Reliability and Association with Injury of Movement Screens: A Critical Review

Article *in* Sports Medicine · June 2016
DOI: 10.1007/s40279-015-0453-1

5 authors, including:

Robert McCunn
Heart of Midlothian Football Club

Karen Aus der Fünten
Universität des Saarlandes

Hugh Fullagar
University of Technology Sydney

Ian McKeown
University of Adelaide

Sports Med
DOI 10.1007/s40279-015-0453-1

REVIEW ARTICLE

Reliability and Association with Injury of Movement Screens: A Critical Review

Robert McCunn[1] · **Karen aus der Fünten**[1] · **Hugh H. K. Fullagar**[1,2] · **Ian McKeown**[3] · **Tim Meyer**[1]

Abstract Subjective assessment of athletes' movement quality is widely used by physiotherapists and other applied practitioners within many sports. One of the beliefs driving this practice is that individuals who display 'poor' movement patterns are more likely to suffer an injury than those who do not. The aim of this review was to summarize the reliability of the movement screens currently documented within the scientific literature and explore the evidence surrounding their association with injury risk. Ten assessments with accompanying reliability data were identified through the literature search. Only two of these ten had any evidence directly related to injury risk. A number of methodological issues were present throughout the identified studies, including small sample sizes, lack of descriptive rater or participant information, ambiguous injury definitions, lack of exposure time reporting and risk of bias. These factors, combined with the paucity of research on this topic, make drawing conclusions as to the reliability and predictive ability of movement screens difficult. None of the movement screens that appear within the scientific literature currently have enough evidence to justify the tag of 'injury prediction tool'.

Electronic supplementary material The online version of this article (doi:10.1007/s40279-015-0453-1) contains supplementary material, which is available to authorized users.

✉ Robert McCunn
bob.mccunn@me.com

1 Institute of Sport and Preventive Medicine, Saarland University, Geb. B8.2, 66123 Saarbrücken, Germany

2 Sport and Exercise Discipline Group, UTS: Health, University of Technology, Sydney, NSW, Australia

3 Port Adelaide Football Club, Adelaide, SA, Australia

Key Points

Subjective assessment of athletes' movement quality is commonplace within professional sport, often in an attempt to predict injury risk.

Of the ten movement screens identified within the scientific literature, only two have had their injury predictive ability investigated via prospective cohort studies.

None of the movement screens present within the scientific literature currently have enough supporting evidence to justify being heralded as 'injury prediction tools'; however, they may well provide practitioners with greater holistic understanding of their athletes' physical capabilities.

1 Introduction

The use of fitness assessments to profile and categorize athletes' physical capabilities is commonplace and a central aspect of many applied practitioners' jobs [1]. The data collected from traditional fitness tests are typically objective in nature, i.e., can be measured in units such as seconds, centimeters, or grams. Movement screening is a type of assessment frequently used within professional soccer as well as other sports and is predominantly a subjective process that aims to measure the 'quality' of a movement pattern [2]. However, for various reasons, including the subjective nature of such assessments and

Published online: 31 December 2015

its relatively recent adoption by practitioners, this practice has received limited attention within the scientific literature. A consensus on what defines movement quality is not available; however, the concept encapsulates the maintenance of correct posture and joint alignment in addition to balance while performing the selected movements. While some sporting institutions measure intuitively related parameters such as strength and joint range of motion, movement quality has been identified as an independent attribute [3, 4]. Therefore, a fitness testing battery that seeks to build a comprehensive profile of an athlete should incorporate an assessment of movement quality. This highlights the need for reliable and valid movement screening tools.

The foundation of a comprehensive injury prevention program is identifying individuals with a high risk of injury [5], and this is one of the key concepts underpinning the practice of movement screening. If athletes who display 'poor' movement patterns have a greater risk of injury than those who display 'good' movement patterns, then screening protocols may be an important component of injury prevention strategies. However, the purpose of movement screens is not to diagnose why a poor movement pattern exists but simply to highlight it [6]. It is up to the judgment of the practitioner as to the course of action, if any, taken in response to the outcome. Furthermore, elite sport is not the only environment in which movement ability is important. Giblin et al. [7] stated that fundamental movement ability (core stability, balance, coordination) is related to perceived competence and confidence associated with physical activity. As such, movement quality is linked to general health as well as sports performance.

Despite movement quality being an important skill for the general population, in addition to athletes, measuring it is problematic due to its subjective nature. A variety of movement screens exist, the most well-known being the functional movement screen (FMS) [6, 8]. The FMS has received attention from researchers, and different aspects of this protocol—such as its reliability and association with injury—have been investigated. However, other screens do exist, with some—but not all—appearing within the scientific literature. No collective critique of the movement screens detailed in the scientific literature, necessary to raise awareness of the available options, currently exists. This would allow practitioners to make informed decisions about which, if any, movement screen is most appropriate for them. Accordingly, the aim of the present review was to summarize the intra- and inter-rater reliability of the available movement screens and discuss the evidence surrounding their ability to determine injury risk.

2 Literature Search

We performed a computerized literature search (Fig. 1) in PubMed, Web of Science, and ScienceDirect for articles published up until 1 July 2015 using the search terms 'movement', 'screen', 'screening', 'reliability', 'injury', 'prediction', 'predicts', 'landing error scoring system', 'tuck jump assessment', 'functional movement screen', 'functional movement screening', 'single leg squat test', 'squat', 'test', 'drop jump', 'drop vertical jump', and 'movement quality' in various combinations. In addition, articles were identified manually from the reference lists of original manuscripts; a total of 51 relevant articles were identified. For the purpose of this review, a movement screen was defined as a protocol designed for use with apparently healthy, uninjured individuals to primarily assess the 'quality' of a movement(s) rather than objective outcomes such as number of repetitions, distance, or time achieved. The movement(s) included should rely on multiple physical qualities to execute correctly, e.g., strength, balance, and flexibility. It is not used to identify specific clinical conditions and does not require interpretation by a medical professional.

3 Reliability of Identified Screens

Ten movement screens that met the definition outlined above and with accompanying reliability data were identified through the literature search (Fig. 2). These screens consisted of the FMS, the Landing Error Scoring System (LESS), single-leg squat screen variations, drop vertical jump screen variations, tuck jump assessment, athletic ability assessment (AAA), conditioning specific movement tasks (CSMT), the netball movement screening tool (NMST), the physical performance measures screen (16-PPM), and the star excursion balance test (SEBT) movement quality screen. A description of the exercises involved in each screen is provided in Table 1.

The reliability of an assessment tool is paramount because it is a pre-requisite for test validity [9]. As such, before any given movement screen can be investigated with respect to injury prediction, it must first be demonstrated that the test is reliable. Throughout the 51 articles identified by the literature search, intra-class correlation coefficients (ICC) were commonly reported. Atkinson and Nevill [10] stated that various qualitative interpretations of ICC values exist, yet none were related to "analytical goals for research" and so it is difficult to say exactly what value constitutes 'good' or 'excellent' reliability. Some of the identified studies classified an ICC value of $\geq$0.75 as good [11], whereas others [12, 13] classified scores of $\geq$0.80 and

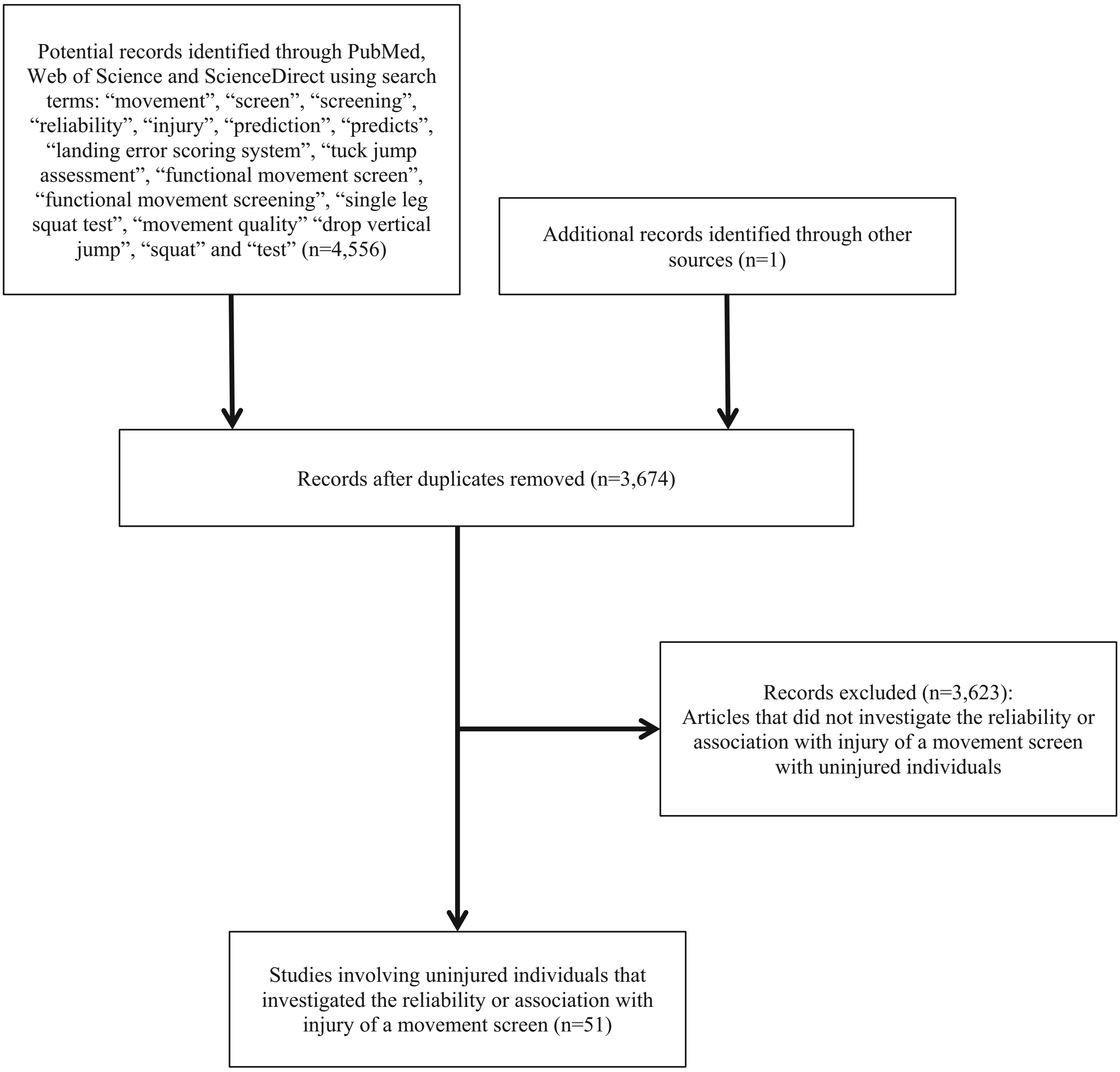

Fig. 1 Flow diagram showing the identification and selection of movement screen studies in the scientific literature for the current review

$\geq$0.90 as good and excellent, respectively. Shultz et al. [14] reported ICC values of 0.40–0.75 as fair to good and >0.75 as excellent. A reasonable consensus as to what can be considered good reliability appears to be an ICC $\geq$0.75; thus, this classification is used throughout this review. In addition to ICCs, kappa values were also often reported. Guidelines presented by Landis and Koch [15] are used to classify these scores, with kappa score indicating strength of agreement as follows: poor <0.00, slight 0.00–0.20, fair 0.21–0.40, moderate 0.41–0.60, substantial 0.61–0.80, and almost perfect 0.81–1.00.

3.1 Methodological Quality Assessment for Reliability Studies

The methodological quality of each paper that reported the reliability of a movement screen was assessed using the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) checklist [16]. This checklist utilizes a four-point scoring system (poor, fair, good, excellent) and contains sub-sections relating to numerous aspects of study design. For the purposes of this critique, only the reliability section (box B), which contains

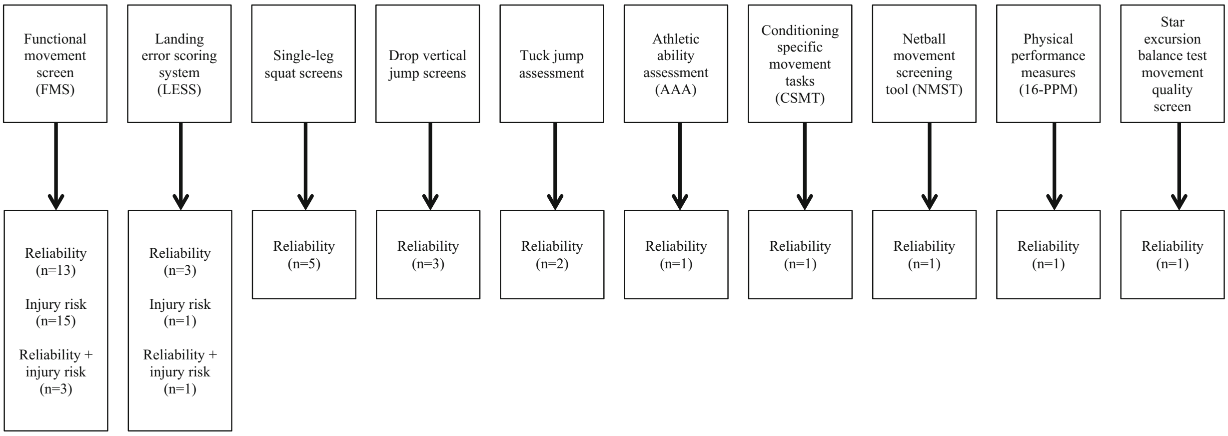

Fig. 2 Breakdown of the individual movement screens identified from the literature search and the number of the articles that investigated their reliability or association with injury

14 questions, was addressed. As per the checklist instructions, if the answer to any of the 14 questions was 'poor' then, based on the 'worst score counts' principle, the study was classified as such. The outcomes of this methodological quality assessment are presented in Table 2 (see the Electronic Supplementary Material [ESM] Appendix S1 for raw data).

3.2 Functional Movement Screen

The FMS comprises seven subtests, including an overhead squat, hurdle step, in-line lunge, shoulder mobility assessment, active straight-leg raise, trunk stability push-up, and a prone two-point rotary stability movement, all performed without external load [6, 8]. As can be seen from Table 3, a total of 16 studies were identified that reported either the intra- or inter-rater reliability of the FMS or related variations [11–14, 17–28]. Seven of the eight studies that investigated the intra-rater reliability reported ICC values $\geq$0.75. In addition, Teyhen et al. [24] reported an ICC of 0.74 (95 % confidence interval [CI] 0.60–0.83). As such, it would appear that the FMS has consistently demonstrated good intra-rater reliability. Of the 16 identified studies, 14 reported inter-rater reliability either in the form of ICC, weighted kappa, or Krippendorff's α. Twelve of these observed an ICC $\geq$0.75, representing good inter-rater reliability. In addition, Minick et al. [28] used the weighted kappa statistic to measure inter-rater reliability; however, rather than use the FMS composite score, they compared raters using the individual subtests. The weighted kappa values ranged from 0.79 to 1.0, with the authors stating that this represented substantial to excellent agreement. In contrast, Shultz et al. [14] used Krippendorff's α statistic to quantify reliability and classified a value $\geq$0.80 as acceptable; however, they reported a score of 0.38 (95 % CI 0.35–0.41). These authors concluded that the FMS demonstrated poor inter-rater reliability and suggested that improved rater training may have resulted in an improved reliability score. The authors also highlighted the difference between years of experience and the number of tests a rater has administered, stating that the latter is likely of greater relevance to improving reliability.

Five studies [11, 18, 20, 21, 29] included information as to the raters' experience (years of clinical practice, number of FMS tests performed, or level of certification), which allowed comparison of test reliability based on these variables. Four studies [11, 18, 20, 29] suggested the experience of the rater did not influence the inter-rater reliability. Additionally, Smith et al. [11] observed good intra-rater reliability (ICCs >0.80) for all raters, regardless of experience. In contrast, Gribble et al. [21] showed that intra-rater reliability did vary depending on the experience of raters. The raters were divided into three groups: athletic training students, athletic trainers who had not previously used the FMS, and athletic trainers who had at least 1 year of experience administering FMS tests. The greater number of raters included in the study by Gribble et al. [21] suggests a stronger experimental foundation, and this may have contributed to their contrasting findings. A clear trend highlighting the importance of rater experience was apparent, with the following ICC values: students, 0.37 (95 % CI −0.79 to 0.78), athletic trainers, 0.76 (95 % CI 0.32–0.92), and experienced athletic trainers, 0.95 (95 % CI 0.68–0.99). However, while Gribble et al. [21] employed a greater number of raters than the other four studies, the ICC values reported were based on the ratings of only three participants. Therefore, the weight of the currently available evidence suggests that the experience of the rater is not a significant factor influencing scoring.

Table 1 Content of the identified movement screens

Screen name	Exercises (*N*)	Name of exercises	Protocol description
Functional movement screen	7	Deep squat	Cook et al. [8]
		Hurdle step	Cook et al. [6]
		In-line lunge	
		Shoulder mobility	
		Active straight-leg raise	
		Trunk stability push up	
		Rotary stability	
Landing error scoring system	1	Drop jump	Padua et al. [34]
Single-leg squat screens			
Single-leg squat task	1	Single-leg half squat	Crossley et al. [36]
Single-leg mini squat	1	Single-leg half squat	Ageberg et al. [38]
Unilateral lower extremity functional tasks	2	Single-leg half squat	Chmielewski et al. [39]
		Lateral step down	
Drop vertical jump screens	1	Drop vertical jump	Nilstad et al. [44]
			Whatman et al. [45]
			Ekegren et al. [46]
Tuck jump assessment	1	Tuck jump	Myer et al. [48]
Athletic ability assessment	9	Prone hold on hands	McKeown et al. [53]
		Lateral hold on hands	
		Overhead squat	
		Single-leg squat off box	
		Walking lunge	
		Single-leg forward hop	
		Lateral bound	
		Push ups	
		Chin ups	
Conditioning specific movement tasks	6	Overhead squat	Parsonage et al. [54]
		Romanian deadlift	
		Single-leg squat	
		Double-leg to single-leg landing	
		Sprint (40 m)	
		Countermovement jump	
Netball movement screening tool	10	Squat	Reid et al. [55]
		Lunge and twist	
		Bend and pull	
		Push up	
		Single-leg squat	
		Vertical jump (land on both legs)	
		Vertical jump (land on one leg)	
		Broad jump	
		Star excursion balance test	
		Active straight leg raise	

Table 1 continued

Screen name	Exercises (*N*)	Name of exercises	Protocol description
Physical performance measures	16	Broad jump	Tarara et al. [56]
		Closed kinetic chain upper extremity stability test	
		Y-balance test	
		In-line lunge for distance	
		Lateral lunge for distance	
		Lumbar endurance	
		Side plank hip abduction	
		Side plank hip adduction	
		Triple hop for distance	
		Nordic hamstring	
		Full squat	
		Downward dog	
		Single-leg squat	
		Shoulder mobility test	
		Active straight leg raise	
		Beighton hypermobility	
Star excursion balance test movement quality screen	1	Anterior reach	Ness et al. (2015) [58]

Another aspect of movement screening pertinent to the reliability of such assessments is whether the rater scores participants in real time or after the test via video recording. One study sought to address this issue [14] and found that the intra-rater reliability was superior when using recorded footage to score participants. One rater assessed individuals live while they were being filmed and retrospectively assessed the footage. They also assessed the participants again 1 week later and scored the tests in real time. The ICCs for the live–live re-test and live–recorded re-test were 0.60 (95 % CI 0.35–0.77) and 0.92 (95 % CI 0.85–0.96), respectively. However, only one rater's scoring was investigated in this manner, hence it is difficult to draw firm conclusions on the respective merits of live and recorded FMS scoring. Nonetheless, the outcome of this study suggests that using recorded footage to assess participants may elicit greater intra- and inter-rater reliability than assessing them in real time.

Recently, it has been shown that participant knowledge of the scoring criteria can influence FMS total score [29]. Participants were assessed prior to and after having the criteria for a perfect score explained to them. Significant improvements in scores were observed simply by providing this information. This finding demonstrates that test reliability may be affected by how the test is delivered by the assessor. If the extent of task instruction and explanation differs between assessors or test occasions it is likely that intra- and inter-rater reliability will be impacted. To ensure that any changes in score are not simply due to familiarization, it is recommended that participants have the scoring criteria clearly explained to them and are allowed practice attempts before being scored. This is not to say that individuals should be coached through the movements, rather, it is imperative they know what is being asked of them without being told how to do it.

Overall, the majority of the identified studies reveal the FMS possesses good intra- and inter-rater reliability, although it should be noted that this conclusion is not unanimous throughout the literature. It should also be noted that the majority of the identified studies were classified as demonstrating poor methodological quality (Table 2). The influence of rater experience on reliability appears negligible. Furthermore, the practice of scoring tests via video footage may aid reliability, yet evidence on this issue is limited, so only tentative conclusions can be drawn. Test reliability is likely influenced by the performer's knowledge of the scoring criteria; as such, it is advisable to provide clear instructions to participants and allow practice attempts to reduce the influence of any learning effect. The depth of research investigating the FMS is much greater than for any other movement screen, yet despite this some organizations choose to use alternative tools [2, 30]. While the specific reasons are not clear, many professional football clubs seemingly do not feel the FMS meets their screening needs. However, a number of other screens

Table 2 Methodological quality (according to COSMIN checklist) of reliability studies[a]

Movement screen	Methodological quality rating		
	Poor	Fair	Good
FMS	Teyhen et al. intra-rater [24]	Shultz et al. [14]	Teyhen et al. inter-rater [24]
	Hotta et al. [17]	Butler et al. [25]	
	Gulgin and Hoogenboom [18]	Minick et al. [28]	
	Letafatkar et al. [19]		
	Parenteau-G et al. [12]		
	Elias [20]		
	Gribble et al. [21]		
	Smith et al. [11]		
	Frohm et al. [22]		
	Klusemann et al. [23]		
	Onate et al. [13]		
	Schneiders et al. [26]		
	Chorba et al. [27]		
LESS	Smith et al. [31]	Padua et al. [32]	Padua et al. [34]
	Onate et al. [33]		
Single-leg squat screens	Örtqvist et al. inter-rater [37]	Örtqvist et al. intra-rater [37]	Junge et al. [35]
	Crossley et al. [36]		
	Ageberg et al. [38]		
	Chmielewski et al. [39]		
DVJ screens	Whatman et al. [45]	Nilstad et al. [44]	
	Ekegren et al. [46]		
Tuck jump assessment	Herrington et al. [51]	Dudley et al. [52]	
AAA	McKeown et al. [53]		
CSMT		Parsonage et al. [54]	
NMST	Reid et al. [55]		
16-PPM	Tarara et al. [56]		
SEBT movement quality screen		Ness et al. [58]	

AAA athletic ability assessment, *COSMIN* consensus-based standards for the selection of health measurement instruments, *CSMT* conditioning specific movement tasks, *DVJ* drop vertical jump, *FMS* functional movement screen, *LESS* Landing Error Scoring System, *NMST* Netball Movement Screening Tool, *PPM* physical performance measures, *SEBT* Star Excursion Balance Test

[a] A quality rating of 'Excellent' is possible but was not achieved by any of the listed studies

appear within the scientific literature, albeit with much less supporting evidence.

3.3 Landing Error Scoring System

The LESS comprises one movement: drop vertical jumps. The accompanying scoring criteria relate to observed errors in technique and result in a potential minimum and maximum score of zero or 19, respectively, with a higher score indicating poorer performance. The four studies reporting either the intra- or inter-rater reliability of the LESS are presented in Table 4 [31–34]. Both Smith et al. [31] and Padua et al. [34] reported excellent intra-rater reliability (ICCs >0.90). When interpreting these findings, it should be noted that, in total, data from only three raters were used to calculate these ICC values. Further research is needed to establish the robustness of such findings. All four of the identified studies measured the inter-rater reliability of the LESS, with the ICC values ranging from 0.72 to 0.92, indicating good repeatability. Again, caution should be employed when analyzing these results, since this conclusion is based on the data from only nine raters. One study investigated the influence of rater experience on LESS scoring and found that novice (<1 year of experience as a certified athletic trainer) and expert (15 years' experience as a certified athletic trainer) raters displayed moderate to perfect agreement on all items [33]. The detailed scoring criteria employed by the LESS likely explain this high level of agreement between raters. The drawback to such a thorough scoring system is inevitably the time it takes to

Table 3 Studies that reported the intra- and/or inter-rater reliability of the functional movement screen

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/ sport		# of raters	Score (95 % CI)	# of raters	Score (95 % CI)
Hotta et al. [17]	10 M	Middle- and long-distance runners (collegiate)	2 Physical therapists	NA	NA	2	ICC 0.98 (0.93–1.00)
Gulgin and Hoogenboom [18]	10 M, 10 F	University students	1 Expert rater, 3 physical therapy students	NA	NA	4	ICC 0.88 (0.77–0.95)
Letafatkar et al. [19]	20	Unknown	2 Physical therapists	NA	NA	2	ICC 0.92
Parenteau-G et al. [12]	28 M	Ice-hockey players (elite youth)	1 Physiotherapist, 3 physiotherapy students	2	Rater 1 ICC 0.96 (0.92–0.98) Rater 2 ICC 0.96 (0.92–0.98)	2	ICC 0.96 (0.92–0.98)
Elias [20]	3 M, 2 F	Squash players (elite)	20 Physiotherapists	NA	NA	20	ICC 0.91
Gribble et al. [21]	2 M, 1 F	University students	16 Students, 15 ATs, 7 expATs	38	All raters ICC 0.75 (0.53–0.87) ExpATs ICC 0.95 (0.68–0.99) ATs ICC 0.76 (0.32–0.92) Student ICC 0.37 (–0.79 to 0.78)	NA	NA
Shultz et al. [14]	18 M, 21 F	NCAA Division 1 varsity athletes	1 Student, 1 physical therapist, 2 ATs, 2 S&C coaches	1	Live test–retest ICC 0.60 (0.35–0.77) Live-recorded test–retest ICC 0.92 (0.85–0.96)	6	Krippendorff's $\alpha = 0.38$ (0.35–0.41)
Smith et al. [11]	10 M, 10 F	University students	2 Students, 1 faculty member, 1 FMS certified instructor	4	Rater 1 ICC 0.90 (0.76–0.96) Rater 2 ICC 0.81 (0.57–0.92) Rater 3 ICC 0.91 (0.78–0.96) Rater 4 ICC 0.88 (0.72–0.95)	4	Occasion 1 ICC 0.89 (0.80–0.95) Occasion 2 ICC 0.87 (0.76–0.94)

Table 3 continued

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score (95 % CI)	# of raters	Score (95 % CI)
Frohm et al. [22]	26 M	Soccer players (elite)	8 Physiotherapists	8	Rater 1 ICC 0.87 Rater 2 ICC 0.77 Rater 3 ICC 0.83 Rater 4 ICC 0.77 Rater 5 ICC 0.79 Rater 6 ICC 0.45 Rater 7 ICC 0.79 Rater 8 ICC 0.75	8	Occasion 1 ICC 0.80 Occasion 2 ICC 0.81
Klusemann et al. [23]	10	Basketball players (elite youth)	8 (unspecified combination of S&C coaches/physiotherapists)	8	ICC 0.82	NA	NA
Onate et al. [13]	12 M, 7 F	University students	1 AT, 1 S&C coach	1	ICC 0.92	2	ICC 0.98
Teyhen et al. [24]	53 M, 11 F	Military personnel	8 Physical therapy students	4	ICC 0.74 (0.60–0.83)	8	ICC 0.76 (0.63–0.85)
Butler et al. [25]	30	Middle school students	1 FMS creator, 1 FMS certified instructor	NA	NA	2	ICC 0.99
Schneiders et al. [26]	10	Recreationally active individuals	2 Academic researchers	NA	NA	2	ICC 0.97
Chorba et al. [27]	3 M, 5 F	University students	2 Physical therapists	NA	NA	2	ICC 0.98
Minick et al. [28]	17 M, 23 F	University students	2 FMS creators, 2 FMS certified instructors	NA	NA	4	Weighted κ values for each test ranged from 0.79 to 1.0 when comparing novice and experienced raters

All values refer to the FMS composite score unless otherwise stated

AT athletic trainer, *CI* confidence interval, *ExpAT* experienced athletic trainer, F female, *FMS* functional movement screen, *ICC* intraclass correlation coefficient, κ kappa, *M* male, *NA* not applicable, *NCAA* National Collegiate Athletic Association, *S&C* strength and conditioning

score each participant. The original LESS protocol requires video recording of tests, with subsequent scoring by assessors from the footage, and this methodology has associated costs, from both a financial and a time perspective. A real-time scoring system to overcome these restrictive issues was developed by Padua et al. [32]. Three raters' real-time scoring of 43 participants was compared, and the resulting ICC values ranged from 0.72 (95 % CI 0.42–0.88) to 0.81 (95 % CI 0.56–0.92), suggesting moderate to good inter-rater reliability for the real-time version. Taken collectively, the initial evidence is promising with regard to the reliability of the test.

3.4 Single-Leg Squat Screens

Five studies were identified that explored either the intra- or inter-rater reliability of movement screens containing various single-leg squat tests (Table 5) [35–39]. The single-leg squat task [36], the single-leg mini squat [35, 37, 38], and another assessment comprising two 'functional tasks'—the single-leg squat and lateral step-down—have been investigated [39]. While these three screens differ slightly in their protocols, the scoring criteria are very similar. For example, all three variations include criteria related to assessment of knee alignment during single leg

Table 4 Studies that reported the intra- and/or inter-rater reliability of the Landing Error Scoring System

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score (95 % CI)
Smith et al. [31]	10	High school/collegiate athletes	2 Raters (occupation unspecified)	2	ICC 0.97	2	ICC 0.92
Padua et al. [32]	19 M, 24 F	Military personnel	3 Athletic trainers	NA	NA	3	Rater 1 vs. 2 ICC 0.81 (0.56–0.92) Rater 1 vs. 3 ICC 0.72 (0.42–0.88) Rater 1 vs. combined 2 and 3 ICC 0.79 (0.64–0.88)
Onate et al. [33]	19 F	NCAA Division 1 soccer players	2 Athletic trainers	NA	NA	2	ICC 0.84
Padua et al. [34]	25 M, 25 F	Military personnel	2 Raters (occupation unspecified)	1	ICC 0.91	2	ICC 0.84

All values refer to the LESS composite score

CI confidence interval, *F* female, *ICC* intraclass correlation coefficient, *LESS* Landing Error Scoring System, *M* male, *NA* not applicable, *NCAA* National Collegiate Athletic Association

squatting. Three of the five identified studies investigated the intra-rater reliability, with kappa values ranging from 0.13 to 0.80, representing poor to substantial agreement. These results are difficult to interpret, since the kappa values within each individual study varied so widely. Differences in the populations observed between studies may explain some of the variance; however, comprehensive participant information was not reported in all instances. Similarly, differences in the precise protocols and scoring criteria for each screen variation may have contributed to the inconsistent results. These values were derived from the data collected by only seven raters in total, and this relatively small evidence source likely contributes to the uncertain findings. Similarly, as can be seen from Table 5, the inter-rater kappa values ranged from 0.00 to 0.92, making conclusions difficult to draw with regard to the reliability of these screening tools.

3.5 Drop Vertical Jump Screens

The drop vertical jump, whereby an individual drops from a raised surface to the floor and immediately jumps vertically as high as possible, is a common screening test performed to identify movement patterns thought to be associated with risk of injury [40–42]. However, quantification of performance on such tests is typically achieved via objective analysis of joint angles and 'separation distance' between the knees and ankles [40–43]. This type of assessment was considered separate to the primarily subjective process of movement screening discussed in this review. However, the literature search identified three studies that described the intra- and inter-rater reliability of drop vertical jump screen variations that conformed to the definition of a movement screen previously outlined here (Table 6) [44–46]. Of the three studies, one used the first-order agreement coefficient (AC1) statistic to analyze reliability [45]. The AC1 values can be interpreted in the same way as described above for kappa values [47]. As with the identified single-leg squat screens, the drop vertical jump screens differ in their protocols; however, the scoring criteria are very similar. For example, all three variations include criteria related to assessment of knee alignment during landing. Intra-rater reliability ranged from moderate to almost perfect; however, much greater variation existed between raters. Across the three studies, AC1 and kappa values ranged from 0.32 to 0.92, representing fair to almost perfect agreement. The poorest intra- and inter-rater reliability was reported by Whatman et al. [45], suggesting that perhaps the protocols and scoring criteria adopted by the other two studies are superior [44, 46]. The results from the identified studies are mixed; therefore, further research utilizing consistent test protocols and scoring criteria are required to elucidate the reliability of these screening tools.

3.6 Tuck Jump Assessment

The tuck jump assessment created by Myer et al. [48] was designed to assess the movement quality associated with repeated jumping and landing and requires an individual to perform repeated tuck jumps for 10 s. The plyometric aspect of this exercise task is relevant, since it has been reported that injury prevention interventions lacking this explosive component have demonstrated limited success in

Table 5 Studies that reported the intra- and/or inter-rater reliability of the identified single-leg squat screens

References	Participation information		Rater information	Intra-rater reliability		Inter-reliability	
	Sample size	Occupation/ sport		# of raters	Score (95 % CI)	# of raters	Score (95 % CI)
Junge et al. [35]	72	Students (children)	2 Physiotherapy students	NA	NA	2	Weighted κ values for each scoring category ranged from 0.54 to 0.86
Crossley et al. [36]	15	Unknown (adult)	3 Physical therapists, 1 'expert panel'	3	Rater 1 κ 0.80 Rater 2 κ 0.70 Rater 3 κ 0.60	4	κ values ranged from 0.60 to 0.80
Örtqvist et al. [37]	33	Students (children)	2 Physiotherapists	1	*K* 0.48 (0.16–0.79)	2	κ 0.57 (0.30–0.85)
Ageberg et al. [38]	8 M, 17 F	Unknown (adult)	2 Physical therapists	NA	NA	2	κ 0.92 (0.75–1.08)
Chmielewski et al. [39]	7 M, 18 F	Unknown (adult)	2 Physical therapists, 1 athletic trainer	3	Weighted κ values for each test and each scoring method ranged from 0.13 to 0.68	3	Weighted κ values for each test and each scoring method ranged from 0.00 to 0.55

CI confidence interval, *F* female, κ kappa, *M* male, *NA* not applicable

Table 6 Studies that reported the intra- and/or inter-rater reliability of the identified drop vertical jump screens

References	Participant information		Rater information	Intra-rater reliability		Inter-rater reliability	
	Sample size	Occupation/sport		# of raters	Score	# of raters	Score
Nilstad et al. [44]	60 F	Soccer players (elite)	3 Physiotherapists	NA	NA	3	κ values ranged from 0.52–0.92
Whatman et al. [45]	12 M, 11 F	Variety of undisclosed sports (youth)	66 Physiotherapists	26	All raters AC1 0.60 (range 0.14–0.92) Raters >14 years' experience AC1 0.65 (range 0.22–0.91) Raters <10 years' experience AC1 0.56 (range 0.20–0.83)	66	All raters AC1 0.34 (95 % CI 0.22–0.47) Raters >14 years' experience AC1 0.36 (95 % CI 0.22–0.50) Raters 10–14 years' experience AC1 0.37 (95 % CI 0.21–0.53) Raters 5–9 years' experience AC1 0.33 (95 % CI 0.33–0.55) Raters <5 years' experience AC1 0.32 (95 % CI 0.19–0.46)
Ekegren et al. [46]	40 F	Soccer players (regional level youth)	3 Physiotherapists	3	Rater 1 κ 0.80 (95 % CI 0.65–1.00) Rater 2 κ 0.85 (95 % CI 0.72–1.00) Rater 3 κ 0.75 (95 % CI 0.58–1.00)	3	Time point 1 κ 0.80 (95 % CI 0.62–0.98) Time point 2 κ 0.77 (95 % CI 0.59–0.95)

AC1 first-order agreement coefficient, *CI* confidence interval, *F* female, κ kappa, *M* male, *NA* not applicable

reducing knee injury [48–50]. Two studies were identified that established the intra- and inter-rater reliability of the tuck jump assessment [51, 52]. Herrington et al. [51] revealed good intra-rater reliability for two raters, with kappa values ranging from 0.81 to 1.00. Similarly, they reported good inter-rater reliability, with a kappa value of 0.88. In contrast, Dudley et al. [52] reported poor to moderate intra- and inter-rater reliability, with ICC values ranging from 0.44 to 0.72. One possible explanation for the discrepancy between these findings—at least with regard to intra-rater reliability—may relate to the differences in sample sizes. Dudley et al. [52] viewed videos of 40 participants, whereas Herrington et al. [51] only assessed ten subjects. This may have resulted in recall bias, with the raters investigated by Herrington et al. [51] potentially remembering the previous scores of the ten participants when scoring their videos for the second time. However, since one of the creators of the tuck jump assessment was a co-author and rater within the Herrington et al. [51] article, more extensive training and experience could also have contributed to the superior reliability values. The tuck jump assessment is unique amongst movement screens in that it requires the participants to perform repeated plyometric movements, with the creators proposing that the increased sport-specificity of the task may aid in highlighting injury risk. However, this assessment currently only demonstrates face validity and this should be remembered when taking this assertion into consideration. The nature of this assessment means that it may be of particular interest to practitioners working in jumping and landing sports such as netball and basketball.

3.7 Athletic Ability Assessment

The recently developed AAA, which consists of nine subtests, is currently used within numerous high-performance environments (unpublished observation) to assess athletes' movement patterns [53]. The nine subtests include a prone hold, lateral hold, overhead squat (with 10-kg bar), single-leg squat off a box, walking lunge (with a 20-kg bar), single-leg forward hop, lateral bound, push-up, and chin-up. McKeown et al. [53] reported excellent intra- and inter-rater reliability [ICC values of 0.97 (90 % CI 0.92–0.99) and 0.96 (90 % CI 0.94–0.98), respectively]. These authors also observed a strong correlation ($r = 0.94$) between athletes' overall AAA scores when assessed live and via video recording, indicating that either method is viable. The AAA maximum score is 117, and a detailed scoring system that stipulates criteria based on body segment is provided, allowing for more precise assessment than some other movement screens since the continuum of possible scores is large.

3.8 Conditioning Specific Movement Tasks

The CSMT screen was developed to aid in the assessment of young rugby union players' readiness to enter elite academies [54]. The CSMT screen comprises six subtests, including an overhead squat (with a 20-kg bar), Romanian deadlift (with a 20-kg bar), single-leg squat, double-leg to single-leg landing, a 40-m sprint, and countermovement jump. A four-point scoring system, similar to that employed by the FMS, was used to rate the quality of the six movements. Intra-rater kappa values ranged from 0.61 to 1.00, indicating substantial to excellent agreement [54]. Similarly, inter-rater kappa values ranged from 0.62 to 1.00. The reliability values are based on the scores given by only two raters, which should be remembered when interpreting the findings. Further investigations utilizing adult populations, greater number of raters, athletes from different sports, and raters of varying experience are required before definitive conclusions can be made with regard to the reliability of the CSMT screen.

3.9 Netball Movement Screening Tool

Another sport-specific screen, the NMST, was designed to assess movement quality in patterns relevant to the sport of netball. The NMST comprises ten subtests, including a squat, lunge with a twist, a bend and pull movement, push-up, single-leg squat, vertical jump (landing on both legs), vertical jump (landing on one leg), broad jump, the SEBT, and an active straight-leg raise. Reid et al. [55] reported intra- and inter-rater ICC values of 0.96 (95 % CI 0.91–0.98) and 0.84 (95 % CI 0.65–0.93), respectively. These values suggest excellent agreement within and between raters; however, the results were based on the scores given by only two examiners. The age of the netball players assessed by Reid et al. [55] ranged between 13 and 17 years, so further reliability studies conducted with adult players are needed to establish the applicability of the results to this population.

3.10 Physical Performance Measures

The 16-PPM is made up of 16 subtests, ten of which are quantitative in nature, e.g., measured in distance or number of repetitions completed [56]. While these ten subtests do not meet the aforementioned definition of movement screening exercises, the 16-PPM also includes six qualitative subtests that do assess how well an athlete performs the required movement. The six qualitatively scored subtests, which are all performed without external load, include an overhead squat, downward dog, single-leg squat, shoulder mobility assessment, active straight-leg raise, and Beighton hypermobility assessment. The following reliability values refer only to the six qualitative subtests. Intra-rater reliability, reported as weighted kappa values, varied between expert and novice raters according to Tarara et al. [56]. Weighted kappa values ranged from 0.32 to 0.81 for the expert rater, representing fair to almost perfect agreement. In contrast, the two novice raters' weighted kappa values ranged from −0.09 to 0.78, indicating poor to substantial agreement between test occasions. As such, it would appear that training is required for raters administering the 16-PPM to ensure consistent scoring. Little information was given as to the occupation or level of qualification of the expert rater, so it is unclear how much training may be required to achieve an acceptable level of consistency. Inter-rater reliability varied widely, with weighted kappa values ranging between 0.24 and 0.93 for individual subtests, representing fair to almost perfect agreement. Taking all the qualitative subtests into account, the 16-PPM appears to be a moderately reliable tool for assessing movement competency if administered by expert raters.

3.11 Star Excursion Balance Test Movement Quality Screen

The SEBT involves the objective measurement of unilateral reach distance of the lower extremity in various directions [57]. One article was identified that applied subjective movement quality criteria to the SEBT [58]. In its original form, the SEBT does not take into account how somebody achieves their score and reports only the objective reach distance in centimeters. Incorporating an assessment of an individual's movement quality during this test may provide additional useful information to practitioners. In the identified study, scoring criteria related to knee, pelvis, and trunk position were used by three physical therapists to score 100 university students [58]. Intra-rater reliability was not assessed, while inter-rater kappa values ranged from 0.18 to 0.60, representing slight to moderate agreement. As information related to within-rater variation was lacking, no judgment can currently be made as to the usefulness of the movement quality version of the SEBT.

4 Injury-Prediction Ability of Movement Screens

Studies that employed a prospective cohort or case–control design and investigated the association between outcome score and injury were identified for two movement screens: the FMS and LESS. Movement screening is widely used by elite sporting organizations in an attempt to detect injury risk [2]. Given this, it is important that the efficacy of movement screens in achieving this goal is understood. That only two of the ten identified screens have any supporting evidence as to their association with injury risk demonstrates that much work is needed to support this practice.

4.1 Methodological Quality Assessment for Injury-Prediction Studies

The methodological quality of each paper that investigated the ability of a movement screen to predict injury was assessed using a previously validated checklist for retrospective and prospective studies [59]. Specifically, an amended version was used as described by McCall et al. [60], since not all of the questions included in the full checklist were relevant for cohort studies. The questions excluded were only appropriate for intervention studies. For the purposes of this review, the questions included were 1, 2, 3, 5, 6, 7, 10, 11, 12, 18, 20, 21, 22, and 25 as previously used [60, 61]. Following the protocol outlined by McCall et al. [60], a percentage score was awarded for each article (see ESM Appendix S2 for raw data). A 'level of evidence' was then awarded based on the procedure outlined by the Scottish Intercollegiate Guidelines Network (SIGN) [62]. Scientific levels of evidence range from one to four according to the type of study. For example, cohort and case–control studies are level two. Levels one and two can score an additional mark of '++', '+', and '−' dependent on the judged quality and risk of bias. Percentage cut-off scores were used to determine if a paper was either of high quality with very low risk of bias (≥75 %), well conducted with low risk of bias (50–74 %), or low quality with high risk of bias (<50 %) [60]. A graded recommendation following the SIGN guidelines was given for each of the two movement screens that have had their injury predictive value investigated. The assignment of the graded recommendation was based on the levels of evidence of the relevant studies and the considered subjective judgment of the present authors. Graded recommendations were as follows: A: strong recommendation, B: moderate recommendation, C: weak recommendation, or D: insufficient evidence to make a specific recommendation [60].

4.2 Functional Movement Screen

A total of 18 articles were included that investigated the link between FMS score and injury risk (Table 7) [17, 19, 27, 63–77]. Ten [19, 27, 64, 70–74, 76, 77] of the 18 studies reported an association between the FMS composite score and injury. It should be noted that one of these studies appears to have reported an incorrect odds ratio (OR) based on the data presented, and the conclusions should be interpreted with caution [19]. Kiesel et al. [77] were the first to investigate the link between FMS score and injury; they followed 46 American Football players over the course of a pre-season (4.5 months). All players completed the FMS at the start of pre-season, and any subsequent injuries that met the defined criteria were recorded. These authors found that the greatest specificity and sensitivity were obtained when a cut-off score of 14 was used. Specificity and sensitivity are measures of the true-negative and true-positive rate, respectively [78]. In the case of this study, the specificity value displayed the proportion of non-injured athletes with a score >14, while the sensitivity value displayed the proportion of injured athletes with a score ≤14. The closer both measures are to a value of 1, the more robust the tool as a predictive instrument. An OR of 11.67 for those scoring ≤14 compared with those scoring >14 was reported by Kiesel et al. [77], and this suggests a significant association between FMS composite score and injury risk. The specificity and sensitivity values were 0.91 and 0.54, respectively. This revealed that, while the proportion of true negatives to false negatives was high, the proportion of true positives to false positives was relatively even. Despite a very large OR of 11.67, only around half of the subsequent injuries were predicted by an FMS score of ≤14. Interestingly, this seminal article by Kiesel et al. [77] is often cited within the scientific literature and explains why a cut-off score of 14 is commonly used when researching the link between FMS and injury risk. Seven articles [27, 64, 70, 71, 73, 74, 76] have since replicated the finding that individuals achieving an FMS composite score of ≤14 have an increased likelihood of experiencing an injury; however, the degree of the relationship varies between studies. Differences in the number of participants, length of follow-up period, and sport/occupation of participants may have contributed to the inconsistencies in strength of relationship between FMS score and injury likelihood. In contrast, eight of the identified studies found no link between FMS composite score and injury risk [17, 63, 65–69, 75]. However, three of these studies [68, 69, 75] utilized very small sample sizes and this may explain the lack of association between FMS score and injury. There may simply have been too few injuries among the participants during the follow-up periods for any association to be observed. As such, the findings of these three studies [68, 69, 75] should carry minimal weight when making any judgment about the predictive value of the FMS.

Due to the inconsistency in findings, the graded recommendation for the FMS is ‘D’. A number of factors contribute to the ambiguity of the collective findings. First, the definition of injury was not consistent among the identified articles. Indeed, this is a common issue in sports medicine at large [79]. Kiesel et al. [77] classified an injury as membership of the injured reserve group and a time-loss of 3 weeks—presumably meaning that only relatively serious injuries were recorded. No details of injured reserve membership criteria or details of the specific injuries experienced were provided. In contrast, O’Connor et al. [76] defined injury as any damage to the body during training that resulted in an individual seeking medical care. This broad definition could have encompassed very minor injuries. McGill et al. [75] only considered back injuries that resulted in missed game play. Such variability in the classification of injuries makes it difficult to compare the results between each study. Similarly, the length of the follow-up period varied widely between studies, with the shortest reported window of observation being 6 weeks and the longest being 2 years [74, 75]. In some instances, the precise length of the injury-tracking period was not specified [19, 27, 67, 70]. It has been previously recommended that epidemiological sports injury studies should follow participants for at least 1 year, as this allows sufficient time for accumulation of exposure and injury events [79, 80]. Unfortunately, most of the identified studies followed participants for less than this time period, and this should be a consideration for future research.

Other relevant considerations that have been ignored by the vast majority of studies are accounting for exposure time and training load. These represent very influential confounding variables that are essential to drawing meaningful conclusions from future prospective studies. Interestingly, a number of populations were investigated by the included studies: athletes, military personnel, elite police officers, and firefighters. For instance, amongst the athlete group, individuals ranged from recreationally active to elite professionals. Given the range of occupations and performance levels of participants, it is perhaps to be expected that an inconsistent relationship between FMS score and injury should be observed when all studies are viewed collectively. The injury patterns between sports and occupations differ [81–83], hence the predictive value of the FMS may not be consistent across all populations. The use of the FMS composite score has been questioned since it is not a unitary construct and, as a result, may be a misleading value [84, 85]. Instead, it has been proposed that using the individual sub-test scores when analyzing FMS performance may be preferable. However, as is shown in Table 7, of the 18 prospective studies, ten reported an association between the composite score and injury likelihood, so it should not be disregarded entirely.

4.3 Landing Error Scoring System

Two studies investigating the link between LESS score and injury were identified through the literature search (Table 7) [31, 86]. Both studies prospectively screened participants before tracking them over the course of a sporting season. Smith et al. [31] did not report any

Table 7 Studies that investigated the relationship between movement screen scores and injury

Movement screen and references	Participant information			Association between scores and injury	Quality score (%)	Level of evidence
	Sample size	Occupation/sport	Age, years[a]			
FMS						
Bardenett et al. [63]	77 M, 90 F	Cross country, American Football, soccer, swimming, tennis and volleyball athletes (high school)	15.2	No association between composite score and injury	87	2++
Garrison et al. [64]	160	Swimming/diving, rugby and soccer athletes (NCAA Division I)	17–22	Score ≤13 = OR 9.52 (95% CI: 4.16–21.79)	80	2++
Hotta et al. [17]	84 M	Middle- and long-distance runners (collegiate)	20.0 ± 1.1	No association between composite score and injury Runners scoring ≤3 on the deep squat and active straight leg raise components = OR of 9.7 (95 % CI 2.1–44.4)	80	2++
McGill et al. [65]	53 M	Elite task force police officers	37.9 ± 5.0	No association between composite score and injury	80	2++
Teyhen et al. [66]	188 M	US army rangers	23.3 ± 3.7	No association between composite score and injury	73	2+
Warren et al. [67]	89 M, 78 F	Basketball, cross country, American Football, golf, T&F, tennis, volleyball, soccer and swimming/diving athletes (NCAA Division I)	20.6 ± 1.6 (injured) 20.0 ± 1.4 (non-injured)	No association between composite score and injury	80	2++
Zalai et al. [68]	20 M	Soccer players (professional)	23.0 ± 3.0	No association between composite score and injury Players who suffered an ankle injury received a lower score for the hurdle step sub-test ($p < 0.05$) Players who suffered a knee injury received a lower score for the deep squat sub-test ($p < 0.05$)	73	2+
Dossa et al. [69]	20 M	Ice-hockey players (elite youth)	16–20	No association between composite score and injury	80	2++
Kiesel et al. [70]	238 M	American Football players (professional)	Unknown	Injured vs. non-injured groups' mean scores 16.9 vs. 17.4 ($p < 0.05$) Score ≤14 = RR 1.87 (95 % CI 1.20–2.96) Players with at least one asymmetry had an RR of 1.80 (95 % CI 1.11–2.74)	60	2+
Knapik et al. [71]	770 M, 275 F	Coast guard cadets	18.1 ± 0.7 (M) 17.9 ± 0.7 (F)	M: score ≤11 = RR 1.64 (95 % CI 1.17–2.32) F: score ≤14 = RR 1.93 (95 % CI 1.27–2.95)	73	2+
Letafatkar et al. [19]	50 M, 50 F	Soccer, handball, and basketball players (recreational)	22.6 ± 3.0	Score <17 = OR 4.7	73	2+
Shojaedin et al. [72]	50 M, 50 F	Soccer, handball and basketball players (recreational)	22.6 ± 3.0	Score ≤17 = OR 4.7	53	2+
Butler et al. [73]	108	Firefighters	Unknown	Score ≤14 = OR 8.31 (95 % CI 3.2–21.6)	60	2+
Lisman et al. [74]	874 M	Marine officer candidates	22.4 ± 2.7	Score ≤14 = OR 2.04 (95 % CI 1.32–3.15)	60	2+
McGill et al. [75]	14 M	Basketball players (collegiate)	20.4 ± 1.6	No association found between composite score and injury	73	2+
O'Connor et al. [76]	874 M	Marine officer candidates	18–30	Score ≤14 = RR 1.5 ($p < 0.05$)	67	2+
Chorba et al. [27]	38 F	Soccer, volleyball, and basketball players (NCAA Division II)	19.2 ± 1.2	Score ≤14 = OR 3.85 (95 % CI 0.98–15.13)	80	2++

Table 7 continued

Movement screen and references	Participant information			Association between scores and injury	Quality score (%)	Level of evidence
	Sample size	Occupation/sport	Age, years[a]			
Kiesel et al. [77]	46 M	American Football players (professional)	Unknown	Injured vs. non-injured groups' mean scores 14.3 vs. 17.4 ($p < 0.05$) Score ≤14 = OR 11.67 (95 % CI 2.47–54.52)	53	2+
LESS						
Padua et al. [86]	348 M, 481 F	Soccer players (elite youth)	13.9 ± 1.8	ACL injured vs. non-injured groups' mean scores 6.2 vs. 4.4 ($p < 0.05$) Score ≥5 = RR 10.7 for indirect and non-contact ACL injury	73	2+
Smith et al. [31]	29 M, 63 F	Lacrosse, soccer, basketball, American Football, field hockey, gymnastics (high school/collegiate)	18.3 ± 2.0	No association between score and non-contact ACL injury	87	2++

ACL anterior cruciate ligament, *CI* confidence interval, *F* female, *FMS* functional movement screen, *LESS* Landing Error Scoring System, *M* male, *NCAA* National Collegiate Athletic Association, *OR* odds ratio, *RR* risk ratio, *T&F* track and field, *2+* well-conducted study with low risk of bias, *2++* high-quality study with very low risk of bias

[a] Data are presented as mean ± standard deviation or range

significant relationship between LESS score and subsequent injury, whereas Padua et al. [86] did. Those ranked by Smith et al. [31] as 'poor' (scoring >6) displayed an OR of 3.62 compared with those ranked as 'excellent' (scoring ≤4) but the 95 % CI crossed 1 (0.87–15.11), indicating that the groups most likely did not differ in their risk of injury. However, Smith et al. [31] only included grade III (complete tear) non-contact anterior cruciate ligament (ACL) injuries in their analysis and, as such, it is not clear if the LESS score was associated with any other type of injury. The LESS protocol involves whole body movement, so the outcome score may potentially display an association with other injury types. The apparent lack of connection between the LESS and ACL injury reported by Smith et al. [31] is surprising, since the screen assesses the degree of knee valgus and flexion during landing, which are both relevant factors to both patellofemoral pain and ACL injury [87]. It was suggested that the narrow range of recorded scores (only 0–11 out of a possible 19) could have contributed to the lack of association with injury [31]. The authors also postulated that the screen may have superior predictive ability with regard to injury among less well-trained or less physically mature individuals undergoing rapid neuromuscular development. This may be due to differences in proprioceptive awareness and strength among these groups compared with more physically mature, well-trained individuals. This theory is somewhat supported by the findings of Padua et al. [86], who observed an almost 11-fold greater risk of ACL injury among individuals with scores of ≥5 compared with those scoring <5. The average age of the participants followed by Padua et al. [86] was 14 years compared with 18 years for the cohort observed by Smith et al. [31]. It may be that the LESS does have some injury-predictive ability but only amongst young populations and in certain sports. Further research is required among both younger and older populations before any firm conclusions can be made regarding that suggestion. Despite a theoretical link between the LESS and lower body injury, especially ACL injury, the evidence is currently ambiguous. As only two studies have prospectively investigated the ability of the LESS to predict injury and they reported conflicting results, the graded recommendation for this movement screen is 'D'.

5 Limitations and Recommendations for Future Research

When interpreting the results of the identified articles, it is important the reader be cognizant of a number of common limitations. The majority of reliability studies were categorized as methodologically poor. While the ICC or kappa scores reported often indicated good to excellent agreement within and between raters, the true value of these findings can be questioned because of the aforementioned methodological quality of the studies. In future studies investigating the reliability of movement screens, rater information such as occupation, years of experience, and number of tests performed should be included to allow for a more thorough interpretation of the results. In addition, larger sample sizes would help improve the methodological quality of future reliability studies. Similarly, future studies investigating the ability of movement screens to predict injury should clearly define what an 'injury' is and state the length of the observation period to allow contextual appraisal of the results. In some sports, such as soccer and rugby, established guidelines for injury reporting already exist [79, 80]. None of the studies investigating the link between movement screening score and injury reported or accounted for the exposure time of the participants. This is a crucial point that must be considered by future studies; without this information, a significant confounding variable is being ignored. All else being equal, the less time a player spends training and playing, then the less opportunity they have to get injured. Readers are not currently able to determine from the current research whether individuals with supposed poorer movement ability are actually at increased risk of injury because of that or simply because their exposure time is greater. Another issue to consider is that if the individual responsible for recording injuries knows the movement screening scores, then an element of bias may exist. Ideally, the individual recording the injury occurrence should be blinded to the outcome of the movement screen.

6 Conclusion

The majority of movement screens identified through the literature search lack a substantial evidence base in relation to both their reliability and their ability to predict injury. However, due to its extensive research base, the FMS is the only movement screen that has consistently demonstrated good intra- and inter-rater reliability. In addition, some studies have suggested possible predictive ability with regard to injury risk for the FMS and LESS; however, this is not a unanimous finding. Based purely on the reported ICCs and kappa values, all identified screens appear to have good reliability with the exception of the various single-leg squat screens and the SEBT movement quality screen. Further research is warranted to verify the initial reliability values for the identified movement screens, since the evidence base is still limited and the majority of the identified reliability studies were classified as methodologically poor. None of the identified movement screens have enough supporting evidence to justify them being heralded as injury prediction tools. Overall, movement screening may be useful for practitioners to enhance their holistic knowledge of an athlete, but it seems the subjectivity of scoring makes it difficult to apply these results to injury prediction with any degree of certainty.

Compliance with Ethical Standards

Funding No funding was provided that contributed to the development of this manuscript. Robert McCunn and Hugh Fullagar are supported by 'Science and Health in Football' scholarships funded by the DAAD (German Academic Exchange Service).

Conflict of interest Robert McCunn, Tim Meyer, Hugh Fullagar, and Karen aus der Fünten declare that they have no conflicts of interest with the content of this review. Ian McKeown declares that he is the primary developer of the athletic ability assessment (AAA) movement screen; however, he does not stand to receive financial gain as a result. Additionally, the AAA is neither a patented commodity nor is it protected by copyright.

References

1. Mendez-Villanueva A, Buchheit M. Football-specific fitness testing: adding value or confirming the evidence? J Sports Sci. 2013;31(13):1503–8.
2. McCall A, Carling C, Nedelec M, et al. Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. Br J Sports Med. 2014;48(18):1352–7.
3. McGill S, Frost D, Andersen J, et al. Movement quality and links to measures of fitness in firefighters. Work. 2013;45(3):357–66.
4. Frost D, Andersen J, Lam T, et al. The relationship between general measures of fitness, passive range of motion and whole-body movement quality. Ergonomics. 2013;56(4):637–49.
5. van Mechelen W, Hlobil H, Kemper HC. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. Sports Med. 1992;14(2):82–99.
6. Cook G, Burton L, Hoogenboom BJ, et al. Functional movement screening: the use of fundamental movements as an assessment of function-part 2. Int J Sports Phys Ther. 2014;9(4):549–63.
7. Giblin S, Collins D, Button C. Physical literacy: importance, assessment and future directions. Sports Med. 2014;44(9):1177–84.
8. Cook G, Burton L, Hoogenboom BJ, et al. Functional movement screening: the use of fundamental movements as an assessment of function—part 1. Int J Sports Phys Ther. 2014;9(3):396–409.
9. Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professionals. Phys Ther Sport. 2003;4(3):122–8.
10. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26(4):217–38.
11. Smith CA, Chimera NJ, Wright NJ, et al. Interrater and intrarater reliability of the functional movement screen. J Strength Cond Res. 2013;27(4):982–7.
12. Parenteau-G E, Gaudreault N, Chambers S, et al. Functional movement screen test: a reliable screening test for young elite ice hockey players. Phys Ther Sport. 2014;15(3):169–75.
13. Onate JA, Dewey T, Kollock RO, et al. Real-time intersession and interrater reliability of the functional movement screen. J Strength Cond Res. 2012;26(2):408–15.
14. Shultz R, Anderson SC, Matheson GO, et al. Test–retest and interrater reliability of the functional movement screen. J Athl Train. 2013;48(3):331–6.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
16. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651–7.
17. Hotta T, Nishiguchi S, Fukutani N, et al. Functional movement screen for predicting running injuries in 18–24 year-old competitive male runners. J Strength Cond Res. 2015. [Epub ahead of print].
18. Gulgin H, Hoogenboom B. The functional movement screening (fms)$^{\mathrm{TM}}$: an inter-rater reliability study between raters of varied experience. Int J Sports Phys Ther. 2014;9(1):14–20.

19. Letafatkar A, Hadadnezhad M, Shojaedin S, et al. Relationship between functional movement screening score and history of injury. Int J Sports Phys Ther. 2014;9(1):21–7.
20. Elias JE. The inter-rater reliability of the functional movement screen within an athletic population using untrained raters. J Strength Cond Res. 2013. [Epub ahead of print].
21. Gribble PA, Brigle J, Pietrosimone BG, et al. Intrarater reliability of the functional movement screen. J Strength Cond Res. 2013;27(4):978–81.
22. Frohm A, Heijne A, Kowalski J, et al. A nine-test screening battery for athletes: a reliability study. Scand J Med Sci Sports. 2012;22(3):306–15.
23. Klusemann MJ, Pyne DB, Fay TS, et al. Online video-based resistance training improves the physical capacity of junior basketball athletes. J Strength Cond Res. 2012;26(10):2677–84.
24. Teyhen DS, Shaffer SW, Lorenson CL, et al. The functional movement screen: a reliability study. J Orthop Sports Phys Ther. 2012;42(6):530–40.
25. Butler RJ, Plisky PJ, Kiesel KB. Interrater reliability of videotaped performance on the functional movement screen using the 100-point scoring scale. Athl Train Sports Health Care. 2012;4(3):103–9.
26. Schneiders AG, Davidsson A, Hörman E, et al. Functional movement screen normative values in a young, active population. Int J Sports Phys Ther. 2011;6(2):75–82.
27. Chorba RS, Chorba DJ, Bouillon LE, et al. Use of a functional movement screening tool to determine injury risk in female collegiate athletes. N Am J Sports Phys Ther. 2010;5(2):47–54.
28. Minick KI, Kiesel KB, Burton L, et al. Interrater reliability of the functional movement screen. J Strength Cond Res. 2010;24(2):479–86.
29. Frost DM, Beach TA, Callaghan JP, et al. FMS™ scores change with performers' knowledge of the grading criteria—are general whole-body movement screens capturing "dysfunction"? J Strength Cond Res. 2013. [Epub ahead of print].
30. McKeown I, Ball N. Current practices of long term athlete development of junior athletes in high performance sport environments. J Aust Strength Cond. 2013;21(1):4–12.
31. Smith HC, Johnson RJ, Shultz SJ, et al. A prospective evaluation of the Landing Error Scoring System (LESS) as a screening tool for anterior cruciate ligament injury risk. Am J Sports Med. 2012;40(3):521–6.
32. Padua DA, Boling MC, Distefano LJ, et al. Reliability of the landing error scoring system-real time, a clinical assessment tool of jump-landing biomechanics. J Sport Rehabil. 2011;20(2):145–56.
33. Onate J, Cortes N, Welch C, et al. Expert versus novice interrater reliability and criterion validity of the landing error scoring system. J Sport Rehabil. 2010;19(1):41–56.
34. Padua DA, Marshall SW, Boling MC, et al. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. Am J Sports Med. 2009;37(10):1996–2002.
35. Junge T, Balsnes S, Runge L, et al. Single leg mini squat: an inter-tester reproducibility study of children in the age of 9–10 and 12–14 years presented by various methods of kappa calculation. BMC Musculoskelet Disord. 2012;13:203.
36. Crossley KM, Zhang WJ, Schache AG, et al. Performance on the single-leg squat task indicates hip abductor muscle function. Am J Sports Med. 2011;39(4):866–73.
37. Örtqvist M, Moström EB, Roos EM, et al. Reliability and reference values of two clinical measurements of dynamic and static knee position in healthy children. Knee Surg Sports Traumatol Arthrosc. 2011;19(12):2060–6.
38. Ageberg E, Bennell KL, Hunt MA, et al. Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. BMC Musculoskelet Disord. 2010;11:265.
39. Chmielewski TL, Hodges MJ, Horodyski M, et al. Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. J Orthop Sports Phys Ther. 2007;37(3):122–9.
40. Noyes FR, Barber-Westin SD, Fleckenstein C, et al. The drop-jump screening test: difference in lower limb control by gender and effect of neuromuscular training in female athletes. Am J Sport Med. 2005;33(2):197–207.
41. Barber-Westin SD, Galloway M, Noyes FR, et al. Assessment of lower limb neuromuscular control in prepubescent athletes. Am J Sport Med. 2005;33(12):1853–60.
42. Dingenen B, Malfait B, Nijs S, et al. Can two-dimensional video analysis during single-leg drop vertical jumps help identify non-contact knee injury risk? A one-year prospective study. Clin Biomech. 2015. [Epub ahead of print].
43. Munro A, Herrington L, Carolan M. Reliability of 2-dimensional video assessment of frontal-plane dynamic knee valgus during common athletic screening tasks. J Sport Rehabil. 2012;21(1):7–11.
44. Nilstad A, Andersen TE, Kristianslund E. Physiotherapists can identify female football players with high knee valgus angles during vertical drop jumps using real-time observational screening. J Orthop Sport Phys. 2014;44(5):358–65.
45. Whatman C, Hume P, Hing W. The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. Phys Ther Sport. 2013;14(3):168–74.
46. Ekegren CL, Miller WC, Celebrini RG, et al. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. J Orthop Sports Phys Ther. 2009;39(9):665–74.
47. Gwet KL. Handbook of inter-rater reliability; the definitive guide to measuring the extent of agreement among multiple raters. 3rd ed. Gaithersburg: Advanced Analytics, LLC; 2012.
48. Myer GD, Ford KR, Hewett TE. Tuck jump assessment for reducing anterior cruciate ligament injury risk. Athl Ther Today. 2008;13(5):39–44.
49. Herman DC, Weinhold PS, Guskiewicz KM, et al. The effects of strength training on the lower extremity biomechanics of female recreational athletes during a stop-jump task. Am J Sport Med. 2008;36(4):733–40.
50. Söderman K, Werner S, Pietilä T, et al. Balance board training: prevention of traumatic injuries of the lower extremities in female soccer players? A prospective randomized intervention study. Knee Surg Sports Traumatol Arthrosc. 2000;8(6):356–63.
51. Herrington L, Myer GD, Munro A. Intra and inter-tester reliability of the tuck jump assessment. Phys Ther Sport. 2013;14(3):152–5.
52. Dudley LA, Smith CA, Olson BK, et al. Interrater and intrarater reliability of the tuck jump assessment by health professionals of varied educational backgrounds. J Sports Med (Hindawi Publ Corp). 2013;2013:Article ID 483503.
53. McKeown I, Taylor-McKeown K, Woods C, et al. Athletic ability assessment: a movement assessment protocol for athletes. Int J Sports Phys Ther. 2014;9(7):862–73.
54. Parsonage JR, Williams RS, Rainer P, et al. Assessment of conditioning-specific movement tasks and physical fitness measures in talent identified under 16-year-old rugby union players. J Strength Cond Res. 2014;28(6):1497–506.
55. Reid DA, Vanweerd RJ, Larmer PJ, et al. The inter and intra rater reliability of the Netball Movement Screening Tool. J Sci Med Sport. 2014;18(3):353–7.
56. Tarara DT, Hegedus EJ, Taylor JB. Real-time test–retest and interrater reliability of select physical performance measures in physically active college-aged students. Int J Sports Phys Ther. 2014;9(7):874–87.
57. Plisky PJ, Rauh MJ, Kaminski TW, et al. Star excusrsion balance test as a predictor of lower extremity injury in high school basketball players. J Orthop Sports Phys Ther. 2006;36(12):911–9.

58. Ness BM, Taylor AL, Haberl MD, et al. Clinical observation and analysis of movement quality during performance on the star excursion balance test. Int J Sports Phys Ther. 2015;10(2):168–77.
59. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52(6):377–84.
60. McCall A, Carling C, Davison M, et al. Injury risk factors, screening tests and preventative strategies: a systematic review of the evidence that underpins the perceptions and practices of 44 football (soccer) teams from various premier leagues. Br J Sports Med. 2015;49(9):583–9.
61. Freckleton G, Pizzari T. Risk factors for hamstring muscle strain injury in sport: a systematic review and meta-analysis. Br J Sports Med. 2013;47(6):351–8.
62. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001;323(7308):334–6.
63. Bardenett SM, Micca JJ, DeNoyelles JT, et al. Functional movement screen normative values and validity in high school athletes: can the FMS be used as a predictor of injury? Int J Sports Phys Ther. 2015;10(3):303–8.
64. Garrison M, Westrick R, Johnson MR, et al. Association between the functional movement screen and injury development in collegiate athletes. Int J Sports Phys Ther. 2015;10(1):21–8.
65. McGill S, Frost D, Lam T, et al. Can fitness and movement quality prevent back injury in elite task force police officers? A 5-year longitudinal study. Ergonomics. 2015. [Epub ahead of print].
66. Teyhen DS, Shaffer SW, Butler RJ, et al. What risk factors are associated with musculoskeletal injury in US army rangers? A prospective prognostic study. Clin Orthop Relat Res. 2015;473(9):2948–58.
67. Warren M, Smith CA, Chimera NJ. Association of the functional movement screen with injuries in division I athletes. J Sport Rehabil. 2015;24(2):163–70.
68. Zalai D, Panics G, Bobak P, et al. Quality of functional movement patterns and injury examination in elite-level male professional football players. Acta Physiol Hung. 2015;102(1):34–42.
69. Dossa K, Cashman G, Howitt S, et al. Can injury in major junior hockey players be predicted by a pre-season functional movement screen—a prospective cohort study. J Can Chiropr Assoc. 2014;58(4):421–7.
70. Kiesel KB, Butler RJ, Plisky PJ. Prediction of injury by limited and asymmetrical fundamental movement patterns in american football players. J Sport Rehabil. 2014;23(2):88–94.
71. Knapik JJ, Cosio-Lima LM, Reynolds KL, et al. Efficacy of functional movement screening for predicting injuries in coast guard cadets. J Strength Cond Res. 2014;29(5):1157–62.
72. Shojaedin SS, Letafatkar A, Hadadnezhad M, et al. Relationship between functional movement screening score and history of injury and identifying the predictive value of the FMS for injury. Int J Inj Contr Saf Promot. 2014;21(4):355–60.
73. Butler RJ, Contreras M, Burton LC, et al. Modifiable risk factors predict injuries in firefighters during training academies. Work. 2013;46(1):11–7.
74. Lisman P, O'Connor FG, Deuster PA, et al. Functional movement screen and aerobic fitness predict injuries in military training. Med Sci Sports Exerc. 2013;45(4):636–43.
75. McGill SM, Andersen JT, Horne AD. Predicting performance and injury resilience from movement quality and fitness scores in a basketball team over 2 years. J Strength Cond Res. 2012;26(7):1731–9.
76. O'Connor FG, Deuster PA, Davis J, et al. Functional movement screening: predicting injuries in officer candidates. Med Sci Sports Exerc. 2011;43(12):2224–30.
77. Kiesel K, Plisky PJ, Voight ML. Can serious injury in professional football be predicted by a preseason functional movement screen? N Am J Sports Phys Ther. 2007;2(3):147–58.
78. Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. BMJ. 1994;308(6943):1552.
79. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. Br J Sports Med. 2006;40(3):193–201.
80. Fuller CW, Molloy MG, Bagate C, et al. Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. Br J Sports Med. 2007;41(5):328–31.
81. Almeida SA, Williams KM, Shaffer RA, et al. Epidemiological patterns of musculoskeletal injuries and physical training. Med Sci Sports Exerc. 1999;31(8):1176–82.
82. Wolf BR, Ebinger AE, Lawler MP, et al. Injury patterns in Division I collegiate swimming. Am J Sports Med. 2009;37(10):2037–42.
83. Bathgate A, Best JP, Craig G, et al. A prospective study of injuries to elite Australian rugby union players. Br J Sports Med. 2002;36(4):265–9.
84. Kazman JB, Galecki JM, Lisman P, et al. Factor structure of the functional movement screen in marine officer candidates. J Strength Cond Res. 2014;28(3):672–8.
85. Li Y, Wang X, Chen X, et al. Exploratory factor analysis of the functional movement screen in elite athletes. J Sports Sci. 2015;33(11):1166–72.
86. Padua DA, DiStefano LJ, Beutler AI, et al. The landing error scoring system as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. J Athl Train. 2015;50(6):589–95.
87. Myer GD, Ford KR, Di Stasi SL, et al. High knee abduction moments are common risk factors for patellofemoral pain (PFP) and anterior cruciate ligament (ACL) injury in girls: is PFP itself a predictor for subsequent ACL injury? Br J Sports Med. 2015;49(2):118–22.