



Development of the Referee Shared Mental Models Measure (RSMMM)

Jorge Sinval^{1,2,3,4*}, João Aragão e Pina⁵, João Sinval⁶, João Marôco²,
Catarina Marques Santos⁷, Sjir Uitdewilligen⁸, M. Travis Maynard⁹ and
Ana Margarida Passos¹

¹ Business Research Unit (BRU-IUL), Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal, ² William James Center for Research, ISPA – Instituto Universitário, Lisbon, Portugal, ³ Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal, ⁴ Faculty of Philosophy, Sciences and Languages of Ribeirão Preto, University of São Paulo, São Paulo, Brazil, ⁵ ISCTE Business School (IBS), Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal, ⁶ Faculty of Economics, University of Porto, Porto, Portugal, ⁷ Organisation, Strategy, and Entrepreneurship, Maastricht University School of Business and Economics, Maastricht University, Maastricht, Netherlands, ⁸ Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands, ⁹ Department of Management, College of Business, Colorado State University, Fort Collins, CO, United States

OPEN ACCESS

Edited by:

Yair Gallily,
Interdisciplinary Center Herzliya, Israel

Reviewed by:

Geoffrey Schweizer,
Heidelberg University, Germany
Asaf Blatt,
Florida State University, United States

*Correspondence:

Jorge Sinval
jorgesinval@gmail.com

Specialty section:

This article was submitted to
Movement Science and Sport
Psychology,
a section of the journal
Frontiers in Psychology

Received: 08 April 2020

Accepted: 28 August 2020

Published: 19 October 2020

Citation:

Sinval J, Aragão e Pina J, Sinval J,
Marôco J, Santos CM,
Uitdewilligen S, Maynard MT and
Passos AM (2020) Development
of the Referee Shared Mental Models
Measure (RSMMM).
Front. Psychol. 11:550271.
doi: 10.3389/fpsyg.2020.550271

The concept of shared mental models refers to the shared understanding among team members about how they should behave in different situations. This article aimed to develop a new shared mental model measure, specifically designed for the refereeing context. A cross-sectional study was conducted with three samples: national and regional football referees ($n = 133$), national football referees and assistant referees and national futsal referees ($n = 277$), and national futsal referees ($n = 60$). The proposed version of the Referee Shared Mental Models Measure (RSMMM) has 13 items that are reflected on a single factor structure. The RSMMM presented good validity evidence both based on the internal structure and based on relations to other variables (presenting positive associations with team work engagement, team adaptive performance, and team effectiveness). Such promising psychometric properties point to an optimistic outlook regarding its use to measure shared mental models in futsal and football referee teams.

Keywords: shared mental models, referees, psychometrics, football, futsal

INTRODUCTION

Shared mental models have been examined in numerous contexts (Resick et al., 2010a; Santos et al., 2015a, 2016; Tomás et al., 2017). However, one context where the role of shared mental models has received relatively little attention is sports referee teams (Filho and Tenenbaum, 2012; Aragão e Pina et al., 2018). This gap is interesting as football and futsal refereeing teams are highly interdependent in conducting their team tasks. Namely, they must coordinate several tasks before, during, and after the game (Samuel, 2015; Hancock et al., 2018); share technical and tactical knowledge to gain an adequate understanding of the task and match game needs (Mascarenhas et al., 2006; Mallo et al., 2012; McEwan and Beauchamp, 2014; Hancock et al., 2018); anticipate and adapt to the needs and actions of other members as well as changing task demands (Cannon-Bowers et al., 1993; Hancock et al., 2018); define a communication protocol to facilitate the team decision-making process (Cunningham et al., 2014; Samuel, 2015; Hancock et al., 2018) and engage in continuous learning together with the other team members (Collina, 2004;

Cunningham et al., 2014; McEwan and Beauchamp, 2014). Nevertheless, to date, only one study has addressed the entire football refereeing team (see, Boyer et al., 2015), and currently there is no shared mental model scale adapted specifically for football and futsal refereeing teams.

Shared mental models refer to an organized and common understanding among team members regarding the essential aspects of work (Klimoski and Mohammed, 1994; Mohammed et al., 2010). Team members hold multiple mental models, about different domains, while they work on a task (Cannon-Bowers et al., 1993; Klimoski and Mohammed, 1994). Cannon-Bowers et al. (1993) proposed four types of models, namely, the equipment model, task model, team interaction model, and team model. The equipment model refers to knowledge about the equipment functioning, technology, and tools with which the team members interact. The task model refers to knowledge about the task procedures, task strategies, contingency plans, and environmental constraints. The team interaction model refers to knowledge about the roles and responsibilities of team members, role interdependencies, interaction patterns, and communication channels. The team model regards knowledge about task-relevant attributes of team members, such as knowledge, skills, abilities, preferences, and tendencies (Cannon-Bowers et al., 1993; Mathieu et al., 2000; Mohammed et al., 2010).

Nevertheless, Mathieu et al. (2000) merged those four models into two domains—task mental models (comprising the equipment and task models) and team mental models (comprising the team interaction and team models). Accordingly, over the years, researchers have most commonly analyzed task mental models and team mental models (Mathieu et al., 2000; Lim and Klein, 2006; Santos and Passos, 2013). Task mental models refer to a similar understanding among team members about work objectives, team resources, task procedures and practices, and task duties. Team mental models refer to a similar understanding among team members about interpersonal interaction, team members' roles and responsibilities, and role interdependencies (Mathieu et al., 2000; Mohammed et al., 2010). In this article, this distinction between task and team mental models was made, as it has received extensive support from empirical studies (Mathieu et al., 2000; Lim and Klein, 2006; Santos and Passos, 2013; Santos et al., 2015a). Another dimension recently proposed by Randall et al. (2011) was also considered, namely, strategy mental models, which refer to “an understanding of strategic priorities, the trade-offs, and relationships among strategic alternatives, and the implications of strategic decisions” (p. 527).

Drawing on research on shared mental models in organizational teams, it is arguable that referee teams that develop shared mental models can anticipate each other's needs and adapt their behaviors to fit tightly to task demands (Cannon-Bowers et al., 1993; Mohammed et al., 2010). When the referee team members develop a common understanding on the tools and technologies they interact with, such as audio communication system, electronic flags, or video assistant referees (VARs), on the task procedures and strategies, as well as on the strategic priorities, they similarly, interpret the cues and make effective and quick decisions on the field (Kellermanns

et al., 2008; Randall et al., 2011). Furthermore, when referee team members develop a shared understanding regarding each other's roles and responsibilities, as well as on the knowledge, skills, and abilities of each other, this enables them to effectively communicate and work in a coordinated way, which allows them to adapt to unexpected events, and perform their tasks successfully (Mathieu et al., 2000; Muponde and Muchemwa, 2011; Santos and Passos, 2013; Boyer et al., 2015; Diotaiuti et al., 2017; Uitdewilligen et al., 2018). Some psychometric instruments to measure mental models have already been proposed: the Shared Mental Model Scale (SMMS; Santos et al., 2015a) also with a shorter unidimensional version (Santos et al., 2015b), the Team-Related Knowledge Measurement Instrument (TeamKMI; Johnson et al., 2007), and the Perceived Mutual Understanding (PMU) scale (Burtscher and Oostlander, 2019). Yet, none of the existing instruments has been tested among referees. Therefore, by developing a shared mental model measure for referees, scholars could begin to examine shared mental models within the context of referee teams and allow further examination of their antecedents and outcomes (Aragão e Pina et al., 2021).

Research Hypotheses

One of the most desirable psychometric properties of an instrument is its dimensionality stability across different samples (Nunnally and Bernstein, 1994). As so, if an instrument maintains its dimensionality with a good fit to the different sample datasets, one can assume that the items and factors proposed are adequate to measure the desired construct/s over different groups of individuals. It is particularly important to have dimensionality evidence when independent samples of the same population are analyzed with the same instrument (Marôco, 2014). Because the perceptions of mental models can vary from context to context, it is expected that the dimensionality of the proposed measure may have a different number of factors in comparison with the instrument in which this measure was initially based (i.e., three factors). However, it is assumed that the proposed dimensionality will present a good fit to the data (H1; i.e., three different samples). It is expected that the proposed dimensionality on the first sample data will be reproduced with a good fit in the two other independent samples. If such dimensionality (i.e., factor structure) holds in all the samples, there will be promising evidence of dimensionality (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014).

Another important psychometric property is the reliability evidence, which can be assessed through internal consistency (Nunnally and Bernstein, 1994). Previous instruments measuring mental models reported acceptable values of internal consistency, as the PMU (Burtscher and Oostlander, 2019) with $\alpha = 0.83$ and $\omega = 0.83$. The TeamKMI reported globally satisfactory internal consistency estimate values (Johnson et al., 2007). Moreover, the SMMS reported satisfactory internal consistency values (Santos et al., 2015b). The second hypothesis (H2) presumes the Referee Shared Mental Models Measure (RSMMM) shows good evidence of the scores' reliability, more specifically in terms of internal consistency (Nunnally and Bernstein, 1994). Such estimates should be desirably high (i.e., ≥ 0.70 ; Iacobucci and Duhachek, 2003). Adequate internal consistency values will indicate that the

items are measuring the same construct, measuring the construct consistently (McDonald, 1999).

The third hypothesis (H3) assumes that the RSMMM will present measurement invariance among referees from different sports. Such property is essential to directly compare groups within the same instrument (Davidov et al., 2014). Measurement invariance has been tested before among referees of different types of sports in a measure of self-efficacy (Myers et al., 2012), also among referees and assistant referees in football (Brandão et al., 2014) and also between elite and non-elite football referees (Johansen et al., 2018).

The extent of the relations of an instrument's scores with external variables constitutes a critical source of validity. This particular source of validity is denominated as validity evidence based on the relation to other variables (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). As such, some related constructs are expected to be associated with shared mental models. Team work engagement is an affective-motivational construct that is expected to be positively related to shared mental models because higher team work engagement means higher team enthusiasm and energy (Costa et al., 2014b). The mental models construct is a cognitive one, which is expected to enhance team members' anticipation of actions and communication, conducting to positive feelings. This is also true regarding team effectiveness, because a higher common understanding of the way the team works will allow predicting behavior patterns that will likely increase the effectiveness of the team (Marks et al., 2002; Mathieu et al., 2009; DeChurch and Mesmer-Magnus, 2010). As such, a positive association between mental models and team effectiveness is expected to be observed. Associated to a higher level of shared mental models is expected to be a higher perception of team adaptive performance. Team members with shared cognitive representations regarding team function will predict the other team members move straightforwardly and consequently improving the ability to react and adjust when necessary (Pulakos et al., 2006). As so, the fourth hypothesis (H4) establishes that the suggested shared mental models measure will present validity evidence based on the relation to other variables, namely, nomological evidence in convergent terms with team work engagement, team effectiveness, and team adaptive performance.

MATERIALS AND METHODS

Sample

This article uses data from three different studies with non-probabilistic convenience samples where data were collected within a cross-sectional survey at the individual level; all participants are Portuguese football referees or assistant referees or futsal referees. Depending on the tournament, football referee teams range from three to seven members, whereas futsal referee teams range from three to four members. Team members within each team are usually the same, with some exceptions (i.e., injuries, not being considered apt in physical or written examinations). However, in the case of the top-class football

national referees (i.e., C1 class), rotation between team members is more frequent.

Study I

The sample I data ($n = 133$) were constituted by national football referees ($n = 67$), with $\text{mean}_{\text{age}} = 30.02$ ($SD_{\text{age}} = 3.16$) years, $\text{mean}_{\text{experience}} = 12.29$ ($SD_{\text{experience}} = 3.15$) years; and football regional referees ($n = 66$) $\text{mean}_{\text{age}} = 26.37$ ($SD_{\text{age}} = 3.21$) years, $\text{mean}_{\text{experience}} = 8.89$ ($SD_{\text{experience}} = 3.15$). All referees completed the questionnaires.

Study II

The sample II data ($n = 277$) were composed by football national referees ($n = 135$) with $\text{mean}_{\text{experience}} = 12.44$ ($SD_{\text{experience}} = 5.06$) years, futsal national referees ($n = 117$) with $\text{mean}_{\text{experience}} = 11.71$ ($SD_{\text{experience}} = 4.88$) years, and football national assistant referees ($n = 25$) with $\text{mean}_{\text{experience}} = 18.44$ ($SD_{\text{experience}} = 4.16$) years. All referees completed the questionnaires.

Study III

The sample III data ($n = 60$) had only futsal national referees with $\text{mean}_{\text{age}} = 34.54$ ($SD_{\text{age}} = 5.52$) years, $\text{mean}_{\text{experience in the current team}} = 2.72$ ($SD_{\text{experience in the current team}} = 3.13$) years.

Measures

All the self-report measures were collected at the individual level, reflecting the perceptions of the subject about the team.

Shared Mental Models

Shared mental models refer to a multidimensional construct. In this article, three dimensions were considered, namely, task mental models, team mental models, and strategy mental models. Referees must develop a similar understanding of the task procedures, practices, and strategies to make decisions, likely scenarios and contingencies (Cannon-Bowers et al., 1993; Mathieu et al., 2000; Aragão e Pina et al., 2019), and contingency plans (Mohammed et al., 2010). Regarding the task mental models, referee team members must develop a similar understanding about the equipment functioning and equipment limitations (Cannon-Bowers et al., 1993), as well as about the technology and tools with which they interact to make decisions (Mathieu et al., 2000; Mohammed et al., 2010). Example of such equipment are the audio communication system, the Video Assistant Referee (VAR), or the goal-line technology (GLT). Referee team members must also develop a similar understanding of the environmental constraints and the aspects of the task environment that affect team performance (Cannon-Bowers et al., 1993).

Concerning the team mental models, referee team members must develop a similar understanding about the roles and responsibilities of each team member, the role interdependencies, and about interaction patterns and communication channels and patterns (Cannon-Bowers et al., 1993; Mathieu et al., 2000; Mohammed et al., 2010). Besides, they must develop a similar understanding about the knowledge, skills, and abilities of each team member and about the team members' preferences to

make decisions during the games (Cannon-Bowers et al., 1993; Mathieu et al., 2000; Aragão e Pina et al., 2019). Regarding the strategy mental models, referee team members must develop a similar understanding of the strategic priorities, as well as the implications of strategic decisions (Randall et al., 2011).

This measure was named as RSMMM (Table 1). Based on relevant literature on shared mental models, namely, in other instruments (Santos et al., 2015a,b), an initial pool of 13 items was developed across the three dimensions: task (e.g., “In my team, the team members have a similar understanding about the technology and tools needed to make decisions during a game”); team (e.g., “In my team, the team members have a similar understanding about the knowledge, skills, and abilities of each other”); and strategy (e.g., “In my team, the team members have a similar understanding about the strategic priorities of the game”). It was ensured that the shared mental models’ items, in particular, the items of the task dimension, were specific to the context of referee teams (Cannon-Bowers and Salas, 2001) by stating that team members have the knowledge needed to make decisions during a game or by providing examples related to the referees’ responsibilities. For instance, “In my team, the team members have a similar understanding about resources needed to make decisions during a game” and “In my team, the team members have a similar understanding about the tasks each team member has to do (e.g., train during the week, prepare the game properly, employ an exemplary behavior, make a difficult decision).” Each item was scored on a seven-point Likert scale (1 = “Totally disagree”, 2 = “Strongly disagree”, 3 = “Disagree”, 4 = “Neither agree, nor disagree”, 5 = “Agree”, 6 = “Strongly agree”, 7 = “Totally agree”).

Team Work Engagement

Team work engagement is defined as an emergent state that develops from team members’ interactions and that cannot be found in individuals being exclusive to teams (Costa et al., 2016). To measure team work engagement, the Team Work Engagement Scale was used (Costa et al., 2014a). This instrument consists of nine items measured in a seven-point Likert scale (1 = “Totally disagree”, 7 = “Totally agree”). Team Work Engagement is seen as a second-order factor (as the individual measure; Sinval et al., 2018b,a) that comprises three first-order dimensions (i.e., vigor, dedication, and absorption). This instrument showed good validity evidence based on the internal structure in previous studies, namely, in terms of reliability, having Cronbach’s α of 0.85 a 0.97 for the vigor factor, 0.88 and 0.95 for the dedication factor; and 0.83 and 0.95 for the absorption factor (Costa et al., 2014a). Examples of items are as follows: “At our work, we feel bursting with energy” (vigor), “We are enthusiastic about our job” (dedication), and “We feel happy when we are working intensely” (absorption).

Team Adaptive Performance

Team adaptive performance is defined as an emergent state that occurs as a consequence of the adaptation process, in which individuals and teams cope with the demands of the context (Maynard et al., 2015). The Team Adaptive Performance Scale was used to measure team adaptive performance (Marques-Quinteiro et al., 2015). This instrument has eight items that were

answered using a seven-point Likert scale (1 = “Totally disagree”, 7 = “Totally agree”). This instrument assumes that team adaptive performance is a second-order latent variable with two first-order latent factors (factor I: problem-solving-oriented factor, six items; and factor II: learning work tasks, technologies, and procedures factor, two items). Examples of items are as follows: “We use creative ideas to manage incoming events” (problem-solving-oriented), and “We remain calm and behave positively under highly stressful events” (learning work tasks, technologies, and procedures).

Team Effectiveness

Team effectiveness is conceived in three criteria: team performance, quality of group experience, and team viability (Aubé and Rousseau, 2005). Team performance has been seen in the function of the assigned team goals (Hackman, 1987). The quality of group experience is defined as the positiveness of the social climate in the team (McGrath, 1991). The team viability consists in the capacity of the team to adapt to external and internal changes and also to the likelihood of team members continuing to work together (Hackman, 1987). The team effectiveness dimension was measured using the Portuguese version of the Scale of Effectiveness of Teams (3Es; Vicente et al., 2014). This instrument has three first-order factors (team performance, quality of group experience, and team viability), which are explained by a hierarchical structure (second-order factor) called effectiveness. The items were scored with a Likert scale from 1 = “Totally disagree” to 7 = “Totally agree”. In the original version with the Canadian sample (Aubé and Rousseau, 2005), the authors studied the internal consistency, and good Cronbach’s α values were evidenced ($\alpha_{\text{team performance}} = 0.82$, $\alpha_{\text{team viability}} = 0.84$, $\alpha_{\text{quality of group experience}} = 0.96$). Examples of items are as follows: “The members of this team attain their assigned performance goals” (team performance); “The social climate in our work team is good” (quality of group experience); and “Team members adjust to the changes that happen in their work environment” (team viability).

Procedure

For samples I and II studies, the institutional review board, and the National Referees’ Committee approved the study. National referees were attending a seminar, and regional referees were attending a promotion seminar compulsory for those wishing to be considered for promotion to the national level. Data were collected at the beginning of each seminar, after providing a brief explanation of the nature of the investigation. The institutional approval of the Portuguese Football Federation was obtained for sample III’s study. All referees participated voluntarily, and written or electronic informed consent was obtained from all participants, and confidentiality for their responses was ensured.

Data Analysis

All statistical analyses were performed with R (R Core Team, 2020) through RStudio (RStudio Team, 2020). The descriptive statistics were obtained with the *skimr* package (McNamara et al., 2018); the coefficient of variation (CV) was calculated through the *sjstats* package (Lüdtke, 2019), and the standard error of the mean (SEM) was estimated by the *plotrix* package

TABLE 1 | Referee Shared Mental Models Measure (RSMMM) items.

Item	English version of RSMMM						Portuguese (Portugal) version of RSMMM							
	Totally disagree	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree	Totally agree	Discordo totalmente	Discordo muito	Discordo em parte	Nem concordo, nem discordo	Concordo em parte	Concordo muito	Concordo totalmente
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Shared Mental Models							Modelos Mentais Partilhados							
1	In my team, members have a similar understanding of the resources that are needed to make decisions during a game.						Na minha equipa, os membros têm um entendimento semelhante sobre os recursos que são necessários para tomar as decisões durante um jogo.							
2	In my team, members have a similar understanding of the technology and tools needed to make decisions during a game.						Na minha equipa, os membros têm um entendimento semelhante sobre a tecnologia e as ferramentas necessárias para tomar as decisões durante um jogo.							
3	In my team, members have a similar understanding of the procedures and practices needed to make decisions during a game.						Na minha equipa, os membros têm um entendimento semelhante sobre os procedimentos e práticas necessários para tomar as decisões durante um jogo.							
4	In my team, even when we are confronted with incidents or problems related to our performance, we have a similar understanding of how to perform our tasks.						Na minha equipa, mesmo quando somos confrontados com incidentes ou problemas relacionados com a nossa atuação, temos um entendimento semelhante sobre como realizar as nossas tarefas.							
5	In my team, members have a similar understanding of what they must do (e.g., train during the week, properly prepare the game, adopt exemplary behavior, make a difficult decision).						Na minha equipa, os membros têm um entendimento semelhante em relação ao que cada um tem que fazer (ex.: treinar durante a semana, preparar adequadamente o jogo, adotar um comportamento exemplar, tomar uma decisão difícil).							
6	In my team, members have a similar understanding of how their roles are related.						Na minha equipa, os membros têm um entendimento semelhante sobre a forma como os papéis de cada um estão relacionados.							
7	In my team, members have a similar understanding of how to interact with each other.						Na minha equipa, os membros têm um entendimento semelhante sobre a forma como interagir uns com os outros.							
8	In my team, members have a similar understanding of what the best methods are for communicating with each other.						Na minha equipa, os membros têm um entendimento semelhante sobre quais os melhores métodos para comunicar uns com os outros.							
9	In my team, members have a similar understanding of each other's knowledge, skills and abilities.						Na minha equipa, os membros têm um entendimento semelhante em relação aos conhecimentos, competências e capacidades de cada um.							
10	In my team, members have a similar understanding of each other's preferences, which are relevant to making decisions during a game.						Na minha equipa, os membros têm um entendimento semelhante em relação às preferências de cada um, que são relevantes para tomar as decisões durante um jogo.							
11	In my team, members have a similar understanding of the game's strategic priorities.						Na minha equipa, os membros têm um entendimento semelhante em relação às prioridades estratégicas do jogo.							
12	In my team, members have a similar understanding of the implications of the strategic decisions that are made.						Na minha equipa, os membros têm um entendimento semelhante em relação às implicações das decisões estratégicas que são tomadas.							
13	In my team, members have a similar understanding of which aspects of the game are most important to team performance.						Na minha equipa, os membros têm um entendimento semelhante sobre quais os aspetos do jogo que são mais importantes para o desempenho da equipa.							

(Lemon, 2006). The mode was calculated with the *DescTools* package (Signorell et al., 2019). Severe univariate normality violations were considered for absolute values of $sk > 3$ and $ku > 7$ (Finney and DiStefano, 2013; Marôco, 2014).

Regarding the exploratory factor analysis (EFA), the Kaiser–Meyer–Olkin (KMO) coefficient was used as a measure of sampling adequacy (Kaiser and Rice, 1974). The Bartlett test (Bartlett, 1951) was chosen to test if the correlation matrix was factorable (i.e., the correlations differ from 0) (Revelle, 2019). KMO values > 0.8 and Bartlett test significance ≤ 0.05 , indicating adequate sampling (Marôco, 2018). The number of factors was determined through the comparison data (CD) approach, as suggested by Ruscio and Roche (2012), which stated that this technique outperforms Parallel Analysis. CD is a variant of Parallel Analysis that reproduces the correlation matrix rather than generating random data (Courtney, 2012). The extraction of the factors was performed using the principal components analysis with a weighted least-squares factoring method on the polychoric correlation (ρ_{PC}) matrix with oblimin rotation and weighted least-squares factoring. The cutoff for items' loadings was 0.40. The CD analysis was conducted using the *RGenData* package (Ruscio, 2018). The Bartlett test, the KMO coefficient, factors' extraction and the ρ_{PC} were produced using the *psych* package (Revelle, 2019). As goodness-of-fit index for the EFA, the RMSR (root mean square of the residual) was used.

Confirmatory factor analysis (CFA) was conducted with the *lavaan* package (Rosseel, 2012) using the weighted least-squares means and variances (WLSMV) estimation method for ordinal variables (Muthén, 1983). As goodness-of-fit indices, the *TLI* (Tucker–Lewis index), *NFI* (normed fit index), χ^2/df (ratio chi-square and degrees of freedom), *CFI* (comparative fit index), the *RMSEA*, and the *SRMR* (standardized root mean square residual) were used. For values of $\chi^2/df < 5$, values of *CFI*, *NFI*, and *TLI* > 0.95 ; values of *SRMR* < 0.08 ; and *RMSEA* < 0.08 , the fit of the model was considered good (Hoyle, 1995; Boomsma, 2000; McDonald and Ho, 2002; Byrne, 2010; Marôco, 2014).

To analyze the convergent validity evidence, the average variance extracted (AVE) was estimated (Fornell and Larcker, 1981). For values of *AVE* ≥ 0.5 (Hair et al., 2019), adequate convergent validity evidence was assumed.

The discriminant validity evidence was tested to verify whether the items that represent a dimension were strongly correlated with other dimensions. To assess such evidence, the Fornell and Larcker's (1981) approach was used: for two factors, x and y , if AVE_x and $AVE_y \geq \rho^2_{xy}$ (squared correlation between the factors x and y), adequate discriminant validity evidence is assumed.

The reliability of the scores was assessed with estimates of internal consistency, α (Cronbach, 1951), and ω (Raykov, 2001), using the *semTools* package (Jorgensen et al., 2019), where higher values were indicative of better internal consistency results. The α coefficient was calculated using the polychoric correlation matrix. The second-order reliability estimates were as follows: the proportion of the second-order factor explaining the total score (ω_{L1}), the proportion of variance explained by second-order factor after partialing the uniqueness of the first-order factor ($\omega_{\text{partial}L1}$), and the variance of the first-order factors explained by the second-order factor (ω_{L2}). Such reliability estimates were

obtained with the *semTools* package (Jorgensen et al., 2019). The confidence intervals (CIs) for the internal consistency estimates were obtained through the *userfriendlyscience* package (Peters, 2018) and the *boot* package (Davison and Hinkley, 1997; Canty and Ripley, 2020) using 1,000 bootstrap replicates. The bias-corrected and accelerated method was used, which tend to provide better coverage in non-normal sampling distributions (Efron and Tibshirani, 1994; Carpenter and Bithell, 2000).

The measurement invariance was assessed and verified using the *lavaan* package (Rosseel, 2012) and the *semTools* package (Jorgensen et al., 2019). A group of five models was compared: (a) configural invariance; (b) first-order factor loadings; (c) thresholds/intercepts of measured variables (depending on if the items are considered or not as categorical); (d) residual variances of observed variables; and (e) latent means. The latent variable means were compared, and Cohen d was used as the effect size (Cohen, 1988).

RESULTS

The presented results refer to three different studies with three different samples. First, the three samples were merged, and the instrument's expected dimensionality analyzed. Subsequently, the samples were individually analyzed to obtain different validity evidence from each of them.

Merge Samples

Validity Evidence Based on the Internal Structure

The dimensionality, reliability of scores, and measurement invariance of the instrument will be tested to verify the robustness of this source of validity evidence.

Items' distributional properties








































As Table 2 shows, none of the items for samples I and II presented severe problems of univariate normality because all of them presented $|sk| < 3$ and $|ku| < 7$ (Finney and DiStefano, 2013; Marôco, 2014). However, some of sample III items' absolute values of ku were greater than 7 (i.e., items 1, 2, 3, 6, and 9; Table 2). Item 5 was the one that presented more variability (i.e., CV) in the answers in all samples.

Following the recommendations of Finney et al. (2016) with categorical items with six or more points, both maximum likelihood estimation with robust (Huber–White) standard errors (MLR) and diagonal weighted least-squares methods (as the WLSMV estimator) can be used. The WLSMV estimator was chosen because it does not require multivariate normality as an assumption. To analyze the validity evidence based on the internal structure of the new measure, several steps were carried (i.e., dimensionality, reliability, and measurement invariance).

Dimensionality

To test the expected three first-order factors of the RSMMM, a CFA was conducted with all the available data from the three collected samples. The CFA is the most appropriate technique to use when there is a definite theory regarding the latent factors and their relationships to the indicators, that is, dimensionality (Brown, 2015; Finch and French, 2015). Items 1 to 4 were used as indicators of the task factor; items 5 to 10 were expected to

TABLE 2 | Items' distributional properties.

Study I items' descriptive statistics (<i>n</i> = 133)											
Item	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Mdn</i>	<i>Max</i>	Histogram	<i>Mode</i>	<i>SEM</i>	<i>CV</i>	<i>sk</i>	<i>ku</i>
Item 1	5.71	0.93	2	6	7		6.00	0.08	0.16	−1.52	3.07
Item 2	5.85	0.82	3	6	7		6.00	0.07	0.14	−0.70	0.59
Item 3	5.89	0.94	2	6	7		6.00	0.08	0.16	−1.42	2.98
Item 4	5.67	0.92	2	6	7		6.00	0.08	0.16	−0.94	1.76
Item 5	5.43	1.27	1	6	7		6.00	0.11	0.23	−1.03	0.87
Item 6	5.93	0.84	2	6	7		6.00	0.07	0.14	−1.18	3.35
Item 7	6.02	0.84	3	6	7		6.00	0.07	0.14	−0.88	1.23
Item 8	5.73	0.86	3	6	7		6.00	0.07	0.15	−0.65	0.80
Item 9	5.71	0.96	2	6	7		6.00	0.08	0.17	−1.10	2.22
Item 10	5.59	0.93	2	6	7		6.00	0.08	0.17	−0.89	1.35
Item 11	5.71	0.89	2	6	7		6.00	0.08	0.16	−0.91	1.89
Item 12	5.70	0.94	1	6	7		6.00	0.08	0.16	−1.35	4.07
Item 13	5.95	0.80	3	6	7		6.00	0.07	0.13	−0.72	0.85
Study II items' descriptive statistics (<i>n</i> = 277)											
Item	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Mdn</i>	<i>Max</i>	Histogram	<i>Mode</i>	<i>SEM</i>	<i>CV</i>	<i>sk</i>	<i>ku</i>
Item 1	6.02	0.87	2	6	7		6.00	0.05	0.14	−1.32	3.04
Item 2	6.03	0.88	2	6	7		6.00	0.05	0.15	−1.34	3.22
Item 3	6.10	0.81	2	6	7		6.00	0.05	0.13	−1.51	4.39
Item 4	5.93	0.83	3	6	7		6.00	0.05	0.14	−0.97	1.67
Item 5	5.81	1.14	1	6	7		6.00	0.07	0.20	−1.54	3.09
Item 6	6.09	0.86	2	6	7		6.00	0.05	0.14	−1.31	2.95
Item 7	6.12	0.92	1	6	7		6.00	0.06	0.15	−1.77	5.40
Item 8	6.03	0.93	2	6	7		6.00	0.06	0.15	−1.47	3.66
Item 9	6.08	0.90	2	6	7		6.00	0.05	0.15	−1.50	3.46
Item 10	5.98	0.90	2	6	7		6.00	0.05	0.15	−1.13	2.08
Item 11	6.04	0.85	2	6	7		6.00	0.05	0.14	−1.43	4.20
Item 12	6.01	0.88	2	6	7		6.00	0.05	0.15	−1.46	3.82
Item 13	6.20	0.86	1	6	7		6.00	0.05	0.14	−1.77	6.38
Study III items' descriptive statistics (<i>n</i> = 60)											
Item	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Mdn</i>	<i>Max</i>	Histogram	<i>Mode</i>	<i>SEM</i>	<i>CV</i>	<i>sk</i>	<i>ku</i>
Item 1	6.10	1.16	1	6	7		6.00	0.15	0.19	−2.74	9.41
Item 2	6.26	1.09	1	6	7		7.00	0.14	0.17	−2.85	10.01
Item 3	6.02	1.07	1	6	7		6.00	0.14	0.18	−2.43	8.42
Item 4	5.67	1.19	1	6	7		6.00	0.15	0.21	−1.85	5.25
Item 5	5.61	1.31	1	6	7		6.00	0.17	0.23	−1.42	2.85
Item 6	6.13	1.02	1	6	7		6.00	0.13	0.17	−2.27	8.31
Item 7	5.92	1.22	1	6	7		6.00	0.16	0.21	−2.20	6.34
Item 8	5.84	1.16	1	6	7		6.00	0.15	0.20	−1.78	4.63
Item 9	6.05	1.22	1	6	7		6.00	0.16	0.20	−2.38	7.19
Item 10	5.62	1.30	1	6	7		6.00	0.17	0.23	−2.04	5.19
Item 11	5.70	1.23	1	6	7		6.00	0.16	0.21	−1.87	4.82
Item 12	5.75	1.30	1	6	7		6.00	0.17	0.23	−1.74	3.82
Item 13	5.83	1.17	1	6	7		6.00	0.15	0.20	−1.76	4.49

be indicators of the dimension team, and items 11 to 13 were developed as potential indicators of the latent variable strategy.

The goodness-of-fit indices were indicative of good fit to the data ($\chi^2(62) = 184.686$, $n = 526$, $\chi^2/df = 2.979$, $CFI = 0.999$, $NFI = 0.998$, $TLI = 0.998$, $SRMR = 0.031$, $RMSEA = 0.061$, $P(rmse \leq 0.05) = 0.032$, 90% CI [0.051; 0.072]). The convergent validity evidence based on the internal structure was good

($AVE_{task} = 0.78$, $AVE_{team} = 0.70$, $AVE_{strategy} = 0.85$). However, the discriminant validity evidence based on the internal structure was not satisfactory, because the latent correlations between the factors were too high ($r_{task \times team} = 0.919$, $p < 0.001$; $r_{task \times strategy} = 0.870$, $p < 0.001$; $r_{team \times strategy} = 0.915$, $p < 0.001$). Comparing the values of the AVE of each pair of factors with their squared correlation value, only one

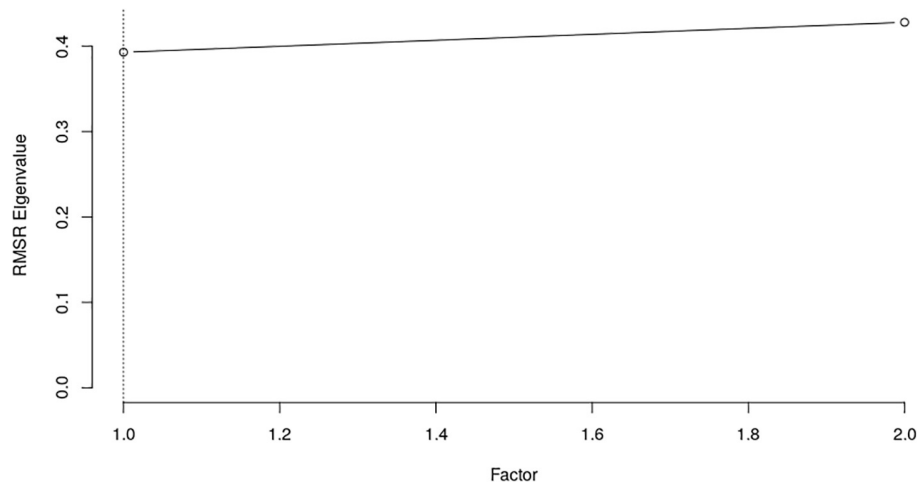


FIGURE 1 | Fit to comparison data ($n = 133$). The CD analysis suggested that the number of factors to retain is one.

of the three pairs (task and strategy) showed evidence of discriminant validity. The $r^2_{\text{task} \times \text{team}} = 0.845$ was greater than $AVE_{\text{task}} = 0.78$ and $AVE_{\text{team}} = 0.70$; the $r^2_{\text{task} \times \text{strategy}} = 0.757$ was smaller than $AVE_{\text{task}} = 0.78$ and $AVE_{\text{strategy}} = 0.85$; and $r^2_{\text{team} \times \text{strategy}} = 0.838$ was greater than $AVE_{\text{team}} = 0.70$, but smaller than $AVE_{\text{strategy}} = 0.85$. Such finding might be indicative of a unidimensional model, which should be investigated through the appropriate analysis (i.e., EFA).

Reliability of the scores: Internal consistency

The merged data of the three different studies revealed good reliability evidence in terms of internal consistency ($\alpha_{\text{task}} = 0.93$, 95% CI [0.91; 0.94]; $\omega_{\text{task}} = 0.87$, 95% CI [0.82; 0.90]; $\alpha_{\text{team}} = 0.93$, 95% CI [0.91; 0.94]; $\omega_{\text{team}} = 0.90$, 95% CI [0.88; 0.91]; $\alpha_{\text{strategy}} = 0.94$, 95% CI [0.92; 0.95]; $\omega_{\text{strategy}} = 0.90$, 95% CI [0.88; 0.92]).

Because the content explained by the three different factors is similar, the dimensionality was investigated using an exploratory approach (EFA), where the EFA's suggested dimensionality from sample I was then tested (through CFA) in samples II and III's data.

When the empirical evidence lacks regarding the construct expected dimensionality, EFA might be most appropriate than CFA (Finch and French, 2015). The EFA attributes a small burden on the researcher concerning the latent factors and their relationships to the indicators, making possible establishing an interval of the number of factors that can emerge from the indicators (Marôco, 2018).

Study I

Validity Evidence Based on the Internal Structure

Dimensionality

Data obtained from study I met the KMO coefficient (0.900) and Bartlett test of sphericity ($\chi^2(78) = 963.521$; $p < 0.001$). The CD suggested that the best solution contains only one factor (Figure 1).

The one-factor solution was adopted, and the results of the correspondent EFA (Table 3) revealed 50.8% of explained variance ($RMSR = 0.086$).

Reliability of the scores: Internal consistency

The study I's data revealed good reliability evidence in terms of internal consistency ($\alpha = 0.93$, 95% CI [0.91; 0.95]; $\omega = 0.93$, 95% CI [0.91; 0.95]).

To test the proposed structure observed in study I's sample and see if it was adequate for a second and third independent samples from the population, a CFA was also performed on study II and study III samples.

Sample II

As previously mentioned, the use of CFA demands strong theoretical and/or empirical evidence regarding the dimensionality of a psychometric instrument. As such, because study I's sample provided empirical evidence supporting the one-factor solution, the CFA will be used to investigate the RSMMM single-factor model (Finch and French, 2015).

Validity Evidence Based on the Internal Structure

Dimensionality

The goodness-of-fit indices were indicative of an acceptable fit of study II's data to the model (Figure 2; $\chi^2(65) = 271.199$, $n = 277$, $\chi^2/df = 4.172$, $CFI = 0.993$, $NFI = 0.991$, $TLI = 0.992$, $SRMR = 0.054$, $RMSEA = 0.107$, $P(\text{rmsea} \leq 0.05) < 0.001$, 90% CI [0.094; 0.121]). In terms of convergent validity based on the internal structure, the estimate of AVE was good ($AVE = 0.67$).

Reliability of the scores: Internal consistency

Regarding study II's internal consistency, the obtained values revealed good validity evidence in terms of reliability ($\alpha = 0.96$, 95% CI [0.95; 0.97]; $\omega = 0.92$, 95% CI [0.89; 0.93]). Both the α and ω coefficients were indicative of good evidence in terms of the reliability of the scores.

TABLE 3 | Exploratory factor analysis loadings and total of explained variance.

Items	Factor 1
Item 1	0.826
Item 2	0.779
Item 3	0.762
Item 4	0.631
Item 5	0.440
Item 6	0.697
Item 7	0.731
Item 8	0.805
Item 9	0.677
Item 10	0.685
Item 11	0.721
Item 12	0.724
Item 13	0.713
Total of variance	0.508

Measurement invariance

Measurement invariance between sports refereed (i.e., football and futsal) was tested using study II's sample. Because there were only 25 football assistant referees, the measurement invariance analysis was performed only with the futsal and football referees. To conduct the measurement invariance considering the ordinal nature of the items, it is required that the items in both groups have the same number of thresholds. Because both groups had a different number of thresholds for some items, it was not possible to use WLSMV. As so, the measurement invariance analysis was performed using the MLR estimator because this method has been shown to work well with categorical data with no severe deviations from the normal distribution (Rhemtulla et al., 2012). As **Table 4** shows, full uniqueness measurement invariance was achieved both by the ΔCFI and $\Delta \chi^2$ criteria (Satorra and Bentler, 2001; Cheung and Rensvold, 2002), which allows establishing comparisons between the shared mental models latent scores among the football and futsal referees.

The shared mental models' latent means presented significant differences among the futsal and football referees ($\Delta \chi^2(1) = 5.495$, $p = 0.019$, $d = 0.168$), with the football referees perceiving higher levels of shared mental models than their futsal counterparts.

Sample III

Validity Evidence Based on the Internal Structure

Dimensionality

Study III's CFA revealed an excellent fit to the data (**Figure 3**; $\chi^2(65) = 74.428$, $n = 60$, $\chi^2/df = 1.145$, $CFI = 0.999$, $NFI = 0.994$, $TLI = 0.999$, $SRMR = 0.059$, $RMSEA = 0.050$, $P(\text{rmsea} \leq 0.05) = 0.482$, 90% CI [0.000; 0.095]). The convergent validity evidence in terms of the internal structure was particularly good ($AVE = 0.74$).

Reliability of the scores: Internal consistency

The internal consistency estimates of study III's sample were like study I's and study II's ones ($\alpha = 0.98$, 95% CI [0.93; 0.99];

$\omega = 0.98$, 95% CI [0.93; 0.99]) and as so were indicative of good evidence in terms of the reliability of the scores.

Validity Evidence Based on the Relation With Other Variables

The validity evidence based on the relation to other variables was investigated using study III's sample. The nomological evidence was verified (i.e., convergent validity with team adaptive performance, team work engagement, and team effectiveness). For such analysis, the structural equation model framework was used.

Measurement model

Because the sample size (i.e., $N_{\text{studyIII}} = 60$) was too small to be used with the WLSMV estimator on this model, the MLR estimator was used in all subsequent analyses. The measurement model of the team work engagement measure revealed an acceptable fit to the data ($\chi^2(25) = 69.366$, $n = 54$, $\chi^2/df = 2.775$, $CFI = 0.932$, $NFI = 0.900$, $TLI = 0.902$, $SRMR = 0.045$, $RMSEA = 0.181$, $P(\text{rmsea} \leq 0.05) < 0.001$, 90% CI [0.131; 0.233]) after constraining the error variance of the first-order factor (dedication) to 0.01 in order to avoid negative variance. The second-order reliability estimates were good ($\omega_{L1} = 0.95$, 95% CI [0.87; 0.99]; $\omega_{\text{partial}L1} = 0.97$, 95% CI [0.92; 0.99]; $\omega_{L2} = 0.98$, 95% CI [0.93; 1.00]).

The hierarchical model of the team adaptive performance scale revealed an acceptable fit to the data ($\chi^2(19) = 44.166$, $p = 0.001$, $n = 54$; $\chi^2/df = 2.325$; $CFI = 0.957$; $TLI = 0.936$; $NFI = 0.928$; $SRMR = 0.034$; $RMSEA = 0.157$; $P(\text{rmsea} \leq 0.05) = 0.004$, 90% CI [0.096; 0.217]) after adding one correlation path between item 5's and item 6's residuals ($r = 0.636$; $p = 0.024$). The variance of the problem-solving-oriented factor was constrained to 0.01 to avoid negative variances. The structural weights (γ) of the two factors were constrained to be equal, to solve the model identification problem of two first-order factors in a hierarchical model. The second-order reliability estimates were good ($\omega_{L1} = 0.96$, 95% CI [0.87; 0.98]; $\omega_{\text{partial}L1} = 0.97$, 95% CI [0.91; 0.99]; $\omega_{L2} = 0.96$, 95% CI [0.85; 0.99]).

The team effectiveness second-order model had a good fit to the data ($\chi^2(32) = 58.072$, $p = 0.003$, $n = 57$; $\chi^2/df = 1.815$; $CFI = 0.964$; $TLI = 0.950$; $NFI = 0.925$; $SRMR = 0.044$; $RMSEA = 0.120$; $P(\text{rmsea} \leq 0.05) = 0.018$, 90% CI [0.068; 0.168]). The second-order reliability estimates were good ($\omega_{L1} = 0.92$, 95% CI [0.69; 0.97]; $\omega_{\text{partial}L1} = 0.95$, 95% CI [0.84; 0.99]; $\omega_{L2} = 0.97$, 95% CI [0.78; 0.98]).

Structural model

Because the used psychometric instruments (i.e., measurement model) presented good validity evidence based on the internal structure, a full structural model was tested for each of the related measures (i.e., team work engagement, team adaptive performance, and team effectiveness). The structural model that related team work engagement and shared mental models revealed an acceptable fit to the data ($\chi^2(206) = 383.091$, $p < 0.001$, $n = 54$; $\chi^2/df = 1.86$; $CFI = 0.900$; $TLI = 0.888$; $NFI = 0.809$; $SRMR = 0.060$; $RMSEA = 0.126$; $P(\text{rmsea} \leq 0.05) < 0.001$, 90% CI [0.106; 0.146]) with a strong and positive latent

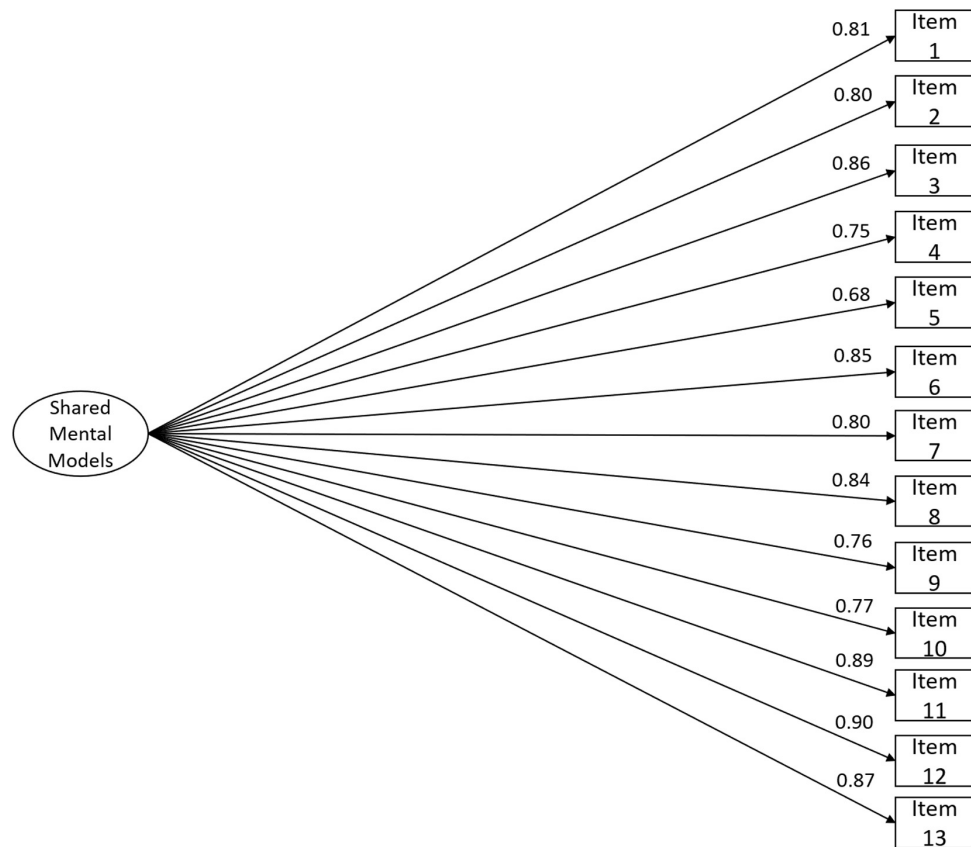


FIGURE 2 | RSMMM one-factor version (13-item) structure fit using study II's sample ($n = 277$). Factor loadings for each item are shown: $\chi^2(65) = 271.199$, $n = 277$, $\chi^2/df = 4.172$, $CFI = 0.993$, $NFI = 0.991$, $TLI = 0.992$, $SRMR = 0.054$, $RMSEA = 0.107$, $P(rmse \leq 0.05) < 0.001$, 90% CI [0.094; 0.121].

correlation (H4; $r_{\text{team work engagement} \times \text{shared mental models}} = 0.764$; $p = 0.099$). The raw correlation between the arithmetic mean of the RSMMM's items and the team work engagement's items was strong and positive ($r = 0.776$; $p < 0.001$). The structural model that correlated shared mental models with team adaptive performance presented an acceptable fit to the data ($\chi^2(185) = 329.269$, $p < 0.001$, $n = 54$; $\chi^2/df = 1.780$; $CFI = 0.916$; $TLI = 0.905$; $NFI = 0.829$; $SRMR = 0.038$; $RMSEA = 0.120$; $P(rmse \leq 0.05) < 0.001$, 90% CI [0.099; 0.141]) showing a positive strong latent correlation (H4;

$r_{\text{adaptive performance} \times \text{shared mental models}} = 0.910$; $p = 0.053$). The raw correlation was also strong and positive between the arithmetic mean of the RSMMM's items and the team adaptive performance's items ($r = 0.888$; $p < 0.001$). Finally, the model that correlated team effectiveness with shared mental models showed an acceptable fit to the data ($\chi^2(226) = 423.832$, $p < 0.001$, $n = 54$; $\chi^2/df = 1.875$; $CFI = 0.895$; $TLI = 0.882$; $NFI = 0.801$; $SRMR = 0.050$; $RMSEA = 0.124$; $P(rmse \leq 0.05) < 0.001$, 90% CI [0.106; 0.142]) revealing positive and strong latent correlation (H4; $r_{\text{effectiveness} \times \text{shared mental models}} = 0.909$; $p = 0.068$). The raw correlation between the arithmetic mean of the team effectiveness' items and the RSMMM's items was strong and positive ($r = 0.857$; $p < 0.001$).

Such correlation values suggest acceptable nomological evidence—particularly in terms of convergent validity evidence—in relation to the team work engagement scores. However, the correlation values between the shared mental models' scores and the team adaptive performance and the team effectiveness seem too high (constructs overlap), providing poor convergent validity evidence.

Some of the presented models had mediocre *RMSEA* values. However, *RMSEA* point estimates depend on sample size, model degrees of freedom, and model misspecification (MacCallum et al., 1996; Chen et al., 2008). To assess the model's fit to the

TABLE 4 | Measurement invariance analysis among futsal and football referees (study II's sample).

Model	χ^2	df	χ^2/df	CFI_{robust}	$\Delta\chi^2$	$\Delta CFI_{\text{robust}}$
Sport refereed						
Configural	436.77	130	3.36	0.888	—	—
Metric	442.68	142	3.12	0.893	4.071 ^{ns}	0.005
Scalar	455.31	154	2.96	0.892	13.485 ^{ns}	−0.001
Full uniqueness	495.30	167	2.97	0.888	16.641 ^{ns}	−0.004
Latent means	499.45	168	2.97	0.887	5.495*	−0.002

^{ns} $p < 0.05$; * $p \leq 0.05$.

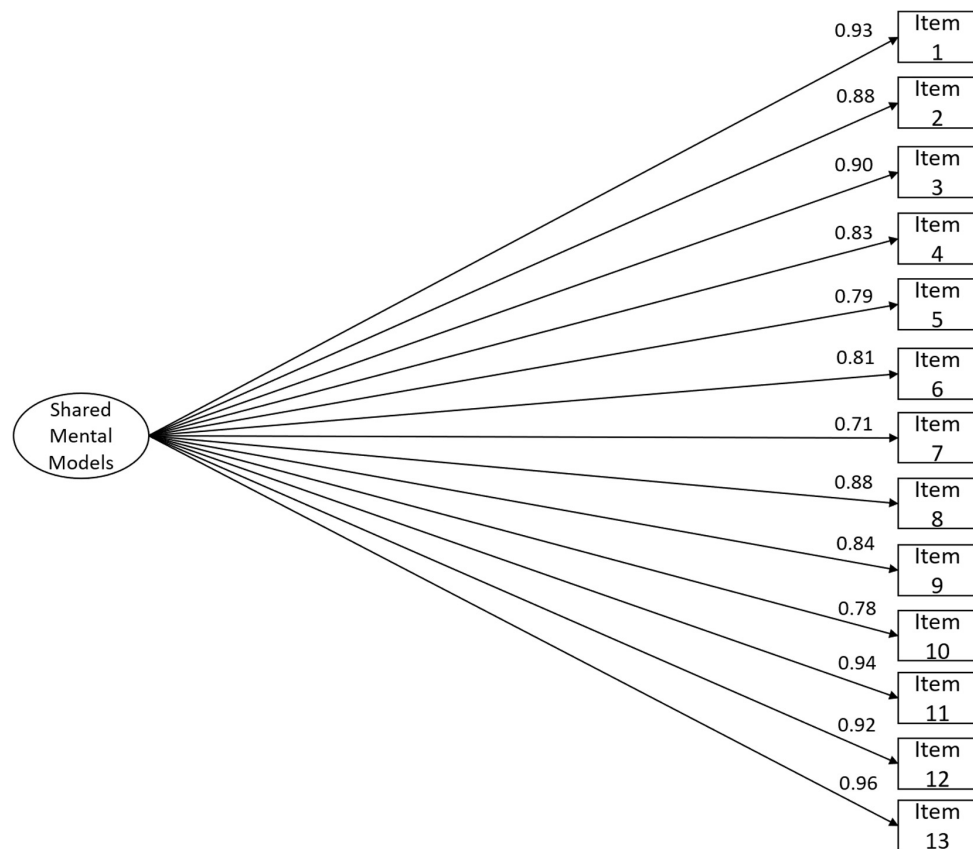


FIGURE 3 | RSMMM one-factor version (13-item) structure fit using study III's sample ($n = 60$). Factor loadings for each item are shown: $\chi^2(65) = 74.428$, $n = 60$, $\chi^2/df = 1.145$, $CFI = 0.999$, $NFI = 0.994$, $TLI = 0.999$, $SRMR = 0.059$, $RMSEA = 0.050$, $P(rmse \leq 0.05) = 0.482$, 90% CI [0.000; 0.095].

data, other goodness-of-fit indices were presented in conjunction, namely, *SRMR*, which showed acceptable to good estimates. The *SRMR* goodness-of-fit index seems to be more robust than *RMSEA* across all conditions (Maydeu-Olivares et al., 2018).

DISCUSSION

There is a need to more fully examine the team dynamics present within professional sport referee teams (Aragão e Pina et al., 2018). In particular, given that shared mental models have been shown to play an important role in shaping team dynamics and performance in other context (Marks et al., 2000; Mathieu et al., 2000; Mascarenhas et al., 2005), there is a need to investigate the impact of referee teams shared mental models on team functioning and adaptability. However, the shared mental model literature suggests that one needs to adapt the measurement of such cognitive structures to the context within which such teams operate. As such, the primary aim of the current study was to develop a measure of shared mental models within the context of professional football and futsal referee teams.

The proposed new measure revealed good psychometric properties. Namely, the shared mental model measure developed in this article presented good validity evidence across the three different samples of football and/or futsal referees presented here.

The RSMMM showed promising validity evidence both based on the internal structure and based on the relation with team work engagement (i.e., nomological evidence). Nomological evidence approaches the network of relations between the constructs. As so, the observed correlations between the latent variables (shared mental models and team work engagement) were aligned with the claims of the literature. Such findings suggest a useful unidimensional measure both for futsal and football referees.

The initial model (three first-order factors) revealed a lack of discriminant validity in terms of internal structure (Fornell and Larcker, 1981), indicating that the content explained by the three factors is similar. The dimensionality that emerged from the CD revealed that the referees on the sample perceive shared mental models as a unidimensional structure. Previous studies in which the RSMMM was based conceptualized it as a three-factor model (Santos et al., 2015a) or as a unidimensional one (Santos et al., 2015b). However, such solutions were not necessarily expected to be found in the referee context. Both the three-factor dimensionality of the SMMS (Santos et al., 2015a) and the unidimensional four-items version (Santos et al., 2015b) were proposed using a sample of teams from diverse contexts that participated on a virtual management challenge. Researchers have identified different dimensions of shared mental models including task, team, and strategy, as these are key aspects of the team work environment (e.g., Mohammed et al., 2010;

Resick et al., 2010b). However, as Mohammed et al. (2000) state, “although the domain of a team model can vary (e.g., individual task work, team task work, team work), it should be viewed as reflecting how team members conceptualize a team-relevant phenomenon” (p. 125). Our study supports Mohammed et al. (2000) argument as our findings consistently suggest (over three different samples) that football and futsal referees have a general understanding of the relevant elements of team work and thereby do not distinguish between the different dimensions. Our findings are following previous studies that analyze the perception of shared mental models (Aubé et al., 2015, 2018; Santos et al., 2015b; Burtscher and Oostlander, 2019). Although conceptually, shared mental models may regard to different aspects of work, practitioners in a domain do not always seem to make this distinction, and results have supported a one-factorial solution (Aubé et al., 2015, 2018; Santos et al., 2015b; Burtscher and Oostlander, 2019). Mental models considerably derive from the occupational context in which they raise (Cannon-Bowers et al., 1993).

Additionally, the referees' tasks are majorly concentrated during the referring of the futsal or football matches. Where everything is interconnected and must be deeply articulated between the team members, such contextual peculiarities might contribute to a perception of mental models as a singular whole. The suggested solution revealed a good fit for the single-factor model in three different samples of referees from two different sports (futsal and football), and as so, the H1 was supported. The RSMMM showed robustness in maintaining its dimensionality even when tested in a different sport other than football. Thus, such stability in the instrument's structure allows for useful perspectives in terms of its implementation within other sports.

The second hypothesis was supported; thus, reliability evidence was good. The internal consistency estimates (i.e., α and ω) values were satisfactory for all the samples, based on the recommended values (Nunnally and Bernstein, 1994). Previous studies that used a similar measure also had good values of internal consistency estimates, namely, the unidimensional shared mental models proposed by Santos et al. (2015b), which had $\alpha = 0.92$, and the PMU (which is another unidimensional measure) had $\alpha = 0.83$ and $\omega = 0.83$ (Burtscher and Oostlander, 2019). The TeamKMI internal consistency values of its five factors ranged from $\alpha_{\text{factor 3}} = 0.75$ to $\alpha_{\text{factor 2}} = 0.89$ (Johnson et al., 2007), whereas the SMMS had not its internal consistency values reported in its original study (Santos et al., 2015a). As such, the obtained results are aligned with previous studies using similar measures.

Measurement invariance among futsal and football referees was obtained. Such kind of psychometric property is essential to establish comparisons between mental model scores. Previous studies with referee samples using other instruments (e.g., Referee Self-Efficacy Scale) have not obtained full uniqueness measurement invariance among the sport referred; in fact, only partial factorial invariance was achieved (Myers et al., 2012). While studies using other instruments among football referees samples obtained different levels of measurement invariance, namely, metric invariance between referees and assistant referees (Brandão et al., 2014), and strong invariance (i.e., scalar

invariance) among elite and non-elite referees (Johansen et al., 2018). An instrument's mean scores should be compared only if scalar measurement invariance is granted (Marôco, 2014). Besides this fact, latent means comparisons should be implemented instead of raw means comparison, because the raw means do not account for measurement error. The established comparisons between shared mental models' latent means were made only after the achievement of full uniqueness measurement invariance. Football referees perceived significantly higher shared mental models' levels than futsal referees, which might be explained by the fact that in football the referee has a higher concentration of responsibilities in comparison with their assistants, whereas in futsal there is a higher sharing of those responsibilities between referees. As so, the perception of the shared mental models can be affected by the number of tasks with shared responsibility among the team members. The different levels of experience might also explain the differences between the shared mental models' levels because the football referees had more years of experience than their futsal counterparts.

Regarding the validity evidence based on the relation to other variables, the RSMMM revealed acceptable nomological validity evidence in terms of convergent evidence with team work engagement. However, the convergent evidence was poor regarding team effectiveness and team adaptive performance. Thus, H4 was partially verified. The correlations among shared mental models with team adaptive performance; and shared mental models with team effectiveness were too high, suggesting some overlap of the constructs. The correlation among shared mental models with team work engagement was more adequate to the extent of convergence expected. Such empirical evidence allows confirming the expected direction of the associations with work engagement, reflecting partial support for the proposed theoretical nomological network (Lissitz and Samuelson, 2007). This source of evidence was analyzed using study III's sample (only composed by futsal referees), which might be a particularity of this small sample. Studies with small samples often report anomalously large effect sizes (Funder and Ozer, 2019), and as such, future replication studies might show that those effect sizes were overestimated with the used small sample ($n_{\text{study III}} = 60$) of futsal referees (Cumming, 2012).

This is the first instrument that explicitly measures shared mental models taking into consideration the specificities of football and futsal referee teams. All psychometric properties were indicative of good validity evidence, revealing a promising instrument for other contexts of referring (e.g., handball, basketball, rugby). The accumulated validity evidence seems to support the intended interpretation of the test scores for the RSMMM (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014).

Limitations and Directions for Future Research

While the current study provided some promising results about the RSMMM within the domain of football and futsal referees, there are some limitations in this study that must be

acknowledged. For starters, this study was solely focused on referee teams within the sports of football and futsal. However, even though these are popular sports, it begs the question of how RSMMM would need to be altered to apply to other professional sport referee teams. Accordingly, it will be pertinent to see future research to examine the RSMMM in other sports and explore how this measure would need to be adjusted to be valuable and useful within other sports contexts.

In the present article, the data were analyzed at the individual level, not aggregated to the team level. Using the data aggregated to the team level could allow for a better understanding of the team's global perceptions of themselves instead of the isolated individuals' perceptions of the team. It is then possible to assess to which extent team members share mental models. For that propose, the level of agreement between team members would have to be considered for the subsequent analyses. It is worth mentioning that some of the referee teams are more stable in their constitution through the season than others (e.g., referees of lower categories tend to have more heterogeneity). The current cross-sectional study only provides a snapshot of the perceptions, which could vary if measured in a time frame (Levin, 2006).

The validity evidence based on the relations to other variables should be deeply investigated in terms of test criterion (e.g., higher team performance). As McNeese et al. (2015) urged, studying team cognition in sport must include a combination of both the shared knowledge and dynamical approaches. Future studies should investigate if shared mental models' levels are associated with performance (e.g., match analysis report ratings or associations' match/season ratings). Regarding the associations' ratings, it might be also interesting to check which of the components of the assessment (there are usually three components: physical performance, performance of the written test regarding rules and the laws of the game, and performance of the match observations attributed by the referees observers) has higher association with shared mental models. The validity evidence based on the relations to other variables should also be investigated in terms of convergent (similar constructs) and discriminant evidence (measures purportedly of different constructs), preferably using different measures of other nature rather than perceptions (i.e., self-report measures). It is challenging to prove that representations exist beyond the boundaries of an individual organism and that such representations can be somehow shared with others. The use of technologies, as multiple eye tracker (Wildman et al., 2014) or hyperbrain networks (Filho et al., 2017), is encouraged. This kind of measures can surpass some of self-report measures limitations (Schwarz, 1999; Baumeister et al., 2007), particularly when it concerns measuring the perception of behaviors instead of behaviors (Lonati et al., 2018). With the robustness of such evidence, the RSMMM might give a step forward in its establishment as a measure of shared mental models among referees.

Study III's sample size is small for structural equation modeling analysis, however, when looking to the number of futsal referees at the national level, it represents a considerable amount (30%) of the population of the Portuguese Football Federation

($N_{\text{season 2018–2019}} = 177$). Nevertheless, future studies should try to increase the number of referees both at the national and regional levels.

Additionally, given that in the collected samples were only a few football assistant referees ($n = 25$), the measurement invariance was not tested among them. Accordingly, given that assistant referees play an essential role within the football referee team, future studies should account for this and explore what impact having assistant referees more represented in future research samples can alter. Finally, within the current study and the underlying data that were used here, measurement invariance across time could not be examined. As a result, no statements regarding the trends that may exist across time can be made. In response, future research should examine this fact and collect the type of data necessary to be able to assess longitudinal measurement invariance. The assessment of validity evidence is an ongoing and never-ending process (Slaney, 2017); thus, the next steps should seem like a natural on the evolution of the RSMMM as an established measure to approach shared mental models within referees.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JorS, JP, and AP contributed to conception and design of the study. JorS, JP and JoãS organized the database. JorS and JM performed the statistical analysis. JorS wrote the first draft of the manuscript. JP, JoãS, and CS wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was supported by Fundação para a Ciência e Tecnologia, grant UIDB/00315/2020. This work was also produced with the support of INCD funded by FCT and FEDER under the project 22153-01/SAICT/2016.

ACKNOWLEDGMENTS

We would like to thank the Portuguese Football Federation for providing access to seminars and for sharing the digital survey with their referees (Study III).

REFERENCES

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Aragão e Pina, J., Passos, A. M., Araújo, D., and Maynard, M. T. (2018). Football refereeing: an integrative review. *Psychol. Sport Exerc.* 35, 10–26. doi: 10.1016/j.psychsport.2017.10.006
- Aragão e Pina, J., Passos, A. M., Carvalho, H., and Travis Maynard, M. (2019). To be or not to be an excellent football referee: different experts' viewpoints. *J. Sports Sci.* 37, 692–700. doi: 10.1080/02640414.2018.1522940
- Aragão e Pina, J., Passos, A. M., Travis Maynard, M., and Sinval, J. (2021). Self-efficacy, mental models and team adaptation: a first approach on football and futsal refereeing. *Psychol. Sport Exerc.* 52:101787. doi: 10.1016/j.psychsport.2020.101787
- Aubé, C., and Rousseau, V. (2005). Team goal commitment and team effectiveness: the role of task interdependence and supportive behaviors. *Gr. Dyn. Theory Res. Pract.* 9, 189–204. doi: 10.1037/1089-2699.9.3.189
- Aubé, C., Rousseau, V., Brunelle, E., and Marques, D. (2018). The relevance of being “on the same page” to succeed as a project team: a moderated mediation model. *Motiv. Emot.* 42, 804–815. doi: 10.1007/s11031-018-9706-2
- Aubé, C., Rousseau, V., and Tremblay, S. (2015). Perceived shared understanding in teams: the motivational effect of being ‘on the same page.’. *Br. J. Psychol.* 106, 468–486. doi: 10.1111/bjop.12099
- Bartlett, M. S. (1951). The effect of standardization on a χ^2 approximation in factor analysis. *Biometrika* 38, 337–344. doi: 10.2307/2332580
- Baumeister, R. F., Vohs, K. D., and Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspect. Psychol. Sci.* 2, 396–403. doi: 10.1111/j.1745-6916.2007.00051.x
- Boomsma, A. (2000). Reporting analyses of covariance structures. *Struct. Equ. Model. A Multidiscip. J.* 7, 461–483. doi: 10.1207/S15328007SEM0703_6
- Boyer, S., Rix-Lièvre, G., and Récopé, M. (2015). L'arbitrage de haut niveau, une affaire d'équipe. *Mov. Sport Sci. Sci. Mot.* 101, 91–101. doi: 10.1051/sm/2014014
- Brandão, M. R. F., Serpa, S., Rosado, A., and Weinberg, R. (2014). Psychometric properties of the burnout inventory for referees. *Mot. Rev. Educ. Física* 20, 374–383. doi: 10.1590/S1980-65742014000400003
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*, 2nd Edn. New York, NY: The Guilford Press.
- Burtscher, M. J., and Oostlander, J. (2019). Perceived mutual understanding (PMU): development and initial testing of a German short scale for perceptual team cognition. *Eur. J. Psychol. Assess.* 35, 98–108. doi: 10.1027/1015-5759/a000360
- Byrne, B. M. (2010). *Structural Equation Modeling with AMOS*, 3rd Edn. New York, NY: Routledge.
- Cannon-Bowers, J. A., and Salas, E. (2001). Reflections on shared cognition. *J. Organ. Behav.* 22, 195–202. doi: 10.1002/job.82
- Cannon-Bowers, J. A., Salas, E., and Converse, S. (1993). “Shared mental models in expert team decision making,” in *Individual and Group Decision Making: Current Issues*, ed. N. J. Castellan (Hillsdale, NJ: Lawrence Erlbaum), 221–246. doi: 10.4324/9780203772744-20
- Canty, A., and Ripley, B. (2020). *boot: Bootstrap R (S-Plus) Functions (R package version 1.3-25) [Computer software]*. Available online at: <https://cran.r-project.org/web/packages/boot/boot.pdf>
- Carpenter, J., and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat. Med.* 19, 1141–1164. doi: 10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J. B., and Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociol. Methods Res.* 36, 462–494. doi: 10.1177/0049124108314720
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model. A Multidiscip. J.* 9, 233–255. doi: 10.1207/S15328007SEM0902_5
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collina, P. (2004). *As Minhas Regras do Jogo - O que o Futebol me Ensinou Sobre a Vida*. Lisboa: Editorial Presença.
- Costa, P. L., Passos, A. M., and Bakker, A. B. (2014a). Empirical validation of the team work engagement construct. *J. Pers. Psychol.* 13, 34–45. doi: 10.1027/1866-5888/a000102
- Costa, P. L., Passos, A. M., and Bakker, A. B. (2014b). Team work engagement: a model of emergence. *J. Occup. Organ. Psychol.* 87, 414–436. doi: 10.1111/joop.12057
- Costa, P. L., Passos, A. M., and Bakker, A. B. (2016). The work engagement grid: predicting engagement from two core dimensions. *J. Manag. Psychol.* 31, 774–789. doi: 10.1108/JMP-11-2014-0336
- Courtney, M. G. R. (2012). Determining the number of factors to retain in EFA: using the SPSS R-Menu v2.0 to make more judicious estimations. *Pract. Assess. Res. Eval.* 18, 1–14.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/BF02310555
- Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York, NY: Routledge.
- Cunningham, I., Simmons, P., Mascarenhas, D. R. D., and Redhead, S. (2014). Skilled interaction: concepts of communication and player management in the development of sport officials. *Int. J. Sport Commun.* 7, 166–187. doi: 10.1123/IJSC.2013-0098
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., and Billiet, J. (2014). Measurement equivalence in cross-national research. *Annu. Rev. Sociol.* 40, 55–75. doi: 10.1146/annurev-soc-071913-043137
- Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.
- DeChurch, L. A., and Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: a meta-analysis. *Gr. Dyn. Theory Res. Pract.* 14, 1–14. doi: 10.1037/a0017455
- Diotiati, P., Falese, L., Mancone, S., and Purromuto, F. (2017). A structural model of self-efficacy in handball referees. *Front. Psychol.* 8:811. doi: 10.3389/fpsyg.2017.00811
- Efron, B., and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall.
- Filho, E., Pierini, D., Robazza, C., Tenenbaum, G., and Bertollo, M. (2017). Shared mental models and intra-team psychophysiological patterns: a test of the juggling paradigm. *J. Sports Sci.* 35, 112–123. doi: 10.1080/02640414.2016.1158413
- Filho, E., and Tenenbaum, G. (2012). “Team mental models in sports: an overview,” in *Athletic Insight's Writings in Sport Psychology (Sports and Athletics Preparation Performance and Psychology)*, ed. R. Schinke (Hauppauge, NY: Nova Science Pub Inc), 329–342.
- Finch, W. H., and French, B. F. (2015). *Latent Variable Modeling with R*. New York, NY: Routledge.
- Finney, S. J., and DiStefano, C. (2013). “Non-normal and categorical data in structural equation modeling,” in *Structural Equation Modeling: A Second Course*, eds G. R. Hancock and R. O. Mueller (Charlotte, NC: Information Age Publishing), 439–492.
- Finney, S. J., DiStefano, C., and Kopp, J. P. (2016). “Overview of estimation methods and preconditions for their application with structural equation modeling,” in *Principles and Methods of Test Construction: Standards and Recent Advances*, eds K. Schweizer and C. DiStefano (Boston, MA: Hogrefe Publishing), 135–165.
- Fornell, C., and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark Res.* 18, 39–50. doi: 10.2307/3151312
- Funder, D. C., and Ozer, D. J. (2019). Evaluating effect size in psychological research: sense and nonsense. *Adv. Methods Pract. Psychol. Sci.* 2, 156–168. doi: 10.1177/2515245919847202
- Hackman, J. R. (1987). “The design of work teams,” in *Handbook of Organizational Behaviour*, ed. J. W. Lorsch (Englewood Cliffs, NJ: Prentice-Hall), 315–342.
- Hair, J. F., Babin, B. J., Anderson, R. E., and Black, W. C. (2019). *Multivariate Data Analysis*, 8th Edn. Andover: Cengage Learning.
- Hancock, D. J., Martin, L. J., Evans, M. B., and Paradis, K. F. (2018). Exploring perceptions of group processes in ice hockey officiating. *J. Appl. Sport Psychol.* 30, 222–240. doi: 10.1080/10413200.2017.1349208
- Hoyle, R. H. (ed.) (1995). *Structural Equation Modeling: Concepts, Issues and Applications*. Thousand Oaks, CA: SAGE Publications.

- Iacobucci, D., and Duhachek, A. (2003). Advancing alpha: measuring reliability with confidence. *J. Consum. Psychol.* 13, 478–487. doi: 10.1207/S15327663JCP1304_14
- Johansen, B. T., Ommundsen, Y., and Haugen, T. (2018). Referee efficacy in the context of Norwegian soccer referees – a meaningful construct? *Psychol. Sport Exerc.* 38, 184–191. doi: 10.1016/j.psychsport.2018.06.013
- Johnson, T. E., Lee, Y., Lee, M., O'Connor, D. L., Khalil, M. K., and Huang, X. (2007). Measuring sharedness of team-related knowledge: design and validation of a shared mental model instrument. *Hum. Resour. Dev. Int.* 10, 437–454. doi: 10.1080/13678860701723802
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2019). *semTools: Useful Tools for Structural Equation Modeling (R package version 0.5-1.933) [Computer software]*. Available online at: <https://cran.r-project.org/package=semTools>
- Kaiser, H. F., and Rice, J. (1974). Little Jiffy, Mark IV. *Educ. Psychol. Meas.* 34, 111–117. doi: 10.1177/001316447403400115
- Kellermanns, F. W., Floyd, S. W., Pearson, A. W., and Spencer, B. (2008). The contingent effect of constructive confrontation on the relationship between shared mental models and decision quality. *J. Organ. Behav.* 29, 119–137. doi: 10.1002/job.497
- Klimoski, R., and Mohammed, S. (1994). Team mental model: construct or metaphor? *J. Manage.* 20, 403–437. doi: 10.1177/014920639402000206
- Lemon, J. (2006). Plotrix: a package in the red light district of R. *R-News* 6, 8–12.
- Levin, K. A. (2006). Study design III: cross-sectional studies. *Evid. Based. Dent.* 7, 24–25. doi: 10.1038/sj.ebd.6400375
- Lim, B., and Klein, K. J. (2006). Team mental models and team performance: a field study of the effects of team mental model similarity and accuracy. *J. Organ. Behav.* 27, 403–418. doi: 10.1002/job.387
- Lissitz, R. W., and Samuelson, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educ. Res.* 36, 437–448. doi: 10.3102/0013189X07311286
- Lonati, S., Quiroga, B. F., Zehnder, C., and Antonakis, J. (2018). On doing relevant and rigorous experiments: review and recommendations. *J. Oper. Manag.* 64, 19–40. doi: 10.1016/j.jom.2018.10.003
- Lüdtke, D. (2019). *sjstats: Statistical Functions for Regression Models (R package version 0.17.3)*.
- MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychol. Methods* 1, 130–149. doi: 10.1037/1082-989X.1.2.130
- Mallo, J., Frutos, P. G., Juárez, D., and Navarro, E. (2012). Effect of positioning on the accuracy of decision making of association football top-class referees and assistant referees during competitive matches. *J. Sports Sci.* 30, 1437–1445. doi: 10.1080/02640414.2012.711485
- Marks, M. A., Sabella, M. J., Burke, C. S., and Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *J. Appl. Psychol.* 87, 3–13. doi: 10.1037/0021-9010.87.1.3
- Marks, M. A., Zaccaro, S. J., Mathieu, J. E., Gessner, D., Klimoski, R., Sanchez, J., et al. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *J. Appl. Psychol.* 85, 971–986. doi: 10.1037//0021-9010.85.6.971
- Marôco, J. (2014). *Análise de Equações Estruturais: Fundamentos Teóricos, Software & Aplicações*, 2nd Edn. Pêro Pinheiro: ReportNumber, Lda.
- Marôco, J. (2018). *Análise Estatística Com o SPSS Statistics*, 7th Edn. Pêro Pinheiro: ReportNumber, Lda.
- Marques-Quinteiro, P., Ramos-Villagrasa, P. J., Passos, A. M., and Curral, L. (2015). Measuring adaptive performance in individuals and teams. *Team Perform. Manag. An Int. J.* 21, 339–360. doi: 10.1108/TPM-03-2015-0014
- Mascarenhas, D. R. D., Collins, D., Mortimer, P. W., and Morris, B. (2005). Training accurate and coherent decision making in rugby union referees. *Sport Psychol.* 19, 131–147. doi: 10.1123/tsp.19.2.131
- Mascarenhas, D. R. D., O'Hare, D., and Plessner, H. (2006). The psychological and performance demands of association football refereeing. *Int. J. Sport Psychol.* 37, 99–120.
- Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., and Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *J. Appl. Psychol.* 85, 273–283. doi: 10.1037/0021-9010.85.2.273
- Mathieu, J. E., Rapp, T. L., Maynard, M. T., and Mangos, P. M. (2009). Interactive effects of team and task shared mental models as related to air traffic controllers' collective efficacy and effectiveness. *Hum. Perform.* 23, 22–40. doi: 10.1080/08959280903400150
- Maydeu-Olivares, A., Shi, D., and Rosseel, Y. (2018). Assessing fit in structural equation models: a monte-carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Struct. Equ. Model. A Multidiscip. J.* 25, 389–402. doi: 10.1080/10705511.2017.1389611
- Maynard, M. T., Kennedy, D. M., and Sommer, S. A. (2015). Team adaptation: a fifteen-year synthesis (1998–2013) and framework for how this literature needs to “adapt” going forward. *Eur. J. Work Organ. Psychol.* 24, 652–677. doi: 10.1080/1359432X.2014.1001376
- McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah, NJ: Routledge.
- McDonald, R. P., and Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychol. Methods* 7, 64–82. doi: 10.1037/1082-989X.7.1.64
- McEwan, D., and Beauchamp, M. R. (2014). Teamwork in sport: a theoretical and integrative review. *Int. Rev. Sport Exerc. Psychol.* 7, 229–250. doi: 10.1080/1750984X.2014.932423
- McGrath, J. E. (1991). Time, interaction, and performance (TIP): a theory of groups. *Small Gr. Res.* 22, 147–174. doi: 10.1177/1046496491222001
- McNamara, A., Arino de la Rubia, E., Zhu, H., Ellis, S., and Quinn, M. (2018). *skimr: Compact and Flexible Summaries of Data (R package version 1.0.3) [Computer software]*. Available online at: <https://cran.r-project.org/package=skimr>
- McNeese, N. J., Cooke, N. J., Fedele, M. A., and Gray, R. (2015). Theoretical and methodical approaches to studying team cognition in sports. *Proc. Manuf.* 3, 1211–1218. doi: 10.1016/j.promfg.2015.07.201
- Mohammed, S., Ferzandi, L., and Hamilton, K. (2010). Metaphor no more: a 15-year review of the team mental model construct. *J. Manage.* 36, 876–910. doi: 10.1177/0149206309356804
- Mohammed, S., Klimoski, R., and Rentsch, J. R. (2000). The measurement of team mental models: we have no shared schema. *Organ. Res. Methods* 3, 123–165. doi: 10.1177/109442810032001
- Muponde, R., and Muchemwa, K. (2011). Dictatorships, disasters, and African soccer: reflections on a moment in Zimbabwean soccer. *African Ident.* 9, 279–290. doi: 10.1080/14725843.2011.591227
- Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data. *J. Econom.* 22, 43–65. doi: 10.1016/0304-4076(83)90093-3
- Myers, N. D., Feltz, D. L., Guillén, F., and Dithurbide, L. (2012). Development of, and initial validity evidence for, the referee self-efficacy scale: a multistudy report. *J. Sport Exerc. Psychol.* 34, 737–765. doi: 10.1123/jsep.34.6.737
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*, 3rd Edn. New York, NY: McGraw-Hill.
- Peters, G.-J. Y. (2018). *userfriendlyscience: Quantitative Analysis Made Accessible (R package version 0.7.2)*.
- Pulakos, E. D., Dorsey, D. W., and White, S. S. (2006). “Adaptability in the workplace: selecting an adaptive workforce,” in *Understanding adaptability: A prerequisite for effective performance within complex environments*, eds C. S. Burke, L. G. Pierce, and E. Salas (Amsterdam: Elsevier), 41–71.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing (version 4.0.0) [Computer software]*. Vienna: R Foundation for Statistical Computing.
- Randall, K. R., Resick, C. J., and DeChurch, L. A. (2011). Building team adaptive capacity: the roles of sensegiving and team composition. *J. Appl. Psychol.* 96, 525–540. doi: 10.1037/a0022622
- Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *Br. J. Math. Stat. Psychol.* 54, 315–323. doi: 10.1348/000711001159582
- Resick, C. J., Dickson, M. W., Mitchelson, J. K., Allison, L. K., and Clark, M. A. (2010a). Team composition, cognition, and effectiveness: examining mental model similarity and accuracy. *Gr. Dyn. Theory Res. Pract.* 14, 174–191. doi: 10.1037/a0018444
- Resick, C. J., Murase, T., Bedwell, W. L., Sanz, E., Jimenez, M., and DeChurch, L. A. (2010b). Mental model metrics and team adaptability: a multi-facet multi-method examination. *Gr. Dyn. Theory Res. Pract.* 14, 332–349. doi: 10.1037/a0018822

- Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research (R package version 1.9.4) [Computer software]*. Available online at: <https://cran.r-project.org/package=psych>.
- Rhemtulla, M., Brosseau-Liard, P. É., and Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* 17, 354–373. doi: 10.1037/a0029315
- Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–21. doi: 10.18637/jss.v048.i02
- RStudio Team (2020). *RStudio: Integrated development for R (version 1.4.672) [Computer software]*. Available online at: <http://www.rstudio.com/>
- Ruscio, J. (2018). *RGenData: Generates Multivariate Nonnormal Data and Determines How Many Factors to Retain (R package version 1.0) [Computer software]*. Available online at: <https://cran.r-project.org/package=RGenData>
- Rusio, J., and Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychol. Assess.* 24, 282–292. doi: 10.1037/a0025697
- Samuel, R. D. (2015). A psychological preparation framework for elite soccer referees: a practitioner's perspective. *J. Sport Psychol. Action* 6, 170–187. doi: 10.1080/21520704.2015.1065938
- Santos, C. M., and Passos, A. M. (2013). Team mental models, relationship conflict and effectiveness over time. *Team Perform. Manag. An Int. J.* 19, 363–385. doi: 10.1108/TPM-01-2013-0003
- Santos, C. M., Passos, A. M., and Uitdewilligen, S. (2016). When shared cognition leads to closed minds: temporal mental models, team learning, adaptation and performance. *Eur. Manag. J.* 34, 258–268. doi: 10.1016/j.emj.2015.11.006
- Santos, C. M., Uitdewilligen, S., and Passos, A. M. (2015a). A temporal common ground for learning: the moderating effect of shared mental models on the relation between team learning behaviours and performance improvement. *Eur. J. Work Organ. Psychol.* 24, 710–725. doi: 10.1080/1359432X.2015.1049158
- Santos, C. M., Uitdewilligen, S., and Passos, A. M. (2015b). Why is your team more creative than mine? The influence of shared mental models on intra-group conflict, team creativity and effectiveness. *Creat. Innov. Manag.* 24, 645–658. doi: 10.1111/caim.12129
- Satorra, A., and Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66, 507–514. doi: 10.1007/BF02296192
- Schwarz, N. (1999). Self-reports: how the questions shape the answers. *Am. Psychol.* 54, 93–105. doi: 10.1037/0003-066X.54.2.93
- Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arppe, A., et al. (2019). *DescTools: Tools for Descriptive Statistics (R package version 0.99.28) [Computer software]*. Available online at: <https://cran.r-project.org/package=DescTools>
- Sinval, J., Marques-Pinto, A., Queirós, C., and Marôco, J. (2018a). Work engagement among rescue workers: psychometric properties of the Portuguese UWES. *Front. Psychol.* 8:2229. doi: 10.3389/fpsyg.2018.02229
- Sinval, J., Pasian, S. R., Queirós, C., and Marôco, J. (2018b). Brazil-Portugal transcultural adaptation of the UWES-9: internal consistency, dimensionality, and measurement invariance. *Front. Psychol.* 9:353. doi: 10.3389/fpsyg.2018.00353
- Slaney, K. (2017). *Validating Psychological Constructs: Historical, Philosophical, and Practical Dimensions*. London: Palgrave Macmillan.
- Tomás, J. M., de los Santos, S., Alonso-Andres, A., and Fernández, I. (2017). Validation of the Maslach burnout inventory-general survey on a representative sample of dominican teachers: normative data. *Span. J. Psychol.* 19:E83. doi: 10.1017/sjp.2016.91
- Uitdewilligen, S., Rico, R., and Waller, M. J. (2018). Fluid and stable: dynamics of team action patterns and adaptive outcomes. *J. Organ. Behav.* 39, 1113–1128. doi: 10.1002/job.2267
- Vicente, Â, Rodrigues, B. S., and D'Oliveira, T. C. (2014). Escala de Eficácia de Equipas (3Es): adaptação para o contexto Português [Scale of Effectiveness of Teams (3Es): adaptation to the Portuguese context]. *Lab. Psicol.* 12, 3–13. doi: 10.14417/lp.884
- Wildman, J. L., Salas, E., and Scott, C. P. R. (2014). Measuring cognition in teams: a cross-domain review. *Hum. Fact. J. Hum. Fact. Ergon. Soc.* 56, 911–941. doi: 10.1177/0018720813515907

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sinval, Aragão e Pina, Sinval, Marôco, Santos, Uitdewilligen, Maynard and Passos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.