

“This is an Accepted Manuscript of an article published by Taylor & Francis Group in Journal of Sports Sciences on 10/08/2020, available online:
<http://www.tandfonline.com/10.1080/02640414.2020.1800371>”

1 **Development and validation of the Referee Training Activity Questionnaire**
2 **(RTAQ): Towards a better understanding of the training practices of**
3 **soccer officials**

4 G. McEwan^a, V. Unnithan^a, C. Easton^a and R. Arthur^a

5 ^aSchool of Health and Life Sciences, University of the West of Scotland, Glasgow, UK

6 Corresponding author:

7 Mr Gary McEwan

8 School of Health and Life Sciences,

9 University of the West of Scotland,

10 Glasgow, UK,

11 G72 0LH

12 Email: Gary.McEwan@uws.ac.uk

13 Co-authors' contact details:

14 Vish.Unnithan@uws.ac.uk

15 Chris.Easton@uws.ac.uk

16 Rosie.Arthur@uws.ac.uk

17 **Abstract Word Count:** 197

18 **Text-only Word Count:** 4885

19 **Number of Tables:** 5

20 **Number of Figures:** 2

21 **Development and validation of the Referee Training Activity Questionnaire**
22 **(RTAQ): Towards a better understanding of the training practices of**
23 **soccer officials**

24 **Abstract**

25 To develop and assess the validity of the *Referee Training Activity Questionnaire (RTAQ)*, a
26 systematic process was employed: 1) item generation; 2) assessments of content and face
27 validity; and 3) assessments of criterion validity. In stage 1, items were generated following
28 semi-structured interviews with an expert panel (n=8). Following content analyses, the
29 *RTAQ* was developed and comprised 3 primary sections (12 sub-sections) assessing: 1)
30 attributes perceived to underpin soccer officiating performance; 2) general training
31 information; and 3) specific training practices. In stage 2, the preliminary *RTAQ* was
32 assessed for content and face validity by a sample of experts (n=6). Based upon the content
33 validity index (CVI), content validity was confirmed for 8 sub-sections ($CVI \geq 0.78$) with 5
34 sub-sections being deemed invalid ($CVI < 0.78$). Various amendments were carried out in
35 accordance with participant feedback. In stage 3, the *RTAQ* was completed by a cohort of
36 officials (n=25) who subsequently recorded a detailed training diary. Negligible mean biases,
37 wide 95% LOA, and significant Pearson correlations were observed between the *RTAQ* and
38 training diaries for most training activities, suggesting the *RTAQ* holds promise as a useful
39 and effective alternative of acquiring insight into the training practices of soccer officials.

40 **Keywords:** *Performance optimisation; developmental activities; football; match officials*

41

42

43 **Introduction**

44 Soccer officiating represents a highly complex and demanding activity characterised by the
45 adjudication of match play within physiologically and psychologically challenging conditions
46 (Helsen & Bultynck, 2004). For instance, in addition to making ~137 observable decisions
47 per match (Helsen & Bultynck, 2004), field referees (FR) typically cover total distances of 9-
48 11 km (Krustrup et al., 2009; Weston, Drust & Gregson, 2011; Castillo et al., 2017), with a
49 corresponding mean match heart rate (HR) of 85-95% of maximal HR (Weston et al., 2006;
50 Castillo et al., 2016). Conversely, although their movement is limited to half of the length of
51 the field, distances of 5-7 km are covered by assistant referees (AR), with mean match HRs
52 of ~79% of maximal HR achieved (Castillo et al., 2016). Furthermore, as matches are often
53 contested in the presence of large partisan crowds, the soccer officials' decisions receive
54 routine scrutiny from players, managers and spectators alike (Webb et al., 2019). Well-
55 developed physical and psychological attributes therefore serve as important requisites of the
56 successful match official (Castagna, Abt & D'Ottavio, 2007; Slack et al., 2013), whilst
57 perceptual-cognitive skills are integral to successful decision-making within dynamic, time-
58 constrained domains (Spitz et al., 2018). To succeed at the top level, soccer officials must
59 also possess sufficient technical knowledge (i.e. knowledge of the Laws of the Game, tactical
60 awareness), whilst game-management skills are necessary for "*managing*" the game
61 effectively (Unkelbach & Memmert, 2008). To acquire expertise within each of these
62 various facets of performance, commitment to training and practice is important (Baker &
63 Young, 2014). Central to the magnitude of the adaptive responses achieved from training are
64 the volume, intensity, frequency and type of practice undertaken (Hawley, 2008).
65 Additionally, as the greatest improvements in performance are achieved when training
66 simulates the demands of competition, an integrated approach to training is advised, whereby
67 the physical, perceptual, and psychological elements of soccer officiating are addressed in

68 unison (Weston et al., 2012). Given the association between training and performance,
69 understanding the habitual training practices of soccer officials is important if researchers and
70 practitioners are to successfully evaluate and adjust training, with the intention of optimising
71 performance (Saw, Main, & Gustin, 2015).

72 In comparison to the volume of literature pertaining to players, very little information
73 is readily discernible on the training practices of soccer officials. Generally, the training of
74 soccer officials is self-led and focuses on physical training activities, with less attention
75 typically paid to the development of other skills pertinent to refereeing (Weston et al., 2012).
76 As such, efficacious and accessible methods of measuring a wider range of training are
77 required in order to monitor and then diversify and enhance the training of soccer officials.
78 Self-report measures such as training diaries have long been utilised within applied sport
79 science research as simple and inexpensive methods of acquiring insight into an athlete's
80 training (Pugliese et al., 2014). However, from a data management and interpretation
81 perspective, their use with large cohorts typically proves problematic given the considerable
82 volume of data generated (Mujika, 2017). Self-administered questionnaires represent an
83 alternative means of collecting such information (Helsen, Starkes, & Hodges, 1998; Ford et
84 al., 2010). Of note is the deliberate practice questionnaire which having been developed for
85 soccer players (Helsen et al., 1998), has previously been implemented with soccer officials
86 (MacMahon et al., 2007). Nonetheless, the appropriateness of the deliberate practice
87 questionnaire for use with soccer officials is questionable. Whilst the items of this
88 questionnaire were adapted to the role of the soccer official following consultation with two
89 referee coaches (MacMahon et al., 2007), a more robust process of item generation appears
90 warranted to ensure their specificity to both FRs and ARs (Saw et al., 2017). Specifically, the
91 inventory of activities listed within the deliberate practice questionnaire were categorised as
92 on-field and off-field, with a large emphasis placed upon physical training activities.

93 Conversely, little detail was provided surrounding activities unique to officials, such as
94 specific decision-making, psychological, technical and game-management skills training.
95 Thus, a clear need exists for the development of a valid measurement tool with the potential
96 to assess these crucial yet often overlooked aspects of a soccer official's training and
97 preparation.

98 In order to develop a valid instrument that sufficiently addresses each aspect of a
99 soccer official's training, a participatory approach is advocated whereby the end-users (i.e.
100 both FR and AR) are actively involved in the development and assessment of its
101 psychometric properties (Shrier et al., 2014). Of the properties suggested to underpin the
102 validity of an instrument are the content validity (the extent to which the questionnaire is
103 deemed to cover all relevant topics of the construct in question) and face validity (the extent
104 to which items appear to provide an adequate reflection of the construct in question)
105 (Mokkink et al., 2010). It is also important that novel instruments be assessed for sufficient
106 levels of criterion validity i.e. the degree to which the results of a questionnaire provide an
107 adequate reflection of a "*gold standard*" (Mokkink et al., 2010). Notably, as questionnaires
108 previously employed to assess the training practices of soccer officials have not been
109 systematically developed and validated for use with soccer FR and AR (MacMahon et al.,
110 2007), scope exists for a new tool to be developed and validated for specific use with such
111 populations.

112 Accordingly, the current investigation aimed to develop and validate a novel
113 measurement tool (the Referee Training Activity Questionnaire) for assessing the training
114 practices of soccer referees.

115 **Methods**

116 *Overview*

117 The Referee Training Activity Questionnaire (*RTAQ*) was developed and validated following
118 a systematic multi-stage process (Artino et al., 2014): 1) item generation and questionnaire
119 development; 2) assessments of content and face validity; and 3) assessments of criterion
120 validity (Figure 1). During each stage, participants were provided with a detailed information
121 sheet outlining the purpose of the study and informed consent was obtained. Institutional
122 ethical approval was obtained for all phases of the study.

123 *****FIGURE 1 NEAR HERE*****

124 ***Stage 1: Item generation and questionnaire construction***

125 In accordance with previous research (MacNamara & Collins, 2011), content domains were
126 defined and an item pool was generated following: 1) a review of the literature; and 2)
127 consultation with experts informed on the attributes and training practices pertinent to soccer
128 officials. To conceptualise the scope of the questionnaire and identify the primary facets of
129 officiating performance, a literature search of PubMed and Google Scholar databases was
130 performed using various combinations of issue-related search terms. Search terms included:
131 (“*referee*” OR “*assistant referee*” OR “*match official*”) AND (“*soccer*” OR “*association*
132 *football*”). Secondary searches included manual searches of the reference lists of the
133 recovered articles (Moher et al., 2015). Following this review of the literature, a thematic
134 analysis of the recovered articles was undertaken whereby articles were arranged into
135 common themes. This process resulted in the identification of five themes deemed to
136 underpin soccer officiating performance, with these categorised as: physical, decision-
137 making, psychological, technical, and game-management. This information was
138 subsequently used to guide semi-structured interviews with experts and build an inventory of
139 items from which to construct the questionnaire.

140 ***Participants***

141 Eight experts, purposefully sampled to incorporate both elite soccer officials and referee
142 managers/coaches, were invited to participate in the preliminary stage of this investigation.
143 Although this cohort were exclusively male, it is important to note that this simply reflects
144 our limited access to elite participants, and that no criteria existed for the exclusion of
145 females. At the time of interviewing, all officials were active elite FR (n=2) and AR (n=2)
146 within Scotland, had officiated within the Scottish Premiership for 12 ± 3 years, and
147 possessed 8 ± 6 years of experience on the FIFA list. A sample of referee managers/coaches
148 (national manager: n=1; development officer: n=1; national referee fitness instructor: n=1;
149 and head of sports medicine: n=1) were also interviewed on the basis that they were
150 responsible for the promotion/demotion of officials between the different professional levels,
151 and for the provision of training support for officials.

152 *Interview procedure*

153 Interviews were semi-structured in nature and were directed by an interview guide developed
154 and revised by the research team following the aforementioned review of the literature.
155 Accordingly, a list of open-ended and non-leading questions were used to assess the training
156 practices and attributes perceived as pertinent to performance, with probes seeking
157 elaboration or clarification implemented to follow up on promising leads (Patton, 2001). All
158 interviews were conducted face-to-face by the lead researcher in a quiet and convenient
159 location and lasted 27 ± 3 min (range: 23–33 min). With the permission of participants,
160 interviews were audio recorded.

161 *Data analysis*

162 Qualitative data were analysed using concurrent deductive and inductive content analysis
163 (Sparkes & Smith, 2013) whereby the analysis was based upon two primary (attributes and
164 training practices) and five secondary (physical, decision-making, psychological, technical,

165 and game-management) *a priori* research themes, whilst remaining open to emergent findings
166 within participants' responses. Firstly, audio-recordings were transcribed verbatim and were
167 subsequently double-checked to ensure accuracy. Close reading of the text was then
168 undertaken by the lead researcher to ensure familiarity with the data. Raw data units were
169 then created from participants' words before being grouped into lower-order themes and then
170 higher-order themes. During analysis, internal homogeneity (that data within a category
171 share clear characteristics) and external heterogeneity (clear differences exist between
172 different categories) was sought (Patton, 2001). To ensure the trustworthiness of the
173 analyses, the authors discussed and confirmed the allocation of raw data units to specific
174 categories through constructive debate (Elo et al., 2014).

175 *Stage 2: Content and face validity*

176 Based upon the findings of stage 1, items were generated directly from participants' wording
177 to ensure that they remained in accordance with participants' perspectives and understanding
178 of the construct in question (Artino et al., 2014). A preliminary draft of the *RTAQ* was
179 subsequently constructed (Table 1). Stage 2 sought to assess the content and face validity of
180 the preliminary *RTAQ*.

181 *Participants*

182 To assess the content and face validity of the *RTAQ*, a purposive sample of 6 participants
183 were recruited (Polit, Beck, & Owen, 2007). This cohort included male soccer officials (FR:
184 n=2; AR: n=1) of varying levels of professional attainment and experience, academics
185 experienced in questionnaire design within sport science (n=2), and an applied sport science
186 practitioner (n=1).

187 *Face and content validation*

188 The preliminary *RTAQ* was distributed electronically and participants were requested to
189 review each section based upon its relevance, readability, redundancy, and ease of
190 interpretation. For each section of the questionnaire, participants used a 4-point Likert scale
191 to rate: the clarity and conciseness of its wording (*1=not clear; 4=very clear*); the relevance
192 of the item to understanding the training practices of officials (*1=not relevant; 4=very*
193 *relevant*); the relevance of the items to officiating (*1=not relevant; 4=very relevant*); the
194 extent to which the items matched the descriptions provided (*1=poor match; 4=good match*);
195 and the degree to which the scales were clear and easy to use (*1=not easy; 4=very easy*)
196 (Dunn, Bouffard, & Rogers, 1999). Participants were also invited to provide any additional
197 comments or suggestions that they may have had (Arnold, Fletcher, & Daniels, 2013).

198 *Data analysis*

199 Likert scale data were dichotomised into “*acceptable*” (rating of 3–4) and “*not acceptable*”
200 (rating of 1–2) and the content validity index (CVI) of each section was calculated (Polit &
201 Beck, 2006). Specifically, the CVI of each section was determined as the number of
202 participants providing a rating of 3–4 (agreed) divided by the total number of participants.
203 As recommended by Polit and Beck (2006), content validity was defined as a $CVI \geq 0.78$.

204 ***Stage 3: Criterion validity of the Referee Training Activity Questionnaire***

205 To assess the *RTAQ* for criterion validity, we compared estimates of training activity
206 involvement (type, frequency, and duration) to that documented during a detailed 7-day
207 training diary (Halson, 2014). During the data collection period, officials maintained their
208 usual training activities and habits. As part of their continued sport science support, officials
209 utilised heart rate monitors (Polar M400, Polar, UK) during their physical training sessions to
210 record measures of internal (heart rate) and external (session duration, total distance covered)
211 load. Specifically, data pertaining to the duration and type of activity was recorded, with

212 heart rate data being used to help identify and confirm the type of session. Consequently,
213 objective data obtained from the heart rate monitors during physical training sessions was
214 used to corroborate the self-report data of the training diaries (Borresen & Lambert, 2009).

215 *Participants*

216 Scottish male FR (n=19) and AR (n=12) spanning various levels of professional attainment
217 and experience agreed to participate in this stage of the investigation (Table 1).

218 *****TABLE 1 NEAR HERE*****

219 *Referee Training Activity Questionnaire (RTAQ)*

220 The *RTAQ* was developed using an online survey tool (Online Surveys, Jisc, UK) and was
221 distributed to participants via a web link towards the end of the 2017-18 competitive season.
222 Participants were permitted two months (April and May) to complete the questionnaire.
223 Although pertinent to consider seasonal variation in training practices (Sæther & Aspvik,
224 2014), the distribution of the questionnaire towards the latter stages of the season was not
225 deemed problematic in the current context given our interest in comparing training at a single
226 time-point.

227 *Training Diary*

228 Following the completion of the *RTAQ*, participants completed a training diary whereby they
229 were asked to record their training activities for a period of seven consecutive days (Halson,
230 2014). Accordingly, seven 24-hour diary sheets were provided to participants who were
231 asked to be as detailed as possible with regards to the type, duration, intensity, and conditions
232 of training. To minimise the risk of risk of report bias, the recommendations of Podsakoff,
233 MacKenzie, and Podsakoff (2012) were followed where participants received clear
234 instructions of how to complete the diary, including examples of the level of detail sought.

235 Participants were asked to complete the diary sheets at the end of each day, and to be
236 consistent with the procedure. Upon completion of the collection period, participants were
237 asked to comment on whether the information they provided was representative of a typical
238 training week. To avoid the confounding influence of seasonal variations in training
239 activities, the maximal time interval between the completion of the *RTAQ* and the training
240 diary was one month (Saether & Aspvik, 2014).

241 *Data reduction and analysis*

242 Thirty-one participants completed the *RTAQ*; however, 6 participants failed to complete the
243 training diary. Consequently, the data of 25 participants were included for analyses. To
244 corroborate the self-report data of the training diaries with the objective data obtained from
245 the heart rate monitors, associations between the total training volumes recorded for each
246 physical training activity were assessed using Pearson product-moment correlations. The
247 relationships between the *RTAQ* and the training diaries, with respect to the total training
248 volumes reported for each distinct training activity listed within the *RTAQ* (Figure 2), were
249 also examined via Pearson correlations, with uncertainty in the estimates expressed as 95%
250 confidence intervals (CI). Correlations of 0.10–0.29, 0.30–0.49, 0.50–0.69, 0.70–0.89, and \geq
251 0.90 were considered small, moderate, large, very large, and nearly perfect (Hopkins et al.,
252 2009). Paired sample t-tests determined whether any systematic bias existed between the
253 outputs of the *RTAQ* and the training diaries, with 95% CI also calculated. Additionally,
254 agreement between the *RTAQ* and training diaries was assessed using the 95% limits of
255 agreement (LOA) approach (Bland and Altman, 2007). Statistical significance was set at $P <$
256 0.05 with statistical procedures completed using Statistical Package for Social Sciences
257 (SPSS 22.0, IBM, USA).

258 **Results**

259 ***Stage 1: Item generation and questionnaire construction***

260 *Thematic analyses*

261 Following content analyses, 12 lower-order themes were identified and were subsequently
262 grouped into three higher order themes: 1) attributes perceived to underpin soccer officiating
263 performance; 2) general training information; and 3) specific training practices. The attributes
264 perceived to underpin soccer officiating performance were categorised as: a) physical; b)
265 decision-making; c) psychological; d) technical; and e) and game-management. General
266 training information comprised: f) domain-specific activities; g) training environment; and h)
267 training objective. Finally, the specific training practices of officials were categorised as: i)
268 physical training; j) decision-making training; k) psychological training; and l) technical
269 training. Please see Figure 2 for an overview of the item pool generated during phase 1.

270 *****FIGURE 2 NEAR HERE*****

271 *Questionnaire construction*

272 An overview of the preliminary structure of the *RTAQ* is presented in Table 2. Briefly, the
273 preliminary questionnaire comprised 2 primary sections and 12 sub-sections. To quantify the
274 frequency and duration apportioned to various general and specific training practices, open-
275 ended response scales were employed. To examine the regularity by which officials partook
276 in training aimed at developing particular skills and attributes, a 5-point Likert scale (*0 = less
277 than once a month; 4 = more than once a week*) was utilised. Response anchors were fully
278 labelled as per previous recommendations (Wade, 2006; Krosnick & Presser, 2010).

279 *****TABLE 2 NEAR HERE*****

280 ***Stage 2: Assessments of content and face validity***

281 *Content validity index (CVI)*

282 When assessed based upon the wording, relevance, and description of questions and items,
283 each section was deemed valid as exhibited by a CVI ≥ 0.78 . Of the 12 sub-sections within
284 the questionnaire, eight received a CVI of 1.00, demonstrating complete agreement amongst
285 experts.

286 When the scales within each sub-section of the questionnaire were assessed based upon their
287 usability, eight were deemed valid (CVI ≥ 0.78) with complete agreement being reached
288 amongst experts (CVI of 1.00) for six sub-sections. Of the five sub-sections pertaining to
289 specific training practices, each received a CVI of 0.50 suggesting them not to be valid.
290 Additional feedback received from participants suggested response anchors to be somewhat
291 ambiguous with suggestions for the implementation of clearer response labels. The CVI for
292 each section of the *RTAQ* is presented in Table 3.

293 *****TABLE 3 NEAR HERE*****

294 *Questionnaire refinement*

295 Following the checks of content and face validity, the questionnaire was revised and refined.
296 Items belonging to sections demonstrating a low CVI were either discarded or improved and
297 retained in accordance with the suggestions of participants. This process resulted in several
298 modifications being made to the initial questionnaire (Figure 1). Once amendments were
299 completed, the questionnaire was re-circulated for approval with no further changes requested
300 at this stage. The final *RTAQ* covered two main themes: 1) general training information (18
301 items) and 2) specific training practices (63 items).

302 **Stage 3: Assessments of criterion validity**

303 *Reliability of the training diary data*

304 Significant Pearson correlations ($r > 0.96$, $P < 0.001$) revealed the estimates of physical
305 training activity derived from the training diary and the heart rate monitors to be strongly
306 related, thus supporting the reliability of this data.

307 *Agreement between the RTAQ and training diary*

308 *Section 1 (general training information):* Negligible mean biases (range: -22 to 6 min), wide
309 95% LOA (range: -155 to 111 min), and significant Pearson correlations ($r > 0.53$, $P < 0.006$)
310 were found between the *RTAQ* and training diaries for most items relating to domain-specific
311 activities, training environment, and training objective. A significant systematic bias ($P =$
312 0.015) revealed the *RTAQ* to provide lower estimations of time spent in classroom-based
313 training by 25 min (95% LOA: -119 to 69 min). A significant systematic bias ($P = 0.001$)
314 also revealed the *RTAQ* to yield lower estimations of time apportioned to skill development
315 sessions by 67 min (95% LOA: -237 to 104 min). The agreement between the *RTAQ* and
316 training diaries for items pertaining to general training information are presented in Table 4.

317 *****TABLE 4 NEAR HERE*****

318 *Section 2 (specific training practices):* Negligible mean biases (range: -10 to 16 min), wide
319 95% LOA (range: -122 to 153 min), and significant Pearson correlations ($r > 0.55$, $P < 0.005$)
320 were found between the *RTAQ* and training diaries for most physical, decision-making,
321 psychological skills, and technical training activities. A significant systematic bias ($P =$
322 0.049) revealed the *RTAQ* to provide higher estimations of time apportioned to game-based
323 activities by 37 min (95% LOA: -138 to 212 min). A significant systematic bias ($P = 0.030$)
324 also revealed the *RTAQ* to yield higher estimations of time apportioned to imagery by 4 min
325 (95% LOA: -14 to 22 min), as well as time apportioned to relaxation ($P = 0.046$) by 5 min
326 (95% LOA: -18 to 28 min). The agreement between the *RTAQ* and training diaries for items
327 relating to specific training practices are presented in Table 5.

328

****TABLE 5 NEAR HERE****

329 ***Final questionnaire***

330 Following assessments of criterion validity, several modifications were carried out (Figure 1).
331 Completion time for the questionnaire was 23 ± 7 min. The final version of the *RTAQ* is
332 available as a supplementary file.

333 **Discussion**

334 The present study aimed to develop and validate a novel measurement tool to capture
335 information pertaining to the training practices of soccer officials. The resulting product is
336 the *RTAQ*, a self-report questionnaire that presents good levels of content, face and criterion
337 validity, and may feasibly be implemented to assess the training practices engaged in by
338 soccer FR and AR. This presents an attractive proposition to the applied practitioner as the
339 longitudinal monitoring of training may permit the successful evaluation, and adjustment, of
340 training (Saw et al., 2015).

341 Although susceptible to report bias, support for self-report measures is growing (Saw
342 et al., 2017). However, to ensure the reliability of the results, it is imperative that an
343 instrument first be assessed for measures of content, face and criterion validity (Mokkink et
344 al., 2010). In contrast to measures previously employed within the literature (MacMahon et
345 al., 2007), items for the *RTAQ* were generated and revised following a systematic process to
346 ensure their specificity and validity to soccer officials. Firstly, an item pool was generated
347 following consultation with a sample of elite soccer officials and their support staff
348 (MacNamara & Collins, 2011). An expert panel subsequently examined items both
349 quantitatively and qualitatively during stage two, with excellent levels of content and face
350 validity being evidenced (Polit & Beck, 2006). In stage three, estimates of training activity
351 involvement reported by the *RTAQ* were compared against those obtained from a detailed

352 training diary (Halson, 2014). For most training activities, the mean bias exhibited between
353 methods was negligible, thus highlighting a good level of agreement. Additionally, large to
354 nearly perfect correlations were evident between the *RTAQ* and training diaries for most
355 items, with such values proving comparable to previous research detailing the associations
356 between the self-reported and actual training duration of physically active participants
357 (Borresen & Lambert, 2006). However, relatively wide LOA were present for various
358 training activities such as high-intensity interval training (-67 to 59 min) and video clip
359 analyses (-122 to 153 min). Such findings are in line with those of Borresen and Lambert
360 (2006). A plausible explanation for the wide LOA presented for certain activities likely
361 relates to potential disparities in the time course in which the *RTAQ* and training diaries were
362 completed. To account for potential seasonal variations, the maximal time interval between
363 completion of the various methods was limited to one month. However, despite these
364 considerations, it is important to acknowledge the variation that may present itself amongst
365 weekly-microcycles (Malone et al., 2015). For instance, training during 1-, 2-, and 3-match
366 weeks is likely to be markedly different, particularly with regards to the volume of high-
367 intensity training performed (Anderson et al., 2016). Moreover, officials in the present study
368 were part-time and combined their officiating careers with additional occupations. It
369 therefore appears reasonable to suggest that extenuating factors will influence and drive an
370 official's training availability from week-to-week. Nevertheless, as participants noted their
371 responses to be representative of a typical training week, these suggestions remain
372 speculative. Given the implications that such variation may have in the ability to effectively
373 prescribe training, it may therefore be important to corroborate with objective data where
374 possible (Borresen & Lambert, 2009).

375 Whilst previous instruments have largely focused on physical training activities, the
376 participatory approach adopted in the current study allowed for the identification of other

377 activities unique to soccer officials, such as specific decision-making, psychological,
378 technical and game-management skills training. Generally, soccer officials are reported to
379 exhibit a propensity for physical conditioning with a weak culture of skill practice (Weston et
380 al., 2012). Moreover, the decision-making training of soccer officials has traditionally
381 focused on the development of declarative knowledge (i.e. Laws of the Game) through
382 lecture-style meetings, with practical experience being limited to the officiating of
383 competitive matches (MacMahon et al., 2007). While physical training may be monitored
384 objectively with relative ease through the use of HR monitors and Global Positioning System
385 (GPS) devices, the monitoring of self-led training activities performed off field proves more
386 challenging. Thus, practitioners and researchers working within the applied setting may wish
387 to implement the *RTAQ* to assess the engagement of soccer officials in the wider spectrum of
388 activities pertinent to performance. Such information may provide important insights into
389 potential shortcomings in current practices, from which action may be taken to promote a
390 greater diversification of training. Notwithstanding, it is also important to acknowledge that
391 whilst the *RTAQ* assesses the total time engaged in various different training activities, it is
392 primarily the intensity and purposefulness of training that govern the magnitude of the
393 adaptive responses achieved (Hawley, 2008). As such, although the *RTAQ* represents a novel
394 method of addressing important areas of training often overlooked, opportunity exists for its
395 implementation alongside current monitoring strategies in order to capture reliable estimates
396 of both training volume and intensity.

397 In the context of elite soccer, the monitoring of training has become common practice,
398 with an abundance of research now available detailing the training practices of the players
399 (Los Arcos, Mendez-Villanueva, & Martínez-Santos, 2017; Weston, 2018). Conversely, very
400 little information is readily discernible on the training of soccer officials, with such
401 discrepancies potentially related to disparities in the resources (financial and personnel)

402 available (Weston et al., 2012; Halson, 2014). Self-report measures such as training diaries
403 are commonly employed within both research and applied practice as simple and inexpensive
404 methods of acquiring insight into an athletes' training (Pugliese et al., 2014). Nonetheless,
405 training diaries often prove problematic when dealing with large cohorts due to the
406 considerable volume of data that can be generated (Mujika, 2017). Furthermore, owing to
407 their self-report nature, training diaries are reliant upon the level of detail that an official
408 provides. In the present study, disparities were revealed in the extent to which participants
409 reported their engagement in psychological skills training, with such information appearing
410 not to be captured by the training diaries. Although we are unable to confirm which method
411 provides the most accurate account, we may speculate that training diaries are likely less
412 effectual in capturing an official's engagement in more casual types of training of higher
413 frequencies and shorter durations. Conversely, whilst not discounting the potential for social
414 desirability response bias, the provision of pre-defined options may facilitate a greater recall
415 of casual training activities. Nevertheless, this remains speculative and warrants further
416 investigation. It is also worth noting the higher completion rates observed for the *RTAQ* in
417 comparison to the training diaries. Indeed, whilst all 31 participants completed the *RTAQ*
418 during the assessments of criterion validity, six officials failed to complete the training diary.
419 These findings provide additional evidence that training diaries may present issues of
420 compliance (Mujika, 2017), and further support the *RTAQ* as a quick and practical alternative
421 of capturing information pertaining to the training practices of soccer officials.

422 Another promising feature of the *RTAQ* concerns its specificity to both FR and AR.
423 Indeed, given their differing roles on the field of play, perspectives of the attributes and
424 training practices pertinent to performance will likely differ. Consequently, to ensure the
425 applicability of the *RTAQ* to each type of official, both FR (n=23) and AR (n=15) were
426 consulted at each stage of its development and validation. Encouragingly, the *RTAQ*

427 presented high levels of content, face and criterion validity amongst both FR and AR, thus
428 emphasising its specificity to these specialised roles. Although a relatively small number of
429 officials were included at each stage, it is important to note that our sample size was
430 restricted to the small numbers of elite participants available – a challenge commonly
431 encountered when conducting research within the applied field. Nonetheless, the sample
432 sizes present remain consistent with those of previous literature (Polit et al., 2007; Heikkilä et
433 al., 2017) and represent a significant progression from previous measures whereby items
434 were amended to the role of the soccer official following consultation with two referee
435 coaches (MacMahon et al., 2007). An additional consequence of our limited access to elite
436 officials was that our cohorts were exclusively male. Although the attributes and training
437 practices identified in the current study are likely applicable across sexes given the
438 comparable demands imposed during match play (Mallo et al., 2010), further research is
439 required to confirm the validity of the *RTAQ* amongst female officials. Finally, as our
440 participants were recruited from categories 1–3, our ability to extrapolate these findings to
441 lower level officials remains limited. Nevertheless, a wide spectrum of officiating levels
442 (regional to intercontinental) was represented in the current investigation. Future research
443 may therefore wish to utilise the *RTAQ* to examine potential differences in the training
444 practices engaged in by soccer officials of varying levels of experience and professional
445 attainment.

446 In conclusion, initial evidence suggests the *RTAQ* to be a valid measure of a soccer
447 official's training practices. The development of a self-report measure in the form of the
448 *RTAQ* therefore represents a valuable addition to the literature and may provide practitioners
449 and researchers with a practical and standardised alternative of obtaining insight into the
450 training of soccer officials. Such information may permit the successful evaluation, and
451 adjustment, of training practices.

452 **Acknowledgements**

453 The authors wish to extend their gratitude to each of the participants for their efforts during
454 the present study as well as Thomas Murphy for his valuable help and support. The authors
455 also wish to thank the Scottish Football Association who provided funding to support a PhD
456 studentship for Gary McEwan.

457 **Disclosure of interest**

458 The authors report no conflict of interest.

459

460

461

462

463

464

465

466

467

468

469

470

471

472 **REFERENCES**

- 473 Anderson, L., Orme, P., Di Michele, R., Close, G. L., Morgans, R., Drust, B., & Morton, J. P.
474 (2016). Quantification of training load during one-, two- and three-game week schedules in
475 professional soccer players from the English Premier League: implications for carbohydrate
476 periodization. *Journal of Sports Sciences*, 34(13), 1250-1259.
- 477 Arnold, R., Fletcher, D., & Daniels, K. (2013). Development and validation of the
478 organisational stressor indicator for sport performers. *Journal of Sport and Exercise*
479 *Psychology*, 35, 180-196.
- 480 Artino, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing
481 questionnaires for educational research: AMEE guide no. 87. *Medical Teacher*, 36(6), 463-
482 474.
- 483 Baker, J., & Young, B. W. (2014). 20 years later: Deliberate practice and the development of
484 expertise in sport. *International Review of Sport and Exercise Psychology*, 7(1), 135-157.
- 485 Bland, J. M., & Altman, D. G. (2007). Agreement between methods of measurement with
486 multiple observations per individual. *Journal of Biopharmaceutical Statistics*, 17(4), 571-
487 582.
- 488 Borresen, J., & Lambert, M. (2006). Validity of self-reported training duration. *International*
489 *Journal of Sports Science & Coaching*, 1(4), 353-359.
- 490 Borresen, J., & Lambert, M. (2009). The quantification of training load, the training response,
491 and the effect on performance. *Sports Medicine*, 39, 779-795.
- 492 Castagna, C., Abt, G., & D'Ottavio, S. (2007). Physiological aspects of soccer refereeing
493 performance and training. *Sports Medicine*, 37(7), 625-646.

494 Castillo, D., Weston, M., McLaren, S. J., Cámara, J., & Yanci, J. (2017). Relationships
495 between internal and external match-load indicators in soccer match officials. *International*
496 *Journal of Sports Physiology and Performance*, 12(7), 922-927.

497 Castillo, D., Yanci, J., Cámara, J., & Weston, M. (2016). The influence of soccer match play
498 on physiological and physical performance measures in soccer referees and assistant referees.
499 *Journal of Sports Sciences*, 34(6), 557-563.

500 Dunn, J. G. H., Bouffard, M., & Rogers, W. T. (1999). Assessing item content-relevance in
501 sport psychology scale-construction research: issues and recommendations. *Measurement in*
502 *Physical Education and Exercise Science*, 3(1): 15-36.

503 Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., Kyngäs, K. (2014). Qualitative
504 content analysis: A focus on trustworthiness. *SAGE Open*, 1-10.

505 Ford, P. R., Low, J., McRobert, A. P., & Williams, A.M. (2010). Developmental activities
506 that contribute to high or low performance by elite cricket batters when recognizing type of
507 delivery from bowlers' advanced postural cues. *Journal of Sport and Exercise Psychology*,
508 32(5), 638-654.

509 Halson, S. L. (2014). Monitoring training load to understand fatigue in athletes. *Sports*
510 *Medicine*, 44(Suppl2), S139-S147.

511 Hawley, J. A. (2008). Specificity of training adaptation: Time for a rethink? *Journal of*
512 *Physiology*, 586(1), 1-2.

513 Heikkilä, M., Valve, R., Lehtovirta, M., & Fogelholm, M. (2017). Development of a nutrition
514 knowledge questionnaire for young endurance athletes and their coaches. *Scandinavian*
515 *Journal of Medicine and Science in Sports*, 28(3), 873-880.

516 Helsen, W. F., & Bultynck, J. B. (2004). Physical and perceptual-cognitive demands of top-
517 class refereeing in association football. *Journal of Sports Sciences*, 22(2), 179-189.

518 Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of
519 deliberate practice. *Journal of Sport and Exercise Psychology*, 20(1), 12-34.

520 Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics
521 for studies in sports medicine and exercise science. *Medicine and Science in Sports and*
522 *Exercise*, 41(1), 3-13.

523 Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. *Handbook of*
524 *Survey Research*, 2(3), 263-314.

525 Krustup, P., Helsen, W., Randers, M. B., Christensen, J. F., MacDonald, C., Rebelo, A. N.,
526 & Bangsbo, J. (2009). Activity profile and physical demands of football referees and assistant
527 referees in international games. *Journal of Sports Sciences*, 27(11), 1167-1176.

528 Los Arcos, A., Mendez-Villanueva, A., & Martínez-Santos, R. (2017). In-season training
529 periodization of professional soccer players. *Biology of Sport*, 34(2), 149-155.

530 MacMahon, C., Helsen, W. F., Starkes, J. L., & Weston, M. (2007). Decision-making skills
531 and deliberate practice in elite association football referees. *Journal of Sports Sciences*, 25(1),
532 65-78.

533 MacNamara, A., & Collins, D. (2011). Development and initial validation of the
534 Psychological Characteristics of Developing Excellence Questionnaire. *Journal of Sports*
535 *Sciences*, 29(12), 1273-1286.

536 Mallo, J., Veiga, S., López de Subijana, C., & Navarro, E. (2010). Activity profile of top-
537 class female soccer refereeing in relation to the position of the ball. *Journal of Science and*
538 *Medicine in Sport*, 13(1), 129-132.

539 Malone, J. J., Di Michelle, R., Morgans, R., Burgess, D., Morton, J. P., & Drust, B. (2015).
540 Seasonal training-load quantification in elite English Premier League soccer players.
541 *International Journal of Sports Physiology and Performance*, 10, 489-497.

542 Moher, D., Shamsee, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P.,
543 Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review
544 and meta-analysis protocol (PRISMA-P) 2015 statement. *Systematic Reviews*, 4:1

545 Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter,
546 L.M., & de Vet, H.C.W. (2010). The COSMIN study reached international consensus on
547 taxonomy, terminology, and definitions of measurement properties for health-related patient-
548 reported outcomes. *Journal of Clinical Epidemiology*, 63, 737-745.

549 Mujika, I. (2017). Quantification of training and competition loads in endurance sports:
550 Methods and applications. *International Journal of Sports Physiology and Performance*,
551 12(Suppl2), S29-S217.

552 Patton, M. Q. (2001). *Qualitative research and evaluation methods* (3rd ed.). Newbury Park,
553 CA, Sage.

554 Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's
555 being reported? Critique and recommendations. *Research in Nursing and Health*, 29(5), 489-
556 497.

557 Polit, D. F., Beck, C. T., Owen, S. V. (2007). Is the CVI an acceptable indicator of content
558 validity? Appraisal and recommendations. *Research in Nursing and Health*, 30(4), 459-467.

559 Pugliese, L., Serpiello, F. R., Millet, G. P., & La Torre, A. (2014). Training diaries during
560 altitude training camp in two Olympic champions: An observational case study. *Journal of*
561 *Sports Science and Medicine*, 13, 666-672.

562 Sæther, S. A., & Aspvik, N. P. (2014). Seasonal variation in objectively assessed physical
563 activity among young Norwegian talented soccer players: A description of daily physical
564 activity level. *Journal of Sports Science and Medicine*, 13(4), 964-968.

565 Saw, A. E., Main, L. C., & Gustin, G. B. (2015). Role of a self-report measure in athlete
566 preparation. *Journal of Strength and Conditioning Research*, 29(3), 685-691.

567 Saw, A. E., Kellmann, M., Main, L. C., & Gustin, P. B. (2017). Athlete self-report measures
568 in research and practice: Consideration for the discerning reader and fastidious practitioner.
569 *International Journal of Sports Physiology and Performance*, 12(Suppl2), S2127-S2135.

570 Slack, L. A., Maynard, I. W., Butt, J., & Olusoga, P. (2013). Factors underpinning football
571 officiating excellence: perceptions of English Premier League referees. *J Appl Sport Psych*,
572 25(3), 298-315.

573 Shrier, I., Raglin, J. S., Levitan, E. B., Mittleman, M. A., Steele, R. J., & Powell, J. (2014).
574 Procedures for assessing psychological predictors of injuries in circus artists: A pilot
575 prospective study. *BMC Medical Research Methodology*, 14, 77.

576 Sparkes, A. C., & Smith, B. (2013). *Qualitative research methods in sport, exercise and*
577 *health: From process to product*. London: Routledge.

578 Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2018). The role of
579 domain-generic and domain-specific perceptual-cognitive skills in association football
580 referees. *Psychology of Sport and Exercise*, 34, 47-56.

581 Unkelbach, C., & Memmert, D. (2008). Game management, context effects, and calibration:
582 the case of yellow cards in soccer. *Journal of Sport and Exercise Psychology*, 30(1): 95-109.

583 Wade, M. V. (2006). Likert-type scale response anchors. Clemson International Institute for
584 Tourism & Research Development, Department of Parks, Recreation and Tourism
585 Management, Clemson University.

586 Webb, T., Dicks, M., Thelwell, R., van der Kamp, J., & Rix-Lievre, G. (2020). An analysis of
587 soccer referee experiences in France and the Netherlands: abuse, conflict, and level of
588 support. *Sport Management Review*, 23(1): 52-65.

589 Weston, M. (2018). Match performances of soccer referees: the role of sports science.
590 *Movement & Sports Science*, 87, 113-117.

591 Weston, M., Bird, S., Helsen, W., Nevill, A., & Castagna, C. (2006). The effect of match
592 standard and referee experience on the objective and subjective match workload of English
593 Premier League referees. *Journal of Science and Medicine in Sport*, 9(3), 256-262.

594 Weston, M., Castagna, C., Impellizzeri, F. M., Bizzini, M., Williams, A. M., & Gregson, W.
595 (2012). Science and medicine applied to soccer refereeing: An update. *Sports Medicine*,
596 42(7), 615-631.

597 Weston, M., Drust, B., & Gregson, W. (2011). Intensities of exercise during match-play in
598 FA Premier League referees and players. *Journal of Sports Sciences*, 29(5), 527-532.

599

Table 1. Demographic information of Referee Training Activity Questionnaire cohort (n = 31)

	Mean (SD)	Range	Number (proportion who completed training diary)
Age (years)	34 ± 7	24 – 49	
Officiating experience (years)	12 ± 6	3 – 31	
Officiating role			
FR			19 (73.7)
AR			12 (91.7)
Category			
Category 1 (FIFA)			4 (75.0)
Category 1			5 (80.0)
Category 1 (Development)			4 (75.0)
Category 2			1 (0.0)
Category 2 (Development)			3 (100.0)
Category 3 SAR (FIFA)			2 (100.0)
Category 3 SAR			9 (88.9)
Category 3 SAR (Development)			1 (100.0)
Category 3			1 (0.0)
Category 3 (Development)			1 (100.0)
Highest level officiated			
Intercontinental			1 (100.0)
Continental			10 (80.0)
National			17 (82.4)
Regional			3 (66.7)

Data are presented as mean (SD), range, or number (proportion who completed training diary). FR, field referee. AR, assistant referee. SAR, specialist assistant referee.

Table 2. Preliminary structure of the Referee Training Activity Questionnaire (RTAQ)

	Items (<i>n</i>)	Question	Response scale
Section 1. General training practices			
1.1. Domain-specific activities	6	Please indicate the number of times (<i>within the past two weeks</i>) and the average duration per session that you have engaged in each of the following activities.	Open-ended
1.2. Training environment	5		
1.3. Training objective	5		
Section 2. Specific training practices			
<i>Section 2.1. Training activities</i>			
2.1.1. Physical	10	Please indicate the number of times (<i>within the past two weeks</i>) and the average duration per session that you have engaged in each of the following activities.	Open-ended
2.1.2. Decision-making	4		
2.1.3. Psychological	9		
2.1.4. Technical	4		
<i>Section 2.2. Attributes</i>			
2.2.1. Physical	10	Please indicate how often you engage in training that is focused on developing each of the following skills and attributes.	5-point Likert
2.2.2. Decision-making	9		
2.2.3. Psychological	7		
2.2.4. Technical	6		
2.2.5. Game-management	6		

Table 3. Content validity index (CVI) for each section of the Referee Training Activity Questionnaire (RTAQ) (n = 6)

	Please rate the degree to which you believe the wording of the above question is clear and concise (1 = not clear; 4 = very clear)		Please rate the degree to which you believe the above question is relevant for understanding the current training practices of soccer officials (1 = not relevant; 4 = very relevant)		Please rate the degree to which you believe the items listed above are of relevance to soccer officials (1 = not relevant; 4 = very relevant)		Please rate the degree to which you believe the items listed above match the descriptions provided (1 = poor match; 4 = good match)		Please rate the degree to which you believe the scales were clear and easy to use (1 = not easy; 4 = very easy)	
	Mean (range)	CVI	Mean (range)	CVI	Mean (range)	CVI	Mean (range)	CVI	Mean (range)	CVI
1. General training practices										
1.1. Domain-specific	3.7 (3–4)	1.00	3.8 (3–4)	1.00	3.8 (3–4)	1.00	3.7 (3–4)	1.00	3.5 (3–4)	1.00
1.2. Environment	3.7 (3–4)	1.00	4.0 (4)	1.00	4.0 (4)	1.00	3.8 (3–4)	1.00	3.5 (3–4)	1.00
1.3. Objective	3.7 (3–4)	1.00	4.0 (4)	1.00	4.0 (4)	1.00	4.0 (4)	1.00	3.5 (3–4)	1.00
2. Specific training practices										
<i>2.1. Training activities</i>										
2.1.1. Physical	3.7 (3–4)	1.00	4.0 (4)	1.00	3.8 (3–4)	1.00	4.0 (4)	1.00	2.3 (1–4)	0.50
2.1.2. Decision-making	3.8 (3–4)	1.00	4.0 (4)	1.00	3.8 (3–4)	1.00	3.7 (3–4)	1.00	2.3 (1–4)	0.50
2.1.3. Psychological	3.7 (3–4)	1.00	3.8 (3–4)	1.00	4.0 (4)	1.00	3.7 (2–4)	0.83	2.5 (1–4)	0.50
2.1.4. Technical	3.7 (3–4)	1.00	3.5 (2–4)	0.83	3.3 (2–4)	0.83	3.3 (2–4)	0.83	2.3 (1–4)	0.50
<i>2.2. Attributes</i>										
2.2.1. Physical	3.7 (3–4)	1.00	3.8 (3–4)	1.00	3.7 (3–4)	1.00	3.7 (3–4)	1.00	3.5 (2–4)	0.83
2.2.2. Decision-making	3.7 (2–4)	0.83	3.8 (3–4)	1.00	3.8 (3–4)	1.00	3.7 (3–4)	1.00	3.5 (2–4)	0.83
2.2.3. Psychological	3.7 (3–4)	1.00	3.7 (3–4)	1.00	3.7 (3–4)	1.00	3.7 (3–4)	1.00	3.7 (3–4)	1.00
2.2.4. Technical	3.5 (3–4)	1.00	3.7 (3–4)	1.00	3.7 (3–4)	1.00	3.5 (2–4)	0.83	3.5 (2–4)	0.83
2.2.5. Game-management	3.7 (3–4)	1.00	3.8 (3–4)	1.00	3.8 (3–4)	1.00	4.0 (4)	1.00	3.7 (3–4)	1.00

Data are presented as mean (range). CVI, content validity index.

Table 4. Systematic bias, 95% limits of agreement (LOA), and Pearson correlations for training volume (n = 25)

	Assessment method		Bias (95% CI)	LOA -	LOA +	r (95% CI)
	RTAQ (min/week)	Training diary (min/week)				
Section 1. Domain-specific activities						
Match officiating	142 ± 63	148 ± 70	-6 (-27 to 15)	-105	93	0.72 (0.45 to 0.87) ***
Coach-led individual training	17 ± 35	11 ± 34	6 (-5 to 18)	-49	61	0.68 (0.39 to 0.85) ***
Coach-led group training	54 ± 55	63 ± 59	-10 (-22 to 2)	-67	47	0.87 (0.73 to 0.94) ***
Self-led individual training	173 ± 91	169 ± 92	5 (-13 to 22)	-77	86	0.90 (0.78 to 0.95) ***
Self-led group training	3 ± 10	5 ± 14	-2 (-5 to 1)	-15	11	0.91 (0.81 to 0.96) ***
Peer-led play	7 ± 26	2 ± 12	5 (-2 to 12)	-28	37	0.89 (0.77 to 0.95) ***
Section 2. Training environment						
Gym-based	131 ± 85	134 ± 90	-3 (-19 to 12)	-79	72	0.91 (0.80 to 0.96) ***
Pitch-based	87 ± 92	83 ± 84	4 (-15 to 23)	-86	95	0.87 (0.72 to 0.94) ***
Track-based	14 ± 27	10 ± 25	3 (-7 to 14)	-46	53	0.53 (0.17 to 0.77) **
Road-based	32 ± 46	35 ± 53	-3 (-12 to 5)	-44	38	0.92 (0.83 to 0.97) ***
Classroom-based	27 ± 47	52 ± 70	-25 (-45 to -5)*	-119	69	0.73 (0.47 to 0.87) ***
Section 3. Training objective						
Skill development session	16 ± 36	82 ± 103	-67 (-102 to -31)**	-237	104	0.59 (0.25 to 0.80) **
Fitness development session	163 ± 72	185 ± 109	-22 (-50 to 6)	-155	111	0.79 (0.58 to 0.91) ***
Injury prevention session	27 ± 28	28 ± 27	-1 (-11 to 9)	-47	45	0.64 (0.32 to 0.83) **
Match preparation session	21 ± 25	21 ± 30	-1 (-9 to 7)	-38	36	0.78 (0.57 to 0.90) ***
Recovery session	35 ± 33	37 ± 34	-2 (-8 to 5)	-32	28	0.90 (0.77 to 0.95) ***

Data are presented as mean (SD). RTAQ, Referee Training Activity Questionnaire. CI, confidence interval. * $P < 0.05$. ** $P < 0.01$. *** $P < 0.001$.

Table 5. Systematic bias, 95% limits of agreement (LOA), and Pearson correlations for training volume (n = 25)

	Assessment method		Bias (95% CI)	LOA -	LOA +	r (95% CI)
	RTAQ (min/week)	Training diary (min/week)				
Physical training activities						
High-intensity interval training	48 ± 50	52 ± 51	-4 (-17 to 9)	-67	59	0.80 (0.59 to 0.91) ***
Continuous aerobic training	36 ± 38	44 ± 43	-9 (-22 to 5)	-73	56	0.68 (0.39 to 0.85) ***
Speed endurance	18 ± 24	10 ± 17	8 (-1 to 16)	30	46	0.58 (0.24 to 0.80) **
Repeated sprint ability	15 ± 23	11 ± 18	4 (-3 to 12)	-31	39	0.65 (0.34 to 0.83) ***
Speed and agility	12 ± 21	8 ± 15	5 (-1 to 11)	-24	33	0.70 (0.43 to 0.86) ***
Resistance training	59 ± 60	68 ± 59	-10 (-22 to 3)	-68	48	0.88 (0.74 to 0.94) ***
Game-based activities	37 ± 89	0 ± 2	37 (1 to 78) *	-138	212	0.04 (-0.36 to 0.43)
Cross training	28 ± 49	38 ± 58	-9 (-22 to 3)	-70	52	0.84 (0.67 to 0.93) ***
Mobility training	16 ± 27	20 ± 28	-4 (-12 to 4)	-41	33	0.76 (0.52 to 0.89) ***
Decision-making training activities						
On-field refereeing simulations	7 ± 22	2 ± 12	5 (-3 to 13)	-32	42	0.51 (0.14 to 0.75) *
Video clip analyses	64 ± 100	49 ± 69	16 (-13 to 45)	-122	153	0.72 (0.45 to 0.87) ***
Interactive video-based decision-making	9 ± 27	8 ± 25	1 (-3 to 4)	-17	19	0.94 (0.87 to 0.98) ***
Other decision-making training	9 ± 21	4 ± 11	4 (-4 to 12)	-32	40	0.49 (0.11 to 0.74) ***
Psychological skills training						
Self-talk	10 ± 32	0 ± 0	10 (-4 to 23)	-54	73	n/a [§]
Emotional control	2 ± 6	0 ± 2	2 (-1 to 4)	-11	14	0.02 (0.38 to 0.41)
Automaticity	5 ± 24	0 ± 0	5 (-5 to 15)	-42	52	n/a [§]
Goal setting	7 ± 14	3 ± 10	4 (-1 to 9)	-17	26	0.63 (0.31 to 0.82) **
Imagery	5 ± 9	0 ± 2	4 (-1 to 8) *	-14	22	0.04 (0.36 to 0.43)
Activation	3 ± 7	0 ± 0	3 (-1 to 6)	-11	17	n/a [§]
Relaxation	7 ± 19	2 ± 9	5 (-1 to 10) *	-18	28	0.90 (0.79 to 0.96) ***
Negative thinking	2 ± 9	0 ± 0	2 (-1 to 6)	-16	20	n/a [§]
Distractibility	0 ± 2	0 ± 0	0 (-1 to 1)	-3	3	n/a [§]
Technical training						
Technical skills training	10 ± 27	5 ± 24	6 (-1 to 11)	-21	32	0.86 (0.71 to 0.94) ***
Tactical research	31 ± 40	19 ± 27	12 (-2 to 26)	-55	78	0.55 (0.19 to 0.73) **
Laws of the game study	20 ± 28	11 ± 25	9 (-1 to 19)	-35	54	0.64 (0.33 to 0.83) **
Game analysis	44 ± 50	47 ± 55	-3 (-22 to 17)	-96	90	0.59 (0.26 to 0.80) **

Data are presented as mean (SD). RTAQ, Referee Training Activity Questionnaire. CI, confidence interval. * $P < 0.05$. ** $P < 0.01$. *** $P < 0.001$. § Unable to be calculated.

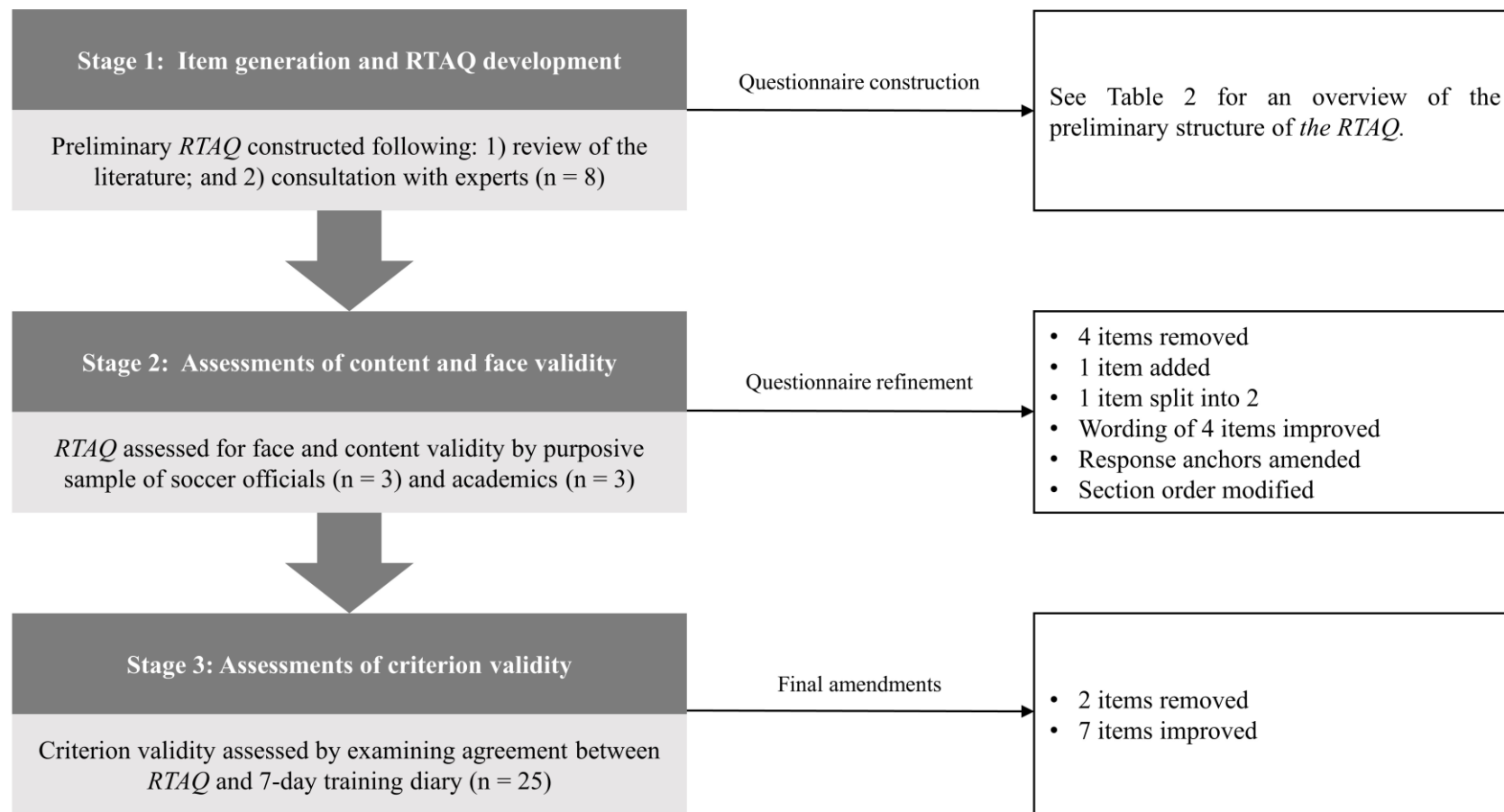


Figure 1. Flow chart of multi-stage process used to develop and validate the Referee Training Activity Questionnaire (*RTAQ*)

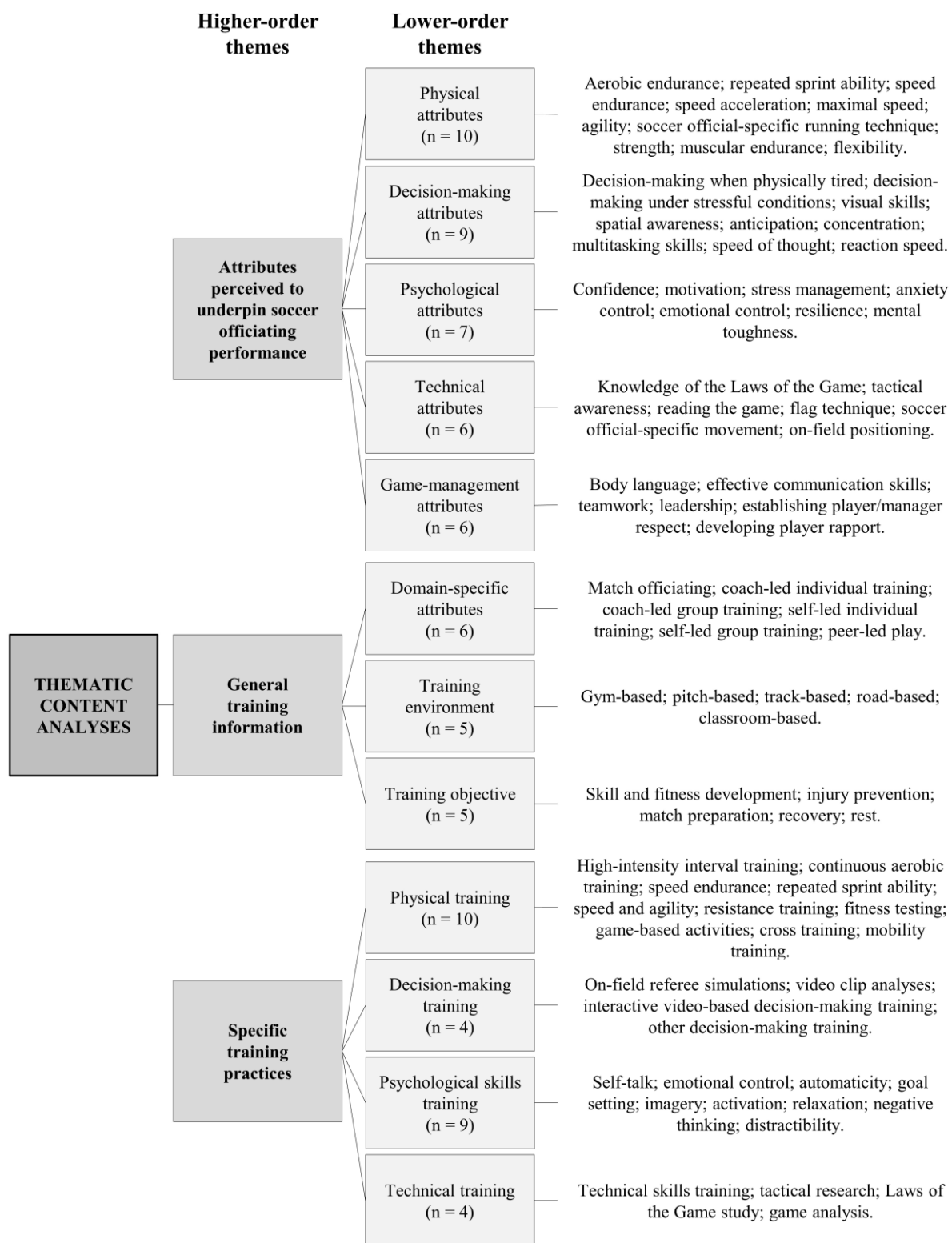


Figure 2. Item pool generated following thematic content analyses of semi-structured interviews