

1 Carmen Barquero-Ruiz, David Kirk & José L. Arias-Estero (2022) Design and Validation of  
2 the Tactical Assessment Instrument in Football (TAIS). *Research Quarterly for Exercise and*  
3 *Sport*, 93:3, 615-632.

4 To link to this article: <https://doi.org/10.1080/02701367.2021.1889457>

5

6

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

8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

### 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.

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

32 In the last two decades, there has been an increased interest from researchers on teaching  
33 games from a tactical perspective (for a review, see Kinnerk, Harvey, MacDonncha, & Lyons,  
34 2018). The importance of pedagogical processes of tactics made necessary the design of  
35 instruments in order to assess these processes appropriately, both in physical education and sport  
36 contexts (e.g., Game Performance Assessment Instrument [GPAI], Oslin, Mitchell, & Griffin,  
37 1998; and Team Sports Assessment Procedure [TSAP], Gréhaigne, Godbout, & Bouthier, 1997).  
38 Football is in general one of the games with the greatest social impact and highly practiced in an  
39 organized form by adults and children around the world (Fédération Internationale de Football  
40 Association, 2017). As a result, the tactical perspective has also impacted the pedagogical  
41 process in youth football (Kinnerk et al., 2018). However, the focus on tactics has made  
42 assessment more difficult for coaches, teachers and researchers, due to the lack of knowledge in  
43 the topic and the limitations of current tactical assessment instruments to report authentic, useful  
44 and valid data of tactical learning outcomes during actual game play in football (Authors a).

45 According to Biggs (1996), learning requires the alignment of the components of the  
46 pedagogical process (i.e., assessment, teachers, students, learning activities, and learning  
47 outcomes, among other). Particularly, assessment is the key element that drives the rest of the  
48 components. In order to be part of the conversation, physical education needs to connect to the  
49 broader views of education. From this perspective, authentic assessment demands connections  
50 with real game experiences and the components of the pedagogical process. As a consequence,  
51 assessment should be authentic by measuring learners' performance in situations as similar as  
52 possible to real game (Wiggins, 2011). This kind of assessment helps teachers, coaches and  
53 learners make sense of play with the context of the game. In relation to the relevance of  
54 assessment, the instruments used to assess learning outcomes could hinder the authenticity of  
55 the assessment and the pedagogical process if they are not valid and context bound. Hence,  
56 assessment instruments should allow the alignment and authenticity of the pedagogical process.

57 To date, four instruments were the most frequently used to assess tactical learning in  
58 youth football (Authors a). On the one hand, there were two generic observation instruments  
59 designed and validated in the 1990s. The first of these, TSAP, was created to assess individual  
60 performance in different team sports, in contexts of pre-assessment and formative assessment  
61 (Gréhaigne et al., 1997). It was intended for peer-assessment although it could be used by  
62 teachers and researchers. This instrument is based on two events, receiving the ball and playing  
63 the ball. From the observation of these two events, a 'global performance index' needs to be  
64 calculated. This index is the result of considering an overall 'volume of play' and 'efficiency  
65 index'. TSAP offers the possibility to measure the on-ball attack, using the individual player as  
66 unit of observation, both in video and in vivo. The second instrument, GPAI, was developed in  
67 the school context to observe 'game performance behaviours that demonstrate tactical  
68 understanding, as well as the player's ability to solve tactical problems by selecting and  
69 applying appropriate skills' (Oslin et al., 1998, p. 231). GPAI was intended to be used by  
70 teachers and students in peer or self-evaluation, although it could be use by researchers. GPAI  
71 includes seven tactical components (base, adjust, decision-made, skill execution, cover and  
72 guard/mark) forming the 'game performance index' and 'game involvement index'. GPAI offers  
73 the possibility to measure both on- and off- ball attack and defence in different sports, using the  
74 individual player as unit of observation, both in video and in vivo. Overall, TSAP and GPAI  
75 allow the discussion of ideas throughout peer assessment procedures, which empowered the  
76 pedagogical process as a form of authentic assessment (Wiggins, 2011).

77 On the other hand, two additional instruments were specifically developed for youth  
78 football. The first of these was the system of tactical assessment in football (FUTSAT), which  
79 was created to assess tactical behaviour of football players. It was intended to be used by  
80 coaches and researchers. FUTSAT is composed of two macro-categories (observation and  
81 tactical principles features) and 76 criteria, although it was not possible to identify all the criteria  
82 (see a description in Costa, Garganta, Greco, Mesquita, & Maia, 2011). Considering all of them,

83 the 'tactical performance index' can be calculated. When using FUTSAT, the unit of  
84 observation is each team attack or defence. FUTSAT offers the possibility to measure both on-  
85 and off- ball attack and defence in video. For example, Borges, Guilherme, Rechenchosky, da  
86 Costa, and Rinadi (2017), showed that U17 players met more frequently the criteria offensive  
87 coverage than U13 players, because they increased their confidence and security in offensive  
88 actions. The second, the Game Performance Evaluation Tool (GPET), was created to analyse  
89 decision-making and skill execution regarding to the tactical problems in relation to which  
90 decisions are made and skills are executed (García-López, González-Víllora, Gutiérrez-Díaz, &  
91 Serra-Olivares, 2013). It was intended to be used by teachers and coaches, although it could be  
92 also used by researchers. GPET sets the analysis of each decision made on tactical problems in  
93 which the players are involved within the game. This instrument is composed of 14 criteria and  
94 no indexes (see a description in García-López et al., 2013). When using GPET, the unit of  
95 observation is the individual player within each tactical problem. GPET offers the possibility to  
96 measure on-ball attackers in video. For example, Práxedes, Del Villar, Pizarro, and Moreno  
97 (2018) analysed the criterion 'pass' as a key game action included in two tactical problems:  
98 maintaining possession of the ball and progressing towards the goal. In both instruments, criteria  
99 refer to game actions (e.g., pass), categories refer to the discrete ways these actions can be  
100 executed throughout the game (e.g., back pass, forward pass, opening pass), and indexes offers  
101 information about the average tactical learning outcomes from formulae that combine the  
102 criteria assessed (e.g., in FUTSAT, tactical performance index is  $\Sigma$  tactical actions / number of  
103 tactical actions, Costa et al., 2011).

104         According to the most recent review regarding assessment practices in tactical learning  
105 in games, both in physical education and sport contexts, these instruments present five main  
106 limitations considering the purposes for which researchers used them (Authors a). First, these  
107 instruments do not consider the interactions among whole team, small groups of players and  
108 individual players when assessing team tactical performance, as TSAP proposed. These

109 interactions can be structured in three organizational levels (Deleplace, 1979; Gréhaigne,  
110 Richard, & Griffin, 2005). The first level, organizational match level, refers to the collective  
111 game actions that imply more than three players from the same team. The second level, partial  
112 forefront organizational level, refers to the game actions developed by at least two players.  
113 Finally, primary organizational level refers to the game actions developed by individual players.  
114 Therefore, organizational match level breaks down into partial opposition relationships forming  
115 the partial forefront organizational level that contains a unit called primary organizational level  
116 (e.g., Deleplace, 1979; Gréhaigne et al., 2005; Kirk, 2017). In practical terms, the levels allow  
117 identification of game actions attending to the number of players involved in such game actions.  
118 Consequently, this identification makes possible the assessment of interrelated game actions  
119 from different levels that have not been measured until now, given that the sum of individual  
120 tactical outcomes does not correspond to team tactical outcomes. However, the three levels  
121 could be applied within a tactical assessment tool nested in the same unit of observation. For  
122 example, considering the attack phase as unit of observation, at organizational match level, a  
123 team could play with ‘amplitude’ moving the ball from one side to the other in order to generate  
124 free spaces. Considering the partial forefront organizational level, this movement of the ball  
125 could be done by giving the ball from one player to other using ‘passes’. Regarding the primary  
126 organizational level, when players are close enough to goal, these passes should result in an  
127 individual shot. According to Kirk (2017), the evaluation of players’ tactical outcomes in each  
128 organizational level favours an authentic tactical assessment.

129         Second, only GPAI and FUTSAT consider all the players’/learners’ roles. Including  
130 information of both attacker and defender roles is also necessary, because defender roles and  
131 decisions made without possession of the ball have a great weight in the total outcomes and are  
132 essential for learning as a team/group (McPhail, Kirk, & Griffin, 2008). Third, all of the  
133 instruments, except for GPET, use indexes or ratios that can hide the nature of the player’s  
134 learning outcome. Moreover, showing the learning outcome in a single datum risks dismissing

135 information that may be of interest in improving learning (Authors a). Nevertheless, there are  
136 indexes, such as game involvement index in GPAI, that have the potential to provide meaningful  
137 information about players' and students' learning because all decisions are counted. Fourth, the  
138 instruments were used without considering the institutional context in which they were validated  
139 (club sport context, community-based football activities and school context), which influences  
140 the outcomes (Rovegno & Kirk, 1995). For example, in the FUTSAT study only participated  
141 players from club sport context and not from physical education (Costa et al., 2011). However,  
142 authors suggest that this instrument can be used in school context. Finally, none of the  
143 instruments include contextual variables, except for GPET and FUTSAT, that presented the  
144 criteria contextualized in tactical principles of play. Tactical outcomes are environment-  
145 dependent, not only considering the tactical principles of play, but requiring the inclusion of  
146 contextual variables that allows setting the assessment in the reality or concrete situation in  
147 which the assessment is done (Sal de Rellán-Guerra, Rey, Kalén, & Lago-Peñas, 2019). For  
148 example, when players are winning, they could make better decisions, because decision making  
149 is affected by game outcomes (Sal de Rellán-Guerra et al., 2019).

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

### 157 **Design**

158 The design of the study was developed in two stages (Figure 1). Stage 1 focused on the  
159 development and design of the instrument while Stage 2 determined the validity and reliability  
160 of the instrument through multiple phases. Stage 1 was subdivided into three phases. In Phase 1,

161 a preliminary list of criteria was determined through a literature review. In Phase 2, the criteria  
162 were delimited through an exploratory observation of several games. Finally, in Phase 3, the  
163 adequacy of the criteria to the aim of the instrument was determined by consulting experts. The  
164 Stage 2 was subdivided into five phases. In Phase 1, content and comprehension validity were  
165 obtained by consulting experts. In Phase 2, content and comprehension validity was confirmed  
166 through a systematic game observation. In Phase 3, construct validity was obtained by  
167 comparing the tactical outcomes from two groups with different skill levels. In Phase 4, criterion  
168 validity was established by comparing the tactical outcome using GPAI and the present  
169 instrument. Finally, in Phase 5, the reliability of the instrument was obtained through inter-rater  
170 reliability.

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

176 \*\*\*\* Figure 1\*\*\*\*

### 177 **Gaining entry and access**

178 Regarding to the selection of the participants, the first author screened all interested  
179 participants for eligibility using a standardised script and email message. These messages and  
180 criteria of eligibility were different depending on the type of participant (experts, observers and  
181 players) and phase. The criteria for experts were (a) at least 10 years of experience in  
182 researching and (b) in research topics related on the aim of the study. Criteria for observers were  
183 (a) at least 3 years of experience in teaching physical education or coaching in different contexts  
184 and (b) having a degree in sport sciences or physical education. Criteria for players were (a)  
185 coming from club sport context, school context and community-based sport context, (b) being  
186 from different skill levels, (c) training with different frequencies and (d) being between eight



187 and 12 years old. Participants who answered the email within 20 days were selected.  
188 Approximately 40% of the people contacted in each phase were selected to participate. Once  
189 they were selected, they were informed of the protocol. Participants were different in all phases,  
190 except for Stage 2 Phase 2 (S2P2), where participants were the same of Stage 1 Phase 2 (S1P2).  
191 The parents of the players signed an informed consent document before the investigation, and  
192 players assented to participate. Players, parents and observers in Stage 2 Phase 3 (S2P3), Stage 2  
193 Phase 4 (S2P4) and Stage 2 Phase 5 (S2P5), were blinded to the study aim, but the experts and  
194 observers from S1P2 and S2P2 were necessarily informed about it. The main author's  
195 University Research Ethics Committee approved the study, which was performed in accordance  
196 with the Helsinki Declaration.

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

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

199 **Procedure.** A review about tactical assessment in youth sport and physical education  
200 was carried out to explore the possible criteria and categories of the instrument. Tactical  
201 assessment was defined as the assessment carried out during a real game, considering techniques  
202 and tactics as two inseparable components of a player's learning. Criteria refer to game actions  
203 (e.g., attack type) and categories refer to the different ways in which these actions can happen  
204 throughout the game (e.g., positional attack, counter-attack; Table 1). For example, 'Attack  
205 type' is determined as the spatial arrangement in the attack. It is considered 'positional attack'  
206 when defence adjust with attackers and attackers take time to reorganize themselves on the pitch.  
207 It is considered 'counter-attack' when attackers do not allow defence to recover their positions  
208 (Figure 2). The bibliographic search was conducted using the following terms: tactical learning,  
209 tactical performance, physical education, observational analysis, tactical assessment instruments,  
210 sport pedagogy and youth games. The quality criteria for the review were: (a) appeared in  
211 journals indexed in the Science Citation Index, Science Citation Index Expanded and Social  
212 Sciences Citation Index; (b) from peer-review journals; (c) both from teaching and coaching

213 contexts and (d) empirical studies that present conclusions or objectives related to the  
214 pedagogical process and assessment of tactical outcomes. As a result of the analysis of the  
215 previous studies, preliminary criteria and categories were established at the three tactical levels  
216 and the contextual level. Two of the authors met for three hours in three consecutive days for  
217 five weeks to decide which criteria had the best fit with the aim of the study, and classified them  
218 into the three tactical levels and contextual level, according to the tactical level proposal that  
219 supports the instrument (e.g., Deleplace, 1979; Gréhaigne et al., 2005; Kirk, 2017) and the  
220 relevance of contextual variables (e.g., Sal de Rellán et al., 2018).

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

229 \*\*\*\*\*Table 1\*\*\*\*\*

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

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

238 football at least two days per week in community-based sport context, and eight only practiced  
239 football in school context two days per week.

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

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

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

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

278 **Results.** When compared to the literature review, the 28 criteria identified in S1P1 were  
279 found in the observation with a frequency of more than 20%. On the contrary, this percentage  
280 was not found in the case of the categories 'clear or catch', 'attack mistake' and 'defensive  
281 mistake' from the criterion 'finalizing type in attack'. Consequently, these categories were  
282 removed from this criterion and changed by the categories 'own goal', 'previous action to goal  
283 kick', 'losing ball' and 'save from goalkeeper' as they appeared with a frequency of more than  
284 20%. Furthermore, the criterion 'support' was found in the observation with a frequency of  
285 more than 20% and consequently added to TAIS. Regarding the criterion 'recovery type', the  
286 categories 'interception or goalkeeper block', 'attackers mistake' and 'end without recovering'  
287 were added as they appeared with a frequency of more than 20%. At the end of this phase, TAIS  
288 was composed by 29 criteria, ten of them included categories.

289 **Phase 3. Adequacy of the instrument through panel of experts**

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

296           **Procedure.** The experts were asked to indicate which criteria and categories, from the  
297 list after S1P2, should be part of the present instrument, considering their tactical and contextual  
298 levels and according to its aim. In addition, they were asked to define each criterion and  
299 category and encouraged to propose new criteria and categories. They were informed about the  
300 nature and objectives of the present instrument. They were contacted by email and provided  
301 with a list of all the criteria and categories in the tactical and contextual levels. They had to  
302 assess the criteria quantitatively using a 5-point Likert-type scale, ranging from 1 (strongly  
303 disagree) to 5 (strongly agree). They were also asked to give explanations regarding to the  
304 scores they gave. Lastly, authors analysed and discussed the experts' answers following the  
305 Bulger and Housner (2007) conditions to remove the criteria: (a) that received a mean rating of  
306 less than three or (b) that were ranked with less than three by more than the 30% of the experts.  
307 After that, the authors made a first proposal of criteria and categories, including their definitions  
308 according to the literature and the information collected from observers and experts.

309           **Results.** As result of the quantitative analysis from expert evaluation, nine criteria were  
310 removed ('change of role control', 'depth of attack', 'retract', 'tempo control', 'wall pass',  
311 'fixing the player', 'centre', 'control' and 'marking'). Those criteria met at least one of the  
312 following conditions, they: (a) obtained an average scored of three or less or (b) were scored  
313 with less than three in content or comprehension by more than 30% of the experts (Table 2). As  
314 a result of the qualitative analysis, eight of the experts suggested changing the category 'ball  
315 divided' from 'situation type' to clarify whether the ball was divided from the point of view of

316 the attackers or the ball was divided from the point of view of the defence. Furthermore, to  
317 criterion ‘clearance’ they suggested to rename it as ‘interception’. As a consequence, the first  
318 version of TAIS was developed. It was composed by 20 criteria, 9 of them categorized. This  
319 version included the definitions of the criteria and categories (Table 3).

320 \*\*\*\*\*Table 2\*\*\*\*\*

321 \*\*\*\*\*Table 3\*\*\*\*\*

## 322 **Stage 2. Validity of the instrument**

### 323 **Phase 1. Content and comprehension validity through panel of experts**

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

331 **Procedure.** The panel of experts checked for each criterion: (a) content, whether the  
332 descriptions of each criteria and its category were adequate to what we wanted to measure and  
333 (b) comprehension, whether the descriptions and its categories were comprehensible and  
334 correctly expressed. In addition, they were asked to assess in general: (a) whether the criteria  
335 classification was appropriate and corresponded to each tactical level and (b) whether they  
336 found the instrument useful. The panel of experts had to assess these aspects quantitatively,  
337 using a 5-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree), and  
338 qualitatively, including explanations or proposals. They were also provided with an open space  
339 to express other suggestions. The panel of experts was informed about the nature and objectives  
340 of the instrument. They were contacted by email and provided with a tool to do this evaluation.  
341 Lastly, authors analysed and discussed the experts’ answers. The criteria scored as less than four

342 were revised. We decided which changes to make considering the aim of the instrument. We  
343 considered all the comments that: (a) met the objective of the present instrument and (b) alluded  
344 to the specific contexts of youth football. After that, the tool was re-sent to the same experts in  
345 order for them to re-evaluate the instrument following the same instructions. This process was  
346 repeated for any of the criteria or categories that were scored less than four, which occurred  
347 twice (Bulger & Housner, 2007). Finally, the Aikens's V coefficient was calculated on the  
348 second evaluation (Aiken, 1985).

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

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

## 361 **Phase 2. Content and comprehension validity through systematic game observation**

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

363 **Procedure.** In this phase, we conducted a systematic observation of the four games from  
364 S1P2 to check if the criteria and categories descriptions were operative. That means, to  
365 substantiate whether it was possible to identify easily the criteria and categories described.  
366 Observers were asked to observe the tactical outcome for each criterion using the instrument.  
367 This task was undertaken during 12 meetings and the observers had to view 26 game phases

368 each meeting. The game phases lasted between five and 15 seconds. They observed a total of  
369 320 game phases. The six observers observed and coded all game footage individually. All  
370 meetings were two hours with each hour having a specific focus. In the first hour, observers  
371 focused on whether the criterion was met or not met. For instance, for the criterion 'shoot', they  
372 identified when a shot on goal was made. Then they registered whether the criteria were  
373 appropriate according to its definition when one of the following circumstances occurred: (a) the  
374 shot resulted in a goal, (b) the shot was directed to goal but there was no score (either it missed  
375 the goal or was saved or cleared) or (c) when the shot missed the goal and resulted in a corner  
376 kick. Criteria were coded in the three tactical levels and the contextual level by observers. They  
377 were nested as the same unit of observation for each game phase. Observers used Virtual Dub  
378 Version 1.10.4 to play the games and an excel sheet to record the information. They were able to  
379 stop the videotapes as many times as necessary and when they had any doubts, the procedure  
380 was to write it down and discuss at the end of the meeting. In the second hour of the meeting, all  
381 the observers and authors discussed together the observers' doubts about the criteria definition,  
382 until an agreement was reached. As a result, the authors modified the definition of criteria and  
383 categories until they achieved a version that allowed the observation of tactical outcomes with  
384 operative criteria.

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

392 **Phase 3. Construct validity through the analysis of tactical outcomes from different skill**  
393 **levels**



394           **Participants.** Participants were 24 players and four observers. Players were aged  
395 between eight and 12. Ten were from a club sport context (club academy programmes),  
396 competing in the regular league and with between four and five years experience. Six were  
397 enrolled in community-based football activities and had one year experience maximum. The  
398 remaining eight were physical education students with no previous experience in football.  
399 Furthermore, regarding the frequency of training, all of them practiced football at least two days  
400 per week.

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

407           **Procedure.** A three-day tournament was conducted and consisted of four 20 minutes  
408 4GK vs 4GK games, based on recommendations from the literature (Machado et al., 2019).  
409 Players were divided into two groups according to their skill level (low or high). They were  
410 classified after a GPAI observation of previous games by expert coaches. Then, the low skill  
411 level group and high skill level groups were randomly subdivided in two subgroups of five  
412 players each. In all groups there were players who came from the three different contexts. The  
413 games were played with official eight to 12 years old laws of the game. The games were  
414 conducted twice between teams with similar skill levels. So that, there were two games between  
415 low skill level groups and two games between high skill level groups. The games were video-  
416 recorded.

417           Observers were asked to observe the tactical outcomes of all players using the instrument  
418 and the recordings of all matches from this three-day tournament (25 frames per second).  
419 Observers coded all criteria defined in TAIS for each game phase. The level of the participants

420 was hidden from the observers. The observation technique was the same described in S2P2, but  
421 in this case there were neither meetings nor final discussion, as they only focused on the  
422 observation. The observers were trained for at least 10 hours in the use of the instrument. The  
423 observation reliability was adequate given the minimum values were set at .70 for Intraclass  
424 Correlation Coefficient (ICC) and Kappa coefficient and 95% for Percentage of Agreement (PA;  
425 Atkinson & Nevill, 1998; Robinson & O'Donoghue, 2007). The four games produced a total of  
426 780 game phases. The observers viewed all the games individually. Wilcoxon test was used to  
427 explore the possible differences between skill levels in all the criteria. Effect size (*ES*) was also  
428 calculated. Statistical significance was set at  $p < .05$ .

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

435 \*\*\*\*\*Table 4\*\*\*\*\*

#### 436 **Phase 4. Criterion validity through the evaluation of the tactical outcome using GPAI and** 437 **the present instrument**

438 **Participants.** Participants in this phase were 30 players and four observers. The players  
439 were aged between eight and 12 years. Seventeen were from a club sport context (club academy  
440 programmes), competing in the regular league and with between three and five years experience.  
441 Seven were enrolled in community-based football activities and had two year experience  
442 maximum. The remaining six were physical education students with no previous experience in  
443 football. Furthermore, regarding the frequency of training, all of them practiced football at least  
444 two days per week.

445 The observers were (a) two coaches graduated in sport sciences with at least one-year  
446 experience as football coach in both club sport context (n=1) and community-based sport (n=1),  
447 and (b) two PhD students in sport sciences (n=1) and physical education (n=1) with experience  
448 in using GPAI in previous studies.

449 **Procedure.** A one-day tournament was organized to this phase. The players were  
450 randomly divided into six teams of five players each. In all groups there were players who came  
451 from the three different contexts and levels. Six games of 20 minutes each were played in  
452 4GKvs4GK according to literature recommendations. The games were played with official eight  
453 to 12 years old laws of the game. The games were video-recorded.

454 The observers were asked to observe the tactical outcome for each criterion using GPAI  
455 (Oslin et al., 1998) and the present instrument. GPAI was selected based on the following. First,  
456 it was the only instrument that allows comparing criterion by criterion using open criteria  
457 description and avoiding the use of indexes. Second, is the most widely extended instrument  
458 used in physical education and youth sports (Authors a). Third, although GPET and FUTSAT  
459 are specific for football, it was impossible to access to their operative criteria descriptions.  
460 However, observers assessed only the primary and partial forefront organizational level, because  
461 organizational match level cannot be assessed with GPAI. Nevertheless, none of the existing  
462 instrument allows assessing the three tactical levels with similar criteria. Two of the observers  
463 evaluated all the footages of the games with GPAI and the other two with the present instrument,  
464 all of them individually. The observation technique for GPAI was systematic because the  
465 observers assessed all the players' game actions. The observers were asked to observe 'cover',  
466 'support', 'decision-making' and 'skill execution' components of GPAI for all the criteria (see  
467 criteria in Table 5). The observers added the number of appropriate and inappropriate decisions,  
468 and correct and incorrect skill executions, according to the definition of a previous study  
469 (Authors b). Adequate decisions included making appropriate choices about what to do during  
470 the game. Correct skill executions corresponded to an efficient performance of the selected skill.

471 The observers training and the observation technique for the present instrument was the same as  
472 described in S2P3. The observation reliability was adequate given the minimum values were  
473 over .70 for ICC/Kappa and over 95% for PA (Atkinson & Nevill, 1998; Robinson &  
474 O'Donoghue, 2007). The six games produced a total of 986 game phases observed with both  
475 instruments.

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

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

487 \*\*\*\*\*Table 5\*\*\*\*\*

#### 488 **Phase 5. Inter-rater reliability**

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

496 two days per week in the community-based sport context, and four only practiced in the school  
497 context two days per week.

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

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

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

513 **Results.** Reliability exceeded .70 according to ICC or Kappa coefficient value (Cohen,  
514 1960), and the 85% according to PA (Brewer & Jones, 2002). Lowest reliability was found for  
515 criterion 'goal difference in favour' (ICC=.70, PA=95%) while highest reliability was found for  
516 criteria 'team', 'score board' and 'period' (Kappa/ICC=1; PA=100%).

## 517 **Discussion**

518 The purpose of this study was to design and validate an instrument to assess tactics in  
519 youth football following an exhaustive, ecological and meticulous process dealing with the  
520 limitations found in other instruments. According to Carretero-Dios and Pérez (2007), design  
521 and validation processes are needed to ensure the accuracy of assessments. Consequently, in the

522 present study we differentiated the design and validation stages with three and five phases,  
523 respectively. At the end of the two stages the instrument included 22 criteria, nine of them with  
524 categories (Figure 2). All of them were carefully selected and validated for the three specific  
525 contexts of football (club sport context, community-based football activities and school context).  
526 In general, none of the current instruments for tactical assessment (FUTSAT, GPET, TSAP and  
527 GPAI) were developed according to the stages and phases outlined in the present study, neither  
528 did they include information about criteria selection and elimination in each phase.

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

544         Similarly to the present work, current instruments for tactical assessment were developed  
545 including content and comprehension validity through experts, except for TSAP (Greháigne et  
546 al., 1997). However, experts in the case of GPET were only teachers and coaches, while in  
547 FUTSAT and TAIS they were also researchers. The fact that GPET was validated by teachers

548 and coaches, could imply that content validity in GPET was useful in practical terms but it was  
549 not connected with research purposes (García-López et al., 2013). Furthermore, none of the  
550 studies, except for the present one, did a second round of panel of experts after modifying the  
551 instrument according to experts' suggestions. Consequently, the present study is the only one  
552 that confirmed the validity with experts after modifying the criteria and categories definitions,  
553 obtaining high Aiken's V values (Aiken, 1985). In addition, in the present study we also  
554 obtained content and comprehension validity through game observation in order to check that  
555 the definitions of criteria and categories made their observation possible. A similar procedure  
556 was followed in the TSAP study, but this was to check that the criteria occur with certain  
557 frequency during game play. However, it is necessary to consider that TSAP is a peer  
558 assessment instrument (Greháigne et al., 1997). So that, at the end of the two phases to obtain  
559 content and comprehension validity, we modified 13 criteria and two categories. This  
560 information is not available in the studies that validated other instruments; although GPET and  
561 GPAI pointed out that they modified criteria according to expert comments (García-López et al.,  
562 2013; Oslin et al., 1998). Therefore, as the content and comprehension validities were obtained  
563 by two different procedures the instrument is stronger, because it was verified both theoretically  
564 by experts and practically by observers (American Educational Research Association [AERA],  
565 American Psychological Association [APA], National Council on Measurement in Education  
566 [NCME], 1999).

567 All the validation studies of the other instruments for tactical assessment, except for  
568 TSAP, developed a phase in which the tactical outcomes were analysed from different skill  
569 levels in real games to obtain construct validity. In terms of results, the GPAI validation study  
570 did not show that GPAI discriminated between skill levels in decision-making and support for  
571 basketball nor adjust and support for volleyball (Oslin et al., 1998). Similarly, GPET validation  
572 study did not find differences in skill execution for passing and dribbling nor decision-making  
573 for dribbling and shooting (García-López et al., 2013). Regarding FUTSAT, there is no

574 information available in the study, despite the fact that authors confirmed its construct validity  
575 (Costa et al., 2011). In comparison, TAIS discriminates between skill levels in all the criteria  
576 except for shoot, due to the low frequency of this criterion in the game. The main difference,  
577 however, resides in the strategy followed to determine the participants' skill levels and in the  
578 participants' contexts. On the one hand, whereas FUTSAT established the skill level using  
579 performance indexes from their own system and GPET determined the level in function of the  
580 participant context, GPAI and TAIS distinguished between individuals previously rated as high  
581 and low in game performance. On the other hand, in GPAI and TSAP studies, there were only  
582 participants from the school context (Greháigne et al., 1997; Oslin et al., 1998), in the FUTSAT  
583 study there were only participants from club sport context (Costa et al., 2011), and in GPET  
584 study the participants came from club sport context and school context (García-López et al.,  
585 2013). However, in the present study, there were participants from the three different  
586 institutional contexts (club sport context, community-based sport context and school context).  
587 Given the strategy followed to distinguish between participants' skill levels and that participants  
588 came from three different institutional contexts, TAIS can be used objectively in the three youth  
589 football contexts, as participants determine the extent of the validity (Boateng et al., 2018).

590 TSAP design study was the only that also included a criteria validity phase as in the  
591 present study (Greháigne et al., 1997). Their correlations amounted to .74 and the lowest value  
592 of the TAIS was .60. Nonetheless, whereas in the TSAP study the reference criterion for  
593 comparison was the agreement of two football experts, we used the assessment of the tactical  
594 outcomes with GPAI as the reference criterion. This validity let external evidence of score  
595 validity, which provides the information about the usefulness or meaning of the test scores  
596 (AERA et al., 1999). However, this external evidence was not complete because TAIS goes  
597 further than any of the other instruments, including the organizational match level criteria,  
598 contextual criteria and without discriminating between technical and tactical components. In  
599 addition, the comparison was done criterion by criterion, instead of using GPAI indexes because



600 TAIS does not include indexes. Similarly to GPET, TAIS followed the literature  
601 recommendations, as using indexes or ratios can mask the results (García-López et al., 2013).  
602 As we noted in the introduction, reflecting the learning outcome in a single datum risks  
603 dismissing information that may be of interest in improving learning (Authors a).

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

618 At the end of the development and validation stages, the result was an instrument  
619 comprising 22 criteria, nine of them with categories, organised in three tactical levels and a  
620 contextual level. In contrast, the other instruments for tactical assessment present between 7  
621 components (GPAI) and 76 criteria (FUTSAT). Moreover, none of them differentiate between  
622 tactical levels or include contextual criteria, although they include criteria from at least one of  
623 the three levels but not nested in the unit of observation. The TAIS tactical level division allows  
624 comparison of the tactical outcomes within each level and according to specific game situations  
625 (Rovegno & Kirk, 1995). Given game situations are context-dependent, contextual criteria

626 enable the reduction of bias caused by their influence (Sal de Rellán-Guerra et al., 2019).  
627 Furthermore, as tactical assessment demands the contextualisation of game situations in a  
628 reference framework, these contextual criteria are crucial for making possible an authentic  
629 assessment. Overall, with the nested unit of observation, we overcome the problem of  
630 considering team tactical outcomes as the sum of individual tactical outcomes, present in most  
631 of the other instruments (GPAI, TSAP and GPET). As a consequence, it is possible to know the  
632 level in which players experience more difficulties and what level or situation correlates with  
633 better learning (Gréhaigne et al., 2005).

634 While GPET only considers criteria from an attacker's role (García-López et al., 2013),  
635 FUTSAT and GPAI include criteria from attack and defence (Costa et al., 2011; Oslin et al.,  
636 1998; both on and off the ball), as in TAIS. However, although GPAI includes four components  
637 for each role, all of them are open description, which demands the adaptation and validation of  
638 criteria for each study (Authors a). On the other hand, though FUTSAT includes 38 criteria from  
639 each role, many of them are difficult to observe in the discrete youth football contexts as they  
640 only validated the criteria in club sport context (Gutiérrez-Díaz, González-Villora, García-López,  
641 & Mitchell, 2011). In contrast, TAIS presents eight closed attacker criteria and four closed  
642 defence criteria validated for all the contexts.

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

652 operatively because they aimed to create a flexible self-constructed instrument. Regarding the  
653 scores generated by instruments, only GPET and the present instrument avoid the use of indexes,  
654 using the sum of each appropriate and inappropriate criterion. On the contrary, FUTSAT, GPAI  
655 and TSAP use indexes including in the same formulae data from all criteria, what could mask  
656 the results (Memmert & Harvey, 2008).

### 657 **Conclusion and practical application**

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

668 Consequently, TAIS offers the possibility of authentic assessment in learning to play games.

669 In practical terms, the present instrument can be applied in the following way. The unit  
670 of observation is each game phase (attack phase and defence phase), represented as a row in an  
671 excel sheet. Given that the three tactical levels are nested in the unit of observation, in each  
672 game phase, evaluators should observe criteria from all the three levels (see Figure 2). It is  
673 recommended to start from criteria of the organizational match level (blue colour criteria in  
674 Figure 2), then criteria from partial forefront organizational level (red colour criteria in Figure 2)  
675 and after that, criteria from primary organizational level (green colour criteria in Figure 2), in  
676 order to assess the interrelated game actions which favours an authentic tactical assessment.

677            Depending on the game phase (attack or defence) the criteria analysed should be those  
678 that correspond to attack or defence roles. On the attack phase, there are criteria related to the  
679 type of attack (attack type) and both on-ball attacker (pass, dribbling, shoot) and off-ball  
680 attacker (amplitude, depth and support). On the defence phase, there are criteria related to the  
681 type of defence (defence type) and both, defender to on-ball attacker (interception and tackle or  
682 charging) and defender to off-ball attacker (defensive coverage).

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

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

697            As the instrument has been designed and validated with football players from eight to 12  
698 years old from three different institutional contexts (club sport context, community-based  
699 football activities and school context), it could be used by coaches from both formal sport and  
700 extra-curricular sport context and teachers in physical education. At the same time, this  
701 instrument could be challenging for teachers since it can be only used for one sport. If the

702 instrument is to be used in other sports, contexts or age groups, it should be validated, for  
703 example, following the stages and phases presented in this manuscript.

704 **What does this Study Add?**

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

720 **References**

721 Authors a

722 Authors b

723 AERA, APA and NCME (1999) *Standards for Educational and Psychological Tests*.

724 Washington DC: APA, AERA, NCME.

725 Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings.

726 *Educational and Psychological Measurement*, 45, 131-142.

727 doi:10.1177/0013164485451012

728 Anguera, M. T. (2003). Observational methods (General). In R. Fernández-Ballesteros (Ed.),

729 *Encyclopedia of psychological assessment, Vol. 2* (pp. 632-637). London: Sage.

730 Arias, J. L., Argudo, F. M. &amp; Alonso, J. (2009). An objective method for analyzing two three-

731 point shapes in mini-basketball. *Revista Internacional de Medicina y Ciencias de la*732 *Actividad Física y del Deporte*, 9, 349-365.

733 Atkinson, G. &amp; Nevill, A. M. (1998). Statistical methods for assessing measurement error

734 (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26, 217-238.

735 doi:10.2165/00007256-199826040-00002

736 Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32,

737 347–364. doi:10.1007/BF00138871

738 Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quíñonez, H. R., &amp; Young, S. L.

739 (2018). Best practices for developing and validating scales for health, social, and

740 behavioral research: A primer. *Frontiers in Public Health*, 6, 149.

741 doi:10.3389/fpubh.2018.00149

742 Borges, P. H., Guilherme, J., Rechenchosky, L., da Costa, L., &amp; Rinadi, W. (2017).

743 Fundamental tactical principles of soccer: A comparison of different age groups. *Journal*744 *of Human Kinetics*, 58, 207-214. doi:10.1515/hukin-2017-0078

- 745 Brewer, C. J., & Jones, R. L. (2002). A five-stage process for establishing contextually valid  
746 systematic observation instruments: The case of rugby union. *The Sport Psychologist, 16*,  
747 138-159. doi:10.1123/tsp.16.2.138
- 748 Brown, E., & O'Donoghue, P. (2007). Relating reliability to analytical goals in performance  
749 analysis. *International Journal of Performance Analysis in Sport, 7*, 28-34.  
750 doi:10.1080/24748668.2007.11868385
- 751 Bulger, S. M. & Housner, L. D. (2007). Modified delphi investigation of exercise science in  
752 physical education teacher education. *Journal of Teaching in Physical Education, 26*, 57-  
753 80. doi:10.1123/jtpe.26.1.57
- 754 Carretero-Dios, H. & Pérez, C. (2007). Standards for the development and review of  
755 instrumental studies: Considerations about test selection in psychological research.  
756 *International Journal of Clinical and Health Psychology, 7*, 863-882.
- 757 Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological*  
758 *Measurement, 20*, 37-46. doi:10.1177/001316446002000104
- 759 Costa, I. T., Garganta, J., Greco, P. J., Mesquita, I., & Maia, J. (2011). System of tactical  
760 assessment in soccer (FUT-SAT): Development and preliminary validation. *Motricidade*,  
761 *7*, 69-83. doi:10.6063/motricidade.7(1).121
- 762 Deleplace, R. (1979). *Rugby de Mouvement – Rugby Total [Rugby in Movement – Total Rugby]*.  
763 Paris, Francia: Éducation Physique et Sports.
- 764 Fédération Internationale de Football Association (2017). "Global Club Football Report 2017."  
765 Report of the FIFA Retrieved from  
766 [https://resources.fifa.com/mm/document/footballdevelopment/proffootballdept/02/90/12/](https://resources.fifa.com/mm/document/footballdevelopment/proffootballdept/02/90/12/72/clubfootballreport_29.6.2017_neutral.pdf)  
767 [72/clubfootballreport\\_29.6.2017\\_neutral.pdf](https://resources.fifa.com/mm/document/footballdevelopment/proffootballdept/02/90/12/72/clubfootballreport_29.6.2017_neutral.pdf)
- 768 García-López, L. M., González-Víllora, S. G., Gutiérrez-Díaz D. C., & Serra-Olivares, J. (2013).  
769 Development and validation of the game performance evaluation tool (GPET) in soccer.  
770 *Sport TK, 2*, 89-99.

- 771 Gréhaigne, J. F., Godbout, P., & Bouthier, D. (1997). Performance assessment in team sports.  
772 *Journal of Teaching in Physical Education, 16*, 500-516. doi:10.1123/jtpe.16.4.500
- 773 Gréhaigne, J. F., Richard, J. F., & Griffin, L. (2005). *Teaching and learning team sports and*  
774 *games*. London: Taylor & Francis.
- 775 Gutiérrez-Díaz, D. C., González-Víllora, S., García-López L. M., & Mitchell, S. (2011).  
776 Differences in decision-making development between expert and novice invasion game  
777 players. *Perceptual and Motor Skills, 112*, 871–888.  
778 doi:10.2466/05.10.11.25.PMS.112.3.871-888
- 779 Kinnerk, P., Harvey, S., MacDonncha, C., & Lyons, M. (2018). A review of the game-based  
780 approaches to coaching literature in competitive team sport settings. *Quest, 70*, 401-418.  
781 doi:10.1080/00336297.2018.1439390
- 782 Kirk, D. (2017). Teaching games in physical education: Towards a pedagogical model. *Revista*  
783 *Portuguesa de Ciências do Desporto, 17*, 17-26. doi:10.5628/rpcd.17.S1A.17
- 784 Machado, G., Padilha, M. B., González, S., Clemente, F. M., & Teoldo, I. (2019). The effects of  
785 positional role on tactical behaviour in a four-a-side small-sided and conditioned soccer  
786 game. *Kinesiology, 51*, 261-270. doi:10.26582/k.51.2.15
- 787 McPhail, A., Kirk, D. & Griffin, L. (2008). Throwing and catching as relational skills in game  
788 play: Situated learning in a modified game unit. *Journal of Teaching in Physical*  
789 *Education, 27*, 100-115. doi:10.1123/jtpe.27.1.100
- 790 Memmert, D., & Harvey, S. (2008). The game performance assessment instrument (GPAI):  
791 Some concerns and solutions for further development. *Journal of Teaching in Physical*  
792 *Education, 27*(2), 220-240. doi:10.1123/jtpe.27.2.220
- 793 Oslin, J. L., Mitchell, S. A., & Griffin, L. L. (1998). The game performance assessment  
794 instrument (GPAI): Development and preliminary validation. *Journal of Teaching in*  
795 *Physical Education, 17*, 231-243. doi:10.1123/jtpe.17.2.231



- 796 Práxedes, A., Del Villar, F., Pizarro, D., & Moreno, A. (2018). The impact of nonlinear  
797 pedagogy on decision-making and execution in youth soccer players according to game  
798 actions. *Journal of Human Kinetics*, 62, 185-198. doi:10.1515/hukin-2017-0169
- 799 Robinson, G., & O'Donoghue, P. (2007). A weighted kappa statistic for reliability testing in  
800 performance analysis of sport. *International Journal of Performance Analysis in Sport*, 7,  
801 12-19. doi:10.1080/24748668.2007.11868383
- 802 Rovegno, I., & Kirk, D. (1995). Articulations and silences in socially critical work on physical  
803 education: Toward a broader agenda. *Quest*, 47, 447-474.  
804 doi:10.1080/00336297.1995.10484169
- 805 Sal de Rellán-Guerra, A., Rey, E., Kalén, A., & Lago-Peñas, C. (2019). Age-related physical  
806 and technical match performance changes in elite soccer players. *Scandinavian Journal*  
807 *of Medicine & Science in Sports*, 29, 1421-1427. doi:10.1111/sms.13463
- 808 Wiggins, G. (2011). A true test: Toward more authentic and equitable assessment. *Phi Delta*  
809 *Kappan*, 92(7), 81-93. doi:10.1177/003172171109200721
- 810

811 Table 1

812 *Preliminary List of Criteria and Categories*


---

|  |
|--|
| Contextual level   |
| Duration   |
| Team   |
| Game phase   |
| Attack phase, Defense phase  |
| Game principle   |
| Ball conservation, Progression, Finalizing, To avoid the goal, To avoid progression, Retrieving ball                   |
| Score board  |
| Winning, Losing, Drawing   |
| Period   |
| First half, Second half, Extra-time  |
| Situation type   |
| Standard game situation, Corner, Penalty, Foul, Ball divided   |
| Finalizing type in attack  |
| Goal, Out of bounds, Clear or catch, Attack mistake, Defensive mistake, Referee decision                               |
| Recovery type  |
| Rival goal, Touch line out, Out, Steal   |
| Organizational match level   |
| Change of role   |
| Offensive structured change, Offensive unstructured change, Defensive structured change, Unstructured defensive change |
| Attack type  |
| Positional attack, Counter-attack  |
| Depth of attack  |
| Depth (offensive progression)  |
| Amplitude  |
| Tempo control  |
| Retract  |
| Defense type   |
| Defense in zone, Individual defense  |
| Partial forefront organizational level   |
| Defensive coverage   |
| Fixing the player  |
| Wall pass  |
| Pass   |
| Center   |
| Primary organizational level   |
| Progressing the ball unopposed   |
| Shoot  |
| Control  |
| Clearance  |
| Marking  |
| Tackle   |

---

813

814

815  
 816 Table 2  
 817 *Average of Scores given by Experts and Percentage of Experts that Scored Each Criterion with*  
 818 *Three or Less to Establish Content and Comprehension Validity*

| Criteria                               | Average score (M) | Percentage of experts that scored the criterion with less than three (%) |
|--|-------------------|--|
| Contextual level                       |                   |  |
| Duration                               | 4.33              | 10   |
| Team                                   | 4.96              | 0  |
| Game phase                             | 4.86              | 0  |
| Game principle                         | 4.25              | 26.66  |
| Score board                            | 4.49              | 10   |
| Period                                 | 4.39              | 10   |
| Situation type                         | 4.31              | 16.66  |
| Finalizing type<br>in attack           | 4.37              | 13.33  |
| Recovery type                          | 4.35              | 16.66  |
| Change of role                         | 3.84              | 36.66  |
| Organizational match level             |                   |  |
| Attack type                            | 4.47              | 10   |
| Depth of attack                        | 4.61              | 36.66  |
| Depth (offensive<br>progression)       | 4.16              | 16.66  |
| Amplitude                              | 4.59              | 16.66  |
| Retract                                | 3.75              | 36.66  |
| Tempo control                          | 4.63              | 16.66  |
| Defense type                           | 4.22              | 23.33  |
| Partial forefront organizational level |                   |  |
| Defensive<br>coverage                  | 4.31              | 26.66  |
| Fixing the player                      | 3.84              | 43.33  |
| Wall pass                              | 3.94              | 46.66  |
| Support                                | 4.63              | 16.66  |
| Pass                                   | 4.10              | 23.33  |
| Center                                 | 4.75              | 53.33  |
| Primary organizational level           |                   |  |
| Progressing the<br>ball unopposed      | 4.35              | 16.66  |
| Shoot                                  | 4.47              | 6.66   |
| Control                                | 4.47              | 36.66  |
| Clearance                              | 4.61              | 10   |
| Marking                                | 4.35              | 46.66  |
| Tackle                                 | 4.68              | 20   |
| Levels'<br>classification              | 5                 | 0  |
| Usefulness of<br>the instrument        | 4.93              | 0  |

819

820

821

822 Table 3

823 *First version of TAIS*

| Level—Criteria—Categories  | Definitions  |
|--|--|
| Contextual level   |  |
| Duration   | The duration of each phase, offensive and defensive  |
| Team   | Team that it is going to be analyzed   |
| Game phase   | Phase from the team that is being assessed, could be attacking phase or defending phase  |
| Attack phase, Defense phase  | Purpose of the attacking or defending phase attending to the game principles. In each phase predominates one game principle  |
| Game principle   |  |
| Ball conservation, Progression, Finalizing, To avoid the goal, To avoid progression, Retrieving ball   | Match score in the phase is being assessed   |
| Score board  | Period of the match, first half, second half or extra-time   |
| Winning, Losing, Drawing   | Context of the phase, can be flow plays situation or set plays situations  |
| Period   |  |
| First half, Second half, Extra-time  | The way in which the attacking team stop having the possession of the ball, finishing the attacking phase  |
| Situation type   | The way in which the defending team recovering the possession of the ball, finishing the defending phase   |
| Standard game situation, Corner, Penalty, Foul, Ball divided from the point of view of the attackers, Ball divided from the point of view of the defenders |  |
| Finalizing type in attack  | Spatial arrangement in the attacking phase   |
| Goal, Out of bounds, Own goal, Previous action to goalkick, Losing ball, Save from goalkeeper, Referee decision  | Action made by the attackers when moving vertically in the field   |
| Recovery type  | Moving width and length, the team place into to the lateral zones  |
| Rival goal, Touch line out, Out, Steal, Interception or goalkeeper block, Attackers mistake, End without recovering  | Attacking direction depending on the verticality. It will be considered direct attack if is more vertical and indirect attack if is less vertical  |
| Organizational match level   |  |
| Attack type  | Action made by the defenders that, independently of their direct opponent, are placed near a teammate in order to be able to do a support action to avoid the opponents maintain the possession of the ball or progressing to their goal. When a defender is placed in the center of an imaginary triangle formed by on-ball attacker, his/her direct opponent and the goal or between on-ball attacker and the goal |
| Positional attack, Counter-attack  | Action made by offball attackers that through a movement, make it easier for on-ball attacker to go forward safely and proceed with the planned strategy   |
| Depth (offensive progression)  | Every time the action of giving the ball to another teammate takes place. A throw-in or a corner will all be considered as a pass  |
| Amplitude  |  |
| Defense type   | Action made by the on-ball attacker when is progressing keeping the control of the ball  |
| Defense in zone, Individual defense  | Action made every time an attacker shot on goal. Does not matter the type or from where it is taken  |
| Partial Forefront organizational level   | Action in which a defender tries to deflect the path of the ball to recover the possession or avoid the progression from the attackers   |
| Defensive coverage   | Action made by the defender in order to steal the ball from the player in possession. When the defender try to stole the ball with his/her leg or approach to the attacker shaking shoulder to shoulder  |
| Support  |  |
| Pass   |  |
| Primary organizational level   |  |
| Progressing the ball unopposed   |  |
| Shoot  |  |
| Interception   |  |
| Tackle   |  |

824

825

826

827 Table 4

828 *Differences between Skill Levels in Each Criterion Assessed with TAIS*

| Criteria                               | Low skill level |      | High skill level |      | Z Wilcoxon | p    | ES  |
|--|-----------------|------|------------------|------|------------|------|-----|
|  | M               | SD   | M                | SD   |            |      |     |
| Organizational match level             |                 |      |                  |      |            |      |     |
| Attack type                            | .01             | .10  | .16              | .36  | -5.243     | .000 | .75 |
| Depth (offensive progression)          | .03             | .17  | .13              | .33  | -3.616     | .000 | .46 |
| Amplitude                              | .05             | .21  | .13              | .33  | -2.857     | .004 | .34 |
| Defense type                           | .17             | .37  | .22              | .41  | -2.184     | .000 | .14 |
| Partial forefront organizational level |                 |      |                  |      |            |      |     |
| Pass                                   | .81             | 1.23 | 1.03             | 1.66 | -4.981     | .000 | .17 |
| Support                                | .31             | .72  | .72              | 1.34 | -3.776     | .000 | .46 |
| Defensive coverage                     | .17             | .43  | .43              | .74  | -2.937     | .003 | .51 |
| Primary organizational level           |                 |      |                  |      |            |      |     |
| Dribbling                              | .22             | .62  | .37              | .72  | -4.414     | .000 | .25 |
| Shoot                                  | .06             | .24  | .19              | .41  | -1.752     | .080 | .46 |
| Interception                           | .27             | .11  | .35              | .37  | -1.785     | .050 | .38 |
| Tackle                                 | .17             | .34  | .31              | .68  | -3.587     | .000 | .32 |

829

830

831

832 Table 5

833 *Correlations between Criteria from TAIS in both Appropriate and Inappropriate Categories and*834 *Criteria Assessed with GPAI Decision-Making, Skill Execution, Cover and Support Components*

| TAIS                             |     |      | GPAI                                   |     |      | Rho | p    |
|----------------------------------|-----|------|--|-----|------|-----|------|
| Criteria                         | M   | SD   | Criteria                               | M   | SD   |     |      |
|                                  |     |      | Partial forefront organizational level |     |      |     |      |
| Appropriate pass                 | .64 | 1.33 | ADM for pass                           | .65 | 1.33 | .93 | .000 |
|                                  |     |      | CSE for pass                           | .62 | 1.25 | .89 | .000 |
| Inappropriate pass               | .34 | .79  | IDM for pass                           | .34 | .68  | .70 | .000 |
|                                  |     |      | ISE for pass                           | .36 | .71  | .64 | .000 |
| Appropriate defensive coverage   | .22 | .52  | Cover appropriate                      | .22 | .52  | .77 | .000 |
| Inappropriate defensive coverage | .08 | .34  | Cover inappropriate                    | .08 | .347 | .67 | .000 |
| Support                          | .53 | 1.10 | Support                                | .53 | 1.10 | .81 | .000 |
|                                  |     |      | Primary forefront organizational level |     |      |     |      |
| Appropriate dribbling            | .22 | .54  | ADM for dribbling                      | .22 | .55  | .85 | .000 |
|                                  |     |      | CSE for dribbling                      | .20 | .51  | .77 | .000 |
| Inappropriate dribbling          | .10 | .32  | IDM for dribbling                      | .11 | .33  | .73 | .000 |
|                                  |     |      | ISE for dribbling                      | .13 | .37  | .60 | .000 |
| Appropriate tackle               | .14 | .35  | ADM for tackle                         | .14 | .36  | .78 | .000 |
|                                  |     |      | CSE for tackle                         | .15 | .37  | .72 | .000 |
| Inappropriate tackle             | .12 | .42  | IDM for tackle                         | .12 | .43  | .89 | .000 |
|                                  |     |      | ISE for tackle                         | .11 | .39  | .82 | .000 |
| Appropriate interception         | .29 | .50  | ADM for interception                   | .29 | .50  | .81 | .000 |
|                                  |     |      | CSE for interception                   | .26 | .48  | .73 | .000 |
| Inappropriate interception       | .14 | .35  | IDM for interception                   | .09 | .35  | .69 | .000 |
|                                  |     |      | ISE for interception                   | .12 | .38  | .60 | .000 |
| Appropriate shoot                | .11 | .33  | ADM for shoot                          | .11 | .33  | .79 | .000 |
|                                  |     |      | CSE for shoot                          | .10 | .32  | .68 | .000 |
| Inappropriate shoot              | .04 | .21  | IDM for shoot                          | .05 | .22  | .86 | .000 |
|                                  |     |      | ISE for shoot                          | .05 | .23  | .65 | .000 |

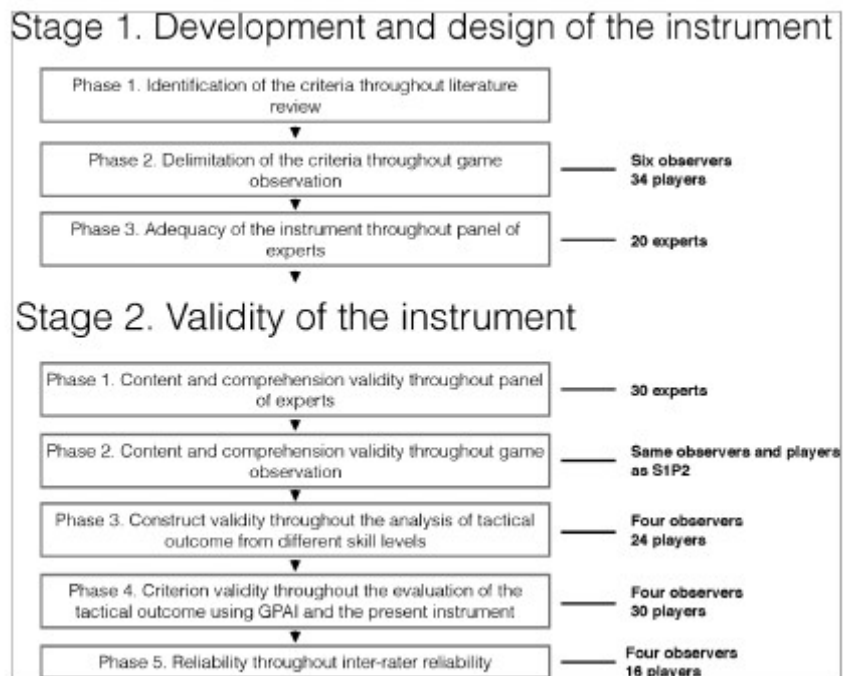
835 ADM: Appropriate decision-making, CSE: Correct skill execution, IDM: Inappropriate decision-making, ISE: incorrect skill execution.

836

837

838 *Figure 1. Stages and phases followed in the development and validation of the instrument. S1P2:*

839 Stage 1 Phase 2.

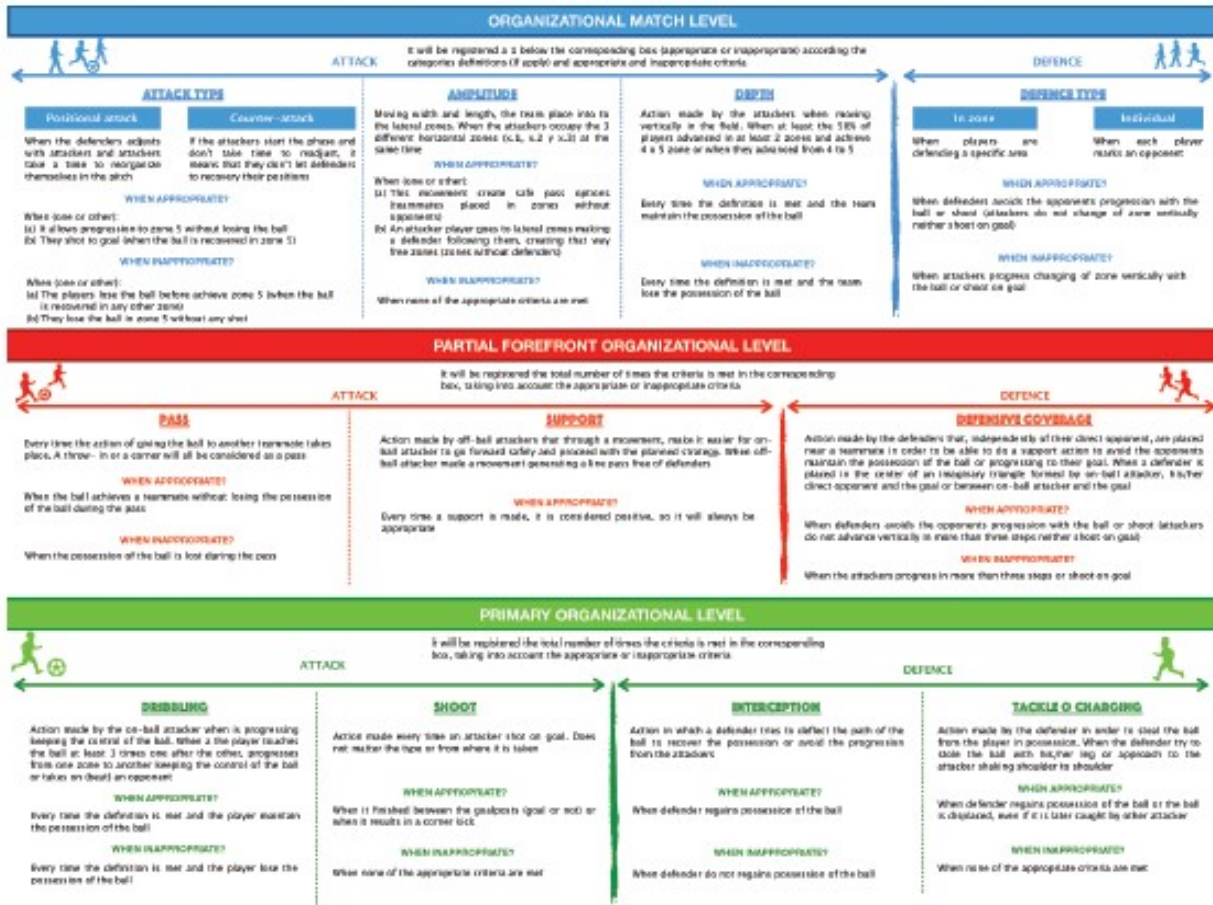


840

841

842 Figure 2. Final version of TAIS.

843



844

845



**CONTEXTUAL LEVEL**

---

**DURATION**  
The duration of each phase, offensive and defensive

---

**TEAM**  
Team that it is going to be analysed

---

**GAME PHASE**  
Phase from the team that is being assessed, could be attacking phase or defending phase

Attack phase      Defence phase

---

**GAME PRINCIPLE**  
Purpose of the attacking or defending phase according to the game principles (Baker, 1992), in each phase predominates one game principle

Ball conservation    Progression    Ending    To avoid the goal    To avoid progression    Recovering the ball

---

**SCORE BOARD**  
Match score in the phase is being assessed

Winning    Loosing    Drawing

---

**GOAL DIFFERENCE IN FAVOUR**  
Goal difference when the team that is being assessed is winning

---

**GOAL DIFFERENCE AGAINST**  
Goal difference when the team that is being assessed is loosing

---

**PERIOD**  
Period of the match, first half, second half or extra-time

First half    Second half    Extra-time

---

**SITUATION TYPE**  
Context of the phase, can be flow play situation or set play situations

Standard game situation    Corner    Penalty    Foul    Ball diverted from the point of view of the attackers    Ball diverted from the point of view of the attackers

---

**FINALISING TYPE IN ATTACK**  
The way in which the attacking team stop having the possession of the ball, finishing the attacking phase

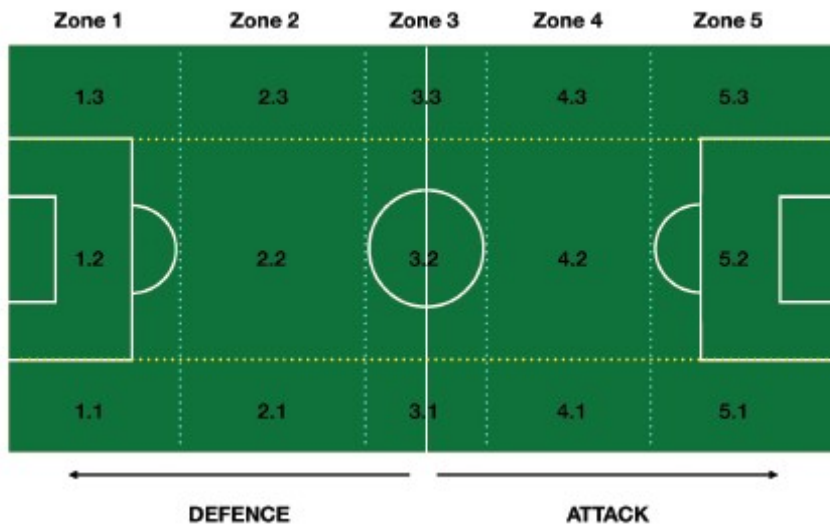
Goal    Out of bounds    Corner    Previous action to goal kick    Loosing ball    Defence foul or Own goal    Save from goalkeeper    Referee decision

---

**FINALISING TYPE IN DEFENCE**  
The way in which the defending team recover the possession of the ball, finishing the defending phase

Goal goal    Touch line out    Out    Steal    Interception or Goalkeeper block    Attackers mistake    End without recovering

846



847

848

849