

Johansen, B. T., Ommundsen, Y., Haugen, T. (2018). Referee efficacy in the context of Norwegian soccer referees – A meaningful construct?. *Psychology of Sport And Exercise*, 38, 184-191.

---

Dette er siste tekst-versjon av artikkelen, og den kan inneholde små forskjeller fra forlagets pdf-versjon. Forlagets pdf-versjon finner du her:

<http://dx.doi.org/10.1016/j.psychsport.2018.06.013>

---

This is the final text version of the article, and it may contain minor differences from the journal's pdf version. The original publication is available here:

<http://dx.doi.org/10.1016/j.psychsport.2018.06.013>

---

1

2 **Referee efficacy in the context of Norwegian soccer referees—a meaningful construct?**

3

4

5 Running head: Scale validation and mediating role of referee self-efficacy

6

7

8 **Bjørn Tore Johansen<sup>#</sup>, Yngvar Ommundsen<sup>\*#</sup>, Tommy Haugen<sup>#</sup>**

9 Faculty of Health and Sport, University of Agder<sup>\*</sup>

10 Norwegian School of Sport Sciences<sup>#</sup>

11 **Abstract**

12 Objectives

13 The purpose of this study was two-fold. Firstly, to examine the measurement/factorial validity  
14 and invariance of the Referee Self-Efficacy Scale (REFS) among Norwegian soccer referees.  
15 Secondly, extending scale validation, we also tested a structural model in which a second-  
16 order version of the REFS was modelled to mediate a set of theoretically informed  
17 antecedents and outcomes.

18 Design

19 Cross-sectional

20 Method

21 One hundred and eleven Norwegian elite referees and 81 non-elite referees completed an  
22 electronic questionnaire measuring expected antecedents and outcomes of perceived self-  
23 efficacy in the role as referees

24 Results

25 Analyses provided support for the first – and second order versions of the REFS scale as well  
26 as for scale invariance. Further, a task goal and experience as referee related positively to  
27 referee positive affect, mediated by referee efficacy. Unexpectedly, physical preparation  
28 related directly to positive affect, whereas no mediation effects or direct effects for mental  
29 preparation and ego goal were observed.

30 Conclusions

31 Measurement/factorial validity and invariance of the REFS were fully supported. Validity of  
32 the scale in mediation model by Guillén and Feltz (2011) was only partly supported.

33 **Keywords:** Referee Self-Efficacy Scale; elite referee; non-elite referee; validation

34 **Highlights**

- 35 • Testing the factorial validity and invariance of the Norwegian REFS scale.
- 36 • Cross-sectional study using latent variable modeling.
- 37 • Support for validity and invariance of the referee efficacy scale.
- 38 • REFS mediated years of experience and task orientation on positive affect.

39

40 **Introduction**

41       Officiating at a game of soccer is both physically and mentally demanding, and it is  
42 probably impossible for referees in any sport to avoid making mistakes during a game. Hence,  
43 from the referees' perspective, the inherent possibility of making mistakes during a game may  
44 lead to loss of confidence, increased anxiety and stress levels among them and, in the worst  
45 case, burnout and more frequent dropout (Guillén & Feltz, 2011).

46       However, Bandura (1997) holds that a person's sense of self-efficacy includes  
47 optimistic self-beliefs, which are defined as the strength of an individual's conviction that he  
48 or she can successfully execute behaviors that are required to achieve certain outcomes. Such  
49 perceptions are predicted to influence task choices, effort expenditure, and resilience to  
50 failure, as well as affective states (Bandura, 1997). Researchers have developed conceptual  
51 frameworks for efficacy beliefs within various performance contexts, such as academic  
52 (Bong, 2001; Federici & Skaalvik, 2012), organizational (Stajkovic & Luthans, 1998), and  
53 sport (Sullivan & Kent, 2003) contexts. Although efficacy beliefs have been well researched  
54 in sports, most of these studies have focused on the self-efficacy beliefs of athletes (Sullivan  
55 & Kent, 2003) and coaches (Myers, Feltz, & Wolfe, 2008) in terms of cognition, affect, and

56 behavior. In contrast, self-efficacy research on sports officiating is less well developed.  
57 Guillén and Feltz (2011) initiated this research agenda with a conceptual model of *referee*  
58 *self-efficacy*, which they termed “referee efficacy” and defined as “the extent to which  
59 referees believe they have the capacity to perform successfully in their job” (p.1). Guillén and  
60 Feltz (2011) emphasized that the concept of referee efficacy in the context of sports is unique.  
61 Thus, the dimensions of referee efficacy and the sources of efficacy information for efficacy  
62 beliefs and performance may differ from those that have been observed in other contexts  
63 (Guillén & Feltz, 2011). They conducted focus group interviews of nine male soccer referees  
64 to collect information for their model of referee efficacy, and six key confidence components  
65 for officiating success emerged: game knowledge, decision-making skills, psychological  
66 skills, strategic skills, communication/control of the game, and physical fitness (Guillén &  
67 Feltz, 2011). They also acknowledged that their model was preliminary and that it probably  
68 contained fewer sources, dimensions, and outcomes pertaining to referee efficacy than they  
69 first suggested. Inspired by this preliminary conceptual framework for referee efficacy,  
70 Myers, Feltz, Guillén and Dithurbide (2012) conducted multiple studies to provide initial  
71 evidence for the validity of measures derived from the Referee Self-Efficacy Scale (REFS,  
72 Myers et al., 2012). The operational definition of referee self-efficacy in REFS is slightly  
73 different than the one Guillén and Feltz (2011) used, and it was oriented more toward the  
74 extent to which a referee believes that he or she can successfully officiate at a  
75 match/competition. The concept of referee self-efficacy was developed for referees in team  
76 sports (Myers et al., 2012), and it was thought to have four first-order dimensions (defined by  
77 just a few items each), including game knowledge (GK), decision-making (DM), pressure  
78 (PR), and communication (CM). The REFS instrument was tested in three different studies  
79 with large sample sizes and preliminary evidence for the internal and external validity of the  
80 measures was provided. Guillén and Laborde (2015) encouraged other scholars in various

81 countries to translate and examine the REFS instrument in order to develop a broader  
82 empirical research base on referee efficacy. Recently, Karacam and Pular (2017) validated a  
83 Turkish version of REFS in a sample of team sport referees in which they expanded the  
84 instrument by integrating a physical dimension. In addition, Guillén, Feltz, Gilson, and  
85 Dithurbide (in press) found suitable properties in terms of the instruments dimensionality and  
86 internal consistency in a study using a Spanish version of REFS on team sport referees. .  
87 Finally, an Italian version of the REFS instrument has been used in a sample of handball  
88 referees (Diotaiuti, Falese, Mancone, & Purromoto, 2017), although this particular paper did  
89 not explicitly address the psychometric properties of the instrument, and used the instrument  
90 as an overall second-order dimension.

91 Despite the abovementioned exceptions, evidence for the ecological/cultural validity  
92 of the REFS instrument may still be considered sparse (Myers et al., 2012).. Thus, the first  
93 purpose of the present study was to examine the measurement/factorial validity of a  
94 Norwegian version of the REFS instrument. In addition, there is evidence of measurement  
95 invariance across relevant subgroups in the sports literature on self-efficacy (Myers et al.,  
96 2012). However, as argued by the same researchers, “it is unclear if it is reasonable to assume  
97 that REFS measures would be comparable across relevant subgroups” (Myers et al., 2012;  
98 p.744). Hence, we also examined the invariance of the scale across two groups of Norwegian  
99 referees: elite and non-elite referees.

100 To extend the work of Myers and his colleagues (Myers et al., 2012), we also tested  
101 the instrument modeled both as first-order dimensions and as an overall second-order  
102 dimension (e.g., Diotaiuti et al., 2017). In the latter case, it would seem possible to facilitate  
103 future tests of more complex structural models using referee efficacy as an antecedent,  
104 mediator, or outcome variable (Byrne, 2012).

105 *Behavioral and cognitive antecedents and outcomes of referee self-efficacy*

106           Following recommendations by Guillén and Feltz (2011), the second purpose of this  
107 study was to extend the validation of the scale by examining the mediating role of referee  
108 efficacy in relation to selected antecedents and outcomes. We took advantage of suggestions  
109 made by Feltz and her colleagues (Feltz, Chase, Moritz, & Sullivan, 1999) and modelled  
110 referees' physical and mental preparation and referees' experience as antecedents. According  
111 to Bandura (1977), preparations and experience are the most dependable for forming efficacy  
112 judgments and they have been found to be strong predictors of coaching efficacy (Feltz et al.,  
113 1999). Of course, the potential importance of physical and mental preparedness is supported  
114 by the consensus that refereeing in soccer is a mental process that requires quick decisions  
115 (Bar-Eli, Plessner, & Raab, 2011; Helsen & Bultynck, 2004; Hoseini, Aslankhani, Abdoli, &  
116 Mohammadi, 2011; Johansen & Haugen, 2013; Plessner & Haar, 2006). Given the  
117 psychological demands required for successful soccer officiating, mental or psychological  
118 preparation is as important as physical preparation (Blumenstein & Orbach, 2014; Giske,  
119 Haugen, & Johansen, 2016; Piffaretti, 2007; Wolfson & Neave, 2007).

120           In accordance with recommendations by Guillén and Feltz (2011), we extended the set  
121 of antecedents by adding referees' achievement goal orientations to the equation. Goal  
122 orientations represent one's cognitive dispositional tendency in evaluating ability and success  
123 in achievement situations (Nicholls, 1989). According to the social cognitive perspective of  
124 achievement goal theory, a task orientation refers to individual effort that is focused on  
125 learning, trying hard, improving, and mastering the task by demonstrating ability. In contrast,  
126 a fixed conception of ability elicits an ego orientation, which is the cognitive dispositional  
127 tendency to want to demonstrate one's superior ability via social comparison. Individuals who  
128 possess a high learning goal orientation are thought to believe that their abilities are malleable  
129 and thus they approach tasks with the intention of developing their skills and abilities. By

130 viewing their ability as fixed, individuals with a high-performance goal orientation approach  
131 tasks with the sole intention of performing well (Dweck & Leggett, 1988; Jourdan, Bandura &  
132 Banfield, 1991). Goal orientations have been shown to influence self-efficacy, with a task  
133 goal orientation facilitating efficacy beliefs, whereas an ego goal orientation has been shown  
134 to diminish self-efficacy (Phillips & Gully, 1997). In sports, a task orientation has  
135 demonstrated a consistent and significant positive association with self-referenced sources of  
136 confidence information about athletic ability, whereas an ego orientation has been positively  
137 associated with the use of normative confidence sources related to ability and expectations for  
138 success (Magyar & Feltz, 2003). Hence, there seems to be evidence for empirical links  
139 between goal orientations and appraisal of sports confidence sources such that different goal  
140 orientations may differentially activate the appraisal of efficacy beliefs.

141         With respect to outcomes of referee self-efficacy, Guillén and Feltz (2011) proposed  
142 referee satisfaction as a positive outcome of high referee efficacy. In the current case, we  
143 emphasized a slightly more affective outcome as represented by positive affective states.  
144 Bandura (1990) argued that when individuals with low self-efficacy expectations perform a  
145 personally important task, they may feel despondent, especially if they anticipate failure. In  
146 contrast, when referee self-efficacy beliefs are strong, individuals are more likely to expect  
147 success, which would be likely to elicit positive emotions as exemplified by positive affect  
148 (Thelwell, Lane, & Weston, 2007). Examining the role of positive affect in refereeing would  
149 seem to be important, given the research findings that a lack of enjoyment while refereeing is  
150 related to intentions to quit and to referees' stay/leave behavior (van Yperen, 1998).

### 151 *Expectations*

152         First, consistent with previous research( Myers et al., 2012; Karacam & Pular, 2017;  
153 Guillén et al., in press), we expected to gain support for the following four first-order



154 dimensions of the Norwegian version of the REFS: game knowledge (GK), decision-making  
155 (DM), pressure (PR), and communication (CM). Second, we expected the first-order  
156 measurement model to be invariant across elite and non-elite referees. Thirdly, we expected to  
157 gain support for the second-order model of the scale.

158 With respect to the structural model, as illustrated in Figure 1, we expected referee  
159 experience and self-reported mental and physical preparation to relate positively to positive  
160 affective states when refereeing, mediated by high referee efficacy. In terms of the referees'  
161 goal orientations, we expected a task goal orientation to relate positively to positive affective  
162 states when refereeing, mediated by high referee efficacy. We expected the opposite to hold  
163 true for an ego goal orientation.

164 Figure 1 about here

165

## 166 **Method**

### 167 *Participants*

168 Two samples of participants were recruited for this study. Sample 1 was composed of  
169 elite Norwegian referees and assistant referees from the top Norwegian men's league,  
170 "*Tippeligaen*," the second-best league, "*Obosliga*," and from the top women's league,  
171 "*Toppserien*." Of the 141 elite referees in Norway, 111 (78.7%) referees (103 males and 8  
172 females) ranging in age from 19 to 49 yrs (mean = 31 yrs,  $SD = 7.00$  yrs) participated. Sample  
173 2 consisted of 81 (74%) of 110 non-elite referees (75 males, 5 females, and one who did not  
174 give his/her gender) ranging in age from 15 yrs to 67 yrs (mean = 8 yrs,  $SD = 15.88$  yrs). The  
175 non-elite referees all officiated in regional and/or lower-level leagues in one referee  
176 administrative region in southern Norway.

177 *Procedure*

178           The Norwegian Social Science Data Services approved this study. The data were  
179 collected using SurveyXact, a web-based program for conducting electronic questionnaires.  
180 The questionnaires were distributed by email before the soccer league season started to all the  
181 referees ranked and listed by the Norwegian Football Federation (NFF). The email provided a  
182 link to the questionnaire and was open for 30 days. The first page of the questionnaire  
183 informed the referees about the purpose of the study and emphasized that participation was  
184 voluntary and anonymous.

185 *Measures*

186           *Referee self-efficacy:* Using a Norwegian translation of the Referee Self-Efficacy  
187 Scale (Myers et al., 2012), referees were asked to rate the extent to which each statement  
188 reflected their perceived self-efficacy as referees. Using a five-point response scale ranging  
189 from “very little” to “very strongly” (Myers, Wolfe, & Feltz, 2005), referees rated themselves  
190 on four first-order dimensions tapping 1) *game knowledge (GK)*, 2) *decision-making (DM)*, 3)  
191 *pressure (PR)*, and 4) *communication (CM)*, respectively. GK was defined as the confidence  
192 that a referee has in his/her knowledge of his/her sport. DM was defined as the confidence  
193 that a referee has in his/her ability to make decisions. PR was defined as the confidence that a  
194 referee has in his/her ability to be uninfluenced by pressure. CM was defined as the  
195 confidence that a referee has in his/her ability to communicate effectively. The internal  
196 consistency coefficients (Raykov’s rho) for the four subdimensions were GK rho = .71, DM  
197 rho = .84, PR rho = .91, and CM rho = .80.

198           *Referee positive affect* was assessed with five items from the short-form version  
199 (Thompson, 2007) of the positive affective states of the original PANAS scale (Watson &  
200 Clark, 1994). The PANAS short form is an example of a scale that shows strong psychometric  
201 properties and can be utilized to capture the distinct facets of positive (as well as negative)

202 affectivity (Thompson, 2007). Referees were presented with a list of five positive emotions  
203 (*happy, excited, content, joyous, enthusiastic*) and asked to “read each item and then indicate  
204 how often you have had these feelings when refereeing during the last month.” Responses  
205 ranged from 1 (seldom/never) to 7 (always). Raykov’s rho for this scale was .78.

206 *Mental preparation for refereeing* was measured by one item that asked, “*Do you practice*  
207 *any form of mental preparation?*” The response choices were “two times a week or more,”  
208 “once a week,” “less than once a week,” and “never”.

209 *Physical preparation for refereeing* was measured by one item that asked, “*How often*  
210 *do you exercise?*” The response choices were “More than once a day,” “Once a day,” “6 days  
211 a week,” “4–5 days a week,” “2–3 days a week,” and “less than 2–3 days a week.”

212 *Experience as referee* was measured by one item that asked, “*How many years have*  
213 *you been acting as referee?*” The response format was number of years as referee.

214 *Achievement goals as a referee* was measured with the Norwegian version of the  
215 Perception of Success Questionnaire (POSQ), which was modified for refereeing by the stem,  
216 “*When I referee a soccer match, I feel most successful when...*” The POSQ was developed as  
217 a sport-specific measure of ego and task goal perspectives (Roberts & Ommundsen, 1996),  
218 and it has been found to be both valid and reliable across various samples when examining  
219 motivational goal perspectives in sport (Roberts, Treasure, & Hall, 1994). For the POSQ, the  
220 referees were asked to think of when they felt the most successful as a referee and to respond  
221 to 12 items, 6 items that reflected a task orientation and 6 that reflected an ego orientation.  
222 The five-category response scale ranged from totally agree to totally disagree. Raykov’s rho  
223 was .87 for the ego-oriented items and .88 for the task-oriented items.

224

225 *Statistical analyses*

226 Means, standard deviations, and bivariate correlations were calculated for all  
227 variables. We tested a research model containing five exogenous variables (achievement  
228 goals, task/ego, mental preparation for refereeing, physical preparation for refereeing, and  
229 experience as a referee) and two endogenous variables. The endogenous variables included  
230 referee efficacy as a second-order construct defined by the four first-order dimensions (GK,  
231 DM, PR, and CM) and positive affect as a referee. Achievement goals, referee self-efficacy  
232 dimensions, and positive affect were all measured as latent constructs, whereas mental  
233 preparation for refereeing (ordinal metric), physical preparation for refereeing (ordinal  
234 metric), and experience as a referee (number of years, interval metric) were observed  
235 variables. We used *Mplus 7.2* statistical software to estimate the measurement models with a  
236 maximum likelihood estimation method with robust standard errors (MLR) because it is  
237 robust to non-normality (Satorra & Bentler, 1994), and we used a bootstrap procedure with  
238 1000 bootstrap samples to examine the structural model.

239 We preferred a CFA approach over ESEM for the REFS scale, the achievement goals,  
240 and the PANAS scale, given that there is sufficient a priori measurement theory for these  
241 constructs (Crawford & Henry, 2004; Myers et al., 2012; Wang, Liu, Chatzisarantis, & Lim,  
242 2010). The overall fit of each measurement model was assessed with the Satorra–Bentler chi-  
243 square (S-B  $\chi^2$ ; Satorra & Bentler, 1994). Although nonsignificant S-B  $\chi^2$  values are deemed  
244 acceptable, they are sensitive to sample size. Hence, they should be inspected alongside the  
245 comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the  
246 standardized root mean square residual (SRMR). A CFI greater than .95, a RMSEA lower  
247 than .06, and a SRMR lower than .08 represent good fit criteria (Hu & Bentler, 1999). We  
248 also provided an overview of any re-specifications of our measurement models, as Byrne  
249 (2012) has recommended. Each latent variable was measured with its respective observed

250 indicators. One indicator per latent variable was fixed at 1.0, allowing us to scale the latent  
251 variables to a common metric.

### 252 *Testing invariance*

253 As previously stated, the study included referees at the international/elite level as well  
254 as non-elite referees (lower national leagues, district leagues). The two samples were  
255 collapsed to achieve the necessary sample size for our main structural analysis. We justified  
256 this because of the importance of testing for invariance to determine whether the two groups  
257 had similar conceptions of the latent variables pertaining to referee self-efficacy (Dimitrov,  
258 2010). Testing invariance is also important because it allows us to use the instrument in future  
259 analyses among coaches refereeing at elite and non-elite levels. In addition to skewness and  
260 kurtosis, estimates of factor loadings, intercepts, variances, residual variances, and z-scores  
261 ( $>1.96$ ) were inspected for direction of association and magnitude.

### 262 *Internal consistency*

263 We used Raykov's rho, a measure of composite reliability, to test internal consistency.  
264 The advantage of Raykov's rho is that it does not require the equal contribution of items to  
265 factorial variance, and it accounts for correlated error variance. Raykov (1998) demonstrated  
266 that Cronbach's alpha may over- or underestimate scale consistency. Because underestimation  
267 is common, Raykov's rho is now preferred, as it may make more accurate estimates of internal  
268 consistency (Yang & Green, 2010).

269

## 270 **Results**

### 271 *Descriptive statistics*

272 Inspection of skewness and kurtosis revealed that all the items generally fell within the  
273 cut-off values of  $\pm 2$  ( DeVellis, 1991). All of the items loaded on their respective latent  
274 constructs (unstandardized estimates ranging from .30 to 1.08, all of which were statistically  
275 significant ( $p < .001$ )). Means, standard deviations, and correlations are presented in Table 1.  
276 The results generally revealed an expected pattern of correlation between the latent constructs  
277 of achievement goals, referee efficacy, and positive affect. Except for the inter-correlation  
278 between physical and mental preparation ( $r = .25, p < .01$ ), there were no significant  
279 relationships between the observed predictor variables, referee efficacy, and positive affect.  
280 Inspection of the intercepts, variances, residual variances, and z-scores ( $>1.96$ ) indicated that  
281 all of the items on the latent constructs (achievement goals, referee efficacy total, and positive  
282 affect) were related to their corresponding latent factors in a satisfactory manner ( $R^2 = 0.261$ –  
283  $0.676$ ). As discussed in the Methods section, Raykov's rho was used as an indicator of  
284 internal consistency and based on .70 as a standard cut-off (DeVellis, 1991), the rho  
285 coefficients were acceptable for all the factors.

286 Table 1 about here

### 287 *Confirmatory factor analysis of referee self-efficacy*

288 The CFA yielded excellent fit indices for the four-dimensional latent referee self-  
289 efficacy construct (S-B  $\chi^2 = 66.35$  [df = 59, N = 192],  $p = .023$ ; CFI = .99; RMSEA = .025  
290 [.00–.05], and SRMR = .04). As shown in Figure 2, standardized loadings were in the range  
291 of .56–.81 for GK; .74–.87 for DM; .86–.90 for PR; and .55–.84 for CM. Figure 2 also shows  
292 that there were positive correlations between the four dimensions, with only one above .60  
293 (.67), and the remaining ones between .42 and .54. As would be expected, the four subscales  
294 have common variance. Nevertheless, the size of the correlations is moderate, indicating quite  
295 strong discriminant validity among the four referee self-efficacy dimensions (Geiser, 2013).

296

Figure 2 about here

297 *Testing the invariance of referee self-efficacy*

298 We used the Satorra–Bentler chi-square (S-B  $\chi^2$ ) difference test, change in CFI, and  
299 RMSEA as goodness of fit indices to test for the invariance of the measurement models  
300 across the two merged samples of referees. As Table 2 shows, for *referee self-efficacy*, the  
301 increase in restrictions reflecting configural, weak, and strong invariance did not result in  
302 significant changes in chi-square values across levels of invariance. The change in fit indices  
303 across the configural, metric, and scalar levels was within accepted limits ( $\Delta\text{CFI} \leq .01$ ,  
304  $\Delta\text{RMSEA} \leq .015$ ), and thus strong invariance was supported (Dimitrov, 2010). Hence, the  
305 invariance analysis of *referee self-efficacy* justifies our tests of the structural model for the  
306 merged sample of referees. Considering these findings, future studies comparing referee self-  
307 efficacy across elite and non-elite referees also seem to be legitimate.

308 *Testing a second-order model of referee self-efficacy*

309 We also examined whether a common second-order factor that captures the four first-  
310 order dimensions could be identified. As previously argued, this is important in the current  
311 case because we want to use a comprehensive structural mediation model to test referee self-  
312 efficacy as a mediator. Such models automatically increase the number of parameter estimates  
313 when using first-order latent variables only, which could easily create non-identifiable  
314 solutions, especially if the balance between sample size and the number of parameters to be  
315 estimated is less than optimal (Geiser, 2013). The result for the second-order measurement  
316 model indicated that the scale can also be used to capture overall referee efficacy as indicated  
317 by good fit indices (S-B  $\chi^2 = 69.19$  [df = 61, N= 191],  $p = .022$ ; CFI = .99; RMSEA = .027  
318 [.00–.05], and SRMR = .04). We used the second-order version of the scale in our structural  
319 model. It should be noted that the second-order model is statistically equivalent to the first-

320 order model, with the same fit indices (Geiser, 2013). However, as there were no inadmissible  
321 parameter estimates for the second-order model (the so-called Heywood case, yielding  
322 negative variance estimates), the modeling of the second-order factor can be trusted (Chen,  
323 Bollen, Curran, Paxton, & Kirby, 2001).

324 *Testing measurement models and invariance of antecedents and outcomes*

325         The initial CFA yielded nonacceptable fit indices for the 6-item version of the task  
326 orientation scale (S-B  $\chi^2 = 42.29$  [df = 9, N= 183],  $p < .001$ ; CFI = .86; RMSEA = .19 [.12–  
327 .20], and SRMR = .10). Inspection of the item loadings revealed that two items (“I do my  
328 best” and “I make a good effort”) did not contribute to the latent construct as indicated by  
329 extremely high residuals. For these referees, these items may not be meaningful for task  
330 orientation (Byrne, 2012). When we removed those items, we found generally good fit indices  
331 for the remaining four items (S-B  $\chi^2 = 8.98$  [df = 2, N= 183],  $p = .011$ ; CFI = .97; RMSEA =  
332 .13 [.05–.26], and SRMR = .003). For ego orientation, the initial CFA yielded nonacceptable  
333 fit indices for the 6-item version of the scale (S-B  $\chi^2 = 143.76$  [df = 9, N= 183],  $p < .000$ ; CFI  
334 = .71; RMSEA = .29 [.25–.33], and SRMR = .09). Inspection of the item loadings revealed  
335 that “I am the best” and “I am completely superior” only contributed modestly to the latent  
336 construct, as indicated by high residuals (.61 and .66, respectively). Apparently, these items  
337 were not central to an ego goal orientation for these referees, who seemed to be focusing more  
338 on whether they fared better than their fellow referees (“I do better than my colleagues when  
339 it comes to refereeing”). With these two items omitted on statistical grounds (as being less  
340 meaningful, see Byrne, 2012), we found a generally good fit for the remaining four ego  
341 orientation items (S-B  $\chi^2 = 8.98$  [df = 2, N = 189],  $p = .011$ ; CFI = .97; RMSEA = .03 [.00–  
342 .05], and SRMR = .04).



343 The initial CFA yielded nonacceptable fit indices for the 5-item PANAS scale (S-B  $\chi^2$   
344 = 56.06 [df = 5, N= 183],  $p < .001$ ; CFI = .75; RMSEA = .25 [.18–.29], and SRMR = .11).  
345 Inspection of the item loadings revealed high covariance between two items (“I am happy”  
346 and “I am content”). Allowing these two items to co-vary (as per recommendations by Byrne,  
347 2012) resulted in good fit indices for the re-estimated model (S-B  $\chi^2 = 7.41$  [df = 4, N = 183],  
348  $p = .012$ ; CFI = .98; RMSEA = .07 [.00–.14], and SRMR = .003).

349 We found equal invariance estimates across the two merged referee samples for task  
350 orientation and referees’ positive affect, whereas the latent construct ego orientation generally  
351 satisfied the requirement for weak or metric invariance (chi-square diff test configural versus  
352 metric  $p = .15$ , CFI = 0.95/0.95; RMSEA = 0.23/0.18, SRMR = 0.05/0.07).

353 Table 2 about here

#### 354 *Structural model*

355 We ran a bootstrap procedure with 1000 bootstrap samples, which produced a  
356 generally satisfactory fit for the hypothesized ANTECEDENTS (*orientations, referee*  
357 *experience, physical preparation for refereeing, mental preparation for refereeing*) →  
358 MEDIATOR (*referee self-efficacy*) → OUTCOME (*positive mood states as referee*)  
359 structural model (S-B  $\chi^2 = 532.55$  [df = 360, N = 190],  $p < .001$ ; CFI = .93; RMSEA = .05  
360 [.04–.06] and SRMR = .08). Figure 3 shows the  $\beta$  coefficients estimated in the full model.  
361 Both significant and nonsignificant paths are shown.

362 Figure 3 about here

363 The individual standardized paths in Figure 3 show two significant positive  
364 antecedent–mediator paths: task orientation ( $\beta = .46$ ,  $p < .001$ ) and experience as a referee ( $\beta$   
365 = .18,  $p < .001$ ). Further, referee self-efficacy (mediator) was positively related to the

366 outcome (positive mood states,  $\beta = .48, p < .001$ ). In addition, a positive antecedent–outcome  
367 path was observed for physical preparation for refereeing ( $\beta = .17, p < .001$ ). The remaining  
368 links were nonsignificant. Table 3 summarizes the total, total indirect, specific indirect, and  
369 direct standardized effects.

370 Table 3 about here

371 A significant total positive effect of *task orientation* on *positive mood states* was  
372 observed (standardized estimate = 0.36, 95% CI<sub>BC</sub>: .15, .56) with a specific indirect effect of  
373 *task orientation* on *positive mood state*, mediated by *referee self-efficacy* (standardized  
374 estimate = 0.22, 95% CI<sub>BC</sub>: .11, .39). There was a significant total effect of *physical*  
375 *preparation for refereeing* on *positive mood states* (standardized estimate = 0.19, 95% CI<sub>BC</sub>:  
376 .04, .33), reflecting a direct effect (standardized estimate = 0.18, 95% CI<sub>BC</sub>: .03, .31). No  
377 direct or indirect effects on *positive mood state* were observed for *ego orientation* (*ego*  
378 *orientation* total effect; standardized estimate = 0.01, 95% CI<sub>BC</sub>:  $-.19, .18$ ) or *mental*  
379 *preparation for refereeing* (*mental preparation for refereeing* total effect; standardized  
380 estimate = 0.09, 95% CI<sub>BC</sub>:  $-0.07, .024$ ). The squared multiple correlations for the  
381 endogenous variables were: *referee self-efficacy* (.26,  $p < .001$ ) and *positive mood states* (.36,  
382  $p < .001$ ).

383

## 384 Discussion

385 The first aim of the study was to investigate the measurement/cultural validity of the  
386 referee efficacy scale, as suggested by Guillén and Laborde (2015), following the line of  
387 research conducted in Turkish (Karacam & Pular, 2017) and Spanish (Guillén et al., in press)  
388 sports officiating contexts. We took a substantive–methodological approach using  
389 confirmatory structural equation modeling to examine the measurement/factorial validity and

390 subgroup invariance (elite versus non-elite referees) for the Norwegian version of the Referee  
391 Self-efficacy Scale.

392 Modeling the four first-order factors as latent variables, the measurement/factorial  
393 validity of the referee self-efficacy dimensions was supported in the Norwegian context. Our  
394 CFA analyses supported the first-order four-factor solution of the scale, as indicated by good  
395 fit indices. We found generally high factor loadings for the latent factor scores across all four  
396 subscales. The discriminant validity of these subscales was also supported by generally  
397 modest cross-correlations between first-order dimensions. We also showed invariance for the  
398 scale across the elite and non-elite referee samples. Taken together, these findings provide  
399 support for the REFS instrument (Myers et. al, 2012) in the context of Norwegian football  
400 refereeing. Our analyses also provided support for the scale conceptualized as a second-order  
401 construct. Our limited sample size precluded using the first-order latent construct in the  
402 current structural model. Nevertheless, confirmation of the validity of the four first-order  
403 dimensions paves the way for future research to test more complex structural models  
404 (including the first-order dimensions) without the risk of exceeding the sample N/parameter  
405 estimates ratio, which could easily lead to non-identifiable solutions (Geiser, 2013).

406 The second aim of this study was to extend the validation of the REFS by examining  
407 whether referee self-efficacy was a mediator for the relationship between sources of efficacy  
408 information for refereeing as antecedents and positive affective experience as a referee as  
409 outcome. The extended validation yielded only partial support for the validity of the scale.

410 Supporting the scale in this respect, the results for our mediation model indicated a  
411 positive relationship between referee self-efficacy as mediator and positive affect. In their  
412 conceptual model, Guillén and Feltz (2011) proposed that referee satisfaction was one of the  
413 positive outcomes of high referee efficacy. When referees feel efficacious, they may make

414 better decisions, communicate more effectively with players, and receive positive feedback  
415 from players, coaches, and spectators, which, in turn, may generate positive mood states while  
416 refereeing (Guillén & Feltz, 2011). This result is consistent with experimental findings from  
417 self-efficacy research on physical activity, suggesting that efficacy may be a key component  
418 of positive affect in physical activity, such as the enjoyment felt at higher intensity levels (Hu,  
419 Motl, McAuley, & Konopack, 2007).

420         With respect to sources of efficacy information for refereeing, in line with  
421 expectations, we found that years of referee experience was a positive antecedent to positive  
422 affect in the role as referee, mediated by referee self-efficacy. The antecedent-mediator link is  
423 consistent with Guillén and Feltz's (2011) conceptual model. It takes years of practice to  
424 develop good skills as a sports official. Further, our findings are in line with those observed  
425 among Italian handball referees in which years of experience rather than age appeared to be a  
426 predictor of self-efficacy (Diotaiuti et al. (2017). At both the non-elite and elite levels of  
427 refereeing, referees typically encounter increasingly challenging tasks as they rise through  
428 their respective soccer leagues and have to officiate at faster and more complex games (Lirgg,  
429 Feltz, & Merrie, 2016). This necessitates the acquisition of an increasingly better grasp of  
430 declarative knowledge (rules and regulations of the game), as well as the acquisition of  
431 improved procedural knowledge, such as perceptual–movement skills to observe and conduct  
432 visual searches while moving on the pitch, combined with fast decision-making skills to meet  
433 the continuously changing game situations. Indeed, the development of expertise in refereeing  
434 takes many years of dedicated practice (Ericsson, Krampe, & Tesch- Römer, 1993). Our  
435 sample of officials reported a mean of 12 years of referee experience. Thus, many of them  
436 may be considered experts at their respective levels.

437         With respect to goal orientations, a task orientation stood out as a positive antecedent  
438 to positive affect, mediated by referee self-efficacy for these referees. The antecedent-

439 mediator link is consistent with Bandura's (1986) argument that the influence of past  
440 performances on self-efficacy depends upon whether the individual regards his or her abilities  
441 as acquired skills or inherent aptitudes (Jourden et al., 1991). As such, it is not surprising that  
442 those with a task orientation are able to gain a sense of task mastery and feel efficacious. In  
443 contrast, an ego orientation, which relies on social comparison information for knowledge  
444 about one's skills, is thought to undermine self-efficacy expectations because typically ego-  
445 oriented individuals believe that skills are less controllable and less flexible (Feltz & Lirgg,  
446 2001). Thus, social comparison information may be used as a source of information for self-  
447 efficacy (Maddux, 1995), which surely leaves ego-oriented individuals more vulnerable to  
448 threats to their sense of self-efficacy. In the current case, being ego involved did not stand out  
449 as an antecedent to self-efficacy beliefs, thus delimiting extended support for the REFS scale.  
450 While this is not readily explainable, it might be due to these referees reporting relatively low  
451 scores on ego orientation. Our results corroborate previous findings in sports research, which  
452 have revealed that the use of self-referenced sources of confidence information is consistently  
453 positively associated with self-efficacy (Magyar & Feltz, 2003). Our finding regarding a task  
454 goal orientation provide support for extending Guillén and Feltz's (2011) conceptual model to  
455 include additional antecedents.

456         Physical preparation for refereeing was not shown to be an antecedent to referee self-  
457 efficacy. At first sight, this finding confounds expectations, given that sources of efficacy also  
458 comprise internal information regarding one's physiological state or condition. Moreover, in  
459 their study Karacam and Pular (2017) included a physical dimension in the REFS instrument  
460 and gained support for this dimension as a source for referee self-efficacy. Among other  
461 things, one's physiological state or condition includes one's level of fitness, vulnerability to  
462 fatigue, and pain tolerance (Bandura, 1997). Hence, we reasoned that referees who prepared  
463 more physically would be more physically fit and that increased level of fitness would fuel

464 their mental energy as well and thus enhance referee self-efficacy related to making valid  
465 decisions, handling pressure, and communicating effectively throughout matches (Lirgg,  
466 Feltz, & Merrie, 2016). Unexpectedly, the findings revealed physical preparation to be a  
467 direct, unmediated antecedent to positive affect. Hence, preparing physically for the role as  
468 referee seems important for positive affect, irrespective of whether such preparations help  
469 raising efficacy beliefs. Indeed, previous research has shown that enhanced aerobic fitness by  
470 taking part in an aerobic running program directly elicits significant increases in positive  
471 affect (Stroth, Hille, Spitzer, & Reinhardt, 2009).

472           Mental preparation for refereeing was not a significant antecedent in our mediation  
473 model and no direct relation to affect was observed. This finding detracts from the scale  
474 validity of the REFS in our structural model. Apparently, although physical preparation for  
475 refereeing matches relates significantly to positive mood states while refereeing, practicing  
476 visualization, self-talk, or concentration exercises for refereeing do not relate to referee self-  
477 efficacy or positive affect. One possible explanation for this difference may be the self-  
478 reported intensity of physical versus mental preparation in our sample of referees. Descriptive  
479 statistics (not reported) revealed that about 85% of the referees exercised more than 4 or 5  
480 days a week, whereas only 30% of them engaged in mental preparation more than twice a  
481 week. Such modest levels of mental preparation may not be enough to influence referee self-  
482 efficacy and thereby positive affect. Plessner and his colleagues (Plessner, Schweizer, Brand,  
483 & O'Hare, 2009) argue that the enhancement of mental skills to meet the psychological  
484 demands inherent in soccer refereeing requires a long-term learning process of intensified  
485 practice using theoretically informed programs. These researchers hold that this would likely  
486 be effective particularly for decisions about fouls and misconduct, which they see as  
487 representing a considerable proportion of all soccer referee decisions. They argue that most of

488 these decisions can be a perceptual–categorization task based on multiple-cue learning in  
489 which referees have to categorize a set of features into two discrete classes (foul/no-foul).

#### 490 *Conclusion*

491 In line with studies performed by Guillén et al. (in press) and Karacam and Pular  
492 (2017) , the results of this study provide good evidence for the measurement/factorial validity  
493 of the Norwegian version of the Referee Self-Efficacy Scale (Myers et al., 2012). . Moreover,  
494 invariance of the scale was confirmed across elite and non-elite soccer referees, and validity  
495 of the second-order model of the scale was obtained. In our mediation model, however,  
496 referee self-efficacy was shown to operate as a significant mediator only in two instances out  
497 of five using an antecedent–mediator–outcome model, as framed by Guillén and Feltz (2011),  
498 with support for the two expected mediated paths comprising task orientation and years of  
499 referee experience as antecedents.

#### 500 *Strengths, limitations and further research*

501 This study benefited from a unique sample of Norwegian referees, consisting of about  
502 79% of the total population of referees at the elite level in Norway and about 74% of the total  
503 population of referees at the non-elite level in one administrative referee region in Norway. In  
504 addition to using structural equation modeling to analyze the measurement and structural  
505 models, we consider the invariance analyses of the REFS scale across the elite and non-elite  
506 referees to be a strong asset of this study. Adding achievement orientations to the equation in  
507 the structural model extends Guillén and Feltz’s (2011) conceptual framework and allows for  
508 additional sources of referee efficacy within the framework. Limitations of this study include  
509 the use of cross-sectional data in the mediation analyses. This precludes any definitive  
510 conclusions about whether the included predictors precede referee self-efficacy as mediators  
511 and, in turn, influence affective states as outcomes. Hence, future studies would do well to use

512 prospective designs. In addition, conducting studies with an intervention–control group design  
513 would help to test the sensitivity of the REFS scale in intervention programs intended to  
514 enhance referee self-efficacy among this important and often neglected group of people in the  
515 sports world.

516



517 **References**

- 518 Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change.  
519 *Psychological Review*, 84, 191–215.
- 520 Bandura, A. (1986). *Social foundation of thought and action: A social cognitive theory*.  
521 Englewood Cliffs, NJ: Prentice-Hall.
- 522 Bandura, A. (1990). Perceived self-efficacy in the exercise of personal agency. *Journal of*  
523 *Applied Sport Psychology*, 2, 128–163.
- 524 Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- 525 Bar-Eli, M., Plessner, H., & Raab, M. (2011). *Judgment, decision-making and success in*  
526 *sport*. John Wiley & Sons: Chichester, West Sussex-UK.
- 527 Blumenstein, B., & Orbach, I. (2014). Development of psychological preparation program for  
528 football referees: Pilot study. *Sport Science Review*, 23(3–4), 113–125.
- 529 Bong, M. (2001). Between- and within-domain relations of academic motivation among  
530 middle and high school students: Self-efficacy, task value, and achievement goals.  
531 *Journal of Educational Psychology*, 93, 23–34.
- 532 Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications,*  
533 *and programming*. New York, NY: Routledge.
- 534 Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in  
535 structural equation models: Causes, consequences, and strategies. *Sociological*  
536 *Methods & Research*, 29, 468–508.

- 537 Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS):  
538 Construct validity, measurement properties and normative data in a large non-clinical  
539 sample. *British Journal of Clinical Psychology*, *43*, 245–265.
- 540 DeVellis, R. F. (1991). *Scale development: Theory and application*. Newbury Park, CA:  
541 Sage.
- 542 Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct  
543 validation. *Measurement and Evaluation in Counseling and Development*, *43*, 121–  
544 149.
- 545 Diotaiuti, P., Falese, L., Mancone, S., & Purromuto, F. (2017). A structural Model of Self-  
546 efficacy in Handball Referees. *Frontiers in Psychology*, *8*, 811.
- 547 Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and  
548 personality. *Psychological Review*, *95*, 256–273.
- 549 Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in  
550 the acquisition of expert performance. *Psychological Review*, *100*, 363–406.
- 551 Federici, R. A., & Skaalvik, E. M. (2012). Principal self-efficacy: relations with burnout, job  
552 satisfaction and motivation to quit. *Social Psychology of Education*, *15*, 295–320.
- 553 Feltz, D. L., Chase, M. A., Moritz, S. E., & Sullivan, P. J. (1999). A conceptual model of  
554 coaching efficacy: Preliminary investigation and instrument development. *Journal of*  
555 *Educational Psychology*, *91*, 765–776.
- 556 Feltz, D. L., & Lirgg, C. D. (2001). Self-efficacy beliefs of athletes, teams, and coaches. In R.  
557 N. Singer, H. A. Hausenblas, & C. Janelle (Eds.), *Handbook of sport psychology*, 2nd  
558 ed. (pp. 340–361). New York: John Wiley & Sons.

- 559 Geiser, C. (2013). *Data analysis with Mplus*. New York: The Guilford Press.
- 560 Giske, R., Haugen, T., & Johansen, B. T. (2016). Training, mental preparation and  
561 unmediated practice among soccer referees: An analysis of elite and sub-elite referees'  
562 reported practice. *International Journal of Applied Sports Sciences*, 28, 31–41.
- 563 Guillén, F., Feltz, D., Gilson, T., & Dithurbide, L. (In press). Analysis of the Psychometric  
564 Properties of the Spanish Version of Referee Self-Efficacy Scale (REFS). *Revista de*  
565 *Psicología del Deporte*,
- 566 Guillén, F., & Laborde, S. (2015). Self-efficacy referee scale: An overview. In *Proceedings of*  
567 *the 14<sup>th</sup> European Congress of Sport Psychology, 14–19 July, Bern, Switzerland*.
- 568 Guillén, F., & Feltz, D. L. (2011). A conceptual model of referee efficacy. *Frontiers in*  
569 *Psychology*, 2, 1–5.
- 570 Helsen, W., & Bultynck, J.-B. (2004). Physical and perceptual-cognitive demands of top-class  
571 refereeing in association football. *Journal of Sports Sciences*, 22, 179–189.
- 572 Hoseini, S. H., Aslankhani, M. A., Abdoli, B., & Mohammadi, F. (2011). The relationship  
573 between number of crowds with anxiety and the function of the soccer premier  
574 league's referees. *Procedia – Social and Behavioral Science*, 30, 2374–2378.
- 575 Hu, L., Motl, R. W., McAuley, E., & Konopack, J. F. (2007). Effects of self-efficacy on  
576 physical activity enjoyment in college-aged women. *International Journal of*  
577 *Behavioral Medicine*, 14, 92–96.
- 578 Hu, L. T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure  
579 analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*,  
580 6, 1–55

581 Johansen, B. T., & Haugen, T. (2013). Anxiety level and decision-making among Norwegian  
582 top-class soccer referees. *International Journal of Sport & Exercise Psychology*, *11*,  
583 215–226.

584 Jourden, F. J., Bandura, A., & Banfield, J. T. (1991). The impact of conceptions of ability on  
585 self-regulatory factors and motor skill acquisition. *Journal of Sport & Exercise*  
586 *Psychology*, *8*, 213–226.

587 Karaçam, A. & Pular, A. (2017). Adaptation study of referee self-efficacy scale (REFS) to  
588 Turkish. *Journal of Physical Education and Sports Sciences*, *11*, 118-128.

589 Lirgg, C. D.; Feltz, D. L., & Merrie, M. D. (2016). Self-efficacy of sports officials: A critical  
590 review of the literature. *Journal of Sport Behavior*, *39*, 39–50.

591 Maddux, J. E. (1995). Self-efficacy theory: An introduction. In J. E. Maddux (Ed.), *Self-*  
592 *efficacy, adaptation, and adjustment: Theory, research, and application* (pp. 3–33).  
593 New York: Plenum.

594 Magyar, T. M., & Feltz, D. L. (2003). The influence of dispositional and situational  
595 tendencies on adolescent girls' sport confidence sources. *Psychology of Sport and*  
596 *Exercise*, *4*, 175–190.

597 Myers, N. D., Wolfe, E. W., & Feltz, D. L. (2005). An evaluation of the psychometric  
598 properties of the coaching efficacy scale for coaches from the United States of  
599 America. *Measurement in Physical Education and Exercise Science*, *9*, 135–160.

600 Myers, N. D., Feltz, D. L., & Wolfe, E. W. (2008). A confirmatory study of rating scale  
601 category effectiveness for the coaching efficacy scale. *Research Quarterly for*  
602 *Exercise & Sport*, *79*, 300–311.

- 603 Myers, N. D., Feltz, D. L., Guillén, F., & Dithurbide, L. (2012). Development of, and initial  
604 validity evidence for, the Referee Self-Efficacy Scale: A multi-study report. *Journal of*  
605 *Sport and Exercise Psychology*, 34, 737–765.
- 606 Nicholls, J. G. (1989). *The competitive ethos and democratic education*. Cambridge, MA:  
607 Harvard University Press
- 608 Phillips, J. M., & Gully, S. M. (1997). Role of goal orientation, ability, need for achievement,  
609 and locus of control in the self-efficacy and goal-setting process. *Journal of Applied*  
610 *Psychology*, 82, 792–802.
- 611 Piffaretti, M. (2007, November). Mental preparation of top-class referees. In *Congress*  
612 *Proceedings, ACAPS Conference – Dijon, France*.
- 613 Plessner, H., & Haar, T. (2006). Sports performance judgments from a social cognitive  
614 perspective. *Psychology of Sport & Exercise*, 7, 555–575.
- 615 Plessner, H., Schweizer, G., Brand, R., & O’Hare, D. (2009). A multiple-cue learning  
616 approach as the basis for understanding and improving soccer referees’ decision  
617 making. *Progress in Brain Research*, 174, 151–158.
- 618 Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated  
619 nonhomogeneous items. *Applied Psychological Measurement*, 22, 375–385.
- 620 Roberts, G. C., Treasure, D. C. and Hall, H. K. (1994). Parental goal orientations and beliefs  
621 about the competitive sport experience of their child. *Journal of Applied Social*  
622 *Psychology*, 24, 631–645.

- 623 Roberts, G. C., & Ommundsen, Y. (1996). Effect of goal orientation on achievement beliefs,  
624 cognition and strategies in team sport. *Scandinavian Journal of Medicine & Science in*  
625 *Sports*, 6, 46–56.
- 626 Satorra, A. & Bentler, P. M. (1994). Corrections to test statistics and standard errors in  
627 covariance structure analysis. In A. von Eye, & C. C. Clogg (Eds.), *Latent variables*  
628 *analysis: Applications for developmental research* (pp. 399–419). London: Chapman  
629 & Hall.
- 630 Stajkovic, A.D., & Luthans, F. (1998). Self-efficacy and work-related performance: a meta-  
631 analysis. *Psychological Bulletin*, 124, 240–261.
- 632 Stroth, S., Hille, K., Spitzer, M., & Reinhardt, R. (2009). Aerobic endurance exercise benefits  
633 memory and affect in young adults. *Neuropsychological Rehabilitation*, 19, 223-243.
- 634 Sullivan, P. J., & Kent, A. (2003). Coaching efficacy as a predictor of leadership style in  
635 intercollegiate athletics. *Journal of Applied Sport Psychology*, 15, 1–11.
- 636 Thelwell, R. C., Lane, A. M., & Weston, N. J. W. (2007). Mood states, self-set goals, self-  
637 efficacy and performance in academic examinations. *Personality and Individual*  
638 *Differences*, 42, 573–583.
- 639 Thompson, E. R. (2007). Development and validation of an internationally reliable short-form  
640 of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural*  
641 *Psychology*, 38, 227–242.
- 642 van Yperen, N. W. (1998). Predicting stay/leave behavior among volleyball referees. *The*  
643 *Sport Psychologist*, 12, 427–439.
- 644 Wang, J. C. K., Liu, W. C., Chatzisarantis, N. L. D., & Lim, C. B. S. (2010). Influence of  
645 perceived motivational climate on achievement goals in physical education: a  
646 structural equation mixture modeling analysis. *Journal of Sport & Exercise*  
647 *Psychology*, 32, 324–338.

648 Watson, D., & Clark, L. A. (1994). Emotions, moods, traits, and temperaments: Conceptual  
649 distinctions and empirical findings. In P. Ekman, & R. J. Davidson (Eds.), *The nature*  
650 *of emotion: Fundamental questions* (pp. 89–93). New York: Oxford University Press.

651 Wolfson, S., & Neave, N. (2007). Coping under pressure: Cognitive strategies for maintaining  
652 confidence among soccer referees. *Journal of Sport Behavior*, *30*, 232–247.

653 Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of  
654 reliability. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 66–81.

655

656

657

658 Table 1. Descriptive statistics and estimated correlation matrix for latent variables.

	M (SD)	1	2	3	4	5	6	7
1 Task orientation	4.36 (.52)	-	.02	.36**	.30**	-.21**	-.22**	-.15*
2 Ego orientation	2.67 (.95)		-	.06	.04	.08	-.09	-.15*
3 Referee self-efficacy	4.35 (.45)			-	.41**	.01	.04	.12
4 Positive mood states	5.38 (.93)				-	.02	.10	.02
5 Mental preparation	2.91 (.96)					-	.25**	-.03
6 Physical preparation	3.98 (1.01)						-	.12
7 Refereeing years	11.96 (8.25)							-

659 Note: \*  $p < .05$ , \*\*  $p < .01$ .

660 Table 2. Testing group invariance of the four dimensions of the Referee Self-Efficacy Scale.

MLR-estimation											
Model	$\chi^2$	df	Model comparison	$\chi^2$ diff	$\Delta$ df	CFI	$\Delta$ CFI	RMSEA (90% CI)	$\Delta$ RMSEA	SRMR	$\Delta$ SRMR
Base model T1	74.34	59				.99		.04 (.000–.060)		.04	
<b>M0</b> Configural/Pattern	170.73	118				.95		.07 (.044–.090)		.06	
<b>M1</b> Weak									–.004	.08	.022
factorial/Metric/Loading	176.47	127	<b>M1-M0</b>	5.74 (n.s)*	9	.95	.003	.06 (.039–.085)			
<b>M2</b> Strong									–.004	.10	.039
factorial/Scalar/Intercept	188.94	136	<b>M2-M0</b>	18.220 (n.s.)*	18	.95	–.001	.06 (.040–.084)			
<b>M2</b> Strong										.10	.017
factorial/Scalar/Intercept	188.94	136	<b>M2-M1</b>	12.479 (n.s.)*	9	.95	–.004	.06 (.040–.084)			

661 Note:  $\chi^2$  = Satorra–Bentler scaled chi-square fit statistic (robust maximum likelihood estimation MLR); CFI = comparative fit index; RMSEA = root mean  
662 square error of approximation; SRMR = standardized root mean square residual; M0 = configural model (no invariance imposed), M1 = invariant factor  
663 loadings; M2 = invariant factor loadings and invariant intercepts. Nonsignificant chi-square diff between models as well as  $\Delta$ CFI  $\geq$  .01;  $\Delta$ RMSEA  $\geq$  .015;  
664  $\Delta$ SRMR  $\geq$  .015 signals invariance by the respective comparison of nested models (Chen, 2007; Cheung & Rensvold, 2002). The chi-square difference test is  
665 sensitive to trivial fluctuations and differences in the context of invariance testing.  $\chi^2$  could be statistically significant even though the absolute differences in



666 parameter estimates are of trivial magnitude and one should therefore rely on the alternative fit indices (Kline, 2011; Little, 2013). In the current case, this is  
667 not a problem with all  $\chi^2$  being nonsignificant (\*). Changes in CFI and RMSEA were less than the cutoffs given signal invariance, even though the chi-square  
668 diff test was statistically significant (Chen, 2007; Cheung & Rensvold, 2002). SRMR and  $\Delta$ SRMR values are included for transparency reasons only

669 Table 3. Bias-corrected confidence intervals and parameter estimates for total, total indirect,  
 670 specific indirect, and direct effects from predictors to mood states.

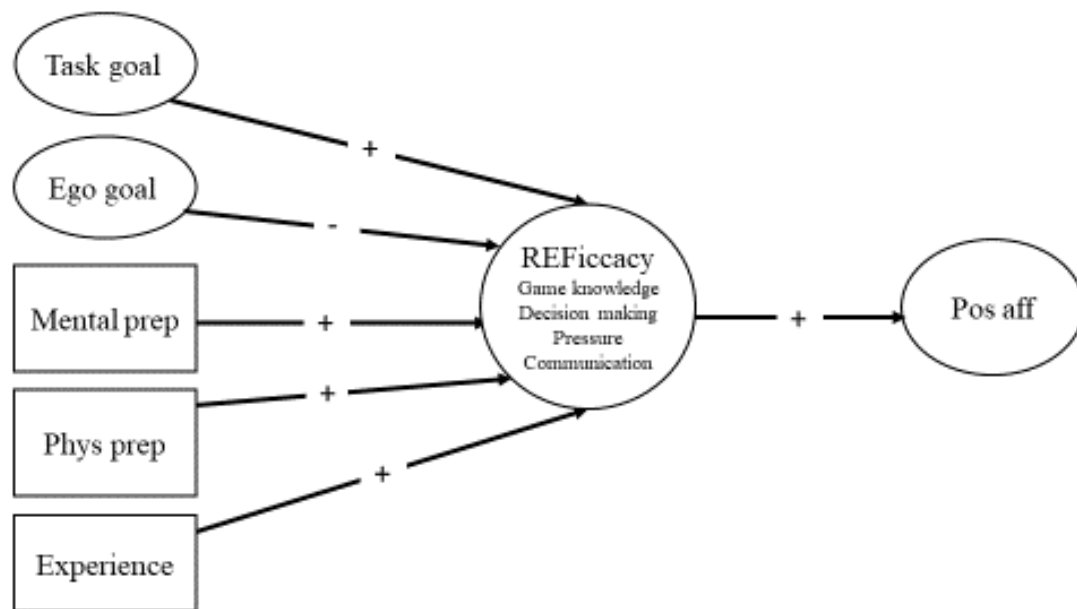
671

<b>Effects from</b>							
<b>predictors to</b>	<b>Lower</b>	<b>Lower</b>	<b>Lower</b>	<b>Standardized</b>	<b>Upper</b>	<b>Upper</b>	<b>Upper</b>
<b>positive</b>	<b>0.5%</b>	<b>2.5%</b>	<b>5%</b>	<b>estimate</b>	<b>5%</b>	<b>2.5%</b>	<b>0.5%</b>
<b>mood state</b>							
<hr/>							
Total							
<i>Task orient.</i>	0.081	0.152	0.185	<b>0.359***</b>	0.532	0.564	0.607
<i>Ego orient.</i>	-0.231	-0.192	-0.153	0.012 n.s	0.146	0.179	0.228
<hr/>							
Total indirect							
<i>Task orient.</i>	0.089	0.114	0.130	<b>0.223***</b>	0.372	0.394	0.444
<i>Ego orient.</i>	-0.124	-0.066	-0.048	0.019 n.s	0.082	0.094	0.110
<hr/>							
Specific indirect							
<i>Task orient.</i>	0.089	0.114	0.130	<b>0.223***</b>	0.372	0.394	0.444
<i>Ego orient.</i>	-0.124	-0.066	-0.048	0.019 n.s	0.082	0.094	0.110
<hr/>							
Direct							
<i>Task orient.</i>	-0.146	-0.110	-0.055	0.136 n.s	0.349	0.385	0.452
<i>Ego orient.</i>	-0.0305	-0.191	-0.155	-0.007 n.s	0.107	0.136	0.214
<hr/>							

<hr/>							
Total							
<i>Experience,</i>	-0.283	-0.209	-0.169	0.020 n.s	0.164	0.184	0.233
<i>Mental prep,</i>	-0.109	-0.065	-0.036	0.091 n.s	0.221	0.242	0.282
<i>Phys prep</i>	-0.042	0.039	0.064	<b>0.191 **</b>	0.312	0.329	0.349
<hr/>							
Total indirect							
<i>Experience,</i>	-0.004	0.016	0.027	0.085 n.s	0.177	0.202	0.232
<i>Mental prep,</i>	-0.011	-0.009	0.003	0.051 n.s.	0.123	0.143	0.173
<i>Phys prep</i>	-0.139	-0.088	-0.069	0.014 n.s	0.093	0.114	0.162
<hr/>							
Specific							
indirect							
<i>Experience,</i>	-0.004	0.016	0.027	0.085 n.s	0.177	0.202	0.232
<i>Mental prep,</i>	-0.028	-0.009	-0.003	0.051 n.s	0.123	0.143	0.173
<i>Phys prep</i>	-0.139	0.088	0.069	0.014 n.s	0.093	0.114	0.162
<hr/>							
Direct							
<i>Experience,</i>	-0.323	-0.268	-0.233	-0.065 n.s	0.090	0.115	0.196
<i>Mental prep,</i>	-0.168	-0.111	-0.085	0.040 n.s	0.165	0.200	0.221
<i>Phys prep</i>	0.012	0.027	0.056	<b>0.177 **</b>	0.293	0.312	0.365
<hr/>							

672 \*\*\*  $p < .001$ ; \*\*  $p < .005$ .

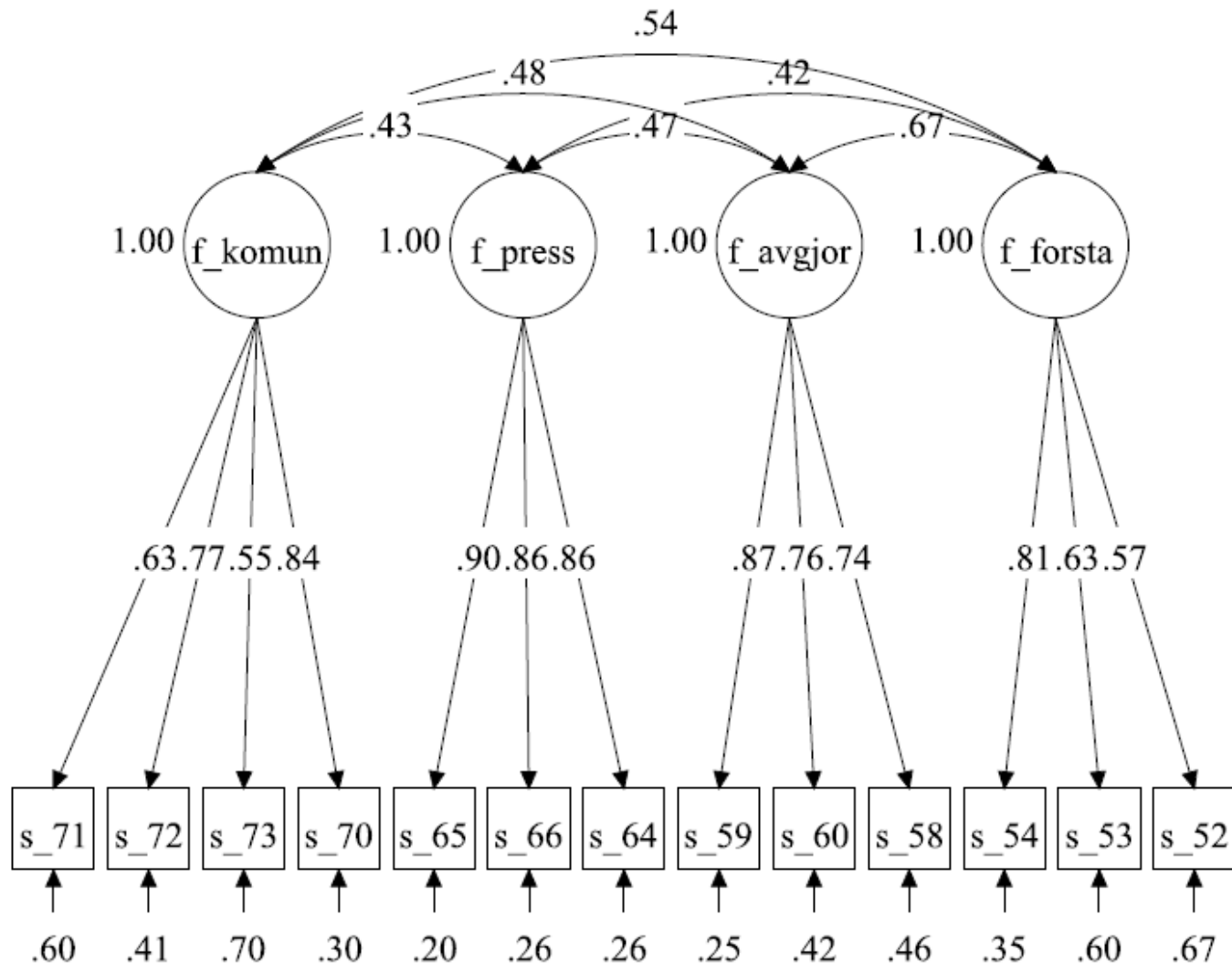
673



674

675 Figure 1. Visual illustration of the hypothesized structural model.

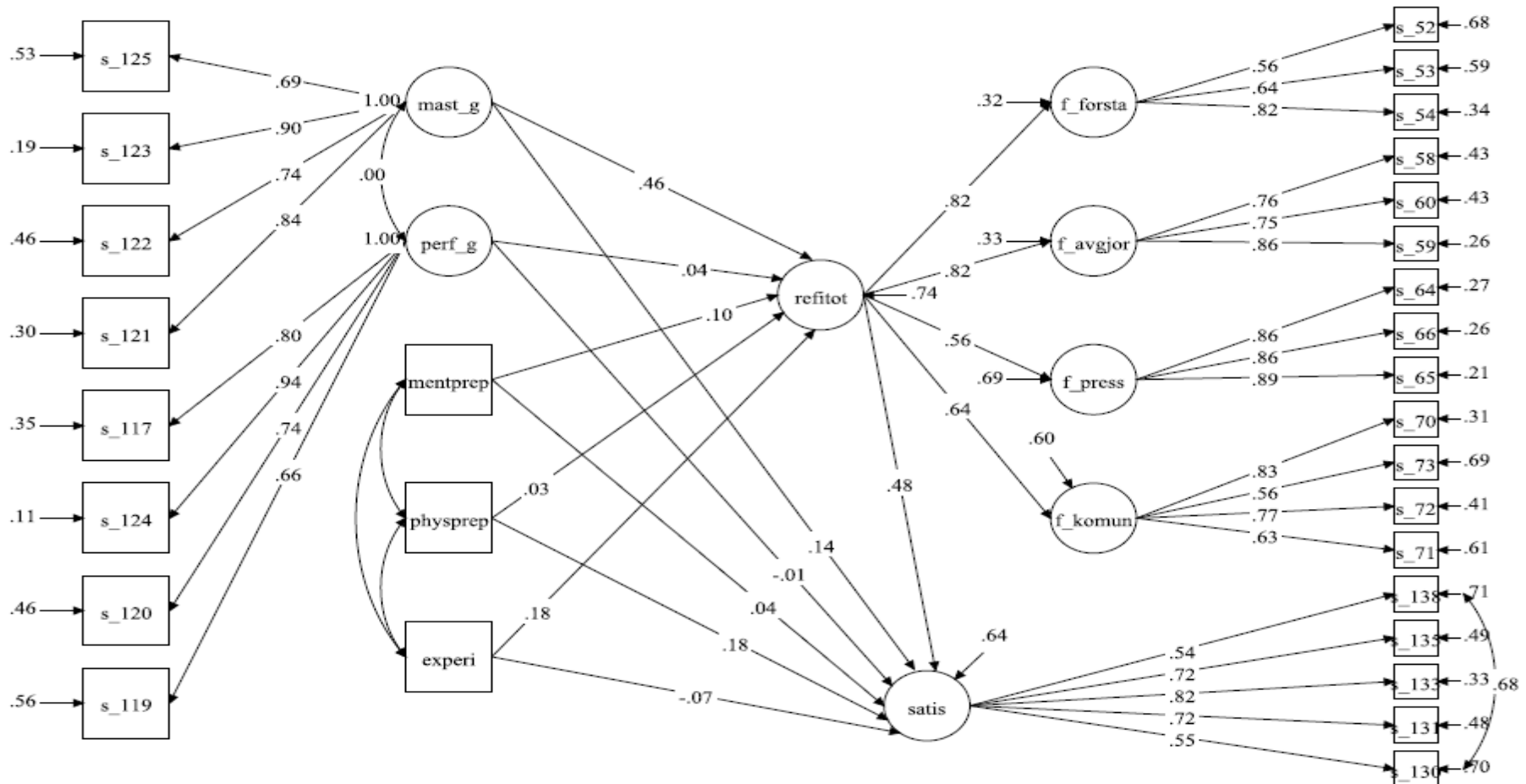
676



677

678 Figure 2. Graphical representation of standardized loading on the four dimensions of referee self-efficacy with respective dimension cross-  
 679 loadings, (f\_komun = Communication), (f\_press = Pressure), (f\_avgjor = Decision-making), (f\_forsta = Game knowledge).

680



681

682 Figure 3. Graphical representation of  $\beta$  coefficients (standardized) as estimated in the full model. All estimated paths shown (mast\_g = task goal  
683 orientation), (perf\_g = ego goal orientation), (mentprep = mental preparation), (physprep = physical preparation), experi = experience), (refitot =  
684 refficacy total – second order), (f\_forsta = game knowledge), (f\_avgjor = decision-making), (f\_press = pressure), (f\_komun = communication),  
685 (satis = referee positive affect).