Design and Validation of the Tactical Assessment Instrument in football (TAIS)
Abstract

The focus on tactics has made assessment more difficult due to the lack of knowledge in the topic and the limitations of current tactical assessment instruments. **Purpose:** To design and validate a tactical assessment instrument in youth football (TAIS) following an exhaustive, ecological and meticulous process and dealing with the limitations found in the literature.

**Method:** The design was divided in two stages related to its development and validation. During the development stage: (a) a preliminary list of criteria was determined through a literature review; (b) the criteria were delimited through an exploratory observation; and (c) the adequacy of the criteria was determined by consulting experts. In the validation stage: (a) content and comprehension validity were obtained by consulting experts and through a systematic game observation; (b) construct validity was obtained by comparing the tactical outcomes from two groups with different skill levels; (c) criterion validity was established by comparing the tactical outcome using Game Performance Assessment Instrument and the present instrument, and (d) the reliability was obtained through inter-rater reliability. **Results:** The research process showed that the instrument is a valid and reliable tool comprised of 22 criteria to assess tactical outcomes in 8-12 years old youth football. **Conclusion:** The TAIS presents several advantages in practical terms with respect to assessment. First, it allows assessment of the three tactical levels nested in the unit of observation. Second, it considers all the player roles. Third, results are presented without general indexes. Fourth, it can be used to assess participants from all the institutional contexts. Finally, it includes contextual variables.

**Keywords:** Tactical learning, sport pedagogy, youth sport, assessment, authentic assessment.
In the last two decades, there has been an increased interest from researchers on teaching games from a tactical perspective (for a review, see Kinnerk, Harvey, MacDonncha, & Lyons, 2018). The importance of pedagogical processes of tactics made necessary the design of instruments in order to assess these processes appropriately, both in physical education and sport contexts (e.g., Game Performance Assessment Instrument [GPAI], Oslin, Mitchell, & Griffin, 1998; and Team Sports Assessment Procedure [TSAP], Gréhaigne, Godbout, & Bouthier, 1997). Football is in general one of the games with the greatest social impact and highly practiced in an organized form by adults and children around the world (Fédération Internationale de Football Association, 2017). As a result, the tactical perspective has also impacted the pedagogical process in youth football (Kinnerk et al., 2018). However, the focus on tactics has made assessment more difficult for coaches, teachers and researchers, due to the lack of knowledge in the topic and the limitations of current tactical assessment instruments to report authentic, useful and valid data of tactical learning outcomes during actual game play in football (Authors a).

According to Biggs (1996), learning requires the alignment of the components of the pedagogical process (i.e., assessment, teachers, students, learning activities, and learning outcomes, among other). Particularly, assessment is the key element that drives the rest of the components. In order to be part of the conversation, physical education needs to connect to the broader views of education. From this perspective, authentic assessment demands connections with real game experiences and the components of the pedagogical process. As a consequence, assessment should be authentic by measuring learners’ performance in situations as similar as possible to real game (Wiggins, 2011). This kind of assessment helps teachers, coaches and learners make sense of play with the context of the game. In relation to the relevance of assessment, the instruments used to assess learning outcomes could hinder the authenticity of the assessment and the pedagogical process if they are not valid and context bound. Hence, assessment instruments should allow the alignment and authenticity of the pedagogical process.
To date, four instruments were the most frequently used to assess tactical learning in youth football (Authors a). **On the one hand**, there were two generic observation instruments designed and validated in the 1990s. **The first** of these, TSAP, was created to assess individual performance in different team sports, in contexts of pre-assessment and formative assessment (Gréhaigne et al., 1997). It was intended for peer-assessment although it could be used by teachers and researchers. This instrument is based on two events, receiving the ball and playing the ball. From the observation of these two events, a ‘global performance index’ needs to be calculated. This index is the result of considering an overall ‘volume of play’ and ‘efficiency index’. TSAP offers the possibility to measure the on-ball attack, using the individual player as unit of observation, both in video and in vivo. **The second** instrument, GPAI, was developed in the school context to observe ‘game performance behaviours that demonstrate tactical understanding, as well as the player’s ability to solve tactical problems by selecting and applying appropriate skills’ (Oslin et al., 1998, p. 231). GPAI was intended to be used by teachers and students in peer or self-evaluation, although it could be use by researchers. GPAI includes seven tactical components (base, adjust, decision-made, skill execution, cover and guard/mark) forming the ‘game performance index’ and ‘game involvement index’. GPAI offers the possibility to measure both on- and off- ball attack and defence in different sports, using the individual player as unit of observation, both in video and in vivo. Overall, TSAP and GPAI allow the discussion of ideas throughout peer assessment procedures, which empowered the pedagogical process as a form of authentic assessment (Wiggins, 2011).

**On the other hand**, two additional instruments were specifically developed for youth football. **The first** of these was the system of tactical assessment in football (FUTSAT), which was created to assess tactical behaviour of football players. It was intended to be used by coaches and researchers. FUTSAT is composed of two macro-categories (observation and tactical principles features) and 76 criteria, although it was not possible to identify all the criteria (see a description in Costa, Garganta, Greco, Mesquita, & Maia, 2011). Considering all of them,
the ‘tactical performance index’ can be calculated. When using FUTSAT, the unit of observation is each team attack or defence. FUTSAT offers the possibility to measure both on-and off-ball attack and defence in video. For example, Borges, Guilherme, Rechenchosky, da Costa, and Rinadi (2017), showed that U17 players met more frequently the criteria offensive coverage than U13 players, because they increased their confidence and security in offensive actions. The second, the Game Performance Evaluation Tool (GPET), was created to analyse decision-making and skill execution regarding to the tactical problems in relation to which decisions are made and skills are executed (García-López, González-Villora, Gutiérrez-Díaz, & Serra-Olivares, 2013). It was intended to be used by teachers and coaches, although it could be also used by researchers. GPET sets the analysis of each decision made on tactical problems in which the players are involved within the game. This instrument is composed of 14 criteria and no indexes (see a description in García-López et al., 2013). When using GPET, the unit of observation is the individual player within each tactical problem. GPET offers the possibility to measure on-ball attackers in video. For example, Práxedes, Del Villar, Pizarro, and Moreno (2018) analysed the criterion ‘pass’ as a key game action included in two tactical problems: maintaining possession of the ball and progressing towards the goal. In both instruments, criteria refer to game actions (e.g., pass), categories refer to the discrete ways these actions can be executed throughout the game (e.g., back pass, forward pass, opening pass), and indexes offers information about the average tactical learning outcomes from formulae that combine the criteria assessed (e.g., in FUTSAT, tactical performance index is Σ tactical actions / number of tactical actions, Costa et al., 2011).

According to the most recent review regarding assessment practices in tactical learning in games, both in physical education and sport contexts, these instruments present five main limitations considering the purposes for which researchers used them (Authors a). First, these instruments do not consider the interactions among whole team, small groups of players and individual players when assessing team tactical performance, as TSAP proposed. These
interactions can be structured in three organizational levels (Deleplace, 1979; Gréhaigne, Richard, & Griffin, 2005). The first level, organizational match level, refers to the collective game actions that imply more than three players from the same team. The second level, partial forefront organizational level, refers to the game actions developed by at least two players. Finally, primary organizational level refers to the game actions developed by individual players. Therefore, organizational match level breaks down into partial opposition relationships forming the partial forefront organizational level that contains a unit called primary organizational level (e.g., Deleplace, 1979; Gréhaigne et al., 2005; Kirk, 2017). In practical terms, the levels allow identification of game actions attending to the number of players involved in such game actions. Consequently, this identification makes possible the assessment of interrelated game actions from different levels that have not been measured until now, given that the sum of individual tactical outcomes does not correspond to team tactical outcomes. However, the three levels could be applied within a tactical assessment tool nested in the same unit of observation. For example, considering the attack phase as unit of observation, at organizational match level, a team could play with ‘amplitude’ moving the ball from one side to the other in order to generate free spaces. Considering the partial forefront organizational level, this movement of the ball could be done by giving the ball from one player to other using ‘passes’. Regarding the primary organizational level, when players are close enough to goal, these passes should result in an individual shot. According to Kirk (2017), the evaluation of players’ tactical outcomes in each organizational level favours an authentic tactical assessment.

Second, only GPAI and FUTSAT consider all the players’/learners’ roles. Including information of both attacker and defender roles is also necessary, because defender roles and decisions made without possession of the ball have a great weight in the total outcomes and are essential for learning as a team/group (McPhail, Kirk, & Griffin, 2008). Third, all of the instruments, except for GPET, use indexes or ratios that can hide the nature of the player’s learning outcome. Moreover, showing the learning outcome in a single datum risks dismissing
information that may be of interest in improving learning (Authors a). Nevertheless, there are
indexes, such as game involvement index in GPAI, that have the potential to provide meaningful
information about players’ and students’ learning because all decisions are counted. Fourth, the
instruments were used without considering the institutional context in which they were validated
(club sport context, community-based football activities and school context), which influences
the outcomes (Rovegno & Kirk, 1995). For example, in the FUTSAT study only participated
players from club sport context and not from physical education (Costa et al., 2011). However,
authors suggest that this instrument can be used in school context. Finally, none of the
instruments include contextual variables, except for GPET and FUTSAT, that presented the
criteria contextualized in tactical principles of play. Tactical outcomes are environment-
dependent, not only considering the tactical principles of play, but requiring the inclusion of
contextual variables that allows setting the assessment in the reality or concrete situation in
which the assessment is done (Sal de Rellán-Guerra, Rey, Kalén, & Lago-Peñas, 2019). For
example, when players are winning, they could make better decisions, because decision making
is affected by game outcomes (Sal de Rellán-Guerra et al., 2019).

In summary, these five limitations highlight that current instruments for tactical
assessment were designed and validated without considering the essence of tactics in youth
football. Consequently, the purpose of this study was to design and validate an instrument to
assess tactics in youth football following an exhaustive, ecological and meticulous process
dealing with the limitations found in the former instruments and taking into account: (a) the
three tactical levels nested in the unit of observation, (b) all the player roles, (c) the results
without general indexes, (d) the institutional context and (e) contextual variables.

**Design**

The design of the study was developed in two stages (Figure 1). Stage 1 focused on the
development and design of the instrument while Stage 2 determined the validity and reliability
of the instrument through multiple phases. Stage 1 was subdivided into three phases. In Phase 1,
a preliminary list of criteria was determined through a literature review. In Phase 2, the criteria were delimited through an exploratory observation of several games. Finally, in Phase 3, the adequacy of the criteria to the aim of the instrument was determined by consulting experts. The Stage 2 was subdivided into five phases. In Phase 1, content and comprehension validity were obtained by consulting experts. In Phase 2, content and comprehension validity was confirmed through a systematic game observation. In Phase 3, construct validity was obtained by comparing the tactical outcomes from two groups with different skill levels. In Phase 4, criterion validity was established by comparing the tactical outcome using GPAI and the present instrument. Finally, in Phase 5, the reliability of the instrument was obtained through inter-rater reliability.

We will now outline the gaining entry and access section and then each one of these stages and phases in detail. Regarding these phases, all the information presented in the stages and phases sections will follow the same structure. At the beginning of each phase there will be information related to the participants and procedures. After that, there will be extended explanations of the results of each phase.

**** Figure 1 ****

**Gaining entry and access**

Regarding to the selection of the participants, the first author screened all interested participants for eligibility using a standardised script and email message. These messages and criteria of eligibility were different depending on the type of participant (experts, observers and players) and phase. The criteria for experts were (a) at least 10 years of experience in researching and (b) in research topics related on the aim of the study. Criteria for observers were (a) at least 3 years of experience in teaching physical education or coaching in different contexts and (b) having a degree in sport sciences or physical education. Criteria for players were (a) coming from club sport context, school context and community-based sport context, (b) being from different skill levels, (c) training with different frequencies and (d) being between eight
and 12 years old. Participants who answered the email within 20 days were selected. Approximately 40% of the people contacted in each phase were selected to participate. Once they were selected, they were informed of the protocol. Participants were different in all phases, except for Stage 2 Phase 2 (S2P2), where participants were the same of Stage 1 Phase 2 (S1P2).

The parents of the players signed an informed consent document before the investigation, and players assented to participate. Players, parents and observers in Stage 2 Phase 3 (S2P3), Stage 2 Phase 4 (S2P4) and Stage 2 Phase 5 (S2P5), were blinded to the study aim, but the experts and observers from S1P2 and S2P2 were necessarily informed about it. The main author’s University Research Ethics Committee approved the study, which was performed in accordance with the Helsinki Declaration.

**Stage 1. Development and design of the instrument**

**Phase 1. Identification of the criteria through a literature review**

**Procedure.** A review about tactical assessment in youth sport and physical education was carried out to explore the possible criteria and categories of the instrument. Tactical assessment was defined as the assessment carried out during a real game, considering techniques and tactics as two inseparable components of a player’s learning. Criteria refer to game actions (e.g., attack type) and categories refer to the different ways in which these actions can happen throughout the game (e.g., positional attack, counter-attack; Table 1). For example, ‘Attack type’ is determined as the spatial arrangement in the attack. It is considered ‘positional attack’ when defence adjust with attackers and attackers take time to reorganize themselves on the pitch. It is considered ‘counter-attack’ when attackers do not allow defence to recover their positions (Figure 2). The bibliographic search was conducted using the following terms: tactical learning, tactical performance, physical education, observational analysis, tactical assessment instruments, sport pedagogy and youth games. The quality criteria for the review were: (a) appeared in journals indexed in the Science Citation Index, Science Citation Index Expanded and Social Sciences Citation Index; (b) from peer-review journals; (c) both from teaching and coaching
contexts and (d) empirical studies that present conclusions or objectives related to the pedagogical process and assessment of tactical outcomes. As a result of the analysis of the previous studies, preliminary criteria and categories were established at the three tactical levels and the contextual level. Two of the authors met for three hours in three consecutive days for five weeks to decide which criteria had the best fit with the aim of the study, and classified them into the three tactical levels and contextual level, according to the tactical level proposal that supports the instrument (e.g., Deleplace, 1979; Gréhaigne et al., 2005; Kirk, 2017) and the relevance of contextual variables (e.g., Sal de Rellán et al., 2018).

**Results.** As a result of the literature review, 52 criteria were identified. Thirteen were eliminated because they referred to game actions that do not take place in football. Those were criteria from net sports (n=6), from games played with an implement (n=3) and from individual sports (n=4). In addition, 11 were eliminated because they did not match with the aim of the study. From those, six were only technical skills criteria and five were specific to elite adult players. After this phase, the preliminary version of the Instrument for Tactical Assessment in Football (TAIS) was created. It was composed of 28 criteria, ten of which included categories, located in the three tactical levels and a contextual level (Table 1).

***Table 1***

**Phase 2. Delimitation of the criteria through exploratory game observation**

**Participants.** Participants in this phase were 34 players and six observers. The players were aged between eight and 12. From them, 16 were football players from club sport context (e.g., club academy programmes), competing in the regular league and with between three and five years experience. Ten were enrolled in community-based football activities and had two years experience maximum. The remaining eight were physical education students with no previous experience in football. Furthermore, regarding to the frequency of training, 16 of the players practiced football at least two days per week in a club sport context, 10 practiced
football at least two days per week in community-based sport context, and eight only practiced
football in school context two days per week.

The observers presented the following characteristics: (a) three were graduates in sport
sciences, with more than three years experience as football coaches in teams from club sport
contexts and (b) three were graduates in physical education, with more than three years
experience as football coaches in teams from community-based sport contexts. All were
postgraduate masters students in sport sciences and had more than one-year experience in sport
pedagogy research (master thesis, doctoral thesis or scientific publication).

**Procedure.** An exploratory observation was completed to delimit the list of criteria and
categories from those identified in Stage 1 Phase 1 (S1P1). Observers were asked to observe
four games. From these four games, two lasted 25 minutes each half and were played with eight
players, including the goalkeepers, according to the category rules. The other two were played
with five players each team, including the goalkeepers (4GKvs4GK form) according to literature
recommendations (Machado, Padilha, González, Clemente, & Teoldo, 2019). The games were
played with official eight-to-12 aged laws of the game. Two of the games were developed in
club sport context. These games were recorded from the regular competition. One game was
from a community-based sport context. It was recorded in an inter-school competition. Finally,
one match was from a school context. It was recorded in a physical education lesson. The
footages were recorded by a high-speed video camera placed diagonally in relation to the goal-
line and the side-line. The video recording procedure was the same for the other phases.

From these four games, observers observed 424 game phases in total during ten meetings.
The game phases lasted between five and 15 seconds. Each meeting lasted two hours. During
the first hour they focused on the observation. Regarding the observation, game phase was set as
unit of observation. It meant each attack or defence from the team observed. The phase changed
each time there was a change in the possession of the ball. Each phase was identified as a row in
an excel sheet. In each phase of game (row) there were registered all the game actions (e.g., pass,
shot, defensive coverage) that they observed and their frequency according to the study aim.

They played the games using Virtual Dub Version 1.10.4. Authors indicated the frames to observe and informed them about the observing rules. Observers were: (a) asked to start with frames selected by authors and observe 40 game phases from these four games each meeting; (b) allowed to observe the game actions with their own order; (c) allowed to stop the videotapes as many times as necessary; and (d) asked to write doubts down and discuss it at the end of the meeting. The six observers viewed all the games individually. In the second hour of the meeting, all the observers and authors shared the results and discussed together the game actions observed. At the end of this phase, authors selected the criteria and categories in function of their frequencies of appearance. More precisely, they selected those that appeared with a frequency higher than 20% (Anguera, 2003; Arias, Argudo, & Alonso, 2009). According to the literature, this meant that they were relevant game actions in football for eight to 12 aged players from three institutional contexts (club sport context, community-based football activities and school context).

Results. When compared to the literature review, the 28 criteria identified in S1P1 were found in the observation with a frequency of more than 20%. On the contrary, this percentage was not found in the case of the categories ‘clear or catch’, ‘attack mistake’ and ‘defensive mistake’ from the criterion ‘finalizing type in attack’. Consequently, these categories were removed from this criterion and changed by the categories ‘own goal’, ‘previous action to goal kick’, ‘losing ball’ and ‘save from goalkeeper’ as they appeared with a frequency of more than 20%. Furthermore, the criterion ‘support’ was found in the observation with a frequency of more than 20% and consequently added to TAIS. Regarding the criterion ‘recovery type’, the categories ‘interception or goalkeeper block’, ‘attackers mistake’ and ‘end without recovering’ were added as they appeared with a frequency of more than 20%. At the end of this phase, TAIS was composed by 29 criteria, ten of them included categories.

Phase 3. Adequacy of the instrument through panel of experts
**Participants.** The participants in this phase were 20 experts. Five were coaches with a sport sciences degree, all of them had over 8 years experience as coaches in youth football (club sport context). The 15 researchers had the following demographics: (a) 10 from Spain, (b) three from the United Kingdom, and (c) two from Canada. These researchers came from the following specialisms: (a) physical education and sport pedagogy (n=7), (b) tactical learning (n=5) and (c) assessment instruments (n=3).

**Procedure.** The experts were asked to indicate which criteria and categories, from the list after S1P2, should be part of the present instrument, considering their tactical and contextual levels and according to its aim. In addition, they were asked to define each criterion and category and encouraged to propose new criteria and categories. They were informed about the nature and objectives of the present instrument. They were contacted by email and provided with a list of all the criteria and categories in the tactical and contextual levels. They had to assess the criteria quantitatively using a 5-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree). They were also asked to give explanations regarding to the scores they gave. Lastly, authors analysed and discussed the experts’ answers following the Bulger and Housner (2007) conditions to remove the criteria: (a) that received a mean rating of less than three or (b) that were scored with less than three in content or comprehension by more than 30% of the experts.

**Results.** As result of the quantitative analysis from expert evaluation, nine criteria were removed (‘change of role control’, ‘depth of attack’, ‘retract’, ‘tempo control’, ‘wall pass’, ‘fixing the player’, ‘centre’, ‘control’ and ‘marking’). Those criteria met at least one of the following conditions, they: (a) obtained an average scored of three or less or (b) were scored with less than three in content or comprehension by more than 30% of the experts (Table 2). As a result of the qualitative analysis, eight of the experts suggested changing the category ‘ball divided’ from ‘situation type’ to clarify whether the ball was divided from the point of view of
the attackers or the ball was divided from the point of view of the defence. Furthermore, to criterion ‘clearance’ they suggested to rename it as ‘interception’. As a consequence, the first version of TAIS was developed. It was composed by 20 criteria, 9 of them categorized. This version included the definitions of the criteria and categories (Table 3).

**Table 2**

**Table 3**

Stage 2. Validity of the instrument

Phase 1. Content and comprehension validity through panel of experts

Participants. Participants were 30 experts, researchers with over 10 years experience in teaching sport sciences (n=21) and physical education (n=9). They had the following demographics: (a) 18 from Spain, (b) three from United Kingdom, (c) five from United States of America, (d) two from Canada, (e) one from Australia and (f) one from Ireland. These researchers came from the following specialisms: (a) physical education and sport pedagogy (n=5), (b) tactical learning (n=17), (c) coaching in football (n=4) and (d) assessment instruments (n=4).

Procedure. The panel of experts checked for each criterion: (a) content, whether the descriptions of each criteria and its category were adequate to what we wanted to measure and (b) comprehension, whether the descriptions and its categories were comprehensible and correctly expressed. In addition, they were asked to assess in general: (a) whether the criteria classification was appropriate and corresponded to each tactical level and (b) whether they found the instrument useful. The panel of experts had to assess these aspects quantitatively, using a 5-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree), and qualitatively, including explanations or proposals. They were also provided with an open space to express other suggestions. The panel of experts was informed about the nature and objectives of the instrument. They were contacted by email and provided with a tool to do this evaluation. Lastly, authors analysed and discussed the experts’ answers. The criteria scored as less than four
were revised. We decided which changes to make considering the aim of the instrument. We considered all the comments that: (a) met the objective of the present instrument and (b) alluded to the specific contexts of youth football. After that, the tool was re-sent to the same experts in order for them to re-evaluate the instrument following the same instructions. This process was repeated for any of the criteria or categories that were scored less than four, which occurred twice (Bulger & Housner, 2007). Finally, the Aikens’s V coefficient was calculated on the second evaluation (Aiken, 1985).

Results. A total of ten criteria were modified. Concretely, two new criteria were established, ‘goal difference in favor’ and ‘goal difference against’, from the criterion ‘score board’. The criteria ‘progressing the ball unopposed’, ‘tackle’ and ‘recovery type’ were renamed as ‘dribbling’, ‘tackle or charging’ and ‘finalizing type in defence’, respectively. Finally, four criteria were redefined ‘attack type’, ‘defence type’, ‘support’ and ‘dribbling’ (Figure 2). Regarding the categories, for the criterion ‘game principle’ the categories were renamed ‘finalizing’ for ‘ending’ and ‘retrieving’ for ‘recovering’.

After the second round of expert evaluation, all the criteria were scored as more than four in the quantitative analysis and no changes were suggested according to qualitative analysis. As a result, TAIS was comprised of 22 criteria, nine of them with categories. The values of Aikens’s V were between .92 (‘defence type’) and 1 (‘goal difference in favour’, ‘attack type’, ‘defence type’, ‘amplitude’, ‘support’, ‘shoot’, ‘tackle or charging’).

Phase 2. Content and comprehension validity through systematic game observation

Participants. Participants were the same observers and players described in S1P2.

Procedure. In this phase, we conducted a systematic observation of the four games from S1P2 to check if the criteria and categories descriptions were operative. That means, to substantiate whether it was possible to identify easily the criteria and categories described. Observers were asked to observe the tactical outcome for each criterion using the instrument. This task was undertaken during 12 meetings and the observers had to view 26 game phases
each meeting. The game phases lasted between five and 15 seconds. They observed a total of 320 game phases. The six observers observed and coded all game footage individually. All meetings were two hours with each hour having a specific focus. In the first hour, observers focused on whether the criterion was met or not met. For instance, for the criterion ‘shoot’, they identified when a shot on goal was made. Then they registered whether the criteria were appropriate according to its definition when one of the following circumstances occurred: (a) the shot resulted in a goal, (b) the shot was directed to goal but there was no score (either it missed the goal or was saved or cleared) or (c) when the shot missed the goal and resulted in a corner kick. Criteria were coded in the three tactical levels and the contextual level by observers. They were nested as the same unit of observation for each game phase. Observers used Virtual Dub Version 1.10.4 to play the games and an excel sheet to record the information. They were able to stop the videotapes as many times as necessary and when they had any doubts, the procedure was to write it down and discuss at the end of the meeting. In the second hour of the meeting, all the observers and authors discussed together the observers’ doubts about the criteria definition, until an agreement was reached. As a result, the authors modified the definition of criteria and categories until they achieved a version that allowed the observation of tactical outcomes with operative criteria.

**Results.** Observers found some issues regarding to the operative description of three criteria. On the one hand, for ‘amplitude’ and ‘depth (offensive progression)’ the specific zones of the pitch were included in order to operationalize the terms lateral zones and vertical advance, respectively. On the other hand, for ‘dribbling criteria’ to operationalize the term ‘clearly has control’ it was established that a player had control when he/she made a minimum of three touches with the control of the ball. As a consequence of this phase, TAIS allowed the observation of tactical outcomes through the operative criteria (Figure 2).

**Phase 3. Construct validity through the analysis of tactical outcomes from different skill levels**
**Participants.** Participants were 24 players and four observers. Players were aged between eight and 12. Ten were from a club sport context (club academy programmes), competing in the regular league and with between four and five years experience. Six were enrolled in community-based football activities and had one year experience maximum. The remaining eight were physical education students with no previous experience in football. Furthermore, regarding the frequency of training, all of them practiced football at least two days per week.

Observers consisted of: (a) three coaches graduated in sport sciences with at least one-year experience as a football coaches in both club sport context (n=2) and community-based sport (n=1) and (b) one graduated in physical education with at least three years experience in teaching physical education. All of them were postgraduate master students in sport sciences and had more than one-year experience in sport pedagogy research (master thesis, doctoral thesis or scientific publications).

**Procedure.** A three-day tournament was conducted and consisted of four 20 minutes 4GK vs 4GK games, based on recommendations from the literature (Machado et al., 2019). Players were divided into two groups according to their skill level (low or high). They were classified after a GPAI observation of previous games by expert coaches. Then, the low skill level group and high skill level groups were randomly subdivided in two subgroups of five players each. In all groups there were players who came from the three different contexts. The games were played with official eight to 12 years old laws of the game. The games were conducted twice between teams with similar skill levels. So that, there were two games between low skill level groups and two games between high skill level groups. The games were video-recorded.

Observers were asked to observe the tactical outcomes of all players using the instrument and the recordings of all matches from this three-day tournament (25 frames per second). Observers coded all criteria defined in TAIS for each game phase. The level of the participants
was hidden from the observers. The observation technique was the same described in S2P2, but in this case there were neither meetings nor final discussion, as they only focused on the observation. The observers were trained for at least 10 hours in the use of the instrument. The observation reliability was adequate given the minimum values were set at .70 for Intraclass Correlation Coefficient (ICC) and Kappa coefficient and 95% for Percentage of Agreement (PA; Atkinson & Nevill, 1998; Robinson & O’Donoghue, 2007). The four games produced a total of 780 game phases. The observers viewed all the games individually. Wilcoxon test was used to explore the possible differences between skill levels in all the criteria. Effect size (ES) was also calculated. Statistical significance was set at \( p < .05 \).

**Results.** The results showed statistically significant differences for all the criteria after the comparison by skill level, except for shooting (Table 4). However, we decided not to remove it, as the low number of shots performed could explain the absence of statistically significant differences. This decision was ratified by the ES for such criterion with regard to the rest of criteria. Contextual criteria were not considered, as it had not sense to compare contextual criteria between skill levels.

****Table 4****

**Phase 4. Criterion validity through the evaluation of the tactical outcome using GPAI and the present instrument**

**Participants.** Participants in this phase were 30 players and four observers. The players were aged between eight and 12 years. Seventeen were from a club sport context (club academy programmes), competing in the regular league and with between three and five years experience. Seven were enrolled in community-based football activities and had two year experience maximum. The remaining six were physical education students with no previous experience in football. Furthermore, regarding the frequency of training, all of them practiced football at least two days per week.
The observers were (a) two coaches graduated in sport sciences with at least one-year experience as football coach in both club sport context (n=1) and community-based sport (n=1), and (b) two PhD students in sport sciences (n=1) and physical education (n=1) with experience in using GPAI in previous studies.

**Procedure.** A one-day tournament was organized to this phase. The players were randomly divided into six teams of five players each. In all groups there were players who came from the three different contexts and levels. Six games of 20 minutes each were played in 4GKvs4GK according to literature recommendations. The games were played with official eight to 12 years old laws of the game. The games were video-recorded.

The observers were asked to observe the tactical outcome for each criterion using GPAI (Oslin et al., 1998) and the present instrument. GPAI was selected based on the following. First, it was the only instrument that allows comparing criterion by criterion using open criteria description and avoiding the use of indexes. Second, is the most widely extended instrument used in physical education and youth sports (Authors a). Third, although GPET and FUTSAT are specific for football, it was impossible to access to their operative criteria descriptions. However, observers assessed only the primary and partial forefront organizational level, because organizational match level cannot be assessed with GPAI. Nevertheless, none of the existing instrument allows assessing the three tactical levels with similar criteria. Two of the observers evaluated all the footages of the games with GPAI and the other two with the present instrument, all of them individually. The observation technique for GPAI was systematic because the observers assessed all the players’ game actions. The observers were asked to observe ‘cover’, ‘support’, ‘decision-making’ and ‘skill execution’ components of GPAI for all the criteria (see criteria in Table 5). The observers added the number of appropriate and inappropriate decisions, and correct and incorrect skill executions, according to the definition of a previous study (Authors b). Adequate decisions included making appropriate choices about what to do during the game. Correct skill executions corresponded to an efficient performance of the selected skill.
The observers training and the observation technique for the present instrument was the same as described in S2P3. The observation reliability was adequate given the minimum values were over .70 for ICC/Kappa and over 95% for PA (Atkinson & Nevill, 1998; Robinson & O’Donoghue, 2007). The six games produced a total of 986 game phases observed with both instruments.

Spearman’s rho was used to explore the correlations between GPAI and the instrument. Statistical significance was set at \( p < .05 \). Given the instrument did not discriminate between ‘decision-making’ and ‘skill execution’ as GPAI, each instrument criterion was tested in correlation to both GPAI components. However, ‘defensive coverage’ in the instrument was compared with ‘cover’ in GPAI. In addition, both instrument criteria were compared considering their appropriation, except for ‘support,’ because it is always considered as appropriate with TAIS.

**Results.** The results showed significant rho values, higher than .60 in all the criteria between GPAI and TAIS, accepted as a high level of correlation (Atkinson & Nevill, 1998; Table 5). This meant that the instrument was valid according to GPAI primary and partial forefront organizational levels.

****Table 5****

**Phase 5. Inter-rater reliability**

**Participants.** Participants were 16 players and four observers. The players were aged between eight and 12 years of age. Six were participants from the club sport context, competing in the regular league, and had between three and four years experience. Five were enrolled in community-based sport activities and had two years experience maximum. The remaining four were physical education students with no previous experience. Furthermore, regarding the frequency of training, four players practiced at least two days per week in a club sport context, two of them practiced at least
two days per week in the community-based sport context, and four only practiced in the school context two days per week.

The observers had the following characteristics: (a) two graduated in sport sciences with at least two years’ experience as football coaches in both, club sport context (n=2) and community-based sport (n=1) and (b) two graduated in physical education, with at least five years experience in teaching physical education. All of them were postgraduate master in sport sciences students. None of them had previous experience in evaluating tactical outcomes with the instrument.

**Procedure.** Two matches were organized to this phase. The games were of 20 minutes each in 4GKvs4GK according to literature recommendations. The games were played with official eight to 12 years old laws of the game. The games were video-recorded. Observers were asked to assess the tactical outcomes using TAIS. The observers training and the observation technique was the same described in S2P3. All the observers viewed both games, individually. They had to observe 235 game phases in total.

According to Brown and O’Donogue (2007), instrument reliability was obtained by an inter-rater reliability. Two different analyses were used according to the type of criteria: (a) ICC or Kappa coefficient and (b) PA (Atkinson & Nevill, 1998; Robinson & O’Donoghue, 2007).

**Results.** Reliability exceeded .70 according to ICC or Kappa coefficient value (Cohen, 1960), and the 85% according to PA (Brewer & Jones, 2002). Lowest reliability was found for criterion ‘goal difference in favour’ (ICC=.70, PA=95%) while highest reliability was found for criteria ‘team’, ‘score board’ and ‘period’ (Kappa/ICC=1; PA=100%).

**Discussion**

The purpose of this study was to design and validate an instrument to assess tactics in youth football following an exhaustive, ecological and meticulous process dealing with the limitations found in other instruments. According to Carretero-Dios and Pérez (2007), design and validation processes are needed to ensure the accuracy of assessments. Consequently, in the
present study we differentiated the design and validation stages with three and five phases, respectively. At the end of the two stages the instrument included 22 criteria, nine of them with categories (Figure 2). All of them were carefully selected and validated for the three specific contexts of football (club sport context, community-based football activities and school context). In general, none of the current instruments for tactical assessment (FUTSAT, GPET, TSAP and GPAI) were developed according to the stages and phases outlined in the present study, neither did they include information about criteria selection and elimination in each phase.

None of the validation studies from the current instruments for tactical assessment presented a literature review phase in order to identify criteria as the present study. FUTSAT and GPET determined their criteria based on the principles of play, TSAP distinguished between when the player gained possession of the ball and how the player disposed of the ball, and GPAI included the game components that determine game performance (Oslin et al., 1998). Similarly, none of the instruments were developed verifying whether the criteria identified were observed in real games of youth football. However, GPAI components were initially developed through consultation with teachers and coaches, while in the present study experts adjusted the criteria after their identification from the literature review and delimitation through game observation. Therefore, the criteria of TAIS were: (a) supported by previous studies considering their relevance in youth sports; (b) obtained from real situations in youth football and (c) agreed by coaches and researchers in physical education and sport pedagogy, tactical learning, and assessment instruments. This process implied that TAIS was designed from inductive and deductive points of view, as the literature recommends (Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018).

Similarly to the present work, current instruments for tactical assessment were developed including content and comprehension validity through experts, except for TSAP (Greháigne et al., 1997). However, experts in the case of GPET were only teachers and coaches, while in FUTSAT and TAIS they were also researchers. The fact that GPET was validated by teachers
and coaches, could imply that content validity in GPET was useful in practical terms but it was not connected with research purposes (García-López et al., 2013). Furthermore, none of the studies, except for the present one, did a second round of panel of experts after modifying the instrument according to experts’ suggestions. Consequently, the present study is the only one that confirmed the validity with experts after modifying the criteria and categories definitions, obtaining high Aiken’s V values (Aiken, 1985). In addition, in the present study we also obtained content and comprehension validity through game observation in order to check that the definitions of criteria and categories made their observation possible. A similar procedure was followed in the TSAP study, but this was to check that the criteria occur with certain frequency during game play. However, it is necessary to consider that TSAP is a peer assessment instrument (Greháigne et al., 1997). So that, at the end of the two phases to obtain content and comprehension validity, we modified 13 criteria and two categories. This information is not available in the studies that validated other instruments; although GPET and GPAI pointed out that they modified criteria according to expert comments (García-López et al., 2013; Oslin et al., 1998). Therefore, as the content and comprehension validities were obtained by two different procedures the instrument is stronger, because it was verified both theoretically by experts and practically by observers (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME], 1999).

All the validation studies of the other instruments for tactical assessment, except for TSAP, developed a phase in which the tactical outcomes were analysed from different skill levels in real games to obtain construct validity. In terms of results, the GPAI validation study did not show that GPAI discriminated between skill levels in decision-making and support for basketball nor adjust and support for volleyball (Oslin et al., 1998). Similarly, GPET validation study did not find differences in skill execution for passing and dribbling nor decision-making for dribbling and shooting (García-López et al., 2013). Regarding FUTSAT, there is no
information available in the study, despite the fact that authors confirmed its construct validity (Costa et al., 2011). In comparison, TAIS discriminates between skill levels in all the criteria except for shoot, due to the low frequency of this criterion in the game. The main difference, however, resides in the strategy followed to determine the participants’ skill levels and in the participants’ contexts. On the one hand, whereas FUTSAT established the skill level using performance indexes from their own system and GPET determined the level in function of the participant context, GPAI and TAIS distinguished between individuals previously rated as high and low in game performance. On the other hand, in GPAI and TSAP studies, there were only participants from the school context (Greháigne et al., 1997; Oslin et al., 1998), in the FUTSAT study there were only participants from club sport context (Costa et al., 2011), and in GPET study the participants came from club sport context and school context (García-López et al., 2013). However, in the present study, there were participants from the three different institutional contexts (club sport context, community-based sport context and school context). Given the strategy followed to distinguish between participants’ skill levels and that participants came from three different institutional contexts, TAIS can be used objectively in the three youth football contexts, as participants determine the extent of the validity (Boateng et al., 2018).

TSAP design study was the only that also included a criteria validity phase as in the present study (Greháigne et al., 1997). Their correlations amounted to .74 and the lowest value of the TAIS was .60. Nonetheless, whereas in the TSAP study the reference criterion for comparison was the agreement of two football experts, we used the assessment of the tactical outcomes with GPAI as the reference criterion. This validity let external evidence of score validity, which provides the information about the usefulness or meaning of the test scores (AERA et al., 1999). However, this external evidence was not complete because TAIS goes further than any of the other instruments, including the organizational match level criteria, contextual criteria and without discriminating between technical and tactical components. In addition, the comparison was done criterion by criterion, instead of using GPAI indexes because
TAIS does not include indexes. Similarly to GPET, TAIS followed the literature recommendations, as using indexes or ratios can mask the results (García-López et al., 2013). As we noted in the introduction, reflecting the learning outcome in a single datum risks dismissing information that may be of interest in improving learning (Authors a).

All the validation studies of other instruments for tactical assessment showed their reliability through an inter-observer procedure. However, the observers were different because in TSAP they were high school students, in GPAI teachers, and in FUTSAT and GPET researchers. In the present study, observers included both teachers and coaches from club sport and community-based sport context, with previous experience as researchers. In addition, similarly to FUTSAT, observers had not participated before in any of the designing and validation phases in order to avoid the risk of bias (Costa et al., 2011). Furthermore, the tests used for analysing reliability were also different between studies. Whereas studies of other instruments showed the use of just ICC (TSAP), Kappa coefficient (FUTSAT), PA (GPAI) or analysis of variance (GPET), in the present study we used Kappa coefficient or ICC according to the nature of the criteria (discrete or continuous) and PA for all of them, following the literature recommendations (Boateng et al., 2018). Despite the differences pointed out, the reliability of TAIS was between .70 (95%) and 1 (100%), similar to those reported in the previous studies, which ranged between .79 (FUTSAT) and 73% (GPAI).

At the end of the development and validation stages, the result was an instrument comprising 22 criteria, nine of them with categories, organised in three tactical levels and a contextual level. In contrast, the other instruments for tactical assessment present between 7 components (GPAI) and 76 criteria (FUTSAT). Moreover, none of them differentiate between tactical levels or include contextual criteria, although they include criteria from at least one of the three levels but not nested in the unit of observation. The TAIS tactical level division allows comparison of the tactical outcomes within each level and according to specific game situations (Rovegno & Kirk, 1995). Given game situations are context-dependent, contextual criteria...
enable the reduction of bias caused by their influence (Sal de Rellán-Guerra et al., 2019).

Furthermore, as tactical assessment demands the contextualisation of game situations in a reference framework, these contextual criteria are crucial for making possible an authentical assessment. Overall, with the nested unit of observation, we overcome the problem of considering team tactical outcomes as the sum of individual tactical outcomes, present in most of the other instruments (GPAI, TSAP and GPET). As a consequence, it is possible to know the level in which players experience more difficulties and what level or situation correlates with better learning (Gréhaigne et al., 2005).

While GPET only considers criteria from an attacker’s role (García-López et al., 2013), FUTSAT and GPAI include criteria from attack and defence (Costa et al., 2011; Oslin et al., 1998; both on and off the ball), as in TAIS. However, although GPAI includes four components for each role, all of them are open description, which demands the adaptation and validation of criteria for each study (Authors a). On the other hand, though FUTSAT includes 38 criteria from each role, many of them are difficult to observe in the discrete youth football contexts as they only validated the criteria in club sport context (Gutiérrez-Díaz, González-Villora, García-López, & Mitchell, 2011). In contrast, TAIS presents eight closed attacker criteria and four closed defence criteria validated for all the contexts.

From a practical point of view, in using TAIS it is crucial to identify attack and defence phases, as they are the unit of observation, as in FUTSAT (Costa et al., 2011). In the case of GPAI, GPET and TSAP, the unit of observation is the player (García-López et al., 2013; Greháigne et al., 1997; Oslin et al., 1998). Nevertheless, registering in TAIS is less complex because the number of criteria included is lower than in other instruments. Furthermore, the present study shows the operative definitions of each appropriate and inappropriate criterion and its categories, including useful information about how to use the instrument in practical terms (Figure 2). This information cannot be found in the validation studies of FUTSAT nor GPET.

On the other hand, in the validation study of GPAI, authors did not define the criteria
operatively because they aimed to create a flexible self-constructed instrument. Regarding the scores generated by instruments, only GPET and the present instrument avoid the use of indexes, using the sum of each appropriate and inappropriate criterion. On the contrary, FUTSAT, GPAI and TSAP use indexes including in the same formulae data from all criteria, what could mask the results (Memmert & Harvey, 2008).

**Conclusion and practical application**

In conclusion, TAIS is a valid and reliable instrument comprised of 22 criteria to assess tactical outcomes in 8-12 years old youth football. TAIS presents the following advantages in practical terms. First, the instrument can be used by researchers, teachers and coaches to evaluate participants from school, club sport and community-based sport contexts. Second, it allows the assessment of criteria from the individual player, small groups to the whole team, at the same time, although the criteria can be chosen according to the assessment purpose. Third, it makes possible the assessment of both attack (on and off the ball) and defence (of the attacker on and off the ball) roles. Finally, it contextualizes each criterion according to the specific situation in which the assessment is carried out. In short, this instrument allows the alignment of the components of the pedagogical process in relation to objectives and assessment. Consequently, TAIS offers the possibility of authentic assessment in learning to play games.

In practical terms, the present instrument can be applied in the following way. The unit of observation is each game phase (attack phase and defence phase), represented as a row in an excel sheet. Given that the three tactical levels are nested in the unit of observation, in each game phase, evaluators should observe criteria from all the three levels (see Figure 2). It is recommended to start from criteria of the organizational match level (blue colour criteria in Figure 2), then criteria from partial forefront organizational level (red colour criteria in Figure 2) and after that, criteria from primary organizational level (green colour criteria in Figure 2), in order to assess the interrelated game actions which favours an authentic tactical assessment.
Depending on the game phase (attack or defence) the criteria analysed should be those that correspond to attack or defence roles. On the attack phase, there are criteria related to the type of attack (attack type) and both on-ball attacker (pass, dribbling, shoot) and off-ball attacker (amplitude, depth and support). On the defence phase, there are criteria related to the type of defence (defence type) and both, defender to on-ball attacker (interception and tackle or charging) and defender to off-ball attacker (defensive coverage).

The recording of the criteria in the excel sheet implies the identification of the category corresponding to attack type or defence type. Then, evaluators should record the frequency of appearance of appropriate and inappropriate criteria, according to their descriptions (Figure 2). Once the game phases have been recorded, results must be shown as a sum of each appropriate and inappropriate criterion. In doing so, it is prevented masking the results as occurs when using indexes.

Furthermore, results of each game phase could be contextualized, for example depending on the goal different in favour, as the instrument includes 11 contextual criteria (black colour in Figure 2). Considering these contextual criteria, the instrument provides evaluators useful information that allows to be more precise when designing the tactical pedagogical process. Nevertheless, teachers, coaches or researchers could choose the criteria depending on the focus of the lessons/unit. The fact that teachers can choose the criteria according to the pedagogical aims, implies that TAIS can be both formative and summative. However, we recommend evaluating all of them because they are interrelated to show players’ tactical learning.

As the instrument has been designed and validated with football players from eight to 12 years old from three different institutional contexts (club sport context, community-based football activities and school context), it could be used by coaches from both formal sport and extra-curricular sport context and teachers in physical education. At the same time, this instrument could be challenging for teachers since it can be only used for one sport. If the
instrument is to be used in other sports, contexts or age groups, it should be validated, for example, following the stages and phases presented in this manuscript.

**What does this Study Add?**

This article is significant in that it shows an instrument that presents several advantages in practical terms regarding the assessment of: (a) the three tactical levels nested in the unit of observation, (b) all the player roles, (c) the results without general indexes, (d) participants from all the institutional contexts, and (e) the contextual variables. The present instrument adds to the existing ones in a significant way. For example, the instrument allows to be aware of what tactical level need to be improved as a group and not only individually, whether teacher/coach should focus on attack or defence actions (with or without the ball) and considering specific game situations, such as ‘Ball divided from the point of view of defence players’ (which is very frequent in youth sports). This instrument offers the opportunity to align the pedagogical components and assist the teacher/coach in teaching and the student/players in understanding their learning/performance. This instrument considers multiple aspects of sport-related games that are socially dynamic and complex in nature, particularly assessment as it relates to tactics. Furthermore, the design of this study is also unique in that it differentiated the design and validation stages with three and five exhaustive phases, respectively, considering participants from the three specific contexts of football in all of these phases.
References

Authors a


Washington DC: APA, AERA, NCME.


doi:10.1177/0013164485451012


doi:10.2165/00007256-199826040-00002


doi:10.3389/fpubh.2018.00149


Table 1

Preliminary List of Criteria and Categories

Table 2

Average of Scores given by Experts and Percentage of Experts that Scored Each Criterion with Three or Less to Establish Content and Comprehension Validity

Table 3

First version of TAIS

Table 4

Differences between Skill Levels in Each Criterion Assessed with TAIS

Table 5

Correlations between Criteria from TAIS in both Appropriate and Inappropriate Categories and Criteria Assessed with GPAI Decision-Making, Skill Execution, Cover and Support Components

Figure 1. Stages and phases followed in the development and validation of the instrument. S1P2: Stage 1 Phase 2.

Figure 2. Final version of TAIS.