

This is a peer-reviewed, accepted author manuscript of the following research article: Barquero, C., Arias, J., & Kirk, D. (Accepted/In press). Design and validation of the Tactical Assessment Instrument in football (TAIS). *Research Quarterly for Exercise and Sport*, 1-34

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

2

3

4

5

Abstract

6
7 The focus on tactics has made assessment more difficult due to the lack of knowledge in the
8 topic and the limitations of current tactical assessment instruments. **Purpose:** To design and
9 validate a tactical assessment instrument in youth football (TAIS) following an exhaustive,
10 ecological and meticulous process and dealing with the limitations found in the literature.
11 **Method:** The design was divided in two stages related to its development and validation.
12 During the development stage: (a) a preliminary list of criteria was determined through a
13 literature review; (b) the criteria were delimited through an exploratory observation; and (c) the
14 adequacy of the criteria was determined by consulting experts. In the validation stage: (a)
15 content and comprehension validity were obtained by consulting experts and through a
16 systematic game observation; (b) construct validity was obtained by comparing the tactical
17 outcomes from two groups with different skill levels; (c) criterion validity was established by
18 comparing the tactical outcome using Game Performance Assessment Instrument and the
19 present instrument, and (d) the reliability was obtained through inter-rater reliability. **Results:**
20 The research process showed that the instrument is a valid and reliable tool comprised of 22
21 criteria to assess tactical outcomes in 8-12 years old youth football. **Conclusion:** The TAIS
22 presents several advantages in practical terms with respect to assessment. First, it allows
23 assessment of the three tactical levels nested in the unit of observation. Second, it considers all
24 the player roles. Third, results are presented without general indexes. Fourth, it can be used to
25 assess participants from all the institutional contexts. Finally, it includes contextual variables.

26 *Keywords:* Tactical learning, sport pedagogy, youth sport, assessment, authentic
27 assessment.

28

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

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.

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

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

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

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

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

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

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

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

155 **Design**

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

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

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

174 **** Figure 1****

175 **Gaining entry and access**

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

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

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

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

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

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

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

227 *****Table 1*****

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

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

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

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

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

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

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

276 **Results.** When compared to the literature review, the 28 criteria identified in SIP1 were
277 found in the observation with a frequency of more than 20%. On the contrary, this percentage
278 was not found in the case of the categories 'clear or catch', 'attack mistake' and 'defensive
279 mistake' from the criterion 'finalizing type in attack'. Consequently, these categories were
280 removed from this criterion and changed by the categories 'own goal', 'previous action to goal
281 kick', 'losing ball' and 'save from goalkeeper' as they appeared with a frequency of more than
282 20%. Furthermore, the criterion 'support' was found in the observation with a frequency of
283 more than 20% and consequently added to TAIS. Regarding the criterion 'recovery type', the
284 categories 'interception or goalkeeper block', 'attackers mistake' and 'end without recovering'
285 were added as they appeared with a frequency of more than 20%. At the end of this phase, TAIS
286 was composed by 29 criteria, ten of them included categories.

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

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

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

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

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

318 ****Table 2****

319 ****Table 3****

320 **Stage 2. Validity of the instrument**

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

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

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

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

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

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

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

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

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

366 each meeting. The game phases lasted between five and 15 seconds. They observed a total of
367 320 game phases. The six observers observed and coded all game footage individually. All
368 meetings were two hours with each hour having a specific focus. In the first hour, observers
369 focused on whether the criterion was met or not met. For instance, for the criterion 'shoot', they
370 identified when a shot on goal was made. Then they registered whether the criteria were
371 appropriate according to its definition when one of the following circumstances occurred: (a) the
372 shot resulted in a goal, (b) the shot was directed to goal but there was no score (either it missed
373 the goal or was saved or cleared) or (c) when the shot missed the goal and resulted in a corner
374 kick. Criteria were coded in the three tactical levels and the contextual level by observers. They
375 were nested as the same unit of observation for each game phase. Observers used Virtual Dub
376 Version 1.10.4 to play the games and an excel sheet to record the information. They were able to
377 stop the videotapes as many times as necessary and when they had any doubts, the procedure
378 was to write it down and discuss at the end of the meeting. In the second hour of the meeting, all
379 the observers and authors discussed together the observers' doubts about the criteria definition,
380 until an agreement was reached. As a result, the authors modified the definition of criteria and
381 categories until they achieved a version that allowed the observation of tactical outcomes with
382 operative criteria.

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

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

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

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

405 **Procedure.** A three-day tournament was conducted and consisted of four 20 minutes
406 4GK vs 4GK games, based on recommendations from the literature (Machado et al., 2019).
407 Players were divided into two groups according to their skill level (low or high). They were
408 classified after a GPAI observation of previous games by expert coaches. Then, the low skill
409 level group and high skill level groups were randomly subdivided in two subgroups of five
410 players each. In all groups there were players who came from the three different contexts. The
411 games were played with official eight to 12 years old laws of the game. The games were
412 conducted twice between teams with similar skill levels. So that, there were two games between
413 low skill level groups and two games between high skill level groups. The games were video-
414 recorded.

415 **Observers were asked to observe the tactical outcomes of all players using the instrument**
416 **and the recordings of all matches from this three-day tournament (25 frames per second).**

417 Observers coded all criteria defined in TAIS for each game phase. The level of the participants

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

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

433 *****Table 4*****

434 **Phase 4. Criterion validity through the evaluation of the tactical outcome using GPAI and** 435 **the present instrument**

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

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

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

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

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

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

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

485 *****Table 5*****

486 **Phase 5. Inter-rater reliability**

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

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

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

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

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

511 **Results.** Reliability exceeded .70 according to ICC or Kappa coefficient value (Cohen,
512 1960), and the 85% according to PA (Brewer & Jones, 2002). Lowest reliability was found for
513 criterion 'goal difference in favour' (ICC=.70, PA=95%) while highest reliability was found for
514 criteria 'team', 'score board' and 'period' (Kappa/ICC=1; PA=100%).

515 **Discussion**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

655 **Conclusion and practical application**

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

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

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

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

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

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

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

702 **What does this Study Add?**

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

718 **References**

719 Authors a

720 Authors b

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

722 Washington DC: APA, AERA, NCME.

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

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

725 doi:10.1177/0013164485451012

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

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

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

729 point shapes in mini-basketball. *Revista Internacional de Medicina y Ciencias de la*730 *Actividad Física y del Deporte*, 9, 349-365.

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

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

733 doi:10.2165/00007256-199826040-00002

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

735 347–364. doi:10.1007/BF00138871

736 Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L.

737 (2018). Best practices for developing and validating scales for health, social, and

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

739 doi:10.3389/fpubh.2018.00149

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

741 Fundamental tactical principles of soccer: A comparison of different age groups. *Journal*742 *of Human Kinetics*, 58, 207-214. doi:10.1515/hukin-2017-0078

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

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

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

809 Table 1

810 *Preliminary List of Criteria and Categories*

811

812 Table 2

813 *Average of Scores given by Experts and Percentage of Experts that Scored Each Criterion with*

814 *Three or Less to Establish Content and Comprehension Validity*

815

816 Table 3

817 *First version of TAIS*

818

819 Table 4

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

821

822 Table 5

823 *Correlations between Criteria from TAIS in both Appropriate and Inappropriate Categories and*

824 *Criteria Assessed with GPAI Decision-Making, Skill Execution, Cover and Support Components*

825

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

827 *Stage 1 Phase 2.*

828

829 *Figure 2. Final version of TAIS.*

830

831