Original research article

International Journal of Sports Science & Coaching

International Journal of Sports Science & Coaching
1–26

DOI: 10.1177/17479541221085236
journals.sagepub.com/home/spo
SAGE

A systematic review of the criterion validity and reliability of technical and tactical field-based tests in soccer

Filipe Manuel Clemente[1,2], Gibson Praça[3], Rafael Oliveira[4,5,6], Rodrigo Aquino[7], Rui Araújo[8,9], Rui Silva[1], Hugo Sarmento[10], and José Afonso[8]

Abstract

Background: Field-based tests are important since they can guarantee ecological validity. Although most such tests are designed to assess physical fitness, tactical and technical dimensions are also determinants in a full battery of tests.

Objectives: This systematic review aimed to (a) summarize the criterion validity and reliability of technical and tactical field-based tests applied in soccer and (b) list valid and reliable technical and tactical field-based tests.

Methods: A systematic review of EBSCO, PubMed, Scielo, SPORTDiscus, and Web of Science databases was performed according to PRISMA guidelines. The eligibility criteria included (i) population (soccer players from any age group, competitive level, or sex); (ii) exposure (exposure to technical and/or tactical field-based tests covering individual technical skills, tactical behaviours, tactical knowledge, and decision-making in the field); (iii) comparator (a laboratory test and/or a criterion field-based test); (iv) outcome (measures of reliability); and (v) no restrictions concerning study design. The synthesis of results focused on the outcomes considered in the eligibility criteria.

Results: A total of 5008 titles were identified, of which 21 met the eligibility criteria. Eight studies' analyses centred on tactical tests. Of those, six studies tested the same instrument (FUT-SAT), one study examined divergent thought, and one study used a soccer-specific skills test Among the 10 included studies involving technical tests, only two repeated the same test (the Loughborough Soccer Passing Test). Three studies included tests concurring with physical ones. According to the main findings, the eight studies that used tactical tests revealed that the outcomes obtained from the tests and observers were reliable. The 10 studies that used technical tests, as well as the three concurrent studies, revealed similarly high reliability. Of the eight studies that used tactical tests, seven involved youth participants. Furthermore, three of the 10 studies that conducted technical tests were done on adults. Of the concurrent tests, three involved youth athletes. Only one study including a tactical analysis presented values for validity, while three others provided information about validity. The four studies provided information about test validity.

Conclusions: Overall, the selected technical and tactical tests, most of which were carried out on youth players, revealed their validity and reliability. Although most studies focused on reliability (which was confirmed regardless of

Reviewers: Stefan Altmann (Karlsruhe Institute of Technology, Germany)
Matthias Kempe (University of Groningen, The Netherlands)

[1]Escola Superior Desporto e Lazer, Instituto Politécnico de Viana do Castelo, Viana do Castelo, Portugal
[2]Instituto de Telecomunicações, Delegação da Covilhã, Lisboa, Portugal
[3]Sports Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[4]Sports Science School of Rio Maior–Polytechnic Institute of Santarém, Rio Maior, Portugal
[5]Life Quality Research Centre, Rio Maior, Portugal
[6]Research Center in Sport Sciences, Health Sciences and Human Development, Quinta de Prados, Edifí-cio Ciências de Desporto, Vila Real, Portugal
[7]LabSport, Department of Sports, Center of Physical Education and Sports, Federal University of Espírito Santo, Vitória, Espírito Santo, Brazil
[8]Centre for Research, Education, Innovation and Intervention in Sport, Faculty of Sport of the University of Porto, Porto, Portugal
[9]University Institute of Maia – ISMAI
[10]Faculty of Sport Sciences and Physical Education, University of Coimbra, Research Unit for Sport and Physical Activity, Coimbra, Portugal

Corresponding author:
Filipe Manuel Clemente, Escola Superior Desporto e Lazer, Instituto Politécnico de Viana do Castelo, Rua Escola Industrial e Comercial de Nun'Álvares, 4900-347 Viana do Castelo, Portugal.
Email: filipe.clemente5@gmail.com

the test included), criterion validity was also observed in the few studies that analysed such factors. Thus, the tests currently proposed for assessing technical and tactical behaviours possess acceptable criterion validity and reliability.

Keywords
Agility, association football, skill, youth sport

Introduction

Exercise tests are often used in soccer training to assess players' physiological, physical, tactical, and technical functions.[1,2] Testing can provide information about the current status of players, identify evolution patterns using repeated measures, and provide useful information to enhance individualized training.[3] Despite the importance of multifactorial testing in soccer, physiological and physical testing is still the most common approach.[4] This fact has been confirmed in several publications,[5,6] as well as narrative[7–9] and systematic reviews[10–12] on physiological and physical testing.

Although soccer performance and technical skills are clearly related to tactical behaviours,[13] there is still a paucity of literature reviewing the characteristics of existing tests used to assess soccer players' technical skills[14,15] and tactical behaviours.[16] One of the few studies related to technical skills was published in 2011, which fundamentally summarized tests related to passing, dribbling, controlling, and shooting the ball.[14] More recently, a systematic review dedicated to skill-related performance in soccer[15] was published. This study primarily described methodological concerns regarding the use of scarce representativeness analysis to assess the skill-related performance of players. It also addressed the absence of information regarding the definitions of technical variables and the context of data collection, which significantly compromises the reproducibility of the studies. Regarding tactical behaviours, a review published in the present year also disclosed some concerns regarding the quality of the criteria used in most articles to evaluate tactical tests.[16]

One of the difficulties in guaranteeing the repeatability and quality of technical skills tests and tactical behaviours tests is that they are executed in real scenarios to improve ecological validity. For example, physiological and physical tests can be executed in the laboratory and in controlled scenarios and environments,[1] thus increasing their repeatability (i.e. the agreement between the results of successive measurements in similar scenarios)[17] and reproducibility (i.e. the degree of agreement between the results of experiments conducted by different players at different locations using different instruments).[18] However, applying laboratory conditions to ensure repeatability and reproducibility decreases a test's ecological validity (i.e. how well a test predicts behaviours in real game scenarios).[19] Even in physical and physiological field-based soccer tests (which are conducted in the field of a game, thus having high ecological validity), they can be executed in scenarios in which repeatability and reproducibility can be ensured based on the low representativeness of the game.

Ensuring the representativeness of the game while increasing ecological validity may be important to coaches, specifically if they intend to alter the training process based on the outcomes and, ultimately, improve players' on-field performance.[20] Thus, field-based tests are needed to characterize each player's status while considering the importance of ensuring trustworthy data and outcomes.

This situation presents a challenge, as the reviews mentioned above utilized either technical skill tests[15] or tactical behaviours tests.[16] Typically, technical skill tests are assessed in two different ways: (i) using analytical drills in which the number of elements is counted based on the time of execution or efficacy of the actions (e.g. the Loughborough Soccer Passing Test or the Loughborough Soccer Shooting Test[21]) and (ii) using observational instruments to classify and quantify technical actions executed during matches (e.g. soccer-specific behaviours measurement tools[22]). Tactical tests usually depend on instruments that classify and quantify behaviours primarily focused on fundamental tactical principles that should be applied while playing, as opposed to those established by experts (e.g. system of tactical assessment in soccer; FUT-SAT[23]).

The benefits of employing these observation-based instruments to test technical and tactical performance are the ability to observe behaviours in match or simulated-match scenarios. This benefit is evident from a recent scoping review disclosing that small-sided games allow coaches to differentiate players of different skill levels.[24] However, one of the challenges associated with such instruments is ensuring the reliability of the data extracted (i.e. the consistency with which results are obtained).[25]

Moreover, when using these observational-based instruments, inter-rater reliability (i.e. the degree to which different raters give consistent scores for the same event), intra-rater reliability (i.e. the degree of consistency for each observer at different moments of observation), and test-retest reliability (i.e. the consistency of a measure from one time to another, depending on the construct of the test) should be ensured.[26] Thus, one of the critical steps when presenting original tests is reporting their

reliability, as doing so provides some consistency for supporting the use of the test.[27]

Typical reports of reliability consider systematic bias and random error as measurements of error.[28] Criterion validity is another factor that can be used when a gold-standard test or instrument is present to help determine the level of agreement between the outcomes of a new test and a well-established test Although many reviews related to technical skills and tactical tests have been carried out,[14–16] no review has exclusively focused on the reliability levels of the abovementioned tests in soccer. However, because the ecological validity of field-based tests could improve variability, reliability should be ensured, as otherwise, the quality of the data could be compromised.[29,30]

Meanwhile, analytical-based technical skills assessments are controllable and, thus, relatively easy to implement.[14] Therefore, lower bias is expected when applying such tests. Considering that talent identification and long-term development require the integration of technical skills and tactical behaviours tests,[31] this systematic review aims to (a) summarize the criterion validity and reliability of technical and tactical field-based tests applied in soccer and (b) list valid and reliable technical and tactical field-based tests relevant to players of different age groups.

Methods

This systematic review followed the Cochrane Collaboration guidelines.[32] The systematic review strategy was conducted according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines.[33] The PICOS (Population, Intervention/exposure, Comparator, Outcome, Study design) was defined as follows: (P) soccer players from any age-group, sex or competitive level; (I) exposed to a technical and/or tactical field-based tests; (C) laboratorial methods and/or concurrent technical/tactical field-based tests; (O) performance measures in the technical and/or tactical field-based tests; (S) no restrictions were placed in terms of study design. The protocol was published in INPLASY (International Platform of Registered Systematic Review and Meta-analysis Protocols) with the identification number of INPLASY202130101 and DOI10.37766/inplasy2021.3.0101.

Eligibility criteria

The inclusion and exclusion criteria can be found in Table 1.

The screening of the title, abstract and reference list of each study to locate potentially relevant studies was independently performed by the two authors (FMC and JA). Additionally, they reviewed the full version of the included papers in detail to identify articles that met the selection criteria. An additional search within the reference lists of the included records was conducted to retrieve additional relevant studies. A discussion was made in the cases of discrepancies regarding the selection process with a third author (RO). Possible errata for the included articles were considered.

Information sources and search

Electronic databases (Cochrane Library, EBSCO, PubMed, Scielo, Scopus, SPORTDiscus, Web of Science) were searched for relevant publications on April 15, 2021. Keywords and synonyms were entered as free text terms in title and/or abstract: [(soccer OR football) AND ("Validity" OR "Accuracy" OR "Reliability" OR "Precision" OR "Variability" OR "Repeatability" OR "Reproducibility" OR "Consistency") AND ("techni*" OR "tactic*" OR "skill*" OR "decision*" OR "psychomotor"). Additionally, the reference lists of the studies retrieved were manually searched to identify potentially eligible studies not captured by the electronic searches. In Scielo, all code lines were searched for all fields, instead of being limited to title or abstract.

Data extraction

A data extraction was prepared in Microsoft Excel sheet (Microsoft Corporation, Readmon, WA, USA) in accordance with the Cochrane Consumers and Communication Review Group's data extraction template.[34] The Excel sheet was used to assess inclusion requirements and subsequently tested for all selected studies. The process was independently conducted by the two authors (JA and FMC). Any disagreement regarding study eligibility was resolved in a discussion. Full text articles excluded, with reasons, were recorded. All the records were stored in the sheet.

Data items

The following information was extracted from the included original articles: (i) criterion validity measure (e.g. typical error, absolute mean error); and (ii) reliability measure (e.g. intraclass correlation test, coefficient of variation). In the case of reliability, we considered: (i) inter- or intra-raters reliability (which represents the consistency of scores among raters and for the same observer in different moments of the same observation); (ii) test-retest reliability (which represents the consistency of the outcome measured in different period for the same population); (iii) internal consistency (for the case of correlation between multiple items in a test that are intended to measure the same construct).

Additionally, the following data items were extracted: (i) number of participants (n), age-group (youth, adults or both), sex (men, women or both), training level (untrained, trained); (ii) characteristics of the experimental approach to the problem, procedures and settings of each study and the

Table 1. Elegibility criteria.

	Inclusion Criteria	Exclusion Criteria
Population	Soccer players from any age group, competitive level or sex	Sports other than soccer (e.g. futsal, beach football, basketball, handball), physical education students, general population.
Intervention/ exposure	Exposed to technical and/or tactical field-based tests covering the: (i) individual technical skills (e.g. pass, dribble); (ii) tactical behaviour; (iii) tactical knowledge; (iv) decision-making in the field.	Physical tests (e.g. aerobic performance, anaerobic performance); laboratory-based tests (e.g. room, simulated games, watching videos). Tests were performed indoor, not in soccer field.
Comparator	A laboratorial test and/or a concurrent field-based test	No laboratorial tests and/or concurrent technical/ tactical field-based tests (e.g., match analysis, laboratory decision-making analysis, SSGs or other drills if GPS or similar devices are merely used to quantify distances
Outcome	Measures of criterion validity (e.g. typical error; mean absolute error) and/or reliability (e.g. intraclass correlation test; coefficient of variation).	No statistics related to criterion validity and reliability. Tactical-technical tests performed under non-comparable conditions (e.g. fatigued vs non-fatigued, pre- post some intervention, etc.). Just one tactical and/or technical test, without reliability calculation and/or without comparison with another tactical/technical test Just 1 tactical and/or technical test, without reliability assessments, compared with physical tests. Relevant tests are not compared among them; they are only used to compare different age groups, or different playing levels.
Study design	No restrictions with regard to study design. In case of more than one time-point, both (pre and post) will be considered.	—
Other	Only original and full-text studies written in English, Portuguese, Spanish, Italian and French.	Written in languages other than English, Portuguese, Spanish, Italian and French. Other article types than original (e.g. reviews, letters to editors, trial registrations, proposals for protocols, editorials, book chapters and conference abstracts).

details of the tests and implementation. Additionally, information about ecological validity, if available, will be described. The instruments were divided into tactical and technical ones. Instruments focused on tactical assessment were defined as those in which the measure was taken in game-based scenario as tactics can be understood as the result of ongoing interactions between the two opposing teams.[35] We have only included those situations occurring in real-match or simulated match scenarios. Thus, situations developed in the laboratory (e.g. virtual reality, screen-based decisions, eye-tracker in the laboratory) were excluded since the focus was on field- based tests. On the other hand, instruments in which the assessment was based on movements outside the game context were classified as technical instruments.

Methodological assessment

The methodological assessment was conducted using the Downs and Black checklist[36] with a modified version.[37] The instrument contains 27 items evaluating 5 dimensions: (1) reporting; (2) external validity; (3) bias (intervention and outcome measurement); (4) confounding (cohort selection bias); and (5) power. Each question is scored as 0 (poor quality) or 1 (good quality), with exception of question 5 ("clear description of principal confounders") that is scored from 0 (not satisfying) to 2 (fully satisfying).[37] Therefore, a maximum of 28 points can be scored for each article. The quality of the article was classified based on the following thresholds[38]: (i) poor (< 14 points); (ii) fair (14–18 points); (iii) good (19–23 points); and (iv) excellent (24–28 points). Three of the authors (JA, RA and HS) assessed the articles.

Results

Study identification and selection

The searching of databases identified a total of 5008 titles (Cochrane Library: 142; EBSCO: 670; PubMed: 467; Scielo: 1; Scopus: 1371; SPORTDiscus: 793; Web of Science: 1564). These studies were then exported to

reference manager software (EndNote™ 20.0.1, Clarivate Analytics, Philadelphia, PA, USA). Duplicates (2245 references) were subsequently removed either automatically or manually. The remaining 2763 articles were screened for their relevance based on titles and abstracts, resulting in the removal of a further 2730 studies. Following the screening procedure, 33 articles were selected for in depth reading and analysis. After reading full texts, a further 14 studies were excluded due to not meeting one or more of the eligibility criteria. In specific, the following studies were excluded based on the fact of being interventions with no particular focus in testing field-based tests and/or criterion validity and/or reliability of the tests.[39–42] The following studies were excluded based on the fact of not being tested in specific field-based scenario, thus being conducted in laboratorial and/or controlled indoor scenarios.[43–50] One study was excluded based on the fact of not presenting clear information about the population, which was not possible to classify as soccer players.[51] One paper did not presented the measures related with criterion validity and/or reliability.[52]

Therefore, 19 studies were considered eligible for inclusion.[21,22,52–68] Analysis of the reference lists of the 19 included studies revealed five potentially relevant titles that had not emerged during the initial searches. Of these, two were deemed eligible for inclusion[69,70] (Figure 1).

Characteristics of the included studies

The characteristics of the studies included in the systematic review can be found in Tables 2 and 3 for tactical and technical tests, respectively. From the total number of 21 studies, seven studies analysed tactical,[61,64–68] and 14 analyse technical tests.[21,23,53–60,62,63,69]

Methodological quality

The overall methodological quality of the studies can be found in Table 4. The 21 included studies were classified as fair (14–18 points). Almost all studies failed in to score in the following items: 8; 14; 15; 23; 24; 25; 27. Those items are associated with the interventions, thus explaining the possibility of scoring.

Results of individual studies

The results concerning the reliability of tactical tests are summarized in Table 5. One study analysed both the validity and reliability of the tests,[23] while six studies only analysed reliability.[61,64–67,70] One study analysed professional athletes,[66] while the other six studies analysed young athletes.[23,61,64,65,67,70] Information concerning the validity and reliability of the studies was extracted according to the conclusions of the studies analysed.

The most commonly reported instrument was the system of tactical assessment in soccer (FUT-SAT). This instrument was introduced by Israel Teoldo in 2011[23] to classify the behaviours of players based on fundamental tactical principles. The FUT-SAT is an observational-based instrument that classifies behaviours based on their execution and the efficacy of their execution. Five offensive and five defensive principles are observed in two-game phases. The observations are made in 3 versus 3 scenarios.

One study[67] also assessed game-related skills in a 5 versus 5 scenario in which the quality of individual offensive and defensive skills, as well as the quality of cooperative offensive and defensive skills with teammates, were assessed using an observational instrument. These items are classified by observers based on specific criteria; these can be observed in the original article.[67]

Finally, the game-test[66] was employed, in which open opportunities are evaluated using specific criteria defined to classify tactical thinking as convergent or divergent. This classification is rated by experts based on the players' solutions.[66]

The results concerning the validity and reliability of technical tests can be found in Table 6. One study analysed professional and non-professional adults athletes,[21] while the study of Mirkov et al. also analysed professional athletes.[60] The remaining nine studies analysed young participants.[53–59,63,69] In one of these studies, young recreational participants were also analysed.[59] Another one of these studies also analysed non-professional adults.[54] Another two studies analysed technical tests but were not included in Table 5 because validity and reliability were not tested.[64,70] The validity and reliability of the studies were determined according to the conclusions offered in the analysed studies.

Three studies utilized concurrent technical tests in young participants[55,57,62] (Table 7). The concurrent tests were based on the interaction between technical skills and physical quality, making it impossible to split both. For example, in one study,[62] the change-of-direction abilities determined by different tests (zig-zag, ajax test, Illinois) were combined with progression with the ball. In another study,[55] repeated sprint ability was combined with dribbling. Finally, another study[57] employed a change-of-direction test with the ball and a specific reactive test with the ball.

Discussion

The technical and tactical assessments of soccer players are important to improving talent identification and long-term development in youth soccer players, as well as managing soccer training according to teams' technical and tactical needs.[71] Therefore, tests that measure the technical and tactical capabilities of soccer players have been gaining interest among researchers and practitioners. However, the tests used to assess youth and adult soccer players' technical skills and tactical behaviours must be valid and reliable.[27] Most of the 21 studies included in this systematic review were primarily concerned with testing the reliability of

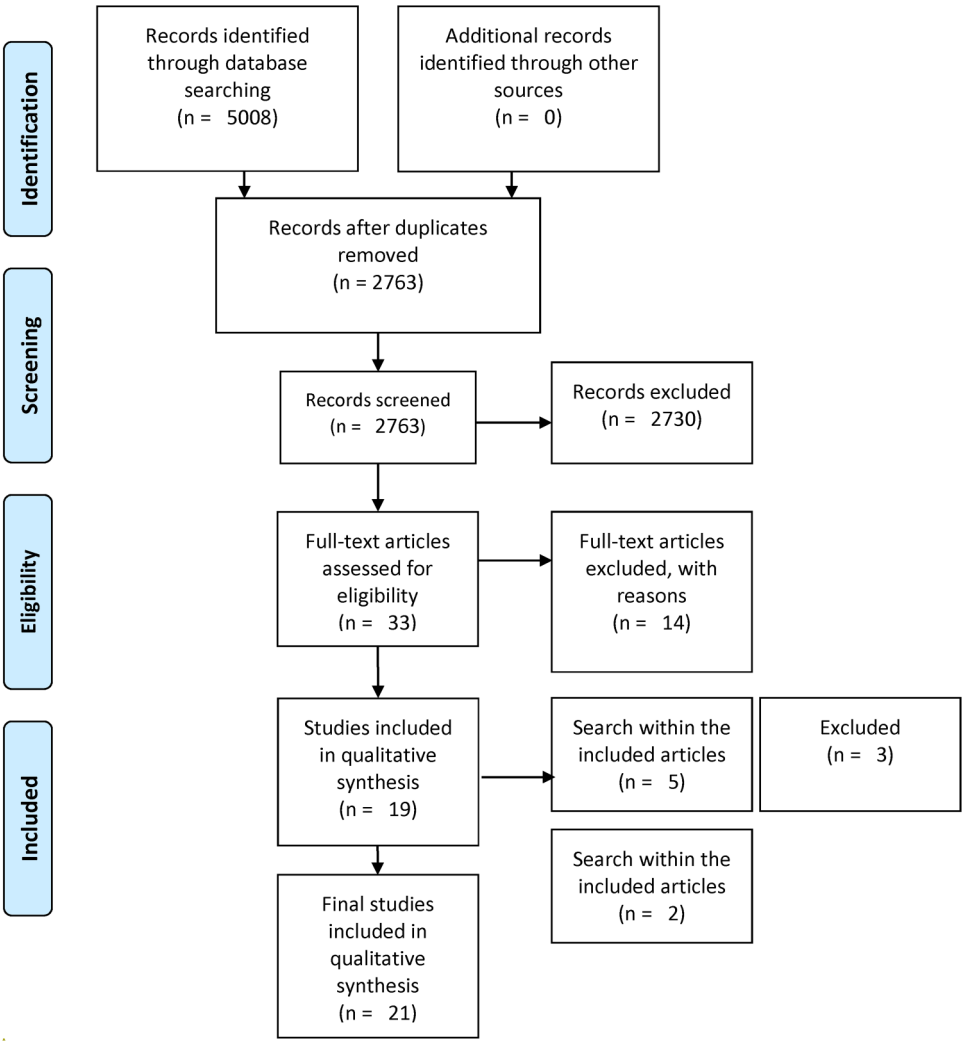

Figure 1. PRISMA flow diagram.

technical and tactical tests and instruments. Table 8 summarizes the main evidence provided by these studies and provides an overview of the findings of this systematic review.

Tactical tests

Tactical tests included in the current systematic review relied on observational analyses, namely FUT-SAT[23,61,64,65,68,70] and the game-related skills and game-test.[66,67] Because observational analyses depend on intra- and inter-observer reliability, most of these studies reported the inter-rater reliability of observers; however, they did not properly evaluate the reliability of the test (i.e. test-retest reliability).

The FUT-SAT's validity was confirmed, and this test consistently reported adequate levels of inter-rater reliability.[23,61,64,65,68,70] Values of inter-observer reliability varied from 0.76 and 0.99,[23,66,67] indicating that different observers consistently provide similar results regardless of the period of analysis. Despite the confidence of evidence regarding the inter-rater reliability, research should focus on confirming internal consistency since the test contains different items. Moreover, inter-method reliability should be researched further, mainly considering that the methods may vary in practical scenarios. Although research on different instruments has included youth populations[61,65,68,70] and adult professional soccer players,[66] more attention should be given to different experimental conditions to test how consistent scores are when there is variation in the scenarios observed.

Two studies[66,67] used observational instruments instead of FUT-SAT to classify players from a tactical point of view. One study used a game-based assessment while applying a 5 versus 5 scenario in which the observers classified the quality of individual offensive and defensive skills and the quality of cooperation with their teammates in attacking and defensive moments.[67] This instrument follows an observational approach to classify behaviours, such as scoring goals, creating goal-scoring opportunities,

Table 2. Study characteristics of tactical tests.

Study	N | Age | Sex	Competitive level	Test environment	Familiarization with the test as part of study design	Days between tests (if assessed)	Test Objectives and/or outcomes extracted	Comparison with the other test (if assessed)
[64]	U12, N: 15 Age: 10.7 ±0.5 Sex: M	Young Amateur	Week post-season training period; natural grass field and by following the usual training schedule of the team (05:00 PM; temperatures from 27 ∘C (80.6 ∘F) to 33 ∘C (91.4 ∘F), without rain or wind)	ND	1 week	FUT-SAT: Aim: Penetration, Offensive Coverage, Depth Mobility, Width and Length, Offensive Unity) Delay, Defensive Coverage, Balance, Concentration, Defensive Unity Outcomes: Offensive Index of Tactical Performance; Tactical Ranking; Sum of the Ranking of the Specific Soccer Skills Tests; Technical Ranking. Soccer Specific Motor Skills Tests: Ball control; Shooting; Passing; Dribbling	Test-retest
[23]	U11, N= 144 Age: ND Sex: M U12, N= 224 Age: ND Sex: M U15, N= 36 Age: NDSex: M U17, N= 36 Age: ND Sex: M	Young Amateur	ND	No familiarization was applied.	3 weeks	FUT-SAT: Aim: Penetration, Offensive Coverage, Depth Mobility, Width and Length, Offensive Unity) Delay, Defensive Coverage, Balance, Concentration, Defensive Unity Outcomes: Offensive Index of Tactical Performance; Tactical Ranking; Sum of the Ranking of the Specific Soccer Skills Tests; Technical Ranking.	NA
[55]	U17, N: 31 Age: 16.1 ±1.1 Sex: M	Young Amateur	Same weekday and at the same period. Test sessions were performed on artificial grass under similar conditions (air temperature 24.8-30.1 °C; humidity 30-41%).	31 players were assessed twice to examine reliability of the RDA test;	1 week	RDA: total time; ideal time; worst trial	NA
[65]	U14, N: 10 Age: ND Sex: M U15, N: 10	Young Amateur	ND	ND	NA	FUT-SAT: Aim: Penetration, Offensive Coverage, Depth Mobility, Width and Length, Offensive	NA

(continued)

Table 2. (continued)

Study	N | Age | Sex	Competitive level	Test environment	Familiarization with the test as part of study design	Days between tests (if assessed)	Test Objectives and/or outcomes extracted	Comparison with the other test (if assessed)
	Age: ND Sex: M					Unity) Delay, Defensive Coverage, Balance, Concentration, Defensive Unity Outcomes: Offensive Index of Tactical Performance; Tactical Ranking; Sum of the Ranking of the Specific Soccer Skills Tests; Technical Ranking.	
[66]	U21, N: 36 Age: 19–21 Sex: ND	Professional	ND	ND	NA	Test: Evaluate open opportunities Aim: Technical activities; Convergent creative actions; Divergent creative actions; Attackers creativity: Divergent Thought (I) and (II); Convergent Thought (I) and (II)	NA
[57]	U13,N: 29 Age: 11–12 Sex: M U15, N: 30 Age: 13–14 Sex: M	Young Amateur	All tests were performed in the morning hours between 10:00 and 12:00	Participants were familiarized with FS_RAG and FS_CODS a day before testing, and performed several testing trials at submaximal intensity and 1–2 at maximum	7–9 days	Tests: Sprint 10m; Sprint 20m; CMJ; Strength index; COD20Yards; FS_RAG; FS_CODS	Test-retest
[61]	U13/U15/ U17/U20, N:48 Age: ND Sex: M	Young Amateur	ND	ND	20 days	FUT-SAT: Aim: Penetration, Offensive Coverage, Depth Mobility, Width and Length, Offensive Unity) Delay, Defensive Coverage, Balance, Concentration, Defensive Unity Outcomes: Offensive Index of Tactical Performance; Tactical Ranking; Sum of the Ranking of the Specific Soccer Skills Tests; Technical Ranking.	Test-retest
[70]	U15, N: 24 Age: 14–15 Sex: M	Young Amateur	Grass surface during the same session of data collection in a random order	ND		FUT-SAT (tactical) Soccer Ability Skill Test Battery (dribble, pass and shot on goal-technical tests)	NA

(continued)

Table 2. (continued)

Study	N | Age | Sex	Competitive level	Test environment	Familiarization with the test as part of study design	Days between tests (if assessed)	Test Objectives and/or outcomes extracted	Comparison with the other test (if assessed)
[67]	U12, N:30 Age: 11.87 ± 0.23 U15, N:30 Age: 14.76 ± 0.42 Sex: M	Young Amateur	Artificial turf	ND	one-day research session (separate for each group)	Dribble Test (turns with the ball) Turning Test (control the ball along a marked route) Assessment of game-related skills in a 5 vs 5 small-sided game: quality of individual skills in attack; quality of individual skills in defense; quality of cooperation skills in attack and defense	NA
[62]	1st league, N: 50 Age: 19–33 Sex: M	Professional	ND	ND	NA	zig-zag running without a ball zig-zag running with a ball Ajax test 5 × 10 metres without a ball Ajax test 5 × 10 metres with a ball Illinois test without a ball Illinois test with a ball	NA
[68]	U11, N:60 U13, N:60 U15, N:60 U17, N:60 U20, N:60 Age: ND Sex: M	Young Amateur	ND	ND	NA	FUT-SAT: Aim: Penetration, Offensive Coverage, Depth Mobility, Width and Length, Offensive Unity) Delay, Defensive Coverage, Balance, Concentration, Defensive Unity Outcomes: Offensive Index of Tactical Performance; Tactical Ranking; Sum of the Ranking of the Specific Soccer Skills Tests; Technical Ranking.	NA

M: male; F: female; NA: non-applicable; ND: non-described; U: underFUT-SAT: System of Tactical Assessment in Soccer; RDA: repeated dribbling ability; CODS: change of direction speed; RAG: reactive agility; FS_CODS: football specific change of direction speed; FS_RAG: football reactive agility; m: metres.

Table 3. Study characteristics of technical tests.

Study	N | Age | Sex	Competitive level	Test environment	Familiarization with the test as part of study design	Days between tests (if assessed)	Test Objectives and/or outcomes extracted	Comparison with the other test (if assessed)
[22]	U12, N: 16 Age: 11.4 ± 0.5 Sex: ND	Young	Artifical grass The test occurred in a 8vs.8 game situation by two repetitions of 15 min (60 × 40m)	ND	4 weeks	S-SBMT: Transition (Positive zonal transition; no zonal transition; negative zonal transition); regaing (tackle; interception; loose ball); goal attempt (on target; off target; blocked)	Test-retest
[21]	N: 24 Age:20.2 ± 1.5 Sex: M N: 24 Age:19.9 ± 0.8 Sex: M	Professional Semi professional and ex-professional All from University Association	ND	3 attempts before the trial	1 day	LSPT: Time taken; Penalty time; Performance time. Aim: pass the ball LSST: Time taken; Shot speed; Points per shot. Aim: shoot the ball	Test-retest
[53]	U10, N = 125 Age: ND Sex: M U12, N = 125 Age: ND Sex: M U14, N = 125 Age: ND Sex: M U16, N = 125 Age: ND Sex: M	Young Amateur	All tests were conducted at the same field, at the same time of the day, after the same 20-min warm up and at an environmental temperature of 27 ± 2°C and humidity of 47 ± 3%.	ND	3-day	Zigzag test: Dribbling technique DAT: Dribbling agility	2 trails
[54]	U11, N:101 Age: ND Sex: M U14, N: 100 Age: ND Sex: M U17, N:118 Age: ND Sex: M Adults, N: 43	Young Amateur Young Amateur Young Amateur Adults Amateur	Outdoor field with natural grass after a 20-min typical warm-up. All tests were performed during the same time of the day to avoid any chronobiological effect at an environmental	players performed 2 times the tests in a random order	2-day	PAT: Number of points PVR: Total score; Number of Lights	Test-retest

(continued)

Table 3. (continued)

Study	N | Age | Sex	Competitive level	Test environment	Familiarization with the test as part of study design	Days between tests (if assessed)	Test Objectives and/or outcomes extracted	Comparison with the other test (if assessed)
	Age: ND Sex: M		temperature of 26 ± 2°C and humidity of 46 ± 3%.				
[56]	U11, N: 20 Age: 10 ± 1.0 Sex: ND U12, N: 12 Age: 11 ± 0.6 Sex: ND	Young Amateur	Same time period of the testing day and U the same climate conditions (8–10 p.m., 24.6 ± 0.1°C temperature and 30.3 ± 1.9% relative humidity) All tests were performed on artificial grass.	ND	NA	Test: MT5M; MT20M; MT15ML outcome: sprint speed Slalom run with a ball zig-zag run with a ball 20-metre sprint with a ball outcome: agility with ball	NA
[58]	U15, N: 13 Age:14.23 ± 0.75 Sex: M	Young Amateur	All tests were performed during 09:00–11:00 in the morning with 25° of temperature with constant humidity.	Each trial was repeated three times	NA	Tests: Slalom running with the ball between the markers; Sidesteps; Agility in the air; Ball handover; Rolling dribbling; Zigzag dribbling; Dribbling; Shooting with the dominant leg	NA
[59]	U13competitive N: 26 Age: 11.49 ± 0.6 Sex: M U13recreational N: 26 Age: 11.16 ± 0.9 Sex: M U19 N: 25 Age: ND Sex: M	Young Amateur Young Recreational Young Amateur	ND	During both trials, the participants were given 5 opportunities to accustom themselves to the LSPT protocol.	1 week	LSPT Aim: To complete the test as quickly as possible while incurring as few penalties as possible. Outcomes: Execution time Penalty time Total time	NA
[60]	1st league, N: 20 Age: 20.4 ± 1.8 Sex: M	Professional	The tests were conducted on a high-quality natural grass soccer field, Without precipitation, with temperature: 24–26°C.	1 practice trial was allowed for habituation.	2 separate days	Throwing-In: Power of the upper body Standing-kick: Power of the lower body and the kicking skill 10-m Sprint: ability to rapidly accelerate Flying 20-m Sprint:	NA

(continued)

Table 3. (continued)

Study	N | Age | Sex	Competitive level	Test environment	Familiarization with the test as part of study design	Days between tests (if assessed)	Test Objectives and/or outcomes extracted	Comparison with the other test (if assessed)
						sprinting ability over a short distance Running 10 × 5 m: running agility from required rapid changes in direction zigzag test: Change of direction zigzag test with the ball: Controlling the ball while changing direction Skill Index: The ratio of the results obtained from the zigzag test without and with the ball was calculated.	
[69]	U18, N: 20 Age: 17.0 ± 0.9 Sex: M	Young Amateur	testing sessions were performed between 10–12 am. S_CODS and S_RAG testing was performed on plastic turf grass	1 familiarization session	NA	FS_CODS FS_RAG	NA
[63]	N: 40 Age: 16.0 ± 1.7 (range: 13.6–18.5) Sex: M	Young Amateur	ND	60s to familiarize with each technique and rules.	NA	University of Queensland Football Skill Assessment Protocol: Passing accuracy over 20m (player's ability to pass a ball accurately) Lofted passing accuracy over 35m (player's ability to play a lofted-pass) Shooting accuracy over 20 m (shot using the in-step) Wall-passing test (player's ability to make two successive accurate passes) Dribbling speed (player's ability to dribble the ball rapidly through a marked circuit) Juggling ability (player's ability to maintain delicate control over a ball by juggling) Short-passing performance with	NA

(continued)

Table 3. (continued)

Study	N | Age | Sex	Competitive level	Test environment	Familiarization with the test as part of study design	Days between tests (if assessed)	Test Objectives and/or outcomes extracted	Comparison with the other test (if assessed)
						rebound boards at 90° (player's ability to receive a pass by bringing the ball under rapid control) Short-passing performance with rebound boards at an angle of 135 degrees (ability to receive a ball from a teammate, turn with the ball through a wide angle and play a subsequent pass to another teammate) Overall performance Maximum performance	

M: male; F: female; NA: non-applicable; ND: non-described; U: under; S-SBMT: soccer-specific behaviour measurement tool; LSPT: Loughborough Soccer Passing Test; LSST: Loughborough Soccer Shooting Test; DAT: Dribbling agility test; PAT: passing accuracy test; PVR: passing and visual recognition test; MT5M: 5 metre sprint from a standing start, MT20M: 20 metre sprint from a standing start, M15ML–flying 15 metre sprint; CMJ: counter movement jump; CODS: change of direction speed; RAG: reactive agility; FS_CODS: football specific change of direction speed; FS_RAG: football reactive agility; m: metres.

Table 4. Methodological assessment using downs and black checklist.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Score
57	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1	0	1	1	0	0	0	0	1	0	16
68	1	1	0	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	17
65	1	1	1	1	0	1	0	0	1	1	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	15
58	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	1	1	1	1	0	0	1	0	0	0	1	0	15
61	1	1	1	1	0	1	0	0	1	0	1	1	1	0	0	1	1	1	1	1	0	0	0	0	0	1	0	15
53	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	1	0	16
54	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	1	1	1	1	1	0	1	0	0	0	1	0	18
60	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	16
62	1	1	1	0	1	1	0	0	1	0	1	1	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	16
69	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	1	1	0	0	0	1	0	18
66	0	0	0	1	0	1	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	14
56	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	17
63	1	1	0	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	16
70	1	1	1	1	1	1	0	0	1	1	0	0	1	0	0	1	1	1	1	1	1	1	0	0	0	1	0	17
64	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	1	1	0	0	0	1	0	18
67	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	1	1	0	1	1	1	0	0	0	1	0	15
59	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0	1	1	1	1	1	0	0	1	0	0	1	0	15
21	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	1	0	1	1	1	1	0	0	0	0	1	0	15
55	1	1	0	1	0	1	1	0	1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	1	0	14
23	1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	0	1	0	0	0	0	0	1	0	16
22	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0	1	0	16

conquering free space, or counteracting the opponent's goal-scoring opportunities. Verbal and non-verbal communication are also assessed.[67] Although the item descriptions are not presented in the document, the reliability levels are acceptable when considering the correlation coefficient.[67] However, the coefficient of variation is above 35% for all the items. Naturally, this raises questions about repeatability, particularly because the descriptors of the items are not presented in the article,[67] making it harder to use in practice.

All studies reporting tactical assessments were based on observational analyses. However, tactical behaviours could be assessed by measuring the validity and reliability of tactical measures obtained by tracking devices. Indeed, previous studies have shown that the data provided by GPS devices seem to be reliable.[72,73] Also, different studies have proposed analysing training game-based activities (usually similar to those adopted in the testing procedures) using positional metrics.[74,75] Therefore, analysing the validity and reliability of such tactical measures in standardized testing contexts could reduce bias by eliminating the need for experts to judge players' behaviours. Also, this kind of analysis would increase the facility to gather data, as these systems provide positional data automatically and, thus, do not require time-consuming observational procedures.

Technical tests

Common methods used to report the validity of technical tests are (i) the ability to discriminate the skill levels of players of different ages[53,54] and competitive levels[14] and (ii) factor[62] and pragmatic analyses using a statistical approach.[56] However, none one the examined studies verified the relationships between technical tests' outcomes and in-game technical performance (i.e. criterion validity), highlighting the need for further research on this topic.

This analysis can provide important information about the specificity of the technical tests performed outside of game contexts. For example, a previous cohort study showed that the Loughborough Soccer Passing Test has insufficient criterion validity in elite youth soccer contexts.[50] Therefore, valid instruments that are reproducible in game contexts might be necessary to obtain a more representative and specific assessment of technical actions in soccer. One positive example is the Movement Awareness and Technical Skills (MATS) instrument.[76] This protocol includes offensive and defensive game-based performance indicators (e.g. first touch control/pass/shot; passing; shooting; running with the ball; evading opponents; tackling; intercepting the ball; forward/backward/sideways/stationary), and its face/content validity and reliability have been established.[76]

Concerning the reliability analysis, the time taken to complete the Loughborough Soccer Passing and Shooting Tests was the most repeatable metric, both in professional players and non-professional players.[21] In contrast, the reliability of the Loughborough Soccer Passing Test in the U13 recreational group was not acceptable,[59] suggesting that this test is more repeatable in older competitive players than younger recreational players. In general, the main findings of the other

Table 5. Criterion validity and/or reliability of tactical tests.

Study	Group or Competitive level	Test	Within-test ICC, r-Pearson or similar and 95%CI	Within-test CV (%) and 95%CI	Valid/Reliable Main finding
[64]	Young U12 Amateur	FUT-SAT	Correlation matrix: 0.91–0.99; ICC: 0.95–0.99	-	Reliable After one week, FUT-SAT were repeated to test reliability.
[23]	Young U11/U12/U15/ U17 Amateur	FUT-SAT	Intra observer (Kappa) Tactical Principles = 0.85, 0.97 Performance Indicators = 0.79, 0.96 Place of Action in the Game Field = 0.92, 0.98 Action Outcome = 0.86, 0.99 Inter observer Tactical Principles = 0.82, 0.99 Performance Indicators = 0.79, 0.98 Place of Action in the Game Field = 0.87, 0.99 Action Outcome = 0.86, 0.99	-	Valid/Reliable Measures used in FUT-SAT are valid and its observations for assessment of tactical behaviour taught within the process of teaching and training in soccer through their formation is reliable
[65]	Young U14/U15 Amateur	FUT-SAT	Inter (Kappa) Tactical principle = 0.988 Location of action = 1 Result of action = 1 Intra (Kappa) Tactical principle = 0.985 Location of action = 1 Result of action = 0.998	-	Reliable Both inter and intra correlation were proved to be reliable.
[66]	U21 Professional	Evaluate open opportunities	Inter-rater ICC Divergent Thought (I) = 0.88 Divergent Thought (II) = 0.83 Convergent Thought (I) = 0.97 Convergent Thought (II) = 0.93	-	Reliable Inter-rater ICC showed good reliability for all variables.
[67]	Young U12 Amateur	Soccer specific skill test	Quality of Individual Skill in Offence: ICC = 0.885; r = 0.630 Quality of Individual Skill in Defense: ICC = 0.786; r = 0.438 Quality of Cooperation Skill in Offence and Defense: ICC = 0.812; r = 0.512	Quality of Individual Skill in Offence: CV = 37.86 Quality of Individual Skill in Defense: CV = 34.12 Quality of Cooperation Skill in Offence and Defense CV = 35.03	Reliable The reliability of game-related skills assessments can be considered moderate in the case of evaluation of all judges (ICC>0.7).
	Young U15 Amateur	Soccer specific skill test	Quality of Individual Skill in Offence: ICC = 0.796; r = 0.450 Quality of Individual Skill in Defense: ICC = 0.791; r = 0.444	Quality of Individual Skill in Offence: CV = 35.12 Quality of Individual Skill in Defense: CV = 34.92	

(continued)

Table 5. (continued)

Study	Group or Competitive level	Test	Within-test ICC, r-Pearson or similar and 95%CI	Within-test CV (%) and 95%CI	Valid/Reliable Main finding
			Quality of Cooperation Skill in Offence and Defense: ICC = 0.759; r = 0.394	Quality of Cooperation Skill in Offence and Defense: CV = 34.40	
[68]	Young U11/U13/U15/ U17/ U20 Amateur	FUT-SAT	Inter-observer's agreement coefficients between = 0.81–0.91 Intra-observer's agreement coefficients between = 0.85–0.92.	-	Reliable Data reliability was assured by the Kappa of Cohen index for FUT-SAT.
[70]	Young U15 Amateur	FUT-SAT	Cronbach's alpha coefficient for the tactical principles place of action and action outcome: Inter-observer's agreement > 0.995 Intra-observer's agreement > 0.989	-	Reliable Tactical principles place of action and action outcome showed reliability.
[61]	Young U13/U15/U17/ U20 Amateur	FUT-SAT	Kappa of Cohen index Inter-observer's agreement = 0.81 Intra-observer's agreement = 0.79–0.88	-	Reliable Data reliability was assured by the Kappa of Cohen index.

U: under; ICC: intraclass correlation; r: Person correlation; CV: coefficient variation; FUT-SAT: System of Tactical Assessment in Soccer.

Table 6. Criterion validity and/or reliability of technical tests.

Study	Group or Competitive level	Test	Within-test ICC, r-Pearson or similar and 95%CI	Within-test CV (%) and 95%CI	Valid/Reliable Main finding
[22]	Young U12 Amateur	S-SBMT	*Inter-observer* Forwards zonal transition (Yule's Q = 0.753); no zonal transition (Yule's Q = 0.753); backwards zonal transition (Yule's Q = 0.975); tackle (Yule's Q = 0.704); interception (Yule's Q = 0.630); loose ball (Yule's Q = 0.481); on target (Yule's Q = 0.917); off target (Yule's Q = 0.917); blocked (Yule's Q = 0.917) *Intra-observer* (week 1) Forwards zonal transition (Yule's Q = 0.901); no zonal transition (Yule's Q = 0.877); backwards zonal transition (Yule's Q = 0.976); tackle (Yule's Q = 0.929); interception (Yule's Q = 0.852); loose ball (Yule's Q = 0.852); on target (Yule's Q = 0.917); off target (Yule's Q = 0.917); blocked (Yule's Q = 0.917) (week 4) Forwards zonal transition (Yule's Q = 0.802); no zonal transition (Yule's Q = 0.778); backwards zonal transition (Yule's Q = 0.975); tackle (Yule's Q = 0.929); interception (Yule's Q = 0.778); loose ball (Yule's Q = 0.778); on target (Yule's Q = 0.917); off target (Yule's Q = 0.917); blocked (Yule's Q = 0.917)	-	Reliable Higher levels of percentage Agreement (>85%) were found for passing = 95.1%; running with the ball = 92.6%; goal attempts = 100%; and tackles = 100% after 1 and 4 weeks of re-test
[21]	Adults Professional	LSPT	Time taken: r = 0.75; ICC = 0.75; SEM = ±1.4; CI = ±2.8 Penalty time: r = 0.37; ICC = 0.36; SEM = ±3.1; CI = ±6.2 Performance time: r = 0.43; ICC = 0.42; SEM = ±3.6; CI = ±7.1	Time taken: CV = 4.7 Performance time: CV = 11.2	Valid/Reliable Performance on both tests was more repeatable in professional players that non-professional players (with at least 1 day interval).
	Adults Non-professional	LSPT	Time taken: r = 0.70; ICC = 0.65; SEM = ±2.5; CI = ±4.9 Penalty time: r = 0.38; ICC = 0.38; SEM = ±4.7; CI = ±9.4	Time taken: CV = 8.0 Performance time: CV = 16.0	

(continued)

Table 6. (continued)

Study	Group or Competitive level	Test	Within-test ICC, r-Pearson or similar and 95%CI	Within-test CV (%) and 95%CI	Valid/Reliable Main finding
			Performance time: r = 0.51; ICC = 0.51 SEM = ±6.2; CI = ±12.3		
	Professional	Loughborough Soccer Shooting Test	Time taken: r = 0.68; ICC = 0.64; SEM = ±0.19 CI = ±0.39 Shot speed: r = 0.31; ICC = 0.30; SEM = ±4.9 CI = ±9.4 Points per shot: r = 0.32; ICC = 0.31; SEM = ±0.46 CI = ±0.93	Time taken: CV = 3.5 Shot speed: CV = 8.4 Points per shot: CV = 49.4	
	Adults; non-professional	Loughborough Soccer Shooting Test	Time taken: r = 0.51; ICC = 0.50; SEM = ±0.30 CI = ±0.59 Shot speed: r = 0.07; ICC = 0.06; SEM = ±5.6 CI = ±11.0 Points per shot: r = 0.24; ICC = 0.23 SEM = ±0.59 CI = ±1.19	Time taken: CV = 5.1 Shot speed: CV = 10.7 Points per shot: CV = 57.8	
[53]	Young U10, U12, U14 Amateur	Dribbling agility test (DAT)	Time: ICC = 0.90 RLOA = 0.969	-	Valid/Reliable DAT has good repeatability (with a 20 min interval) and can detect differences in dribbling skill between young soccer players of different ages
[54]	Young, U11 Amateur	Passing accuracy test	SEM = 0.8; LOA = -4.8; 5.6; ICC = 0.76	-	Valid/Reliable Both tests could be valuable tools to assess the accuracy of ball passing and the ability to scan the field in players of different ages.
	Young, U11 Amateur	Passing and visual recognition test	SEM = 0.6; LOA = -4.5; 4.9; ICC = 0.76	-	
	Young, U14 Amateur	Passing accuracy test	SEM = 1; LOA = -5.3; 5.7; ICC = 0.53	-	
	Young, U14 Amateur	Passing and visual recognition test	SEM = 0.5; LOA = -5.8; 5.7; ICC = 0.83	-	
	Young, U17 Amateur	Passing accuracy test	SEM = 1.2; LOA = -5.6; 4.9; ICC = 0.41	-	
	Young, U17 Amateur	Passing and visual recognition test	SEM = 0.6; LOA = -6.7; 8.2; ICC = 0.77	-	
	Adults; non-professional	Passing accuracy test	SEM = 1.8; LOA = -2.2; 3.6; ICC = 0.21	-	
	Adults; non-professional	Passing and visual recognition test	SEM = 1.5; LOA = -2.2; 3.6; ICC = 0.33	-	
[56]	Young U11, U12 Amateur	Slalom run with a ball zig-zag run with a ball 20-m sprint with a ball	Slalom run with a ball: Ilr = 0.66; α = 0.85 zig-zag run with a ball: Ilr = 0.52; α = 0.75 20-m sprint with a ball: Ilr = 0.65; α = 0.83	- - -	Valid/Reliable Motor tests assessing speed and specific agility had high level of reliability and pragmatic validity.
[58]	Young		ICC range for all variables = 0.85–0.97	-	Reliable

(continued)

Table 6. (continued)

Study	Group or Competitive level	Test	Within-test ICC, r-Pearson or similar and 95%CI	Within-test CV (%) and 95%CI	Valid/Reliable Main finding
	U15 Amateur	All test are described in table 1			Reliability was found to be excellent in 3 consecutive days.
[59]	Young, U13 Recreational	LSPT	Execution time: ICC = 0.44; CI = 0.07–0.70 Penalty time: ICC = 0.41; CI = -0.03–0.69 Total time: ICC = 0.41; CI = 0.04–0.69	-	Reliable for U19 The U19 competitive group showed the greatest number of acceptable (>0.70) ICC values for within-day and between-day reliability of the LSPT while the U13 recreational group showed less than acceptable reliability values (2 daý test – retest).
	Young, U13 Competitive	LSPT	Execution time: ICC = 0.89; CI = 0.77–0.95 Penalty time: ICC = 0.52; CI = 0.17–0.75 Total time: ICC = 0.81; CI = 0.61–0.91	-	
	U19 Competitive Amateur	LSPT	Execution time: ICC = 0.93; CI = 0.84–0.97 Penalty time: ICC = 0.70; CI = (0.43–0.86) Total time: ICC = 0.89; CI = 0.76–0.95	-	
[60]	Professional	Different tests	ICC (CI) Throwing-In = 0.87 (0.74–0.95) Standing kick = 0.76 (0.53–0.83) 10-m sprint = 0.81 (0.64–0.92) 20-m sprint = 0.93 (0.85–0.97) Running 10 × 5 m = 0.94 (0.88–0.98) Zigzag = 0.84 (0.56–0.89) Zigzag with the ball = 0.81 (0.64–0.91) Skill index = 0.89 (0.73–0.96)	CV (CI) Throwing-In = 7.1 (5.6–9.4) Standing kick = 9.2 (7.2–12.6) 10-m sprint = 3.2 (2.6–4.3) 20-m sprint = 2.1 (1.7–2.8) Running 10 × 5m = 1.2 (0.9–1.5) Zigzag = 2.5 (2.0–3.2) Zigzag with the ball = 3.3 (2.6–4.3) Skill index = 3.9 (3.1–5.4)	Reliable The results obtained suggest a high reliability for all tests (> 0.75, on 2 separate days).
[69]	Young U18 Amateur	 FS_CODS FS_RAG	ICC S_CODS = 0.92 S_RAG = 0.7–0.88	CV S_CODS = 5.85 S_RAG = 3.66–4.94	Reliable This study confirmed the high reliability of the newly developed soccer-specific tests of RAG and CODS in youth players on 2 separate days.
[63]	Young (range: 13.6–	University of Queensland Football Skill Assessment	20 m pass: ICC = 0.90; r(CI) = 0.59 (0.33–0.76) 35 m pass: ICC = 0.93; r (CI) = 0.65 (0.42–0.80) 20 m shot: ICC = 0.88; r(CI) = 0.54 (0.28–0.73)	-	Reliable This paper identifies a reliable scientific protocol that can be used to measure individual and

(continued)

Table 6. (continued)

Study	Group or Competitive level	Test	Within-test ICC, r-Pearson or similar and 95%CI	Within-test CV (%) and 95%CI	Valid/Reliable Main finding
	18.5) Amateur		Wall-pass: ICC = 0.91; r(CI) = 0.60 (0.38–0.77) Juggles: ICC = 0.95; r(CI) = 0.72 (0.53–0.84) Dribble: ICC = 0.83; r(CI) = 0.47 (0.18–0.68) Rebound 90: ICC = 0.88; r(CI) = 0.54 (0.27–0.73) Rebound 135: ICC = 0.94; r(CI) = 0.67 (0.44–0.81) Overall performance: ICC = 0.98;r(CI) = 0.87 (0.76–0.93)		composite skill performance in soccer players in a within session.

U: under; ICC: intraclass correlation; r: Person correlation; CI: confidence interval; CV: coefficient variation; RDA: repeated dribbling ability; TEM: technical error of measurement; S-SBMT: soccer-specific behaviour measurement tool; FS_CODS: football specific change of direction speed; FS_RAG: football reactive agility; CMJ: counter movement jump; m: metres; MT5M: 5 metre sprint from a standing start; MT20M: 20-metre sprint from a standing start; M15ML: flying 15-metre sprint; SEM: standard error of measurement; DAT: Dribbling agility test; RLOA: ratio limits of agreement; LOA: limits of agreement; α: Cronbach's alpha coefficient; m: metres; LSPT: Loughborough Soccer Passing Test

Table 7. Concurrent technical tests.

Study	Competitive level	Tests	Concurrent test intercorrelation matrix and coefficient correlations	Valid/Reliable Main finding
[62]	Professional	zig-zag running without a ball zig-zag running with a ball Ajax test 5 × 10 metres without a ball Ajax test 5 × 10 metres with a ball Illinois test without a ball Illinois test with a ball	Range values of intercorrelation matric for all tests = 0.25–0.63 Principal component factor analysis: agility Correlation coefficients of all the agility tests: Zig-zag running without a ball = 0.721 Zig-zag running with a ball = 0.812 Ajax test 5 × 10 metres without a ball = 0.645 Ajax test 5 × 10 metres with a ball = 0.671 Illinois test without a ball = 0.896 Illinois test with a ball = 0.751	Valid All tests have a single subject of measure, which can be defined as agility. All the tests have high projections, and with that a diagnostic validity of the first main component. The highest projection and diagnostic validity of all the applied 6 tests for evaluation of agility is shown through the Illinois test without a ball.
[55]	Youth U17 Amateur	Repeated dribbling ability	Correlation between sessions RDA total time = 0.924; RDA ideal time = 0.913; RDA worst trial = 0.813 ICC RDA total time = 0.841–0.963; RDA ideal time = 0.820–0.958; RDA worst trial = 0.611–0.910 %CV RDA total time = 3.84; RDA ideal time = 3.90; RDA worst trial = 7.53	Reliable The dribbling protocol presented reasonable reproducibility properties.
[57]	Youth U11 Amateur	Football-specific reactive agility Football-specific change-of-direction test	Inter-testing reliability FSRAG (ICC = 0.79; CV = 0.05; SEM = 0.24) FSCODS (ICC = 0.82; CV = 0.04; SEM = 0.16)	Reliable Results showed appropriate intra-testing and inter-testing reliability of the FS_RAG (ICC: 0.79) and FS_CODS as well as higher reliability was showed in U15 (ICC: 0.82–0.85) than in U13 (ICC: 0.78–0.80).

U: under; RDA: Repeated dribbling ability ; ICC: intra-class correlation test; CV: coefficient of variation; FSRAG: Football-specific reactive agility; FSCODS:.
Football-specific change-of-direction test; SEM: standard error of the mean.

Table 8. Summary of evidence regarding reliability levels of the tests.

	Studies confirming validity	Studies confirming reliability	Young	Adults
Tactical				
FUT-SAT	23	23,61,64,65,68,70	Yes	Yes
Game-related skills		67	Yes	-
Game-test (divergent and convergent)		66	-	Yes
Technical				
S-SBMT		22	Yes	-
LSPT	21	21,59	-	Yes
LSST				
Dribbling agility test	53	53	Yes	-
Passing accuracy test	54	54	Yes	No
Slalom run with a ball	56	56	Yes	-
Zig-zag run with a ball	56	56	Yes	-
20-m sprint with a ball	56	56	Yes	-
Slalom running with the ball between the markers		58	Yes	-
Sidesteps		58	Yes	-
Agility in the air		58	Yes	-
Ball handover		58	Yes	-
Rolling dribbling		58	Yes	-
Zigzag dribbling		58	Yes	-
Dribbling		58	Yes	-
Shooting with the dominant leg		58	Yes	-
Throwing-In		60	Yes	-
Standing kick		60	Yes	-
10-m sprint		60	Yes	-
20-m sprint		60	Yes	-
Running 10 × 5 m		60	Yes	-
Zigzag		60	Yes	-
Zigzag with the ball		60	Yes	-
Skill index		60	Yes	-
FS_CODS		69	Yes	-
FS_RAG		69	Yes	-
University of Queensland Football Skill Assessment		63	Yes	-
zig-zag running with a ball	62	62	-	Yes
Ajax test 5 × 10 metres with a ball	62	62	-	Yes
Illinois test with a ball	62	62	-	Yes
Repeated dribbling ability		55	Yes	-
Football-specific reactive agility		57	Yes	-
Football-specific change-of-direction test		57	Yes	-

-: not tested; S-SBMT: soccer-specific behaviour measurement tool; FS_CODS: football specific change of direction speed; FS_RAG: football reactive agility; FUT-SAT: System of Tactical Assessment in Soccer; LSPT: Loughborough Soccer Passing Test; LSST: Loughborough Soccer Shooting Test.

included protocols indicated good reliability (ICC > 0.70) regarding the outcomes of the technical tests.[53,54,56,58,60,63,69]

In contrast to tactical tests, evaluations of the reproducibility of technical tests outside game contexts do not require judges' analyses.[77] However, the researchers and coaches should ensure that the equipment (e.g. hand-held stopwatches; video-cameras) used to measure test performance (e.g. time taken in the LSPT; shot speed in the LSST) is reliable,[21,77] as this is essential to reducing bias. In addition, biological aspects (e.g. motivation, fatigue, nutritional intake, sleep quality) should be considered and controlled to avoid noise in the measurements of technical tests' outcomes.[29]

A limitation shared by several of the included studies that used technical tests is the absence of a measure of within-subject standard deviation (i.e. standard error of measurement – SEM). Only two studies reported these values.[21,54] From a practical point of view, the SEM reported for each test could help monitor individual changes. The threshold of 1.5 to 2.0 times the SEM appears to be realistic for indicating likely changes in performance outcomes.[27] For example, the SEM of the time taken to complete the Loughborough Soccer Passing Test in professional players was ± 1.4 s;[21] therefore, an observed change of at least two to three seconds in the time

professional players take during the test would indicate that a real change occurred. Further studies could provide the SEM values for other technical tests reported in the current systematic review.

Despite the abovementioned limitations, the present systematic review has some useful practical applications. For instance, the overall technical skill tests selected in this systematic review were considered valid and reliable for assessing a wide range of individual technical actions, both in youth and adult soccer players. This matter represents a step forward for practitioners who work in youth and adult soccer contexts, as it allows them to conduct soccer technical tests with confidence that they measure what they are intended to measure. They can also be sure that they are obtaining consistent and stable data on players' technical skills. Although accurate and repeatable tactical tests are lacking (thus limiting their practical use in soccer contexts), the FUT-SAT instrument was revealed as a valid and reliable instrument for assessing soccer players' tactical behaviours. However, more studies using the FUT-SAT need to be conducted on young and adult populations; so far, only one study has tested its validity and reliability in youth soccer players.

Study limitations

Our systematic review has some limitations. For instance, we focused only on the validity and reliability of tactical and technical soccer tests, which might have influenced the capacity of this work to provide information about typical values obtained by the included tests. Moreover, we have included only articles written in English, Portuguese, Spanish, Italian, and French, thus excluding potentially valuable reports published in other languages. Additionally, only original articles were included, meaning that tests reported in theses, conferences, or reports were excluded. In most cases, the tests used in the included articles were not tested for content validity, criterion validity, or parallel forms (reliability). Moreover, few studies tested different instruments for the same population, mainly to identify the sensitivity of the tests to discriminate players based on their skill level, competitive level, or age group. Additionally, no study included elite professional players, which makes it hard to identify how these tests will be handled in scenarios involving these players. Thus, future research should find ways to make tactical and technical tests easier to apply while evaluating their capacity to discriminate performance, both in youth players and professionals.

Conclusions

This systematic review assessed the criterion validity and reliability levels of various technical and tactical field-based tests applied in soccer. Most of the tests presented in the current systematic review were assessed for their reliability in youth players. However, the main evidence indicates that all tests performed by youth and adult players were reliable. Reliability was the psychometric property most often analysed across the included studies. Meanwhile, more studies confirming the content validity and criterion validity of these tests are required.

Acknowledgements

This study made part of one curricular unit of Master in Sports Training at Escola Superior de Desporto e Lazer, Instituto Politécnico de Viana do Castelo, Portugal.

Authorship Contributions

FMC lead the project, wrote and revised the original manuscript. RRC analysed and interpreted the data, wrote the statistical report and revised the original manuscript. HS and JA run the data search, performed the methodological assessment, conducted the data extraction, wrote and revised the original manuscript. RO and ZA wrote and revised the original manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Filipe Manuel Clemente: This work is funded by Fundação para a Ciência e Tecnologia/ Ministério da Ciência, Tecnologia e Ensino Superior through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020. No other specific sources of funding were used to assist in the preparation of this article.

ORCID iDs

Filipe Manuel Clemente https://orcid.org/0000-0001-9813-2842
Gibson Praça https://orcid.org/0000-0001-9971-7308
Rafael Oliveira https://orcid.org/0000-0001-6671-6229
Rodrigo Aquino https://orcid.org/0000-0002-4885-7316

References

1. Barker AR and Armstrong N. Exercise testing elite young athletes. In: *The elite young athlete [internet]*. Basel: KARGER, 2010, pp.106–125. Available from: https://www.karger.com/Article/FullText/320642.
2. Noakes TD. Implications of exercise testing for prediction of athletic performance: a contemporary perspective. *Med Sci Sport Exerc [Internet]* 1988 Aug; 20: 319–330. Available from: http://journals.lww.com/00005768-198808000-00001.
3. Myers JN and Froelicher VF. Exercise testing and prescription. *Phys Med Rehabil Clin N Am [Internet]* 1995 Feb; 6: 117–151. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1047965118304819.
4. Weldon A, Duncan MJ, Turner A, et al. Contemporary practices of strength and conditioning coaches in professional

soccer. *Biol Sport [Internet]* 2021; 38: 377–390. Available from: https://www.termedia.pl/doi/10.5114/biolsport.2021.99328.
5. Grazette N, McAllister S, Ong CW, et al. Reliability of a musculoskeletal profiling test battery in elite academy soccer players. *Mirkov D, editor. PLoS One [Internet]* 2020 Jul 23; 15: e0236341. Available from: https://dx.plos.org/10.1371/journal.pone.0236341.
6. Dugdale JH, Arthur CA, Sanders D, et al. Reliability and validity of field-based fitness tests in youth soccer players. *Eur J Sport Sci [Internet]* 2019 Jul 3; 19: 745–756. Available from: https://www.tandfonline.com/doi/full/10.1080/17461391.2018.1556739.
7. Kelly AL and Williams CA. Physical characteristics and the talent identification and development processes in male youth soccer: a narrative review. *Strength Cond J [Internet]* 2020 Dec; 42: 15–34. Available from: https://journals.lww.com/10.1519/SSC.0000000000000576.
8. Murr D, Raabe J and Höner O. The prognostic value of physiological and physical characteristics in youth soccer: a systematic review. *Eur J Sport Sci [Internet]* 2018 Jan 2; 18: 62–74. Available from: https://www.tandfonline.com/doi/full/10.1080/17461391.2017.1386719.
9. Turner A, Walker S, Stembridge M, et al. A testing battery for the assessment of fitness in soccer players. *Strength Cond J [Internet]* 2011 Oct; 33: 29–39. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00126548-201110000-00003.
10. Grgic J, Oppici L, Mikulic P, et al. Test–retest reliability of the Yo-Yo test: a systematic review. *Sport Med [Internet]* 2019 Oct 3; 49: 1547–1557. Available from: http://link.springer.com/10.1007/s40279-019-01143-4.
11. Schmitz B, Pfeifer C, Kreitz K, et al. The Yo-Yo intermittent tests: a systematic review and structured compendium of test results. *Front Physiol* 2018; 9: 870.
12. Grgic J, Lazinica B and Pedisic Z. Test-retest reliability of the 30–15 intermittent fitness test: a systematic review. *J Sport Heal Sci [Internet]* 2020 May; 10(4): 0–20. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2095254620300570.
13. Kannekens R, Elferink-Gemser MT and Visscher C. Tactical skills of world-class youth soccer teams. *J Sports Sci [Internet]* 2009 Jun; 27: 807–812. Available from: http://www.tandfonline.com/doi/abs/10.1080/02640410902894339.
14. Ali A. Measuring soccer skill performance: a review. *Scand J Med Sci Sports [Internet]* 2011 Apr; 21: 170–183. Available from: http://doi.wiley.com/10.1111/j.1600-0838.2010.01256.x.
15. Aquino R, Puggina EF, Alves IS, et al. Skill-related performance in soccer: a systematic review. *Hum Mov [Internet]* 2017 Dec 20; 18: 3–24. Available from: http://www.degruyter.com/view/j/humo.2017.18.issue-5/humo-2017-0042/humo-2017-0042.xml.
16. Rechenchosky L, Menegassi VM, Jaime MdO, et al. Scoping review of tests to assess tactical knowledge and tactical performance of young soccer players. *J Sports Sci [Internet]* 2021 Sep 17; 39: 2051–2067. Available from: https://www.tandfonline.com/doi/full/10.1080/02640414.2021.1916262.
17. Harper DGC. Some comments on the repeatability of measurements. *Ringing Migr [Internet]* 1994 Aug; 15: 84–90. Available from: http://www.tandfonline.com/doi/abs/10.1080/03078698.1994.9674078.
18. Goodman SN, Fanelli D and Ioannidis JPA. What does research reproducibility mean? *Sci Transl Med [Internet]* 2016 Jun; 8: 341ps12. Available from: https://www.science.org/doi/10.1126/scitranslmed.aaf5027.
19. Kihlstrom JF. Ecological validity and "ecological validity". *Perspect Psychol Sci [Internet]* 2021 Mar 16; 16: 466–71. Available from: http://journals.sagepub.com/doi/10.1177/1745691620966791.
20. Fernandes-Da-Silva J, Castagna C, Teixeira AS, et al. Ecological and construct validity of a repeated sprint test in male youth soccer players. *J Strength Cond Res [Internet]* Jan; 37(7): 2000–2009. Available from: https://journals.lww.com/00124278-900000000-94989.
21. Ali A, Williams C, Hulse M, et al. Reliability and validity of two tests of soccer skill. *J Sports Sci [Internet]* 2007 Nov; 25: 1461–1470. Available from: http://www.tandfonline.com/doi/abs/10.1080/02640410601150470.
22. Cobb NM, Unnithan V and McRobert AP. The validity, objectivity, and reliability of a soccer-specific behaviour measurement tool. *Sci Med Footb [Internet]* 2018 Jul 3; 2: 196–202. Available from: https://www.tandfonline.com/doi/full/10.1080/24733938.2017.1423176.
23. Costa ITD, Garganta J, Greco PJ, et al. System of tactical assessment in soccer (FUT-SAT): development and preliminary validation. *Motricidade* 2011; 7: 69–84.
24. Klingner FC, Huijgen BC, Den Hartigh RJ, et al. Technical–tactical skill assessments in small-sided soccer games: a scoping review. *Int J Sports Sci Coach [Internet]* 2021 Nov 22. Available from: http://journals.sagepub.com/doi/10.1177/17479541211049532.
25. Andrade C. Internal, external, and ecological validity in research design, conduct, and evaluation. *Indian J Psychol Med [Internet]* 2018 Sep 1; 40: 498–499. Available from: http://journals.sagepub.com/doi/10.4103/IJPSYM.IJPSYM_334_18.
26. Heale R and Twycross A. Validity and reliability in quantitative studies. *Evid Based Nurs [Internet]* 2015 Jul; 18: 66–67. Available from: https://ebn.bmj.com/lookup/doi/10.1136/eb-2015-102129.
27. Hopkins WG. Measures of reliability in sports medicine and science. *Sport Med [Internet]* 2000; 30: 1–15. Available from: http://link.springer.com/10.2165/00007256-200030010-00001.
28. Atkinson G and Nevill AM. Statistical methods For assessing measurement error (reliability) in variables relevant to sports medicine. *Sport Med [Internet]* 1998; 26: 217–238. Available from: http://link.springer.com/10.2165/00007256-199826040-00002.
29. Hopkins WG. How to interpret changes in an athletic performance test *Sportscience* 2004; 8: 1–7.
30. González-Víllora S, Serra-Olivares J, Pastor-Vicedo JC, et al. Review of the tactical evaluation tools for youth players, assessing the tactics in team sports: football. *Springerplus [Internet]* 2015 Dec 2; 4: 663. Available from: http://www.springerplus.com/content/4/1/663.
31. Koopmann T, Faber I, Baker J, et al. Assessing technical skills in talented youth athletes: a systematic review. *Sport Med* 2020; 50: 1593–1611.
32. Green S and Higgins J. *Cochrane handbook for systematic reviews of interventions*. Hoboken, NJ, USA: John Wiley & Sons, 2005.

33. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med [Internet]* 2009 Jul 21; 6, (7):e1000097. Available from: https://dx.plos.org/10.1371/journal.pmed.1000097.
34. Collaboration C. Data extraction template for included studies [Internet]. 2016 [cited 2021 Jan 2]. Available from: https://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/public/uploads/det_2015_revised_final_june_20_2016_nov_29_revised.doc.
35. Gréhaigne JF and Godbout P. Tactical knowledge in team sports from a constructivist and cognitivist perspective. *Quest* 1995; 47: 490–505.
36. Downs SH and Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Heal [Internet]* 1998 Jun 1; 52: 377–384. Available from: https://jech.bmj.com/lookup/doi/10.1136/jech.52.6.377.
37. Simic M, Hinman RS, Wrigley TV, et al. Gait modification strategies for altering medial knee joint load: a systematic review. *Arthritis Care Res (Hoboken) [Internet]* 2010; 63: 405–426. Available from: http://doi.wiley.com/10.1002/acr.20380.
38. O'Connor SR, Tully MA, Ryan B, et al. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Res Notes [Internet]* 2015 Dec 6; 8: 224. Available from: http://www.biomedcentral.com/1756-0500/8/224.
39. Altmann S, Neumann R, Ringhof S, et al. Soccer-specific agility: reliability of a newly developed test and correlates of performance. *J Strength Cond Res [Internet]* 2020 Apr 21: Ahead-of-p. Available from: https://journals.lww.com/10.1519/JSC.0000000000003635.
40. Bullock W, Panchuk D, Broatch J, et al. An integrative test of agility, speed and skill in soccer: effects of exercise. *J Sci Med Sport [Internet]* 2012 Sep; 15: 431–436. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1440244012000291.
41. Knoop M, Fernandez-Fernandez J and Ferrauti A. Evaluation of a specific reaction and action speed test for the soccer goalkeeper. *J Strength Cond Res [Internet]* 2013 Aug; 27: 2141–2148. Available from: https://journals.lww.com/00124278-201308000-00013.
42. Ljac V, Witkowski Z, Gutni B, et al. Toward effective forecast of professionally important sensorimotor cognitive abilities of young soccer players. *Percept Mot Skills [Internet]* 2012 Apr; 114: 485–506. Available from: http://journals.sagepub.com/doi/10.2466/05.10.25.PMS.114.2.485-506.
43. Bacvarevic BB, Pazin N, Bozic PR, et al. Evaluation of a composite test of kicking performance. *J Strength Cond Res [Internet]* 2012 Jul; 26: 1945–1952. Available from: https://journals.lww.com/00124278-201207000-00028.
44. Kelly A, Wilson MR, Jackson DT, et al. Technical testing and match analysis statistics as part of the talent development process in an English football academy. *Int J Perform Anal Sport [Internet]* 2020 Nov 1; 20: 1035–1051. Available from: https://www.tandfonline.com/doi/full/10.1080/24748668.2020.1824865.
45. Kutlu M, Yapici H and Yilmaz A. Reliability and validity of a new test of agility and skill for female amateur soccer players. *J Hum Kinet [Internet]* 2017 Feb 25; 56: 219–227. Available from: https://www.sciendo.com/article/10.1515/hukin-2017-0039.
46. Kutlu M, Yapıcı H, Yoncalık O, et al. Comparison of a new test For agility and skill in soccer With other agility tests. *J Hum Kinet [Internet]* 2012 Jun 1; 33: 143–150. Available from: https://www.sciendo.com/article/10.2478/v10078-012-0053-1.
47. Le Moal E, Rué O, Ajmol A, et al. Validation of the loughborough soccer passing test in young soccer players. *J Strength Cond Res [Internet]* 2014 May; 28: 1418–1426. Available from: https://journals.lww.com/00124278-201405000-00031.
48. Padrón-Cabo A, Rey E, Pérez-Ferreirós A, et al. Test–retest reliability of skill tests in the F-MARC battery for youth soccer players. *Percept Mot Skills [Internet]* 2019 Oct 30; 126: 1006–1023. Available from: http://journals.sagepub.com/doi/10.1177/0031512519866038.
49. Russell M, Benton D and Kingsley M. Reliability and construct validity of soccer skills tests that measure passing, shooting, and dribbling. *J Sports Sci [Internet]* 2010 Nov; 28: 1399–1408. Available from: http://www.tandfonline.com/doi/abs/10.1080/02640414.2010.511247.
50. Serpiello FR, Cox A, Oppici L, et al. The loughborough soccer passing test has impractical criterion validity in elite youth football. *Sci Med Footb [Internet]* 2017 Jan 2; 1: 60–64. Available from: https://www.tandfonline.com/doi/full/10.1080/02640414.2016.1254810.
51. Larkin P and O'Connor D. Talent identification and recruitment in youth soccer: recruiter's perceptions of the key attributes for player recruitment. *Sampaio J, editor. PLoS One [Internet]* 2017 Apr 18; 12: e0175716. Available from: http://dx.plos.org/10.1371/journal.pone.0175716.
52. Pedersen AV, Lorås H, Norvang OP, et al. Measuring soccer technique with easy-to-administer field tasks in female soccer players from four different competitive levels. *Percept Mot Skills [Internet]* 2014 Dec; 119: 961–970. Available from: http://journals.sagepub.com/doi/10.2466/03.30.PMS.119c31z2.
53. Bekris E, Gissis I and Kounalakis S. The dribbling agility test as a potential tool for evaluating the dribbling skill in young soccer players. *Res Sport Med [Internet]* 2018 Oct 2; 26: 425–435. Available from: https://www.tandfonline.com/doi/full/10.1080/15438627.2018.1492395.
54. Bekris E, Kounalakis S, Ispirlidis I, et al. Evaluation of ball passing and space detection skill in soccer: implementation of two new soccer tests. *Res Sport Med [Internet]* 2020 Oct 1; 28: 518–528. Available from: https://www.tandfonline.com/doi/full/10.1080/15438627.2020.1789133.
55. Duarte JP, Tavares Ó, Valente-dos-Santos J, et al. Repeated dribbling ability in young soccer players: reproducibility and variation by the competitive level. *J Hum Kinet [Internet]* 2016 Dec 1; 53: 155–166. Available from: https://www.sciendo.com/article/10.1515/hukin-2016-0019.
56. Kovacevic Z, Zuvela F and Kuvacic G. Metric characteristics of tests assessing speed and agility in youth soccer players. *Sport Mont [Internet]* 2018 Oct 1; 16: 9–14. Available from: http://www.sportmont.ucg.ac.me/?sekcija = article&artid = 1418.
57. Krolo A, Gilic B, Foretic N, et al. Agility testing in youth football (soccer)players; evaluating reliability, validity, and correlates of newly developed testing protocols. *Int J Environ Res Public Health [Internet]* 2020 Jan 1; 17: 294.

Available from: https://www.mdpi.com/1660-4601/17/1/294.
58. Majic F, Hraste M and Jelaska I. Canonical relations between coordination and technical skills in young football players. *Acta Kinesiol* 2020; 14: 88–91.
59. McDermott G, Burnett AF and Robertson SJ. Reliability and validity of the loughborough soccer passing test in adolescent males: implications for talent identification. *Int J Sports Sci Coach [Internet]* 2015 Jun; 10: 515–527. Available from: http://journals.sagepub.com/doi/10.1260/1747-9541.10.2-3.515.
60. Mirkov D, Nedeljkovic A, Kukolj M, et al. Evaluation of the reliability of soccer-specific field tests. *J Strength Cond Res [Internet]* 2008 Jul; 22: 1046–1050. Available from: https://journals.lww.com/00124278-200807000-00003.
61. Müller E, Garganta J, Santos RMM, et al. Comportamento e Desempenho Táticos: Estudo Comparativo entre Jogadores de Futebol e Futsal. *Rev Bras Ciência e Mov [Internet]* 2016 Jun 30; 24: 100–109. Available from: http://www.bibliotekevirtual.org/index.php/2013-02-07-03-02-35/2013-02-07-03-03-11/2092-rbcm/v24n02/21473-comportamento-e-desempenho-taticos-estudo-comparativo-entre-jogadores-de-futebol-e-futsal.html.
62. Tasevski Z, Gashi AI, Markovski N, et al. Factorial validity of agility tests for soccer players. *Res Phys Educ Sport Heal* 2018; 7: 35–39.
63. Wilson RS, James RS, David G, et al. Multivariate analyses of individual variation in soccer skill as a tool for talent identification and development: utilising evolutionary theory in sports science. *J Sports Sci [Internet]* 2016 Nov 26; 34: 1–13. Available from: http://www.tandfonline.com/doi/full/10.1080/02640414.2016.1151544.
64. Aquino R, Marques R, Petiot G, et al. Relationship between procedural tactical knowledge and specific motor skills in young soccer players. *Sports [Internet]* 2016 Nov 17; 4: 52. Available from: http://www.mdpi.com/2075-4663/4/4/52.
65. Sousa RB, Soares VOV, Praça GM, et al. Tactical behavior assessment in soccer: fundamental tactical principles between U-14 and U-15 players. *Rev Bras Ciência e Mov [Internet]* 2015 Jun 30; 23: 59–65. Available from: http://www.bibliotekevirtual.org/index.php/2013-02-07-03-02-35/2013-02-07-03-03-11/1125-rbcm/v23n02/11560-avaliacao-do-comportamento-tatico-no-futebol-principios-taticos-fundamentais-nas-categorias-sub-14-e-sub-15.html.
66. Kiliç Ç, Acar MF, Tok Mİ, et al. Measuring tactical creativity with the help of divergent thought: Example of TRNC U-21 football teams. In: J Hum Sport Exercise - 2020 - Winter Conferences of Sports Science [Internet]. Universidad de Alicante; 2020. Available from: http://hdl.handle.net/10045/107031.
67. Rubajczyk K and Rokita A. Relationships between results of soccer-specific skill tests and game-related soccer skill assessment in young players aged 12 and 15 years. *Trends Sport Sci* 2015; 22: 197–206.
68. Costa Id, Garganta J, Greco PJ, et al. Assessment of tactical principles in youth soccer players of different age groups. *Rev Port Ciências do Desporto [Internet]* 2010; 10: 147–157. Available from: http://www.fade.up.pt/rpcd/_arquivo/artigos_soltos/vol.10_nr.1/1.3.pdf.
69. Pojskic H, Åslin E, Krolo A, et al. Importance of reactive agility and change of direction speed in differentiating performance levels in junior soccer players: reliability and validity of newly developed soccer-specific tests. *Front Physiol [Internet]* 2018 May 15; 9: 506. Available from: http://journal.frontiersin.org/article/10.3389/fphys.2018.00506/full.
70. Praça GM, Soares VV, Matias CdS, et al. Relação entre desempenhos tático e técnico em jovens jogadores de futebol. *Rev Bras Cineantropometria e Desempenho Hum [Internet]* 2015 Feb 28; 17: 136. Available from: https://periodicos.ufsc.br/index.php/rbcdh/article/view/1980-0037.2015v17n2p136.
71. Low B, Coutinho D, Gonçalves B, et al. A systematic review of collective tactical behaviours in football using positional data. Vol. 50, Sports Medicine. Springer International Publishing; 2020. 343–385 p.
72. Akyildiz Z, Yildiz M and Clemente FM. The reliability and accuracy of Polar Team Pro GPS units. *Proc Inst Mech Eng Part P J Sport Eng Technol [Internet]* 2020 Dec 7: 175433712097666. Available from: http://journals.sagepub.com/doi/10.1177/1754337120976660.
73. Vickery WM, Dascombe BJ, Baker JD, et al. Accuracy and reliability of gps devices for measurement of sports-specific movement patterns related to cricket, tennis, and field-based team sports. *J Strength Cond Res* 2014; 28: 1697–1705.
74. Praça GM, Andrade AGP, Bredt SdG, et al. Progression to the target vs. regular rules in Soccer small-sided Games. *Sci Med Footb* 2021. https://doi.org/10.1080/24733938.2020.1869811
75. Clemente FM, Afonso J, Castillo D, et al. The effects of small-sided soccer games on tactical behavior and collective dynamics: a systematic review. *Chaos, Solitons & Fractals [Internet]* 2020 May; 134: 109710. Available from: https://doi.org/10.1016/j.chaos.2020.109710.
76. Larkin P, O'Connor D and Williams AM. Establishing validity and reliability of a movement awareness and technical skill (MATS) analysis instrument in soccer. *Int J Perform Anal Sport [Internet]* 2016 Apr 3; 16: 191–202. Available from: https://www.tandfonline.com/doi/full/10.1080/24748668.2016.11868880.
77. Filipe A, Clemente M, Aquino R, et al. Variability of internal and external loads and technical / tactical outcomes during small-sided soccer games : a systematic review. 2022; (September 2021).