Yes
R01	W_P01	Squat	Work	Knee Control	0	0	Yes
R01	W_P01	Squat	Work	Back Control	0	0	Yes
R01	W_P01	Lunge	Reps	Knee Control	1	1	Yes
R01	W_P01	Lunge	Reps	Back Control	1	1	Yes
R01	W_P01	Lunge	Tempo	Knee Control	1	1	Yes
R01	W_P01	Lunge	Tempo	Back Control	1	0	No
R01	W_P01	Lunge	Work	Knee Control	0	1	No
R01	W_P01	Lunge	Work	Back Control	1	1	Yes
R01	W_P01	Hinge	Reps	Knee Control	1	1	Yes
R01	W_P01	Hinge	Reps	Back Control	1	1	Yes
R01	W_P01	Hinge	Tempo	Knee Control	1	1	Yes
R01	W_P01	Hinge	Tempo	Back Control	1	1	Yes
R01	W_P01	Hinge	Work	Knee Control	0	1	No
R01	W_P01	Hinge	Work	Back Control	1	1	Yes
R01	W_P01	Push	Reps	Back Control	1	1	Yes
R01	W_P01	Push	Reps	Shoulder Control	0	1	No
R01	W_P01	Push	Tempo	Back Control	0	0	Yes
R01	W_P01	Push	Tempo	Shoulder Control	0	1	No
R01	W_P01	Pull	Reps	Back Control	0	1	No
R01	W_P01	Pull	Reps	Shoulder Control	1	1	Yes
R01	W_P01	Pull	Tempo	Back Control	1	1	Yes
R01	W_P01	Pull	Tempo	Shoulder Control	1	1	Yes
R01	W_P01	Pull	Work	Back Control	1	1	Yes
R01	W_P01	Pull	Work	Shoulder Control	0	0	Yes

Appendix E

Rater Completion Time Summary

Table 4. Average Completion Time per Rater Across Both Scoring Sessions

Rater ID	Average Completion Time (minutes)
R01	29.28
R02	24.59
R03	32.70
R04	4.79
R05	28.80
R06	20.20
R07	15.68
R08	34.71
R09	25.17
R10	27.13
R11	10.38
R12	13.83
R13	32.25
R14	17.45
R15	23.67
R16	49.23
R17	30.95
R18	58.75
R19	36.35
R20	46.82
R21	48.64
R22	17.32
Mean	28.68