Joska: Isori Kampar Jurnal | Indonesian Sport & Physical Scientist Association Journal
Vol 2 (1), 01-08, 2025.

LITERATURE REVIEW

Evaluating Functional Capacity in Athletes: A Systematic Review of Measurement Tools and Techniques

Nimrot Manalu[1*], Mei Chintia Haerani[1], Yusuf Naibaho[1], Esran Polin Harefa

Universitas Negeri Medan, Indonesia.

ABSTRACT

Objectives: This systematic review aimed to evaluate and compare the reliability, validity, and clinical utility of functional capacity assessment tools and techniques used in athletic populations across various sports disciplines.
Methods: A comprehensive search of electronic databases (PubMed, CINAHL, SPORTDiscus, Web of Science, and Cochrane Library) was conducted for studies published between January 2000 and September 2024. Studies evaluating functional capacity measurement tools in athlete populations were included. Two independent reviewers assessed methodological quality using the COSMIN checklist, and data extraction was performed using a standardized form.
Results: From 1,742 initially identified studies, 87 met the inclusion criteria, evaluating 43 distinct functional capacity assessment tools. Field-based tests demonstrated higher ecological validity but lower standardization than laboratory assessments. Sport-specific tools showed greater sensitivity for detecting performance deficits compared to generic assessments. Reliability was highest for technology-assisted measurements (ICC>0.85) compared to observational tools (ICC 0.62-0.79). The Y-Balance Test, Functional Movement Screen, and sport-specific performance batteries emerged as the most comprehensively validated tools across multiple sports.
Conclusion: While several valid and reliable tools exist for evaluating functional capacity in athletes, selection should be guided by sport-specific requirements, available resources, and assessment purpose. Multi-dimensional assessment batteries that combine quantitative performance metrics with qualitative movement pattern analysis provide the most comprehensive evaluation of athletic functional capacity. Future research should focus on establishing minimal clinically important differences and developing sport-specific normative data.

Key Words: functional assessment, athletic performance, movement screening, return-to-sport, measurement properties, performance testing, sport-specific evaluation.

Received: September 18, 2025 | Accepted: January 10, 2025 | Published: January 27, 2025

INTRODUCTION

Functional capacity in athletes represents a multidimensional construct encompassing the integrated function of neuromuscular, cardiorespiratory, and biomechanical systems that support sport-specific performance demands (Verhagen & Gabbett, 2019). The assessment of functional capacity serves numerous purposes within sports medicine and performance contexts, including injury risk identification, rehabilitation progression monitoring, return-to-sport decision-making, and performance optimization (Bompa & Buzzichelli, 2019; Manske & Reiman, 2021). With the evolution of sports science, there has been a proliferation of assessment methodologies ranging from simple field tests to sophisticated laboratory-based measurement systems, creating challenges for practitioners in selecting appropriate evaluation tools.

The concept of functional capacity extends beyond isolated physical qualities (e.g., strength, power, endurance) to include movement quality, neuromuscular control, and the integration of fundamental movement patterns that underpin sport performance (Cook et al., 2014). This holistic conceptualization necessitates evaluation approaches that can capture both quantitative performance outcomes and qualitative movement characteristics. The distinction between functional capacity assessment and traditional physical fitness testing lies in the emphasis on movement patterns and neuromuscular control that specifically relate to athletic performance and injury risk (Hewett et al., 2016). Previous reviews have examined specific aspects of functional assessment in athletic populations, including those focused on particular measurement tools (e.g., Functional Movement Screen, Y-Balance Test) or specific athletic populations (e.g., team sports, overhead athletes; Bonazza et al., 2017; Butler et al., 2013). Systematic reviews by McCall et al. (2015) and Hegedus et al. (2015) have provided valuable insights into the measurement properties of selected functional tests, but these have typically been restricted to specific categories of assessment tools or limited athletic populations.

Despite the significant body of research investigating various functional capacity assessment methodologies, existing reviews have not comprehensively synthesized evidence across the spectrum of available tools and techniques. Notably, the relative merits of field-based versus laboratory-based assessments, technology-assisted versus observational measurements, and generic versus sport-specific evaluation tools remain incompletely characterized (Bishop, 2008; Robertson et al., 2017). Furthermore, the translation of assessment

*Corresponding Author: Mei Chintia Haerani; email: meichintia0@gmail.com

findings into practical applications for training program design, rehabilitation progression, and return-to-sport decision-making remains challenging for practitioners (Meeuwisse et al., 2007). The rapid advancement of technology has introduced novel measurement approaches, including wearable sensors, force plate systems, and three-dimensional motion analysis, expanding the range of available assessment options (Camomilla et al., 2018). However, the relative advantages of these technological approaches compared to traditional field-based assessments have not been systematically evaluated across diverse athletic populations. Several notable gaps exist in the current literature regarding functional capacity assessment in athletes: 1. Limited systematic comparison of measurement properties (reliability, validity, responsiveness) across the spectrum of available assessment tools and methodologies. 2. Incomplete understanding of the contextual factors influencing the selection and interpretation of functional capacity assessments, including sport-specific considerations, developmental level, and sex differences. 3. Insufficient evidence regarding the predictive validity of functional capacity measures for injury risk, rehabilitation outcomes, and performance enhancement. 4. Inadequate standardization of test protocols, normative data, and interpretation guidelines for many functional assessment tools across different athletic populations. 5. Limited investigation of the integration of technology-enhanced measurements with traditional assessment approaches to optimize the evaluation of functional capacity. 6.Unclear guidelines for selecting appropriate assessment batteries based on specific purposes (e.g., injury screening versus performance enhancement) and resource constraints in applied settings.

The proliferation of functional capacity assessment methodologies presents both opportunities and challenges for sports medicine practitioners, strength and conditioning professionals, and researchers working with athletic populations. A comprehensive synthesis of evidence regarding the measurement properties, practical utility, and contextual considerations of available assessment tools is essential to guide evidence-based practice in this domain. By systematically evaluating the strengths and limitations of different functional capacity assessment approaches, this review aims to provide a framework for selecting appropriate evaluation strategies based on specific purposes, available resources, and contextual factors.

The integration of findings from this systematic review has significant implications for enhancing the precision of injury risk screening, optimizing rehabilitation progression, informing return-to-sport decision-making, and guiding performance enhancement interventions in athletic populations. Furthermore, identifying key gaps in the current evidence base will direct future research efforts toward addressing critical questions regarding functional capacity assessment in athletes.

The primary objectives of this systematic review were to: 1. Identify and categorize the range of tools and techniques used to assess functional capacity in athletic populations across various sports disciplines. 2. Evaluate and compare the measurement properties (reliability, validity, and responsiveness) of identified functional capacity assessment methodologies. 3. Examine the practical utility, feasibility, and implementation considerations of different assessment approaches in applied sports settings. 4. Assess the evidence regarding the predictive validity of functional capacity measures for injury risk, rehabilitation outcomes, and performance enhancement. 5. Synthesize available evidence to develop recommendations for selecting appropriate functional capacity assessment strategies based on specific purposes, available resources, and contextual factors. 6.Identify critical gaps in the current evidence base to guide future research efforts in athletic functional capacity assessment.

METHODOLOGY

Search Strategy

Web of Science, and Cochrane Library. Studies published between January 2000 and September 2024 were included to capture contemporary assessment methodologies while maintaining historical perspective on established techniques. The search strategy was developed in consultation with a health sciences librarian and included a combination of Medical Subject Headings (MeSH) terms and free-text keywords related to functional capacity, athletic populations, and measurement properties.

The complete search strategy for PubMed was as follows: ((("Athletes"[Mesh] OR athlete*[tiab] OR player*[tiab] OR sportsperson*[tiab] OR "Sports"[Mesh] OR sport*[tiab]) AND ("Physical Functional Performance"[Mesh] OR "functional capacity"[tiab] OR "functional performance"[tiab] OR "functional assessment"[tiab] OR "functional testing"[tiab] OR "functional measurement"[tiab] OR "performance test*"[tiab] OR "movement quality"[tiab] OR "movement assessment"[tiab] OR "movement screen*"[tiab])) AND ("Reproducibility of Results"[Mesh] OR valid*[tiab] OR reliab*[tiab] OR responsive*[tiab] OR "measurement properties"[tiab] OR psychometric*[tiab] OR clinimetr*[tiab] OR "measurement characteristics"[tiab] OR utility[tiab] OR feasib*[tiab])). Similar search strategies were adapted for other databases following their specific syntax requirements. Reference lists of eligible studies and relevant review articles were manually searched to identify additional studies. Citation tracking of key articles was also performed using Google Scholar to ensure comprehensive coverage of the literature.

Eligibility Criteria

Studies were included if they met the following criteria: 1. Population: Studies involving athletic populations (competitive athletes at any level or individuals regularly participating in organized sports) aged 15 years or older. 2. Intervention/Exposure: Studies evaluating one or more tools or techniques designed to assess functional capacity, defined as the integrated function of neuromuscular, cardiorespiratory, and biomechanical systems supporting sport-specific performance. 3. Outcomes: Studies reporting at least one measurement property (reliability, validity, or responsiveness) of the assessment tool(s) or comparing multiple assessment methodologies.

4. Study Design: Original research using observational (cross-sectional or longitudinal) or experimental designs. Methodological studies specifically designed to evaluate measurement properties were included. 5. Language: Studies published in English.

Studies were excluded based on the following criteria: 1. Studies focusing exclusively on isolated physical qualities (e.g., strength, power, endurance) without integration into functional movement patterns or sport-specific contexts. 2. Studies involving exclusively non-athletic populations or patients with specific medical conditions. 3. Studies evaluating assessment tools without reporting measurement properties. 4. Case reports, conference abstracts, narrative reviews, or commentaries without original data. 5. Studies published before January 2000 or after September 2024.

Study Selection Process

The study selection process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). Two independent reviewers (initials blinded for review) screened titles and abstracts of identified studies against the eligibility criteria. Full-text articles of potentially eligible studies were retrieved and independently assessed by the same reviewers. Discrepancies in study selection were resolved through discussion, with a third reviewer (initials blinded for review) consulted when consensus could not be reached.

A standardized screening form was developed based on the eligibility criteria to ensure consistency in the selection process. The form included fields for population characteristics, assessment methodology, reported measurement properties, and reasons for exclusion when applicable. The selection process was documented using a PRISMA flow diagram detailing the number of studies identified, screened, assessed for eligibility, and included in the final analysis.

Data Extraction

Data extraction was performed independently by two reviewers (initials blinded for review) using a standardized form developed specifically for this review. The form was piloted on a sample of ten studies to ensure comprehensiveness and usability before full implementation. Extracted data included: 1. Study characteristics: Author(s), publication year, country, study design, sample size, funding source. 2. Population characteristics: Sport type, competitive level, age, sex, training experience, injury history (if reported). 3. Assessment tool characteristics: Name/classification of tool, measurement domain, assessment procedure, equipment requirements, administration time, scoring method. 4. Measurement properties: Reliability (intra-rater, inter-rater, test-retest), validity (content, construct, criterion), responsiveness, measurement error, minimal detectable change, minimal clinically important difference (if reported). 5. Practical utility aspects: Feasibility, acceptability, training requirements for administrators, cost, interpretation guidelines. 6. Main findings and authors' conclusions.

When multiple publications reported on the same study population, data were extracted as a single study with the most comprehensive information used and supplemented by additional publications as necessary. Authors were contacted for clarification or additional information when essential data were missing or unclear.

Quality Assessment

Methodological quality of included studies was assessed using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) risk of bias checklist (Mokkink et al., 2018). This tool was specifically designed to evaluate the methodological quality of studies on measurement properties and has been widely used in systematic reviews of assessment tools. The checklist evaluates methodological quality across multiple domains including reliability, measurement error, content validity, structural validity, hypothesis testing, cross-cultural validity, criterion validity, and responsiveness. Two independent reviewers (initials blinded for review) assessed each study, with disagreements resolved through discussion or consultation with a third reviewer. Each measurement property was rated on a four-point scale (very good, adequate, doubtful, or inadequate) following COSMIN guidelines. The overall quality rating for each measurement property was determined by the lowest score across all relevant items ("worst score counts" principle).

Data Synthesis: Given the anticipated heterogeneity in assessment methodologies, athletic populations, and reporting of measurement properties, a narrative synthesis approach was adopted. The synthesis was structured around the following elements: 1. Categorization of assessment tools: Assessment tools were categorized based on measurement domain (e.g., balance/stability, movement quality, power/agility, sport-specific performance), methodology (field-based vs. laboratory-based, technology-assisted vs. observational), and sports application (general vs. sport-specific). 2. Synthesis of measurement properties: For each assessment tool or category, evidence regarding reliability, validity, and responsiveness was synthesized considering the methodological quality of contributing studies. When multiple studies evaluated the same measurement property for a given tool, findings were summarized with consideration of study quality, consistency of results, and sample characteristics. 3. Practical utility assessment: Evidence regarding feasibility, acceptability, and implementation considerations was synthesized to inform practical recommendations. 4. Contextual considerations: Findings were synthesized considering contextual factors such as sport type, competitive level, sex, and assessment purpose to develop nuanced recommendations for specific applications.

Where sufficient homogeneous data were available, meta-analysis of measurement properties (e.g., pooled reliability coefficients) was planned using random-effects models. However, this was anticipated to be limited by heterogeneity in study methodologies and reporting formats. The strength of evidence for each measurement property of identified assessment tools was evaluated using criteria adapted from Terwee et al. (2007), considering the number of studies, methodological quality, and consistency of findings. Evidence was classified as strong, moderate, limited, conflicting, or unknown for each measurement property of each assessment tool or category.

RESULTS

Study Selection

The initial database search yielded 1,742 unique records after duplicate removal. After title and abstract screening, 312 articles were selected for full-text review. Following detailed assessment against eligibility criteria, 87 studies met inclusion criteria and were included in the final analysis. The primary reasons for exclusion during full-text review were: studies not reporting measurement properties (n=98), non-athletic populations (n=56), assessment of isolated physical qualities rather than functional capacity (n=43), and inappropriate study design (n=28).

Characteristics of Included Studies

The 87 included studies were published between 2000 and 2024, with 53 (61%) published after 2015, reflecting the growing interest in functional capacity assessment in recent years. Studies were conducted across 24 countries, with the majority from the United States (n=26), Australia (n=12), the United Kingdom (n=9), and Canada (n=8). Sample sizes ranged from 12 to 867 participants (median=46, interquartile range=28-84). Male athletes were more frequently studied (present in 78 studies) than female athletes (present in 52 studies), with 43 studies including both sexes. The included studies encompassed diverse sporting disciplines: team sports (n=42), including soccer, basketball, rugby, and volleyball; individual sports (n=32), including running, swimming, tennis, and gymnastics; and mixed sport samples (n=13). Competitive levels ranged from recreational to elite professional athletes, with collegiate/university athletes being the most commonly studied population (n=31). Study designs included methodological studies specifically designed to evaluate measurement properties (n=49), longitudinal cohort studies (n=21), cross-sectional studies (n=15), and experimental studies evaluating interventions with functional capacity assessments as outcomes (n=2).

Characteristics of Included Studies

The review identified 43 distinct assessment tools or methodologies across the included studies. These were categorized into the following domains:

Table 1. Functional Capacity Assessment Tools Categorized by Domain

No.	*Domain*	*Assessment Tools / Methods*	*Number of Studies*
1	**Movement Quality Assessment Tools** (n=12)	- Functional Movement Screen (FMS)	17
		- Movement Competency Screen (MCS)	5
		- Athletic Ability Assessment (AAA)	3
		- Performance Matrix	2
		- Sport-specific movement screening protocols	8
2	**Balance and Neuromuscular Control** (n=9)	- Y-Balance Test (YBT) & Star Excursion Balance Test (SEBT)	14
		- Balance Error Scoring System (BESS)	5
		- Dynamic Postural Stability Index (DPSI)	4
		- Force plate stability measures	6
		- Single-leg hop stabilization tests	4
3	**Multi-directional Speed and Agility** (n=8)	- Change of direction tests (505, T-test, pro-agility)	11
		- Reactive agility tests	7
		- Illinois Agility Test and modifications	4
		- Sport-specific agility batteries	5
4	**Jump and Landing Assessments** (n=7)	- Drop vertical jump with biomechanical analysis	9
		- Landing Error Scoring System (LESS)	6
		- Single-leg hop test battery	10
		- Countermovement jump with biomechanical analysis	7
5	**Sport-Specific Performance Batteries** (n=7)	- Sport-specific movement competency tests	8
		- Sport-specific conditioning tests	6
		- Technical skill assessment with physiological load	3

Measurement approaches varied considerably, with 28 tools classified as field-based assessments and 15 as laboratory-based assessments. Twenty tools incorporated technology-assisted measurement (e.g., force plates, motion capture, inertial sensors), while 23 relied primarily on observational assessment or basic timing/measurement equipment.

Reliability

Reliability was the most commonly reported measurement property, assessed in 73 studies. Intra-rater reliability was reported for 31 tools, inter-rater reliability for 28 tools, and test-retest reliability for 37 tools. The methodological quality of reliability studies was variable, with 15 studies rated as "very good," 24 as "adequate," 27 as "doubtful," and 7 as "inadequate" based on COSMIN criteria. Technology-assisted measurements generally demonstrated superior reliability compared to observational assessments. Intraclass correlation coefficients (ICCs) for technology-assisted measures ranged from 0.78 to 0.97 (median=0.88), while observational measures showed ICCs ranging from 0.62 to 0.91 (median=0.79). Field-based tests showed comparable test-retest reliability (ICC range: 0.72-0.94, median=0.84) to laboratory assessments (ICC range: 0.75-0.97, median=0.87). Among specific tools, the highest reliability was reported for force plate-derived measures (ICC>0.90), followed by the Y-Balance Test (ICC range: 0.85-0.93), and sport-specific performance batteries (ICC range: 0.78-0.92). The Functional Movement Screen showed more variable reliability (ICC range: 0.74-0.91 for total score; 0.62-0.86 for individual items), with inter-rater reliability generally lower than intra-rater reliability. Standard error of measurement (SEM) and minimal detectable change (MDC) values were reported for 23 tools, providing important clinical interpretation guidelines. However, only 12 studies explicitly addressed measurement error in relation to minimal clinically important differences.

Reliability

Validity was assessed in 67 studies, with construct validity (n=54) more frequently evaluated than criterion validity (n=23) or content validity (n=12). The methodological quality of validity studies was generally lower than reliability studies, with 9 studies rated as "very good," 22 as "adequate," 29 as "doubtful," and 7 as "inadequate." Content validity was primarily established through expert consensus methods, with 8 of 12 studies employing formal content validation procedures. Construct validity was assessed through hypothesis testing with related measures, known-groups comparison, or factor analysis. Criterion validity was typically evaluated against performance outcomes, injury occurrences, or high-fidelity laboratory measures.

Sport-specific assessment batteries demonstrated the strongest evidence for criterion validity in relation to performance outcomes (correlation coefficients ranging from 0.62 to 0.87). The Y-Balance Test, force plate stability measures, and biomechanical jump assessments showed moderate correlations with injury risk (relative risk ratios ranging from 1.6 to 4.3 for identified cut-points). The Functional Movement Screen showed inconsistent relationships with both performance outcomes (correlation coefficients ranging from -0.08 to 0.59) and injury risk (relative risk ratios ranging from 1.2 to 3.1 for various cut-points), with methodologically stronger studies generally reporting weaker associations. Movement quality assessments generally demonstrated stronger associations with injury risk than with performance outcomes. Laboratory-based biomechanical assessments showed stronger criterion validity for identifying movement deficits, while field-based tests demonstrated greater ecological validity in relation to sport performance. Measures incorporating sport-specific movements or demands generally showed stronger validity for relevant performance outcomes compared to generic assessment tools.

Responsiveness

Responsiveness—the ability to detect meaningful change over time—was the least frequently reported measurement property, assessed in only 28 studies for 18 assessment tools. Methodological quality was predominantly "doubtful" (n=16) or "adequate" (n=9), with only 3 studies rated as "very good." Effect sizes for detecting change following training interventions or during rehabilitation ranged from small (0.2-0.5) for general movement quality assessments to large (>0.8) for sport-specific performance measures. The minimal clinically important difference (MCID) was established for only 7 assessment tools, highlighting a significant gap in the literature regarding interpretation of change scores. The most responsive tools were sport-specific performance batteries and technology-assisted jump/landing assessments, which consistently detected training-induced changes with moderate to large effect sizes. The Functional Movement Screen and other observational movement quality assessments showed limited responsiveness, with small to moderate effect sizes typically reported.

Practical Utility and Implementation Considerations

Feasibility and implementation aspects were explicitly addressed in 49 studies. Equipment requirements ranged from minimal (basic observational tools) to extensive (laboratory-based biomechanical assessment), with associated cost implications. Administration time varied from 5 minutes (single-test measures) to over 60 minutes (comprehensive assessment batteries). Training requirements for test administrators were substantial for observational movement quality assessments (typically 8-20 hours of training reported for reliable administration), moderate for complex field tests, and high for laboratory-based assessments requiring technical expertise. Several studies (n=14) reported on acceptability and athlete experience, with generally positive feedback for field-based and sport-specific assessments. The most feasible tools for large-scale implementation in resource-limited settings were identified as the Y-Balance Test, modified versions of the Functional Movement Screen, and selected sport-specific field tests. These tools balanced reasonable equipment requirements, moderate training needs, and acceptable administration time while maintaining adequate measurement properties.

Contextual Factors Influencing Assessment Selection and Interpretation

Contextual factors influencing the selection and interpretation of functional capacity assessments were explicitly addressed in 38 studies. Sport-specific considerations were most commonly discussed (n=31), followed by competitive level (n=18), sex differences (n=15), and assessment purpose (n=22). Sport-specific normative data were established for 19 assessment tools across various sports, enabling more meaningful interpretation of results for specific athletic populations. Sex-specific normative values were provided for 12 tools, with significant differences observed in movement patterns and performance metrics between male and female athletes in most studies examining this factor. The influence of competitive level was documented for 14 assessment tools, with higher-level athletes typically demonstrating better movement quality, greater stability, and superior performance on sport-specific measures compared to lower-level competitors. However, the discriminative ability of assessment tools varied considerably across competitive levels and sporting disciplines.

Assessment purpose emerged as a critical factor in tool selection, with different tools recommended for injury screening (prioritizing movement quality and stability measures), rehabilitation monitoring (emphasizing responsive measures with established MDCs), return-to-sport decision-making (focusing on sport-specific functional tests), and performance enhancement (utilizing measures closely related to sport performance)

DISCUSSION

This systematic review synthesized evidence from 87 studies evaluating 43 distinct functional capacity assessment tools across diverse athletic populations. The findings reveal a complex landscape of measurement options with varying strengths, limitations, and contextual considerations. Several key themes emerged from this comprehensive analysis that have important implications for both

research and practice in athletic assessment. First, the trade-off between standardization and ecological validity represents a fundamental consideration in functional capacity assessment. Laboratory-based assessments utilizing sophisticated technology demonstrated superior measurement precision and reliability but often lacked direct relevance to sport-specific demands. Conversely, field-based and sport-specific assessments showed greater ecological validity and practical utility but typically exhibited more variable measurement properties. This highlights the importance of purposeful selection based on the intended application, with laboratory assessments potentially more suitable for detailed analysis of movement mechanics and field tests more appropriate for assessing functional performance in context. Second, the integration of quantitative and qualitative assessment approaches emerged as a valuable strategy for comprehensive functional capacity evaluation. Quantitative measures (e.g., time, distance, force) provided objective performance metrics with generally superior reliability, while qualitative assessments of movement quality offered insights into underlying movement patterns and potential injury risk factors. The strongest evidence supported multi-dimensional assessment batteries that combined both quantitative performance outcomes and qualitative movement evaluation, particularly when contextualized to specific sporting demands. Third, the findings revealed substantial variation in the robustness of measurement properties across assessment tools and methodologies. Reliability was generally well-established for most tools, particularly for technology-assisted measurements, but validity evidence was more variable and often limited by methodological weaknesses in study design. Responsiveness was notably under-investigated, with limited evidence regarding minimal clinically important differences for most assessment tools. This presents challenges for interpreting change scores in both research and clinical contexts, particularly for rehabilitation progression and return-to-sport decision-making. The current findings both confirm and extend previous reviews in this domain. Consistent with systematic reviews by McCall et al. (2015) and Hegedus et al. (2015), we found variable evidence supporting the measurement properties of commonly used functional assessment tools, with stronger evidence for reliability than for validity or responsiveness. However, by comprehensively evaluating the spectrum of available assessment approaches across diverse athletic populations, the present review provides a more nuanced understanding of the relative strengths and limitations of different measurement strategies.

Previous reviews have typically focused on specific assessment tools (e.g., Bonazza et al., 2017, on the Functional Movement Screen) or particular athletic populations (e.g., Fox et al., 2014, on team sport athletes). While these focused reviews provided valuable insights into specific aspects of functional assessment, the present review offers a broader perspective by systematically comparing different assessment approaches across diverse sporting contexts. This comparative analysis revealed important patterns in the relative utility of different measurement strategies for specific purposes and populations.

The current findings extend earlier work by Moran et al. (2017) and Tarara et al. (2016) regarding the contextual factors influencing functional capacity assessment. By systematically examining the influence of sport type, competitive level, sex, and assessment purpose across multiple studies, this review provides a more comprehensive framework for considering these contextual factors in assessment selection and interpretation. The importance of sport-specific adaptations and sex-specific normative values emerged more prominently than in previous reviews, highlighting the limitations of generic assessment approaches in diverse athletic populations.

The findings of this systematic review have several important implications for research and practice in athletic assessment. For researchers, the identified gaps in measurement properties—particularly regarding responsiveness and minimal clinically important differences—highlight critical areas for future investigation. The limited evidence for predictive validity in relation to injury risk and performance outcomes underscores the need for well-designed prospective studies with appropriate sample sizes and methodological rigor. For practitioners, the synthesis of evidence across multiple assessment tools and contexts provides a framework for selecting appropriate measurement strategies based on specific purposes, available resources, and athlete characteristics. The findings support a multi-dimensional approach to functional capacity assessment that combines quantitative performance metrics with qualitative movement evaluation, ideally contextualized to the specific demands of the sport. The superior reliability of technology-assisted measurements suggests potential advantages for objective monitoring, while the ecological validity of field-based and sport-specific assessments highlights their value for practical applications. The variability in measurement properties across different contexts emphasizes the importance of establishing sport-specific and sex-specific normative values for meaningful interpretation of assessment results. The limited transferability of generic cut-points or normative values across different athletic populations suggests caution in applying standardized criteria without contextual consideration. Furthermore, the findings highlight the need for careful selection of assessment tools based on the specific purpose, with different approaches potentially more suitable for injury screening, rehabilitation monitoring, return-to-sport decision-making, and performance enhancement.

Several limitations should be considered when interpreting the findings of this systematic review. First, despite comprehensive search strategies, some relevant studies may have been missed, particularly those published in non-English languages or outside traditional academic journals. The focus on studies reporting measurement properties may have excluded potentially valuable assessment approaches that have not undergone formal psychometric evaluation. Second, the heterogeneity in study designs, athlete populations, and reporting methods limited the potential for quantitative synthesis of measurement properties. While narrative synthesis allowed for comprehensive consideration of contextual factors and practical applications, the inability to conduct meta-analysis for many assessment tools limits the precision of conclusions regarding relative measurement properties. Third, the methodological quality of included studies was variable, with many studies demonstrating significant limitations in design or reporting. This was particularly evident for validity and responsiveness studies, where methodological weaknesses may have influenced the reported findings. The application of the COSMIN risk of bias tool helped to account for methodological quality in evidence synthesis, but the predominance of studies with "doubtful" or

"adequate" ratings highlights the need for more rigorous methodological approaches in future research. Fourth, publication bias may have influenced the available evidence, with studies demonstrating favorable measurement properties potentially more likely to be published than those with negative or null findings. While the inclusion of a broad range of assessment tools and approaches may have mitigated this bias to some extent, the potential influence on the overall evidence synthesis cannot be discounted. Finally, the focus on athletes aged 15 years and older limits the generalizability of findings to younger athletic populations, where developmental considerations may significantly influence functional capacity assessment. The underrepresentation of female athletes in many included studies also limits conclusions regarding sex-specific aspects of functional capacity assessment, despite the identified importance of this contextual factor.

CONCLUSION

This systematic review provides a comprehensive synthesis of evidence regarding the measurement properties, practical utility, and contextual considerations of functional capacity assessment tools in athletic populations. The findings demonstrate that while numerous valid and reliable assessment tools exist across diverse sporting contexts, the selection of appropriate measurement strategies should be guided by specific purposes, available resources, and athlete characteristics. The strongest evidence supports multi-dimensional assessment approaches that combine quantitative performance metrics with qualitative movement evaluation, ideally contextualized to the specific demands of the sport. Technology-assisted measurements generally demonstrate superior reliability and precision, while field-based and sport-specific assessments offer greater ecological validity and practical utility in applied settings. The Y-Balance Test, sport-specific performance batteries, and technology-enhanced jump assessments emerged as the most comprehensively validated tools across multiple contexts.

Critical gaps in the current evidence base include limited investigation of responsiveness properties, insufficient establishment of minimal clinically important differences, and inadequate development of sport-specific and sex-specific normative values for many assessment tools. The predictive validity of functional capacity measures for injury risk and performance outcomes remains incompletely characterized, highlighting the need for well-designed prospective studies in this domain.

Future research should prioritize: (1) establishing minimal clinically important differences and responsiveness properties for commonly used assessment tools; (2) developing comprehensive sport-specific and sex-specific normative databases; (3) investigating the predictive validity of functional capacity measures through prospective cohort studies; (4) examining the optimal integration of technology-enhanced and observational assessment approaches; and (5) determining the most effective assessment batteries for specific purposes across diverse athletic contexts.

By addressing these research priorities and implementing evidence-based assessment strategies informed by the findings of this review, sports medicine practitioners, strength and conditioning professionals, and researchers can enhance the precision and utility of functional capacity assessment in athletic populations. This, in turn, has the potential to improve injury risk identification, optimize rehabilitation progression, inform return-to-sport decision-making, and guide performance enhancement interventions across the spectrum of athletic participation.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

REFERENCES

Bompa, T. O., & Buzzichelli, C. A. (2019). Periodization: Theory and methodology of training (6th ed.). Human Kinetics.

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. American Journal of Sports Medicine, 45(3), 725-732. https://doi.org/10.1177/0363546516641937

Bishop, D. (2008). An applied research model for the sport sciences. Sports Medicine, 38(3), 253-263. https://doi.org/10.2165/00007256-200838030-00005

Butler, R. J., Southers, C., Gorman, P. P., Kiesel, K. B., & Plisky, P. J. (2013). Differences in soccer players' dynamic balance across levels of competition. Journal of Athletic Training, 48(6), 804-809. https://doi.org/10.4085/1062-6050-48.5.02

Camomilla, V., Bergamini, E., Fantozzi, S., & Vannozzi, G. (2018). Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review. Sensors, 18(3), 873. https://doi.org/10.3390/s18030873

Cook, G., Burton, L., Hoogenboom, B. J., & Voight, M. (2014). Functional movement screening: The use of fundamental movements as an assessment of function-part 1. International Journal of Sports Physical Therapy, 9(3), 396-409.

Fox, A. S., Bonacci, J., McLean, S. G., Spittle, M., & Saunders, N. (2014). What is normal? Female lower limb kinematic profiles during athletic tasks used to examine anterior cruciate ligament injury risk: A systematic review. Sports Medicine, 44(6), 815-832. https://doi.org/10.1007/s40279-014-0168-8

Hegedus, E. J., McDonough, S. M., Bleakley, C., Baxter, D., & Cook, C. E. (2015). Clinician-friendly lower extremity physical performance tests in athletes: A systematic review of measurement properties and correlation with injury. British Journal of Sports Medicine, 49(10), 642-648. https://doi.org/10.1136/bjsports-2014-094094

Hewett, T. E., Myer, G. D., Ford, K. R., Paterno, M. V., & Quatman, C. E. (2016). Mechanisms, prediction, and prevention of ACL injuries: Cut risk with three sharpened and validated tools. Journal of Orthopaedic Research, 34(11), 1843-1855. https://doi.org/10.1002/jor.23414

Isnandar, M., Alhumary, F. M., Krzykarist, K., Zega, R., Hutajulu, R. B., Putri, R. J. A., … & Hofmeister, M. (2025). Integrating sports massage into pre-training routines can enhance athletic preparedness and performance capacity in kabaddi players. INSPIREE: Indonesian Sport Innovation Review, 6(01), 11-18. https://doi.org/10.53905/inspiree.v6i01.137

Manske, R., & Reiman, M. (2021). Functional testing in human performance (2nd ed.). Human Kinetics.

McCall, A., Carling, C., Davison, M., Nedelec, M., Le Gall, F., Berthoin, S., & Dupont, G. (2015). 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. British Journal of Sports Medicine, 49(9), 583-589. https://doi.org/10.1136/bjsports-2014-094104

Meeuwisse, W. H., Tyreman, H., Hagel, B., & Emery, C. (2007). A dynamic model of etiology in sport injury: The recursive nature of risk and causation. Clinical Journal of Sport Medicine, 17(3), 215-219. https://doi.org/10.1097/JSM.0b013e3180592a48

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Quality of Life Research, 27(5), 1171-1179. https://doi.org/10.1007/s11136-017-1765-4

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

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Robertson, S., Kremer, P., Aisbett, B., Tran, J., & Cerin, E. (2017). Consensus on measurement properties and feasibility of performance tests for the exercise and sport sciences: A Delphi study. Sports Medicine - Open, 3(1), 2. https://doi.org/10.1186/s40798-016-0071-y

Tarara, D. T., Hegedus, E. J., & Taylor, J. B. (2016). Real-time test-retest and interrater reliability of select physical performance measures in physically active college-aged students. International Journal of Sports Physical Therapy, 11(6), 945-954.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 60(1), 34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Verhagen, E., & Gabbett, T. (2019). Load, capacity and health: Critical pieces of the holistic performance puzzle. British Journal of Sports Medicine, 53(1), 5-6. https://doi.org/10.1136/bjsports-2018-099819