Assessment for Tactical Learning in Games: A Systematic Review

Carmen Barquero-Ruiz

Facultad de Deporte, UCAM Universidad Católica San Antonio de Murcia, Spain.

José Luis Arias-Estero

Facultad de Deporte, UCAM Universidad Católica San Antonio de Murcia, Spain

David Kirk

School of Education, University of Strathclyde, UK; University of Queensland, St Lucia, Australia

Abstract

The assessment of tactics is a subject of great interest in physical education and sport pedagogy. However, the lack of knowledge of the topic and the variety of assessment instruments makes the assessment of tactics difficult. This study aimed to describe assessment in relation to tactical learning outcomes through an analysis of assessment instruments, based on variables that must be considered when using an instrument: (a) criteria definitions; (b) tactical levels; (c) indexes; (d) units of observation; (e) player/learner roles and (f) institutional contexts. Hence, the following instruments were found: Game Performance Assessment Instrument, spatial location instruments, Game Performance Evaluation Tool, Team Sport Assessment Procedure, and System of Tactical Assessment in Soccer. Building on the review's purpose, the following issues were found. First, some studies reviewed used non-validated criteria. Second, not all studies considered the three tactical levels (match level, partial forefront level and primary level). Third, the majority of the studies used indexes that masked the results. Four, the individual unit of observation was widely used to assess global tactical learning outcomes. Five, many instruments were used in contexts for which they were not validated. According to these limitations, general recommendations are proposed. First, researchers should use validated instruments as long as the characteristics of the instruments are aligned with the nature of the study. Second, it is recommended when validating an instrument to consider the following general guidelines: (1) only use validated criteria descriptions; (2) include all three tactical levels; (3) do not use indexes; (4) use the team as the unit of observation; (5) assess both defender and attacker roles; (6) develop the instrument in the same institutional context as the study context; (7) include context variables if applicable.

Keywords

Games analysis instruments, tactical learning assessment, tactical awareness, evaluation, sport pedagogy, youth sport.

Introduction

In the last decade, there has been an increase in interest in teaching games through a tactical-technical perspective (Kinnerk et al., 2018). This perspective made necessary a move away from the traditional teaching-learning-assessing approach that focused on sports technique, to another that considers techniques and tactics as two inseparable components of a player's learning (Holt et al., 2002). Considering this new approach, the assessment of tactical-technical learning components is placed in the spotlight (e.g. Catalán-Eslava et al., 2018; Morales-Belando et al., 2018). However, the focus on tactics has made assessment more difficult for coaches, teachers and researchers due to the lack of knowledge of the topic. In addition, the variety and complexity of the tactical assessment instruments increase this difficulty (Arias-Estero and Castejón, 2012; Harvey et al., 2015).

Formerly, most studies related to assessment in sport extracted the data from questionnaires or interviews (Arias-Estero and Castejón, 2014). Due to limitations with these approaches, which failed to capture the contextual factors affecting learning, it was important to adopt a more ecological approach when it came to teaching and assessing players (Holt et al., 2002). Along this line, new assessments were required to obtain information about tactical learning outcomes. According to González-Víllora et al. (2015), quality measurement instruments are required for a proper and effective assessment of tactical learning. Consequently, they summarized the different tactical instruments in soccer in order to show their main characteristics. However, they only described the instruments and their uses to assess football tactics without identifying their limitations and giving recommendations about their design and selection. Arias-Estero and Castejón (2012) highlighted the use of two principal instruments created by researchers to provide themselves and teachers with tools to assess learning outcomes in real game contexts. These were the Game Performance Assessment Instrument (GPAI, Oslin et al., 1998) and the Team Sport Assessment Procedure

(TSAP, Gréhaigne et al., 1997). However, Arias-Estero and Castejón (2012) highlighted the weaknesses of these tactical-technical assessment instruments. Furthermore, Memmert and Harvey (2008) pointed out the difficulties and solutions to be considered in researching with GPAI. The difficulties confronted by coaches, teachers and researchers in relation to assessment of learning give rise to an inaccurate use of the instruments, that could compromise their results (Kirk, 2005).

This paper focuses on assessment as a part of the teaching-learning process in games. Throughout the paper, the term assessment comprises the collection of information about players' learning during the stages of planning and teaching-learning (Veal, 1988). From this perspective, assessment is an integral part of the teaching-learning process and not an add-on. Assessment helps to identify the capacities and weaknesses of players. This information facilitates adjustment of the teaching-learning process to support players, individually and collectively, to improve their performances. Furthermore, it allows teachers and coaches to sum up what has been learned, identifying the problems still to be resolved (Desrosiers et al., 1997).

In order for an assessment instrument to be effective, it is necessary to think about the links between expected learning outcomes and assessment (Biggs, 1996). Aligning these will not only benefit the players by ensuring the validity and reliability of the assessments, but alignment also helps to ensure that the correct skills and knowledge are being assessed. Different assessment instruments measure different skills. Therefore, it could be that the instruments used to assess tactical learning outcomes are not being used appropriately according to the relation between research purposes and instruments' characteristics.

The purpose of this study is, then, to describe assessment in relation to tactical learning outcomes, based on variables that must be considered when using an assessment instrument: (a) criteria definitions - degree of openness; (b) tactical levels; (c) indexes; (d) units of observation; (e) player/learner roles and (f) institutional contexts.

Method

The first author conducted the literature search, collated the abstracts, and applied the initial inclusion criteria. The electronic databases were searched on the 17th of September 2018. They were: PubMed, Web of Science, and Google Scholar. The following terms were used: ‘tactical knowledge’, ‘tactical awareness’, ‘procedural tactical knowledge’, ‘decision making’, ‘skill performance’, ‘skill execution’ and ‘tactical behaviour’. Reference lists of included articles were searched to identify additional relevant studies.

The descriptive data analyzed in the present work were taken from manuscripts that met the following inclusion criteria: (1) research studies published between 1990 and September 2018; (2) written in English; (3) from peer-review journals; (5) that appeared in journals indexed in the Science Citation Index, Science citation Index Expanded and Social Sciences Citation Index; (6) presented conclusions related to teaching-learning process and assessment of tactical learning outcomes in school, extracurricular sport and formal sport contexts in any category of sport. Studies that used instruments designed for students’ self-rating (questionnaires, video-tests, image recognition, etc.) and interviews were excluded to prevent the results of real assessment of participants’ tactical behaviours from being confounded with the assessment of the verbalization of their tactical capabilities. Also, studies performed in electronic-sports or special populations (people with special needs or disabilities) were excluded. Moreover, neither experimental nor non-empirical articles were included. The systematic review was undertaken in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Figure 1). In total, 215 articles were retrieved from the database search and an additional 55 articles were identified through reference lists. Then, 122 articles were excluded due to duplication. Next, 78 articles were excluded because at least one of the exclusion criteria appeared in the abstract. Finally, 20 articles were omitted after full-text examination because the studies did not meet one or more of the inclusion criteria. At the end of the screening procedure, 50 articles remained for the systematic review (Table 1). Disagreements over inclusion and exclusion of articles were resolved by consensus between two investigators (authors 1 and 2).

****Figure 1****

****Table 1****

The findings of the review are discussed in five sections according to the manuscript purpose. Although the review followed an inductive analysis, the sections were determined deductively after the full-text examination as result of their relevance to assessment (criteria definitions - degree of openness, tactical levels, indexes, units of observation, player/learner roles and institutional contexts). Several of the reviewed studies provided evidence that related to multiple sections (i.e. criteria definitions and indexes, tactical levels and unit of observation, unit of observation and player role). Therefore, some of the discussion crossed sections and related some sections with others.

To assess the quality of the appraisal process, many systematic reviews adopt a protocol for assessing the quality of studies using standardized assessments. In this case, the guidelines for healthcare research were followed, in which the Cochrane Handbook for Systematic Reviews discourages the assessment of study quality in favor of assessing the risk of bias within each study. It addresses five types of bias that can occur in research. For the present systematic review, the first two authors assessed the quality of the included studies using the Cochrane risk of bias tool (Higgins & Green, 2011). Following the Cochrane Handbook for Systematic Reviews, the five domains of bias appraised are: (1) selection bias,

(2) detection bias, (3) attrition bias, (4) reporting bias, and (5) other bias. Each article was scored in each item as low (+), high (−), or unclear (?) risk of bias. Studies were considered low risk of bias when all domains were scored as low risk of bias or if one item was scored as high risk or unable to determine. If two domains were scored as high or unable to determine risk of bias, the study received a moderate risk of bias. Finally, when more than two domains were scored as high risk of bias, the study was regarded to possess a high risk of bias.

Results and discussion

Risk of bias assessment

For the systematic review, most of the assessed articles (n = 38, 76%) were at low risk of bias (Table 2). In general, these articles did not present attrition bias or reporting bias. The main weaknesses were the random selection (selection bias) and blinding outcome (detection bias). Twenty percent of the assessed articles (n = 10) were at moderate risk of bias. In general, as occurred in the studies with low risk of bias, the high risk or unclear score were in selection and detection bias. In those cases, detection bias was classified as high risk because inefficient blinding could affect the results. Finally, only two studies were at high risk of bias (4%). One of them was high risk in selection and detection bias, and also was unclear in relation to attrition bias. In contrast, the other high-risk article presented high risk in three categories (selection, attrition and reporting bias).

****Table 2****

Overview of findings

A total of five instruments appeared in the reviewed studies. Most studies focused on GPAI (n=22), spatial location (n=13), Game Performance Evaluation Tool (GPET, n=9), System of tactical assessment in Soccer (FUTSAT, n=4), and TSAP (n=3). Seven of the studies used instruments created and validated for the study itself and/or as adaptations from other instruments (i.e. GPAI and GPET). Consequently, those were included as GPAI or GPET instruments.

The results of the review were presented by mean of counts and percentages. These counts and percentages showed the occurrence of each variable category analysed over the total amount of manuscripts corresponding to each instrument, except for tactical levels and learner role variables. That occurred because in various manuscripts the assessment implied more than one level and role. All the information presented in the next sections followed the same structure. At the beginning there is a short introduction. Second, there are tables to show counts and percentages by each instrument (vertically). After the tables, there are extended explanations of the results and discussion.

Criteria definition – Degree of openness

Criteria demand operational descriptions that allow their observation. Depending on their descriptions, criteria can be classified as closed or open. A criterion is closed when it is predefined and validated in previous studies, whereas is open when it is created for a specific study or purpose.

Most of the reviewed GPAI studies (n=17, 80.95%, Table 3) were classified as open criteria. This could be due to the nature of GPAI, because when its originators created it, they validated six general components that appeared in several games (decision-making, skill execution, support, adjust, cover and base, Oslin et al., 1998). These components did not have a closed description and researchers had to define them for each study. For instance, Whipp et al. (2015: 5) described appropriate decision-making as ‘the player holding the ball up to allow teammates to get free shooting at goal’ whereas Gil-Arias et al. (2016: 4) defined it as ‘the player attacking on a very cohesive block to get block-out’. Furthermore, as decision-making is a very complex term, it is necessary to specify the kind of the appropriate decision-making (i.e. is it about shots, about dribbles, etc.), which did not appear in the mentioned studies.

****Table 3****

Notwithstanding no studies defined the game action involved in the decision-making, four studies were found (19.04%, Table 3) that used closed criteria definitions. This occurred because those four studies were based on criteria descriptions designed and validated in previous studies for specific games. In this case, the games were basketball and football (Blomqvist et al., 2005; Tallir et al., 2007).

Comparable to GPAI, in TSAP there were six general components that appeared in several games (received balls, conquered balls, offensive balls, successful shots, volume of play and lost balls). However, TSAP criteria, unlike GPAI, were predefined in the validation study (Gréhaigne et al., 1997). Despite that, only one study used the validated closed criteria (Brandes and Elvers, 2017); the rest of the TSAP studies used open criteria (n=2, 66.66%). As with GPAI studies, this occurred because authors decided to stipulate the criteria for the specific games, in this case for ultimate and hockey (Hastie, 1998; Nadeau et al., 2008).

Similarly, there were five spatial location studies (38.46%, Table 3) with open criteria definitions, in which the authors created their own criteria for the study or adapted existing criteria. The rest (n=8, 66.66%) had closed criteria definitions. In particular, authors used the tactical actions defined in the Teoldo et al. (2011) and Clemente et al. (2014) studies. However, although all the spatial location studies included the location of players on the pitch (the Approximate Entropy technique -ApEn- and the centroide), there were differences in the criteria they evaluated. Concretely, in order to combine spatial location criteria with tactical actions criteria, tactical criteria definitions were created as in GPAI or TSAP, for ApEn technique and centroide.

In contrast, the majority of GPET studies (n=8, 88.88%, Table 3) used closed criteria. Seven of them utilized the criteria definitions validated in the original study, and one study used the criteria definitions adapted to and validated especially for squash (i.e. Catalán-Eslava and González-Víllora, 2015). Just one study was classified as open criteria, given that authors created the definitions specifically for handball (i.e. García-López and Gutiérrez-Díaz, 2012). The high number of studies with closed criteria was in accordance with GPET, as it has 14 closed criteria definitions for football (García-López et al., 2013). In fact, the only study that used open criteria was the one carried out in handball. As GPET was designed specifically for football, its uses in other games will require its adaptation (Memmert and Harvey, 2008), as in the studies mentioned (Catalán-Eslava et al., 2018; García-López and Gutiérrez-Díaz, 2012). Likewise, all FUTSAT studies used closed criteria. In those studies, authors selected the criteria definitions from the 76 original criteria definitions for football. These criteria were designed and validated specifically for FUTSAT (Teoldo et al., 2011). In this case, there were no studies in which FUTSAT was adapted to other games. That could be due to the exhaustive and operative description of the criteria (Anguera et al., 2017). However, it also could be because in all the FUTSAT studies reviewed, the FUSTAT was used by its creators, who work principally in football.

In conclusion, considering the studies with open criteria definition (n=25, 50%), four articles were found that had no validation process (16%; n=2, 8% from GPAI and n=2, 8% from spatial location). Sixteen studies reported inter/intra-rater reliability scores (64%; n=11, 44% from GPAI, n=2, 8% from TSAP and n=3, 12% from spatial location). Finally, five studies reported content validity and inter/intra-rater reliability scores (20%; n=4, 16% from GPAI and n=1, 16% from GPET). In contrast, in closed criteria studies (n=25, 50%), the validation processes found in the designing and validating studies included face validity, content validity, construct validity and both inter and intra-rater reliability in all the studies (n=4, 16% from GPAI, n=1, 4% from TSAP, n=8, 32% from spatial location, n=9, 36% from GPET, n=4, 16% from FUTSAT). Design and validation processes are needed to ensure the accuracy of observations using open criteria instruments (Memmert and Harvey, 2008).

Therefore, although all the instruments reviewed were validated and well established, it is necessary to know the possible advantages or disadvantages that they present. The studies that used closed criteria did a more complex validation process than those that used open criteria. Furthermore, closed criteria definitions are one of the best strategies to enable replicability (Olsen and Larsen, 1997). However, instruments with open criteria definitions enable their adaptation to a range of contexts, allowing the use of the same instrument in different sports. In sum, the ideal strategy could be to use universal closed criteria definitions that fit in all the contexts that have undergone an exhaustive validation process (Arias-Estero and Castejón, 2012). Nonetheless, it is an impossible ideal because criteria need to be adapted to each context and game, according precisely to what is being assessed (Kirk and MacPhail, 2002). Consequently, there are two options to conduct an assessment process properly. One is to choose an instrument with closed criteria that is aligned to the study context (Biggs, 1996). Another is to perform an optimum validation process to adapt criteria descriptions to the study context (face validity, content validity construct validity, inter and intra-rater reliability).

Tactical levels

Different organizational levels can be identified in invasion games. ‘Match level’ corresponds to the global opposition relationships (Gréhaigne et al., 2005). ‘Partial forefront level’ includes partial opposition relationships involving a few players. Finally, the ‘primary level’ comprises the one-to-one level (Gréhaigne, 1992). None of the instruments used in the studies reviewed classified the criteria according to these three tactical levels (Table 4). However, for the purposes of this paper, criteria from studies reviewed were classified on the three levels, in order to substantiate which levels were actually assessed. The differentiation in tactical levels allows researchers to focus on the specific aspects of learning required in any given site and context (Kirk, 2017).

In GPAI, all studies assessed primary level actions (100%) whereas 14 (66.66%) assessed partial forefront level actions (Table 4). However, 52.38% of GPAI studies (n=11), concluded that global outcomes improved as result of primary and partial forefront levels, but match level was not assessed. Furthermore, one study assessed 'the performance of a team analyzing the actions of the ball carrier in a small-sided game of rugby union' but they only collected and analysed data on primary and partial forefront levels (Llobet-Martí et al., 2016: 5). Similar results were found when the studies used GPET and TSAP. In the case of GPET, nine (100%) studies included primary level actions and eight (88.8%) partial forefront level actions. Otherwise, TSAP studies only assessed primary and partial forefront levels in all the studies (100%). However, most of TSAP and GPET studies concluded that global outcomes improved (n=10, 83.33%; n=2, 16.67% from TSAP, n 8, 66.67% from GPET). As Gréhaigne et al. (1997) have argued, all components from three tactical levels need to be assessed to judge global outcomes. Thus, these studies should not have reported on general improvements on tactical learning outcomes.

****Table 4****

The opposite situation appeared with spatial location studies. Three of them assessed primary level (23.07%, Table 4), five assessed partial forefront level (38.46%), and 11 assessed match level (84.61%). The problem here is that when they were assessing just match level actions, they included in their conclusion primary level improvements in tactical learning outcomes, which was not consistent with the data they obtained (Figueira et al., 2018; Gonçalves et al., 2016; 2017; Ric et al., 2017).

Finally, for FUTSAT, all studies assessed actions from the three levels (Table 4). Consequently, global conclusions were aligned with data collected. That was possible because FUTSAT enables the assessment of game actions as for example, 'penetration' from match level, 'keep possession of the ball with passes' from partial forefront level and 'shoot at goal' from primary level (Teoldo et al., 2011). Nevertheless, the huge number of FUTSAT criteria (76) could make difficult their selection according to the player's level, as they were not validated for a specific age level (from U11 to U17). According to the literature, 12 is a critical age for tactical learning in the teaching-learning process for invasion games (Kirk and McPhail, 2002). As such, the actions proposed must be adapted to this stage. In this respect, some authors showed that players were not able to know the meaning of committing a strategic error to gain possession of the ball, which is a FUTSAT criterion (González-Víllora et al., 2010).

In conclusion, 26 of the reviewed studies used one or another level to assess tactical learning outcomes while distinguishing between the levels (52%; n=11, 22% from GPAI, n=4, 8% from spatial location, n=8, 16%, from GPET, n=1, 2% from FUTSAT, n=2, 4% from TSAP). However, it is not possible to affirm that previous studies assessed the global tactical learning outcomes, because they did not consider game actions from each tactical level (Deleplace, 1979). In short, assessing the global tactical learning outcomes in invasion games means assessing all three levels. Other options could be that a study just focuses, for instance, on a particular level. In this case, it is meaningful to assess and report on just this level. That occurred in the non-invasion game studies, where authors only assessed the primary level and concluded appropriately (n=5, 10% from GPAI and n=1, 2% from GPET).

Indexes

Quantifiable indexes are the transformation of the result of the assessed tactical variables into a number that is representative of the assessment (Arias-Estero and Castejón, 2012). In general, indexes are formulae that combine some of the criteria assessed to obtain information about the average tactical learning outcomes (e.g. in GPAI, Game Performance Index is the sum of the component index divided by the number of components analyzed).

Differing from the results obtained by Arias-Estero and Castejón (2012), in which the main indexes were Decision Making Index and Game Performance Index, the reviewed studies used mainly the Skill Execution Index. This index was used in 13 studies (61.90%, Table 5), while Game Performance Index and Decision Making Index were used in 12 studies (57.14%). Other indexes as such Cover Index and Support Index only appeared twice (9.52%), whereas no studies with Mark Index were found. Those results were logical considering that in the original study of GPAI, Oslin et al. (1998) proposed an index for each component (Decision Made Index, Skill Execution Index, Support Index, Cover Index and Mark Index). Those indexes were the result of the division of number of efficient actions between the number of inefficient actions made. Then there is the Game Involvement Index that is the sum of all the actions made, and finally the Game Performance Index that is the sum of the component index divided by the number of components analyzed. However, for the validated index for GPAI, five problems were exposed related to mathematical and reliability limitations that compromise the results (Memmert and Harvey, 2008). Some of the studies reviewed support these considerations. Four studies (19.04%) used a dichotomous scoring system, comparing appropriate and inappropriate decisions. On the other hand, one study (4.76%) created the Index of Performance. This can be calculated by dividing the actions (considering their weights and frequencies) by the summation of the total frequencies. For instance, when the criterion ‘ball carrier breaks the defense’ was met, it was scored with 3, while the criterion ‘ball carrier scores a try’ was scored with 1 (Llobet-Martí et al., 2016: 3).

****Table 5****

TSAP indexes do not present the limitations indicated for GPAI (Memmert and Harvey, 2008). In this case, the performance score ([volume of play / 2] + [efficiency index x 10]) was computed on the basis of two indexes: Efficiency Index ([conquered balls +

offensive balls + successful shoot or goal] / [10+lost balls]) and Volume of Play (conquered balls + received balls). All the articles reviewed used the two indexes. Similarly, FUSTAT authors created the Tactical Performance Index following Memmert and Harvey (2008). It was used in all the reviewed articles (n=4). It was calculated by the sum of all the actions from each component: Performance of the Principle, Quality of Principle Performance, Place of Action in the Game Field and Action Outcome, divided by the total number of game actions. In general, the indexes from GPAI, TSAP and FUTSAT aimed to combine the criteria analyzed previously to obtain an average outcome, which facilitates the interpretation of the results. That means that although the instruments already provide information regarding tactical learning, the purpose was to combine all these data in order to obtain a single outcome result.

In contrast, indexes were not presented in spatial location studies. Seven studies (53.8%, Table 5) presented the results by means, while six studies (46.15%) used other techniques to combine the raw data. The three specific techniques used were length per width ratio (Lpwratio, e.g. Folgado et al., 2014), Spatial Exploration Index (e.g. Figueira et al., 2018), and Approximate Entropy technique (e.g. Figueira et al., 2018). These techniques have the aim of combining the spatial location data to obtain comprehensible information about tactical learning, which would not be understandable otherwise from a tactical point of view. Compared to other instruments, there were no negative consequences of the use of these indexes found in the literature. This could be due to two reasons. First, the results were based on complex and careful mathematical process (e.g. Gonçalves et al., 2017). Second, the use of this type of study is not widespread, at least with objectives related to the teaching-learning process and the assessment of tactical learning outcomes in youth organized context and school context. Therefore, further studies should focus on their practical application in the real game play context.

Finally, in GPET no specific formulae were used in the studies reviewed. In two studies (22.2%, Table 5) researchers obtained percentages from decision-making and skill execution for each principle of play. In seven studies (77.77%), researchers compared the average score for decision-making and skill execution. In all studies, the results were based on the GPET coding procedure (García-López et al., 2013). Those studies were an example of how to present results without indexes, as researchers obtain as much information as possible from determinate game actions. In fact, researchers analyzed the effects of interventions focusing in a small number of game actions (passing, dribbling, shooting and support). That made possible the easy management of results so that the use of indexes was unnecessary. As a result, it is possible to guarantee the direct applicability of the data and the extrapolation to the real game play context, for example, by suggesting that a tactical approach improves the action of passing (e.g. Práxedes et al., 2018).

In conclusion, 25.49% of the studies that used indexes could have reported imprecise results (n=13, 100% from GPAI, Memmert and Harvey, 2008). The biggest problem of using indexes is skewed data. Using indexes or ratios can mask the nature of the player’s learning outcomes profile. The more actions included in an index, the more masked are the results. Nevertheless, the teaching-learning process involves a huge number of variables (Rink, 2014). As such, trying to reflect the learning outcome in a single datum risks dismissing information that may be of interest in improving learning. Consequently, it is preferable not to use indexes if they are not required, as showed in GPET studies. However, in case of using indexes, a further solution could be including a huge number of indexes, separating the indexes in function of the study interest, for instance, having validated indexes for each game action or game phase. Moreover, it could be interesting to consider the solutions presented by Memmert and Harvey (2008), for example, data from game performance and game involvement could be considered in the same formula.

Unit of observation

Unit of observation, also named unit of measurement, identifies the amount and type of individuals included in the observation and for whom data are collected (Merriam and Tisdell, 2016). Unit of observation is very important as it determines the scope of the obtained results and unit of analysis. Four units of observation appeared most frequently in the studies analysed. The first is individual player per game, recording all the actions from each individual player for the total duration of the game. Second, decision making unit (DMU), recording all the actions from each individual player for four-second time periods for the total duration of the game. Third, team per game, recording the actions from the whole team for the total duration of the game. Finally, team per ball possession unit, recording all the actions from the whole team for each possession for the total duration of the game (Table 6).

****Table 6****

Most of the studies from GPAI (n=18, 85.71%) and TSAP (n=3, 100%) used individual player per game as the unit of observation (Table 6). This was because GPAI and TSAP were created to provide information about individual player outcomes (Gréhaigne et al., 1997; Oslin et al., 1998). The rest of the GPAI studies (n=3, 14.28%) used DMU as the unit of observation because although they used GPAI components, this was done through coding instrument procedures. Nevett et al. (2001) introduced DMU, while García-López et al. (2011) used this unit of observation for the GPET in the original study. Consequently, all the GPET studies reviewed used this unit of observation. Having a unit of observation based on individual player or DMU means that an individual learning outcome is being assessed. However, it has been observed that 24 of the described studies sought to compare the outcomes as if they had considered the relations among teammates and opponents (48.48% from GPAI, n=3, 9.09% from TSAP and n=5, 15.15% from GPET). This implies that they were considering tactics as the sum of the individual outcomes, and missing the point that individual and team outcomes are relational (MacPhail et al., 2008). For example, at match level, the *rapport de force* refers to the antagonist links existing between several players or groups of players confronted with changes in game situation (e.g. from attack to loss of possession to defense, Gréhaigne et al., 1997). The *rapport de force* refers to the configuration of players in both teams at the moment possession is lost by one team and secured by the other. In other words, from these interactions during the game appear different relationships between the players of a team, forming the tactical levels. If all three levels are not analyzed, the actions emerging as reaction to the opposition moves are not registered. Consequently, the assessment will not be complete unless all the levels are measured.

In contrast, the unit of observation from spatial location and FUTSAT studies was the team. In the case of spatial location studies, it was team per game and for FUTSAT studies, it was team per ball possession. Specifically, this was because spatial location studies used a static approach to the observation of game play (Gréhaigne et al., 2005). This is an observational approach based on the distribution of the players on the pitch. FUTSAT was validated considering the team per ball possession as unit of observation (Teoldo et al., 2011). Consequently, all the studies reported results consistent with the unit of observation, as they were able to collect enough data to conclude on global tactical outcomes.

Overall, the individual was the unit of observation most used in the reviewed studies. Indeed, only individual player per game and DMU were extended to more than one instrument (n=33, 66%; n=21, 42% from GPAI, n=9, 18% from GPET; n=3, 6% from TSAP). However, global tactical learning in games is a complex phenomenon that implies different tactical levels, and not all the instruments were designed to assess all of them (Harvey et al., 2010). For this reason, researchers, teachers and coaches should pay special attention to the unit of observation as it delimits what is actually being assessed. Choice of the unit of observation will influence the results. According to O'Donoghue et al. (2012), previous performance analysis research has included examples of statistical inferences being made from small sample sizes, when artificially creating large samples by making the individual the unit of observation. In other words, using a large number of events from a small number of matches is inappropriate because results hide practical effects. This means that the units of observation as individual player per game or DMU are less reliable than units of observation as team per game, or team per ball possession, when the sample sizes are not large. Consequently, considering that the individual unit of observation does not provide the general outcomes and could cause statistical inferences, it is recommended that team per game or team per ball possession as unit of observation be used (O'Donoghue et al., 2012). This said, individual unit of observation could be fine to assess individual tactical learning outcomes providing that the sample size is big enough to allow a reliable data analysis.

Player/Learner role

Player role during games determine game actions that players can make. In general, roles can be classified in general as attacker and defender. Furthermore, each general role can be subdivided into attacker-on-ball, attacker-off-the ball and defender to on-ball attacker, defender to off-ball attacker, respectively.

Attacker roles were assessed in all GPAI (n=21), GPET (n=9) and TSAP (n=3) studies (Table 7). In contrast, defender roles were only assessed in 16 studies (32%; n=13, 26% from GPAI, n=1, 2% from TSAP, n=2, 4% from GPET). This seems reasonable because only GPAI was designed to assess attacker and defender roles, and allows the choice of components according to the objective of the analysis. Furthermore, TSAP and GPET were created to be used only with attacker roles (García-López et al., 2013; Gréhaigne et al., 1997), although three of these studies also assessed defender roles. This occurred because these studies adapted and validated TSAP to specific games (i.e. Nadeau et al., 2008) and GPET (i.e. Catalán-Eslava et al., 2018; Gutiérrez et al., 2014). As such, they had different criteria. As a result, findings were logical considering the nature of these instruments. However, from a utility perspective, it may be unhelpful to assess only attacker roles in the studies carried out in invasion games (football, basketball and ultimate) as defenders highly influenced the team outcomes (Deleplace, 1979; MacPhail et al., 2008).

****Table 7****

All the spatial location and FUTSAT studies assessed both attacker and defender roles in all the studies. In this case, the results obtained were consistent, as the studies were carried out in invasion games (football, hockey and rugby), and the instruments were created to assess global outcomes considering both player roles. The fact that these instruments included defender roles is logical considering that both were validated in formal sport contexts, where the information tends to be as complete as possible. Paradoxically, this point reflects the importance and limitations of assessment in the school context compared to formal sport contexts (Kirk, 2005).

In general, from all the studies reviewed, 66% analyzed both attacker and defender roles at the same time (n=33; n=13, 26% from GPAI; n=13, 26%; n=2, 4% from GPET; n=4, 8% from FUTSAT; n=1, 2% from TSAP). According to Arias-Estero and Castejón (2012), attacker roles appeared more frequently in the literature. This could be due to two reasons. First, attacker roles have been considered more important, as they seem to have a direct relation to the match results (Sarmento et al., 2018). Second, criteria related to defender roles demand intangible movements that make difficult operative and objective observation (Anguera et al., 2017). However, defender roles had a great weight in the total outcomes. Decision and actions made without possession of the ball are essential for team success (McPhail et al., 2008). Thus, the importance of defender roles should not be ignored, as all players (both attacker and defender) can influence the tactical outcomes (Aranda et al., 2019). In short, if global tactical outcomes want to be known, all of them must be considered, especially in invasion games. The new techniques based on spatial location system open a new path to explore the relationship between players and assist the analysis of defenders. For that reason, it could be a good idea to introduce this kind of technique progressively in youth sports and school contexts such as physical education lessons.

Institutional context

According to the literature, the context in which the teaching-learning process is developed influences the learning outcomes (Rovegno and Kirk, 1995). In the present review, institutional context has been differentiated into three main groups: school context (physical education classes), extracurricular sport context (interschool sport programmes, unofficial competitions, etc.) and formal sport context (competitions at any level, from national sport organizations).

Sixty-six percent of the studies from GPAI (n=14) and the 33.33% from TSAP (n=1) were conducted at school, whereas five studies from GPAI (23.80%) and two from TSAP (66.66%) were conducted on the formal sport context (Table 8). Furthermore, two studies from GPAI (9.52%) were carried out on extracurricular sport contexts. Both GPAI and TSAP were created and validated for physical education classes, which explains why most of the studies were used in the school context. Nevertheless, there were some GPAI and TSAP studies carried out in other contexts, as a consequence of the developing use of these instruments in research. However, as these instruments were not designed to be used in other contexts, they did not consider the differences between youth sport contexts and other school physical education characteristics pointed out by Gutierrez-Díaz et al. (2011). For instance, considering skill level, the same criteria definitions will not be equivalent in different contexts.

****Table 8****

In contrast, GPET was used in a formal sport context (n=6, 66.67%), and in a school context (n=3, 33.33%), whereas FUTSAT was just used in a formal sport context (n=4, 100%). However, authors affirmed that these instruments can be used in the three different contexts. These results could be due to GPET validation process included participants from both formal sport and school contexts, concerned about the differences presented in learning according to the context. On the other hand, FUTSAT participants were selected only from the formal sport context. In this regard, it could be difficult to extrapolate this instrument to other contexts (Kirk and MacPhail, 2002).

Finally, 12 spatial location studies (92.30%) were conducted on the formal sport context, whereas just one (8.33%) was carried out in the extracurricular sport context. In contrast, none were used in the school context. In fact, McGarry (2009) highlighted that match performance instruments are normally set in the formal sport context. That is mainly due to two reasons. First, the expense of this kind of technology, which most educational communities cannot afford. Second, the actual design of these instruments does not allow researchers to obtain information as tangible as the rest of the instruments presented. This means that, in practical terms, it would be difficult to use the tactical outcomes information from spatial location to improve the teaching-learning process.

In conclusion, from the reviewed studies, 11 (22%) used an instrument validated in a different institutional context. However, it is well established in the literature that context influences the learning outcomes and, as a consequence, the validity of the instruments (Rovegno and Kirk, 1995). Accordingly, it is necessary to validate the instruments for the context in which they will be used. If not, there will be an inadequate assessment that does not consider the specifications of each context. For instance, Gutierrez-Díaz et al. (2011) showed that, for two groups of the same age but in different institutional contexts, the action of 'pass' presented different levels of ability. Consequently, this implies that the description of 'successful pass' should be different for these groups, as the teaching-learning process should be adapted to the level of expertise, allowing all of the students/players to achieve success. Furthermore, as each environment is determined by contextual variables, it is also recommended to include such variables according to the specific context (Gómez et al., 2013). This occurred in two spatial location studies developed in a formal sport context, where halves and spatial location were included (i.e. Figueira et al., 2018; Ric et al., 2016).

Conclusions and future directions

The purpose of this study was to describe assessment in relation to tactical learning outcomes, based on variables that must be considered when using an assessment instrument: (a) criteria definitions–degree of openness; (b) tactical levels; (c) indexes; (d) units of observation; (e) player/learner roles and (f) institutional contexts. Building on this purpose, the following issues were highlighted. First, some studies used non-validated criteria to make the assessment. Second, not all studies considered the three tactical levels to assess global tactical outcomes. Third, the majority of the studies used indexes that masked the results. Fourth, the individual player unit of observation was widely used to assess global tactical outcomes. Fifth, many instruments were used in contexts for which they were not validated. These limitations denote the lack of alignment when using an assessment instrument, which could be crucial for the accuracy of the results obtained. In other words, conclusions from these studies could be based on inaccurate results as a consequence of an incorrect use of the assessment instrument. As such, the studies could be generating knowledge built on erroneous results.

In further studies, one important variable to consider could be the examination of the researcher or evaluator outcomes. For example: evaluators' training; evaluators' previous experience; instruments' reliability and observers' reliability; observation procedures etc.

Practical application

According to these limitations, general recommendations are proposed in this review. First, researchers should use validated instruments as long as the characteristics of the instruments are aligned with the nature of the study, in order to minimize assessment bias. For this purpose, Figure 2 shows the possibilities of each instrument included on this review according to the analyzed variables (first circle: the different instruments; the second circle: institutional context; the third circle: unit of observation and tactical levels; fourth circle: player/learner roles). Second, it is recommended when validating an instrument to consider the following general guidelines: (1) only use validated criteria descriptions; (2) include all three tactical levels; (3) do not use indexes; (4) use the team per game or the team per ball possession as the units of observation; (5) assess both defender and attacker roles; (6) develop the instrument in the same institutional context as the study context and (7) include context variables if applicable. Third, attending to the need to validate an instrument to assess specific tactical learning outcomes, in Figure 3, a checklist is included to guide researchers in order to ensure assessment is aligned with the intended purposes. This checklist is also developed to assist researchers when adapting or using existing instruments for their studies.

**** Figure 2****

****Figure 3****

Funding

During the development of this work, the first author had a scholarship for the Training of University Teachers from Ministerio de Educación, Cultura y Deporte de España (FPU17/00606).

References

Anguera MT, Camerino O, Castañer M, et al. (2017) The specificity of observational studies in physical activity and sports sciences: Moving forward in mixed methods research and proposals for achieving quantitative and qualitative symmetry. *Frontiers in Psychology* 8: 2196.

Aranda R, González-Ródenas J, López-Bondia I, et al. (2019) "REOFUT" as an observation tool for tactical analysis on offensive performance in soccer: Mixed method perspective. *Frontiers in Psychology* 10: 1476.

Arias-Estero JL and Castejón J (2012) Review of the instruments most frequently employed to assess tactics in physical education and youth sports. *Journal of Teaching in Physical Education* 31: 318-391.

Arias-Estero JL and Castejón F (2014) Using instruments for tactical assessment in physical education and extra-curricular sports. *European Physical Education Review* 20: 525-535.

Biggs J (1996) Enhancing teaching through constructive alignment. *Higher Education* 32: 347-364.

Blomqvist M, Vänttinen T and Luhtanen P (2005) Assessment of secondary school students' decision-making and game-play ability in soccer. *Physical Education and Sport Pedagogy* 10: 107-119.

Brandes M and Elvers S (2017) Elite youth soccer players' physiological responses, time-motion characteristics, and game performance in 4 vs. 4 small-sided games: The influence of coach feedback. *The Journal of Strength and Conditioning Research* 31: 2652-2658.

Catalán-Eslava M, González-Víllora S, Pastor-Vicedo JC, et al. (2018) Analysis of tactical, decisional and executional behaviour according to the level of expertise in squash. *Journal of Human Kinetics* 61: 227-240.

Catalán-Eslava M and Víllora SG (2015) Validation of a wall-net sports measurement instrument: Squash performance evaluation tool (SPET-HERS). *Retos* 27: 73-80.

Clemente FM, Martins FML, Mendes RS, et al. (2014) A systemic overview of football game: The principles behind the game. *Journal of Human Sport and Exercise* 9: 656-667.

Deleplace R (1979) *Rugby de Mouvement – Rugby Total [Rugby in Movement – Total Rugby]*. Paris: Éducation Physique et Sports.

Desrosiers P, Genet-Volet Y and Godbout P (1997) Teachers' assessment practices viewed through the instruments used in physical education classes. *Journal of Teaching in Physical Education* 16: 211-228.

Figueira B, Gonçalves B, Masiulis N, et al. (2018) Exploring how playing football with different age groups affects tactical behaviour and physical performance. *Biology of Sport* 35: 145-153.

Folgado H, Lemmink KA, Frencken W, et al. (2014) Length, width and centroid distance as measures of team's tactical performance in youth football. *European Journal of Sport Science* 14: 487-492.

García-López LM, González-Víllora, SG, Gutiérrez-Díaz DC, et al. (2013) Development and validation of the game performance evaluation tool (GPET) in soccer. *Sport TK* 2: 89-99.

García-López LM and Gutiérrez-Díaz CD (2012) Gender differences in the game behaviour in invasion games. *Physical Education and Sport Pedagogy* 17: 289-301.

Gil-Arias A, Moreno MP, García-Mas A, et al. (2016) Reasoning and action: Implementation of a decision-making program in sport. *The Spanish Journal of Psychology* 19: e60.

Gómez MA, Lago-Peñas C and Pollard R (2013) Situational variables. In: McGarry T, O'Donoghue P and Sampaio J (eds) *Routledge Handbook of Sports Performance Analysis*. New York: Routledge, pp.277-287.

Gonçalves B, Esteves P, Folgado H, et al. (2017) Effects of pitch area-restrictions on tactical behavior, physical, and physiological performances in soccer large-sided games. *Journal of Strength and Conditioning Research* 31: 2398-2408.

Gonçalves B, Marcelino R, Torres-Ronda L, et al. (2016) Effects of emphasising opposition and cooperation on collective movement behaviour during football small-sided games. *Journal Sports Sciences* 34: 1346-1354.

González-Víllora S, García-López LM, Gutiérrez-Díaz DC, et al. (2010) Tactical awareness and decision making in youth football players (12 years): A descriptive study. *Infancia y Aprendizaje* 34: 489-501.

González-Víllora S, Serra-Olivares J, Pastor-Vicedo JC, et al. (2015) Review of the tactical evaluation tools for youth players, assessing the tactics in team sports: Football. *Springerplus* 2: 663.

Greáhaigne JF, Godbout P and Bouthier D (1997) Performance assessment in team sports. *Journal of Teaching in Physical Education* 16: 500–516.

Gréhaigne JF, Bouthier D and David B (1997) Dynamic-system analysis of opponent relationship in collective actions in football. *Journal of Sports Sciences* 15: 137-149.

Gréhaigne JF, Godbout P and Bouthier D (1997) Performance assessment in team sports. *Journal of Teaching in Physical Education* 16:500-516.

Gréhaigne JF, Richard JF and Griffin L (2005) *Teaching and Learning Team Sports and Games*. London: Taylor & Francis.

Gréhaigne JF (1992) *L'Organisation du Jeu en Football [The Organization of Play in Football]*. Paris: ACTIO.

Gutiérrez D, Fisette J, García-López LM, et al. (2014) Assessment of secondary school students' game performance related to tactical contexts. *Journal of Human Kinetics* 42: 223-234.

Gutiérrez-Díaz DC, González-Víllora S, García-López LM, et al. (2011) Differences in decision-making development between expert and novice invasion game players. *Perceptual and Motor Skills* 112(3): 871-888.

Harvey S, Cushion CJ and Sammon P (2015) Dilemmas faced by pre-service teachers when learning about and implementing a game-centred approach. *European Physical Education Review* 21: 238-256.

Harvey S, Cushion CJ, Wegis HM et al. (2010) Teaching games for understanding in american high-school soccer: A quantitative data analysis using the game performance assessment instrument. *Physical Education and Sport Pedagogy* 15: 29-54.

Hastie PA (1998) Skill and tactical development during a sport education season. *Research Quarterly for Exercise and Sport* 64: 368-379

Higgins JPT, and Green S (2011) Cochrane handbook for systematic reviews of interventions. Available at: http://handbook.cochrane.org (accessed 31 October 2018).

Holt N, Strean W and García-Bengochea E (2002) Expanding the teaching games for understanding model: New avenues for future research and practice. *Journal of Teaching in Physical Education* 21: 162-176.

Kinnerk P, Harvey S, MacDonncha C, et al. (2018) A review of the game-based approaches to coaching literature in competitive team sport settings. *Quest* 70: 401-418.

Kirk D (2016) Model based teaching and assessment in physical education: The tactical games model. In: Green K and Hardman K (eds) *Physical Education: Essential Issues*. London: Sage, pp.128-143.

Kirk D (2017) Teaching games in physical education: Towards a pedagogical model. *Revista Portuguesa de Ciências do Desporto* 17: 17-26.

Kirk D and MacPhail A (2002) Teaching games for understanding and situated learning: Rethinking the Bunker-Thorpe model. *Journal of Teaching in Physical Education* 21: 177-192.

Llobet-Martí B, López-Ros V, Barrera-Gómez J, et al. (2016) Assessing novices' game performance in Rugby Union: The rugby attack assessment instrument (RAAI). *Journal of Teaching in Physical Education* 32: 181-186.

McGarry T (2009) Applied and theoretical perspectives of performance analysis in sport: Scientific issues and challenges. *International Journal of Performance Analysis in Sport* 9: 128-140.

McPhail A, Kirk D and Griffin L (2008) Throwing and catching as relational skills in game play: Situated learning in a modified game unit. *Journal of Teaching in Physical Education* 27:100-115.

Memmert D and Harvey S (2008) The game performance assessment instrument (GPAI): Some concerns and solutions for further development. *Journal of Teaching in Physical Education* 27: 220-240.

Merriam SB and Tisdell EJ (2016) *Qualitative Research: A Guide to Design and Implementation (4th ed)*. San Francisco: Jossey-Bass.

Morales-Belando MT, Calderón A and Arias-Estero JL (2018) Improvement in game performance and adherence after an aligned TGfU floorball unit in physical education. *Physical Education and Sport Pedagogy* 23: 657-671.

Nadeau L, Godbout P and Richard JF (2008) Assessment of ice hockey performance in real-game conditions. *European Journal of Sport Science* 8: 379-388.

Nevett M, Rovegno I, Babiarz M, et al. (2001) Changes in basic tactics and motor skills in an invasion-type game after a 12-lesson unit of instruction. *Journal of Teaching in Physical Education* 20: 352-369.

O'Donoghue P, Papadimitriou K, Gourgoulis V, et al. (2012) Statistical methods in performance analysis: An example from international soccer. *International Journal of Performance Analysis in Sport* 12: 144-145.

Olsen E and Larsen O (1997) Use of match analysis by coaches. In: Reilly T, Hughes M and Bangsbo J (eds) *Science and Football III*. London: Taylor & Francis, pp.251-306.

Oslin JL, Mitchell SA and Griffin LL (1998) The game performance assessment instrument (GPAI): Development and preliminary validation. *Journal of Teaching in Physical Education* 17: 231-243.

Práxedes A, Del Villar F, Pizarro D, et al. (2018) The impact of nonlinear pedagogy on decision-making and execution in youth soccer players according to game actions. *Journal of Human Kinetics* 62: 185-198.

Ric A, Hristovski R, Gonçalves B, et al. C (2016) Timescales for exploratory tactical behaviour in football small-sided games. *Journal of Sports Sciences* 34: 1723-1730.

Ric A, Torrents C, Gonçalves B, et al. (2017) Dynamics of tactical behaviour in association football when manipulating players' space of interaction. *PLoS ONE* 12: e0180773.

Rink J (2014) Teacher effectiveness in physical education-consensus? *Research Quarterly for Exercise and Sport* 85: 282-286.

Rovegno I and Kirk D (1995) Articulations and silences in socially critical work on physical education: Toward a broader agenda. *Quest* 47: 447-474.

Sarmento H, Anguera MT, Pereira A, et al. (2018) Talent identification and development in male football: A systematic review. *Sports Medicine* 48: 907-931.

Tallir IB, Lenoir M, Valcke M, et al. (2007). Do alternative instructional approaches result in different game performance learning outcomes? Authentic assessment in varying game conditions. *International Journal of Sport Psychology* 38(3): 263-282.

Teoldo I, Garganta J, Greco PJ, et al. (2011) System of tactical assessment in Soccer (FUT-SAT): Development and preliminary validation. *System* 7(1): 69-83.

Veal ML (1988a) Pupil assessment issues: A teacher educator perspective. *Quest* 40: 151-161.

Whipp PR, Jackson B, Dimmock JA, et al. (2015) The effects of formalized and trained non-reciprocal peer teaching on psychosocial, behavioral, pedagogical, and motor learning outcomes in physical education. *Frontiers in Psychology* 6: 149.

Author biographies

Carmen Barquero-Ruiz is concluding her PhD at the Faculty of Sport, UCAM Universidad Católica San Antonio de Murcia, Spain.

José Luis Arias-Estero is Lecturer at the Faculty of Sport, UCAM Universidad Católica San Antonio de Murcia, Spain.

David Kirk is Professor at School of Education, University of Strathclyde, Scotland.

Table 1. Articles included in the review.

References included in the review

Araujo et al., 2016, EPER, 22, 185-200	Mahedero et al., 2015, JTPE, 34, 626-641
Blomqvist et al., 2005, PESP, 5, 208-229	Mesquita et al., 2012, EPER, 19, 205-218
Borges et al., 2017, JHK, 58, 207-214	Morales-Belando and Arias-Estero, 2017, RQES, 888, 513-523
Brandes and Elvers, 2017, JSCR, 31, 2652-2658	Morales-Belando and Arias-Estero, 2017, JTPE, 36, 209-219
Castelao et al., 2014, IJPAS, 14, 801-813	Morales-Belando et al., 2018, PESP, 23, 657-671
Catalan-Eslava et al., 2018, JHK, 61, 227-240	Moreno et al., 2011, PESP, 16, 251-264
Chatzopoulos et al., 2006, PMS, 103,463-470	Nadeau et al., 2008, EJSS, 8, 379-388
Chen et al., 2013, JTPE, 32, 100-109	Olthof et al., 2015, HMS, 41, 92-102
Correia et al., 2012, JSMS, 12, 244-249	Oslin et al., 1998, JTPE, 17, 231-243
Farias et al., 2015, JTPE, 34, 363-383	Padilha et al., 2017, IJPAS, 17, 721-736
Farias et al., 2018, JSSM, 17, 56-65	Praxedes et al., 2018, JHK, 62, 185-198
Figueira et al., 2018, BS, 35, 145-153	Praxedes et al., 2018, PO, 13, e0190157
Folgado et al., 2014, EJSS, 14, 487-492	Ric et al., 2016, JSS, 34, 1723-1730
French et al., 1996, JTPE, 15, 418-438	Ric et al., 2017, PO, 12, e0180773
Gil-Arias et al., 2016, SJP, 19, e60	Sampaio and Maçãs, 2012, IJSM, 33, 395-40
Gonçalves et al., 2016, JSS, 34, 1346-1354	Serra-Olivares et al., 2015, SAJ, 37, 119-129
Gonçalves et al., 2017, JSCR, 31, 2398-2408	Serra-Olivares et al., 2015, JHK, 46, 251-261
Gray and Sproule, 2011, PESP, 16, 15-32	Serra-Olivares, et al., 2016, JTPE, 35, 208-218
Gutierrez et al., 2011, PMS, 112, 871-888	Silva et al., 2014, JHK, 41, 191-202
Gutiérrez et al., 2014, JHK, 42, 223-334	Tallir et al., 2012, ,IJSP, 43, 425-437
Harvey et al., 2010, PESP, 15, 29-54	Timmerman et al., 2017, IJSC, 12, 588-594
Hastie, 1998, RQES, 66, 368-379	Travassos et al., 2014, IJPAS, 14,594-605
Hastie et al., 2009, EJSS, 9, 133-140	Turner and Martinek, 1999, RQES, 70, 286-296
Lago, 2009, JSS, 17, 1463-1469	Vaz et al., 2012, PMS, 11, 594-604
Llobet-Marti et al., 2016, JTPE, 32, 181-186	Whipp et al., 2015, FP, 6, 149

EPER: European Physical Education Review. PESP: Physical Education and Sport Pedagogy. JHK: Journal of Human Kinetics. IJPAS: International Journal of Performance Analysis in Sport. JSCR: The Journal of Strength and Conditioning Research. PMS: Perceptual and Motor Skills. JTPE: Journal of Teaching in Physical Education. JSMS: Journal of Sciences and Medicine in Sport. BS: Biology of Sport. SJP: The Spanish Journal of Psychology. JSS: Journal of Sports Sciences. RQES: Research Quarterly for Exercise and Sport. EJSS: European Journal of Sport Science. HMS: Human Movement Science. PO: Plos One. SAJ: South African Journal for Research in Sport, Physical Education and Recreation. IJSM: International Journal of Sports Medicine. IJSP: International Journal of Sport and Exercise Psychology. IJSC: International Journal of Sports Science and Coaching. FP: Frontiers in Psychology.

Table 2. Risk of bias assessment.

References	Random sequence generator (selection bias)	Blinding outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Catalan-Eslava et al. (2018)	-	+	+	+	+
Farias et al. (2018)	-	+	+	+	+
Figueira et al. (2018)	-	+	+	+	+
Morales-Belando et al. (2018)	-	+	+	+	+
2018 Praxedes et al. (2018)	-	-	?	+	+
Praxedes et al. (2018)	-	+	+	+	+
Borges et al. (2017)	-	+	+	+	+
Brandes and Elvers (2017)	-	+	+	+	+
Gonçalves et al. (2017)	-	+	+	+	+
Morales-Belando and Arias-Estero (2017)	+	+	+	+	+
Morales-Belando and Arias-Estero (2017)	+	+	+	+	+
Padilha et al. (2017)	-	+	+	+	+
Ric et al. (2017)	-	+	+	+	+
Timmerman et al. (2017)	-	+	+	+	+
Araujo et al. (2016)	-	+	+	+	+
Gil-Arias et al. (2016)	-	+	+	+	+
Gonçalves et al. (2016)	-	+	+	+	+
Llobet-Marti et al. (2016)	?	?	+	+	+
Ric et al. (2016)	-	+	+	+	+
Serra-Olivares et al. (2016)	-	+	+	+	+
Farias et al. (2015)	-	-	+	+	+
Mahedero et al. (2015)	-	+	+	+	+
Olthof et al. (2015)	+	+	+	+	+
Serra-Olivares et al. (2015)	-	+	+	+	+
Serra-Olivares et al. (2015)	-	+	+	+	+
Whipp et al. (2015)	+	+	+	+	+
Castelao et al. (2014)	+	?	+	+	+
Folgado et al. (2014)	-	+	?	+	+
Gutiérrez et al. (2014)	-	+	+	+	+
Silva et al. (2014)	-	+	+	+	+
Travassos et al. (2014)	-	+	+	+	+
Chen et al. (2013)	-	+	-	-	+
Correia et al. (2012)	-	+	+	+	+
Mesquita et al. (2012)	-	?	+	+	+
Sampaio and Maças (2012)	-	+	+	+	+
Tallir et al. (2012)	-	?	+	+	+
Vaz et al. (2012)	-	+	+	+	+

Gutierrez et al. (2011)	+	+	+	+	+
Gray and Sproule (2011)	-	-	+	+	+
Moreno et al. (2011)	-	+	+	+	+
Harvey et al. (2010)	-	+	+	+	+
Hastie et al. (2009)	-	?	+	+	+
Lago (2009)	-	+	+	+	+
Nadeau et al. (2008)	-	+	+	+	+
Chatzopoulos et al. (2006)	-	+	+	+	+
Blomqvist et al. (2005)	-	?	+	+	+
Turner and Martinek (1999)	+	+	+	+	+
Hastie (1998)	?	+	+	+	+
Oslin et al. (1998)	?	?	+	+	+
French et al. (1996)	+	?	+	+	+

Table 3. Counts and percentages of openness criteria definition by instrument.

	GPAI	SPATIAL LOCATION	GPET	FUTSAT	TSAP
	% (n)	% (n)	% (n)	% (n)	% (n)
Open criteria (ad hoc)	80.95 (17)	38.46 (5)	11,11 (1)	-	66.66 (2)
Close criteria (predefined)	19.04 (4)	66.66 (8)	88,88 (8)	100 (4)	33.33 (1)-

Table 4. Counts and percentages of each tactical level by instrument (non-exclusive).

	GPAI	SPATIAL LOCATION	GPET	FUTSAT	TSAP
	% (n)	% (n)	% (n)	% (n)	% (n)
Primary level	100(21)	23.07 (3)	100 (9)	100 (4)	100 (3)
Partial forefront level	71.42 (15)	38.46 (5)	88.88 (8)	75 (3)	100 (3)
Match level	-	84.61 (11)	-	100 (4)	-

Table 5. Counts and percentages of used of indexes by instrument.

	GPAI	SPATIAL LOCATION	GPET	FUTSAT	TSAP
	% (n)	% (n)	% (n)	% (n)	% (n)
Use indexes	66.66 (14)	-	-	100 (4)	100 (3)
No indexes	33.33 (7)	100 (13)	100 (9)	-	-

Table 6. Counts and percentages of each unit of observation by instrument.

	GPAI	SPATIAL LOCATION	GPET	FUTSAT	TSAP
	% (n)	% (n)	% (n)	% (n)	% (n)
Individual player per game	85.71 (18)	-	-	-	100 (3)
Decision – making unit	14.28 (3)	-	100 (9)	-	-
Team per game	-	100 (13)	-	-	-
Team per ball possession	-	-	-	100 (4)	-

Table 7. Count and percentages of each player/learner role by instrument (non-exclusive).

	GPAI	SPATIAL LOCATION	GPET	FUTSAT	TSAP
	% (n)	% (n)	% (n)	% (n)	% (n)
Attackers	100 (21)	100 (13)	100 (9)	100 (4)	66.6 (3)
Defenders	61.90 (13)	100 (13)	22.2 (2)	100 (4)	33.3 (1)

Table 8. Counts and percentages of each institutional context by instrument.

	GPAI	SPATIAL LOCATION	GPET	FUTSAT	TSAP
	% (n)	% (n)	% (n)	% (n)	% (n)
Formal sport context context	23.80 (5)	92.30 (12)	66.67 (6)	100 (4)	66.66 (2)
School context	66.66 (14)	-	33.33 (3)	-	33.33 (1)
Extracurricular sport context	9.52 (2)	7.69 (1)	-	-	-

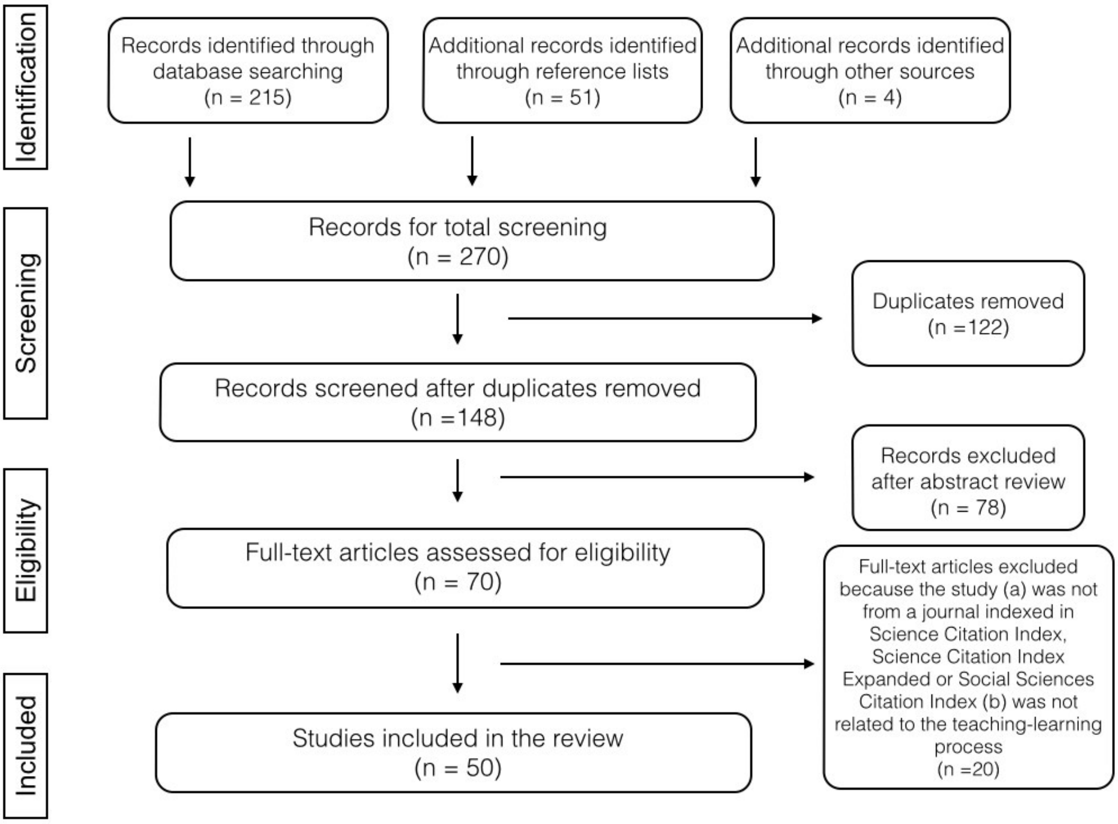

Figure 1. Study selection PRISMA flow diagram.

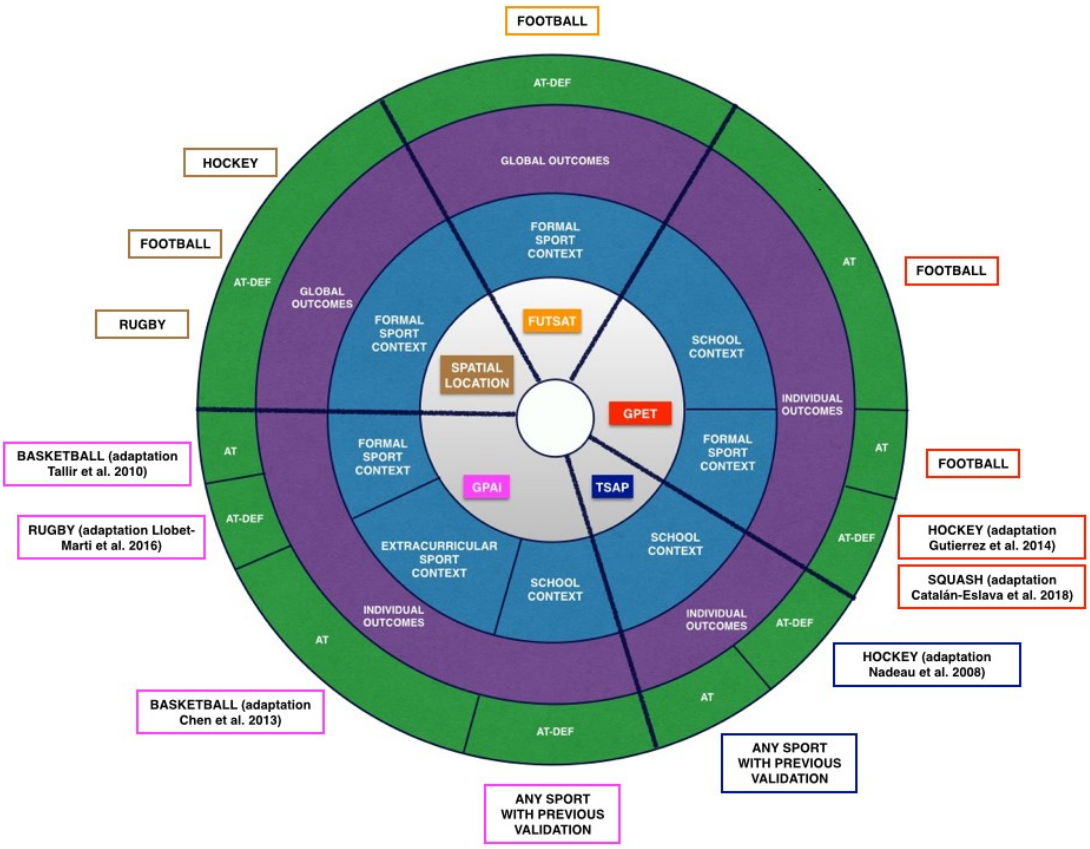

Figure 2. Graphic of instruments to assist the selection for assessment.

AT: attacker roles. AT-DEF: both attacker and defender roles.

	What to consider in relation to study characteristics	if **YES**	if **NOT**
CRITERIA DEGREE OF OPENNESS	*Does it want to compare between different games?*	Use open validated criteria	Use close validated criteria
INDEXES	*Does it need indexes to present the results?*	Separate indexes in function of study interest	Do not present indexes
TACTICAL LEVELS	*Does it assess global outcomes?*	Include actions from three tactical levels	Include actions from the tactical level that influence the study aim
UNIT OF OBSERVATION	*Does it focus on individual player?*	Use Decision-Making Unit or individual player per game	Use Team per ball possession or team per game
INSTITUTIONAL CONTEXT	*Are there other studies developed in the same context?*	Use this instrument if possible	Develop or adapt the instrument according to this institutional context

Figure 3. Checklist to assist in developing and selecting instruments not included in Figure 2.