


Please cite the Published Version

Carrington, Stuart, Turner, Martin , North, Jamie and Brady, Abbe (2025) Development, Validation, and Reliability of the Irrational Beliefs Scale for Sports Officials (IBSSO). *Frontiers in Psychology*, 16. 1571447 ISSN 1664-1078

DOI: <https://doi.org/10.3389/fpsyg.2025.1571447>

Publisher: Frontiers Media S.A.

Version: Published Version

Downloaded from: <https://e-space.mmu.ac.uk/639259/>

Usage rights:  [Creative Commons: Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

Additional Information: This is an open access article which appeared in *Frontiers in Psychology*

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

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



OPEN ACCESS

EDITED BY

Ian Cunningham,
Edinburgh Napier University, United Kingdom

REVIEWED BY

Noel E. Brick,
Ulster University, Ireland
Aydın Karaçam,
Bandırma Onyedi Eylül University, Türkiye

*CORRESPONDENCE

Martin James Turner
✉ m.turner@mmu.ac.uk

RECEIVED 05 February 2025

ACCEPTED 27 March 2025

PUBLISHED 10 April 2025

CITATION

Carrington SC, Turner MJ, North JS and
Brady A (2025) Development, validation, and
reliability of the Irrational Beliefs Scale for
Sports Officials (IBSSO).
Front. Psychol. 16:1571447.
doi: 10.3389/fpsyg.2025.1571447

COPYRIGHT

© 2025 Carrington, Turner, North and Brady.
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.

Development, validation, and reliability of the Irrational Beliefs Scale for Sports Officials (IBSSO)

Stuart C. Carrington¹, Martin James Turner^{2*}, Jamie S. North¹
and Abbe Brady¹

¹Faculty of Sport, Technology and Health Science, St Mary's University, Twickenham, London, United Kingdom, ²Manchester Metropolitan University, Manchester, United Kingdom

Application of Rational-Emotive Behavior Therapy (REBT) within performance environments is increasing, however measures that assess irrational beliefs in specific populations are encouraged. A population that may benefit from REBT is sports officials. This paper reports the development, validation and reliability of the Irrational Beliefs Scale for Sports Officials (IBSSO). Item development was drawn from original items of the Irrational Performance Beliefs Inventory (iPBI), then refined over three stages using an expert panel, novice panel and industry panel. Officials ($N = 402$; 349 male, 50 female, 3 undisclosed) from 11 sports (M years' experience = 13.02; $SD = 10.24$) completed the inventory, with exploratory factor analysis suggesting a 3, 4, and 5-factor model from 22 remaining items. A new sample of 154 officials (140 male, 12 female, 2 undisclosed) representing 9 sports (M years' experience = 14.61, $SD = 11.96$) completed the IBSSO, along with 6 other related measures (e.g., Automatic Thoughts Questionnaire, Affective Reactivity Index) to assess criterion validity. A four-factor model showed acceptable fit, with self-depreciation, peer rejection demands, emotional control demands, and approval identified as subscales, as well as a three-factor model. The IBSSO was positively correlated with the additional measures and negatively correlated with age, demonstrating concurrent validity. To assess convergent validity, 94 new officials (83 male, 10 female, 1 undisclosed; $M_{age} = 36.74$ years, $SD = 15.03$) completed the IBSSO and iPBI simultaneously. The IBSSO was positively correlated with the iPBI, indicating convergent validity. Furthermore, 29 officials (25 male, 4 female, M years' experience = 14.57, $SD = 12.44$) completed the IBSSO over three-time points, with a repeated-measures MANCOVA and Intra-Class Coefficients confirming test-retest reliability. The 16-item four-factor model was accepted based on statistical and theoretical fit. The paper presents a measure of irrational beliefs in sports officials, with investigation into the effectiveness of REBT with this population recommended.

KEYWORDS

confirmatory factor analysis, referees, rational-emotive behavior therapy, exploratory factor analysis, officials

Introduction

Adverse events are an inevitable part of life. Although the impact of adverse events on undesirable emotions (e.g., depression and anxiety) is stable between-persons (i.e., unfortunate events promote undesirable emotions), the relationship within-persons is bidirectional (e.g., depressed individuals are influenced by their environment but also actively shape it; Maciejewski et al., 2021). The bidirectional relationship between the individual and their environment on emotional outcomes is consistent with the fundamental assumption of Rational Emotive Behavior Therapy (REBT; Ellis, 1957), namely that it is not an adverse

situation in isolation that promotes maladaptive behaviors and emotions, but rather the individual's beliefs about that situation. As the elimination of negative events is impractical, unavoidable, and perhaps undesirable (see Seery, 2011), increased discussion and application of strategies to modify undesirable emotional responses to adverse events is worthwhile.

REBT is situated within the psychotherapeutic approach of cognitive behavioral therapy and is focussed on the role played by cognitions in emotional (and behavioral) responses. Consequently, it rejects traditional "Adversity-Consequence" (A-C) models to explain responses to negative situation and adopts an interactive "Goals-Adversity-Beliefs-Consequence" (GABC) model (see Figure 1). In this model, adverse events (A) to one's goals (G) are met with unhealthy or healthy negative emotions (C), depending on one's beliefs (B).

According to REBT the beliefs that create unhelpful emotions (e.g., anxiety) are labeled irrational as the subsequent behaviors they promote (e.g., avoidance or withdrawal) do not assist an individual in reaching their goal (Turner, 2022). Dryden and Branch (2008) identified four irrational beliefs: one primary belief (demands) and three secondary beliefs (awfulizing, self-deprecation and low-frustration tolerance). In contrast, beliefs that promote adaptive emotions and behaviors (e.g., concern and dealing with the perceived threat) are identified as rational as they assist the individual in goal achievement (Ellis and Dryden, 1997). Rational beliefs are identified as preferences (primary belief), anti-awfulizing, self-acceptance and high-frustration tolerance (secondary beliefs; Dryden, 2014; Dryden and Branch, 2008).

It is the promotion of adaptive emotions that has seen REBT successfully applied in a variety of contexts including education (e.g., Caruso et al., 2017), the military (e.g., Jarrett, 2013), clinical treatment (e.g., Hyland et al., 2013), and sport and exercise (e.g., Turner and

Bennett, 2018). For instance, REBT has been credited with improving performance, increasing resilience, decreasing anxiety, and enhancing physical activity levels in elite and non-elite participants (Deen et al., 2017; O'Connor, 2018; Turner et al., 2020; Wood et al., 2017). A recommendation drawn from the successful application of REBT in sporting environments is for researchers to investigate and understand the benefits of REBT on other agents and stakeholders in sport and exercise (Jordana et al., 2020) such as sports officials.

Sports officials (e.g., referees, umpires and judges) have recently been the focus of increased attention from academic researchers (Hancock et al., 2021). However, research on sports officials has typically focussed on physical performance (e.g., Castillo, 2016, 2017; Weston et al., 2011), well-being (see Webb et al., 2021) and the use of video-training to develop decision-making skills (e.g., Samuel et al., 2019; Spitz et al., 2018). While video-training interventions have been shown to improve perceptual-cognitive skills (e.g., Helsen and Bultynck, 2004), such a reductionist approach fails to acknowledge the role of individual differences such as experience, social support and emotional regulation that mediates performance. For instance, Page and Page (2010) reported that the experience of sports officials mediates the impact of contextual constraints (e.g., crowd noise) on performance, while Praschinger et al. (2011) concluded that subjective "thresholds" for game management strategies (defined as the shift from accurate to adequate decisions; Raab et al., 2021) influenced decision-making following verbal abuse from players. Such differences mediate the impact of social constraints, such as crowd influence and player behavior, on officiating performance because of a desire to receive praise or avoid criticism (see Dohmen and Sauermann, 2016). Although research involving REBT and officials is in its infancy, four REBT sessions improved the performance of one rugby union official

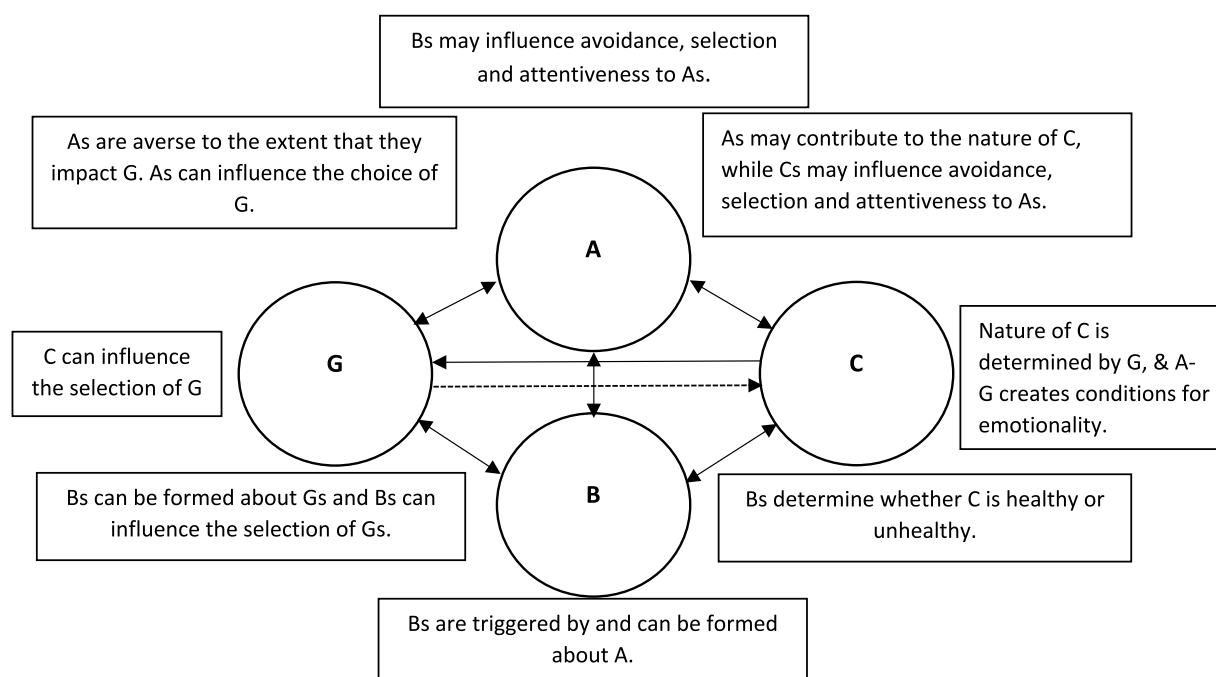


FIGURE 1

The interdependent GABC framework. Dotted line reflects the more distal relationships between G and C due to mediating effect of A that promotes negative emotion, not the G in isolation (adapted from Turner, 2022). G, goal; A, adverse event; B, beliefs; C, consequences (emotional and behavioral).

and decreased anxiety in two rugby union referees, with this improvement maintained over a 12-week period (Maxwell-Keys et al., 2022).

Minimizing unhealthy negative emotions, such as anxiety, may benefit performance by mediating situational appraisals. Irrational beliefs have been associated with increased threat appraisals (e.g., evaluation of future loss or harm), with rational beliefs associated with challenge appraisals (e.g., perceived future gain; Lazarus, 1991) due to an outcome being perceived as incongruent to goal achievement (Chadha et al., 2019; Lazarus, 1999). Promoting challenge appraisals may be especially helpful for sports officials, given the contextual constraints they operate within. Basketball officials, for instance, reported varying level of stress, challenge and threat appraisals depending on the time and score-line of a game (Ritchie et al., 2017). Although these situations were assessed using hypothetical scenarios, limiting the external validity of the findings, the identification of situations that cause officials to experience varying emotional outcomes is consistent with previous research. For example, Neil et al. (2013) demonstrated that adverse events promote maladaptive emotional and behavioral consequences in soccer referees (e.g., avoidance of appropriate disciplinary action for fear of potential conflict or criticism). Consequently, attempts to minimize maladaptive responses to negative events, such as “avoidance” (when an official takes, or does not take, action to achieve a desirable reaction from other agents; see Nevill et al., 2017) and self-depreciation (Mansell, 2021), are justified in pursuit of improved performance of sport officials. The focus of REBT on producing goal-congruent emotions and behaviors that mediate officiating performance (Bruch and Feinberg, 2017; Kostrna and Tenenbaum, 2022) alongside its short-term effectiveness, that is valued within officiating environments (Bowman and Turner, 2022; MacMahon et al., 2015), underlines the potential benefit of application in sports officials.

Irrespective of the population studied, however, there are important limitations of many traditional measures of irrational beliefs. Original measures of irrational beliefs, such as the Rational Behavior Inventory (Shorkey and Whiteman, 1977), contained items that assessed emotional and behavioral outcomes as well as beliefs (Ramanaiah et al., 1987). For instance, Robb and Warren (1990) demonstrated that the item, “I often get excited or upset when things go wrong,” assessed both the emotional outcome and its frequency of occurring following an adverse event. This is inconsistent with REBT theory, which states beliefs are the only cognitions that need to be assessed (Dryden and Branch, 2008). Such conceptual problems prompted Terjesen et al. (2009) to review psychometric measures of irrational beliefs and consequently recommend the need to develop a novel measure as, while some inventories reported good reliability (e.g., the Shortened General Attitude Belief Scale; SGABS, Lindner et al., 1999), all of the measures failed to provide a correlation with other measures of irrationality or cognitive disturbance, which would be expected, along with a lack of reporting around discriminant validity. Consequently, Terjesen et al. (2009) made the following recommendations regarding the development of an irrational beliefs inventory:

- a) items should only reflect the assessment of beliefs and not emotional or behavioral consequences
- b) items not only measure agreement/disagreement with irrational beliefs but also agreement/disagreement with rational beliefs

- c) items must contain roughly the same amount of the core irrational beliefs unless one is more clinically useful than others
- d) single item responses are avoided as they fail to account for the belief strength
- e) that the construct validity of the measure is not only reported but established *a priori* by having experts in REBT agree that items support the theoretical aims
- f) internal reliability should be measured and reported with a minimum Cronbach's α of 0.70 met (Nunnally, 1972)

The development of the Irrational Performance Beliefs Inventory (iPBI; Turner et al., 2018a) aimed to address the recommendations made by Terjesen et al. (2009). The inventory was developed after consultation with an expert panel to generate, eliminate and/or amend any items that were not consistent with REBT theory. Specifically, items were sent to the expert panel in an Excel spreadsheet with drop down menus adjacent to each item. Expert panel members were asked to select which irrational belief and which subscale they believed the item measured. Additionally, there was an opportunity to add notes for each item, should the expert panel feel the item was ambiguous in any sense (theoretically, wording or otherwise). Following this step, a novice panel consisting of psychology graduates with little or no knowledge of REBT theory assessed the clarity, wording, and length of the items. Items measured agreement and disagreement of the four irrational beliefs (demands, low-frustration tolerance, self-depreciation and awfulizing) and account for belief strength by requesting responses on a Likert scale. The scale also reported internal consistency coefficients of between 0.90 and 0.96, exceeding the criteria for excellent test score reliability (Nunnally and Bernstein, 1994; Turner et al., 2018a). Importantly, the iPBI is a context specific measure. In this case, the chosen context is performance, which is relevant as the application of REBT has been shown to be effective in sporting and business environments (e.g., Turner and Barker, 2015; Turner and Bennett, 2018).

While the general context of performance featured within the iPBI is a benefit of the measure, a further, and important, recommendation for inventory development made by Terjesen et al. (2009) is that measures test specific populations. This endorsement is supported by Ziegler and Horstmann (2015), who advised that the construction of measures reflect theoretical advances that, while behaviors are influenced by individual personality, it is that individual's *perception* of a specific situation that needs to be assessed (see Fleenor and Nofle, 2009; Funder, 2006). Therefore, adapting the iPBI for application involving target populations has three benefits. First, it would meet contemporary recommendations for bespoke measurement design, resulting in a higher signal-to-noise ratio and allow for greater awareness of the prevalence of irrational beliefs, a neglected area of research (Pargent et al., 2019; Turner, 2014). Second, it would provide insight into how contextually relevant situations are perceived by those directly involved, assisting populations who may benefit from interventions involving REBT (Jordana et al., 2020; Turner et al., 2018a; Ziegler and Horstmann, 2015). Finally, within the context of sports officials, the development of a measure that assesses how sports officials perceive relevant situations and their likely behavioral and emotional outcomes would assist in the desired holistic development of this population (e.g., Raab et al., 2021; Webb, 2017).

The aim of this study was to develop a version of the iPBI for specific and bespoke application with sport officials to provide REBT

practitioners with a practical and contextually relevant tool for assessing core irrational beliefs with this population. A secondary aim was to establish the criterion and convergent validity of the new measure. As age and experience has been shown to reduce irrational beliefs (Ndika et al., 2012; Turner et al., 2019), it was hypothesized that younger and less experienced officials would report higher levels of irrational beliefs than their older and more experienced counterparts. A final aim was to assess test–retest reliability of the new measure.

Preparatory work—item development

To develop and validate the measure, a three-stage process was undertaken and recorded using a decision-making log to enhance reporting accuracy. The first stage was item generation, the accuracy of which is essential to the development of a good measure (see Turner et al., 2018a). One hundred and thirty-three items from the original iPBI were reviewed and, if necessary, amended for consideration. Items were amended to enhance their relevance to the target population, as recommended by Terjesen et al. (2009), however it was necessary to ensure that items were not too specific. Avoidance of specificity is consistent with REBT theory (see Dryden and Branch, 2008; Ellis, 2005) which emphasizes the need to prevent a conditional “should/demand” to meet a future condition, such as “if I want to perform well as an official, it would be terrible to get important decisions wrong.” “To perform well” is a future condition, which is met, in part, with decision-making accuracy, whereas the scale’s purpose is to measure irrational beliefs. Therefore, an item relating to decision-making, identified as integral to the performance of sports officials (see Bruch and Feinberg, 2017; Samuel et al., 2020), was presented as “it’s terrible to fail by getting important decisions wrong,” measuring awfulizing, a secondary irrational belief. Such amendments increased relevance to sports officials as did the inclusion of two new subscales, control (14 items; 7 regarding self-control and 7 regarding control of others) and responsibility (7 items), reported as valued qualities in sports officials (see Carrington et al., 2022; Morris and O’Connor, 2017; Neil et al., 2013). Two of the authors, one of whom holds the Primary Certificate in REBT and one of whom holds the Advanced Practicum in REBT, reviewed each item. Following the same process that led to the development of the iPBI (Turner et al., 2018a), items were removed if they were judged to measure the behavioral or emotional response, rather than the belief, or if the primary belief and subscale measured could not be agreed upon. Consistent with the view of DeVellis (2012) that conservatism is the optimum approach for scale development, a policy adopted throughout the process, items that appeared repetitive or could not be amended to be more relevant for sports officials were retained, resulting in 90 items progressing to the second stage of development.

Following item development, the scale underwent review from three panels; expert, novice and a panel of experts from the target population (officials’ panel). This approach assists theoretical consistency, content validity and the possibility of refinement (Grant and Davis, 1997; Turner et al., 2018a). Experts ($n = 3$), each qualified REBT practitioners, were asked to assign an irrational belief (e.g., demand) and subscale (e.g., rejection) to each item. Four items were removed (“I cannot stand being snubbed by people I trust”; “I need to avoid failing in making important decisions”; “I need to be in control of others”; “I am no good if I do not succeed in controlling others”) following discrepancy across the expert panel regarding the irrational belief being measured. The remaining 86 items were subjected to review by a novice panel consisting of post-graduate psychology students ($n = 4$), with no training or grounding in REBT theory, to assess item clarity (see DeVellis, 2012). The panel reported that many items appeared to be repeated and that assessment of the “control” subscale was unclear. Following discussion with co-authors, similar items were retained to allow for statistical analysis and some items were refined to provide greater clarity (e.g., “if I am not in control then it means I am useless” became “if I am not in control of my emotions then it means I am useless”). An industry panel of football officials ($n = 3$, male = 2, female = 1, representing Level 1a, Level 3 and Level 5), were asked to review the clarity, appropriateness and relevance of the language used in each item to enhance social validity, defined as the social significance of goals and acceptability of potential intervention procedures (Common and Lane, 2017), within the target demographic. Eight items were reported as unclear. From these eight items, one item (“I would be useless if others threatened my status among my colleagues”) was removed due to incoherence reported by the officials’ panel. The other 7 items were amended where possible (e.g., the clarity of “it’s awful if others do not approve of me” was questioned as the opinion of some was considered more important than others, and so became “it’s awful if others that are important to me do not approve of me”), leaving 85 items, including 22 unchanged items from the iPBI.

The final stage involved repeating the panel process with one member of each panel to evaluate the clarity and validity of the remaining items. Prior to this stage, items assessing demands were amended to include the preference (e.g., “I need others to approve of my actions” became “I would like, and therefore I need, others to approve of my actions”), to maximize internal validity by making explicit the demand being measured, as well as including the positively phrased item, “I want to be, therefore I must be, in control of my emotions.” This was viewed favorably by the REBT practitioner, improving construct validity *a priori* (see Terjesen et al., 2009) and the remaining panels reported no other concerns. Ethical approval to collect data using the 86-item measure was received by the Ethics Committee of the University of the first author.

The final stage involved repeating the panel process with one member of each panel to evaluate the clarity and validity of the remaining items. Prior to this stage, items assessing demands were amended to include the preference (e.g., “I need others to approve of my actions” became “I would like, and therefore I need, others to approve of my actions”), to maximize internal validity by making explicit the demand being measured, as well as including the positively phrased item, “I want to be, therefore I must be, in control of my emotions.” This was viewed favorably by the REBT practitioner, improving construct validity *a priori* (see Terjesen et al., 2009) and the remaining panels reported no other concerns. Ethical approval to collect data using the 86-item measure was received by the Ethics Committee of the University of the first author.

Study 1: exploratory factor analysis

Method

Participants

A final sample of 402 sports officials participated in the Exploratory Factor Analysis stage of the inventory development ($M_{\text{age}} = 41.20$, $SD = 14.56$). Of this sample, 349 were male, 50 were female, and three chose not to disclose their gender. As identification of officiating level is challenging due to varied classification across sports (see Webb et al., 2021), years’ experience of officiating was requested with the sample reporting with an average of 13.02 ($SD = 10.27$) years’ experience as a qualified sports official. The sample of participants comprised sports officials of 31 different nationalities, representing 5 continents and 11 sports (football, rugby union, rugby league, field hockey, netball, handball, cricket, GAA, baseball, badminton and futsal). While sample size for statistical validation analyses is disputed and frequently seen as arbitrary (see Brown, 2015), the final sample exceeded general guidelines of 200–300 observations (Boateng et al., 2018) and the recommendation of 5 participants per new item ($5 \times 65 = 325$; DeVellis, 2012).

Procedure

Data collection took place using the online survey platform Qualtrics (Provo, UT). Links to surveys were distributed via social media (e.g., Twitter, LinkedIn) and personal emails to gatekeepers such as officiating development officers, Referee/Officiating Associations and national governing bodies. After confirming that they had read the participant information sheet, participants were asked to confirm their consent to participate in the research study. Participants completed an 86-item survey by indicating item response on a Likert-scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), endorsed by Terjesen et al. (2009) to ascertain strength of belief.

Preliminary analysis

Data from an original sample of 513 participants was assessed for suitability prior to data analysis. One hundred and five responses were removed due to the absence of a significant amount of data points (e.g., participant dropout after commencing the study). Additionally, a further 6 responses were removed due to evidence of patterned responding, leaving 402 participants. From this sample, isolated cases of missing data were observed. Little's Missing Completely at Random (MCAR) test indicated that items were not entirely missing at random, $\chi^2(2,705) = 2,268, p < 0.001$, however removing incomplete cases was not considered prudent as (a) it would decrease the sample size and (b) the impact of data imputation (e.g., expectation maximization; where average scores are projected) is negligible when the percentage of missing data is lower than 5% (Schäfer, 1999). As the largest percentage of missing data from any item was 1%, and data imputation has been shown to improve efficiency of effect estimates compared to complete case analysis (Madley-Dowd et al., 2019), expectation maximization was used to impute missing data.

Data analysis

Reflecting the development and validation of the iPBI (see Turner et al., 2018a), the initial aim of analysis was to identify the factor structure of the measure and examine internal reliability, avoiding the assumption that the iPBI and new items conformed to the 4-factor structure. This was accomplished through EFA using SPSS Statistics 28 (Microsoft, Albuquerque, NM). Sampling adequacy and suitability of data for factor analysis was assessed using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity. An EFA with oblique rotation (oblimin) was run on all 86 items. Items were retained if pattern matrix reported a primary factor correlation above 0.50 and secondary factor correlation below 0.30, as well as communality exceeding 0.40 (see Costello and Osborne, 2005; Field, 2009), with lowest scoring items removed in an iterative process. A maximum likelihood factor analysis was then conducted, with factors reporting eigenvalues greater than 1 deemed to meet the Kaiser's criterion to determine the optimum factor solution for the data as per previous EFAs (e.g., Werner et al., 2014) and recommendations (e.g., Field, 2009). However, other sources (e.g., Velicer and Jackson, 1990) state that only retaining factors with eigenvalues greater than 1 lacks accuracy. Consequently, to determine the number of factors to extract, the scree test was used, identified as the optimum choice for researchers conducting an EFA (Costello and Osborne, 2005).

Results

Sampling adequacy exceeded the desired value of greater than 0.70 (KMO = 0.940) and was categorized as "marvelous" (Lloret et al., 2017; Kaiser, 1974). Factor analyses was deemed suitable by Bartlett's test of sphericity, $\chi^2(3,655) = 21570.00, p < 0.001$. Fifty-two items were removed for failing to meet the criteria regarding correlation, factor loading and communality scores. Visual analysis of the scree plot, recommended as best practice to determine factor extraction and a strategy used in previous studies (Costello and Osborne, 2005; Turner et al., 2021; Watkins, 2018), suggested items loaded onto six-factors and therefore 6 factors were extracted. Results of the factor analysis for the remaining 22 items can be seen in Table 1.

When items that loaded onto each factor were analyzed, clear dimensions were identified (Table 1). Four dimensions could be clearly identified as self-depreciation (e.g., "If I fail in things that matter to me, it means I am a failure"), peer rejection demands (e.g., "I do not want to be, therefore I must not be, dismissed by my peers"), emotional control demands (e.g., "I want to be, therefore I must be, in control of my emotions") and approval (e.g., "I would like, and therefore I need, others to approve of my actions"). The fifth factor consisted of two items ("I do not want to, so I absolutely should not, fail in things that matter to me" and "It's terrible if the members of my match-day team do not respect me") that did not appear to be related and was labeled as "miscellaneous." Although six factors were extracted, factor loadings indicated five clear factors (ranging from 0.60 to 0.78, 0.51 to 0.92, 0.55 to 0.74, 0.57 to 0.75, and 0.54 to 0.73 respectively).

Discussion

The purpose of the EFA was to identify factors found within the measure. Six factors were extracted based on the scree plot being identified as the most advisable indicator (Costello and Osborne, 2005; Field, 2009), however the sixth factor was not the primary factor for any item. Therefore, the EFA suggested a five-factor model should be the initial model tested at the confirmatory factor analysis (CFA) stage. Although five factors were identified, it was unclear as to what the fifth factor was measuring based on visual analysis of its items ("I do not want to, so I absolutely should not, fail in things that matter to me" and "It's terrible if the members of my match-day team do not respect me") and was therefore categorized as "miscellaneous." Following a "global assessment" (e.g., checking primary loadings and wording of items within each factor; see Costello and Osborne, 2005), the other four factors were clear and, based on the belief the items within the factors were measuring, categorized as: self-depreciation, peer rejection demands, emotional control demands, and approval. With the exception of self-depreciation, these factors differed from those identified in the iPBI. Although this was not hypothesized, identifying contextually relevant factors reflects the benefit of developing population-specific measures. Furthermore, although "approval" contained one item assessing low frustration tolerance and one item assessing awfulizing, their subordinate position to demands in the structure of irrational beliefs (see David et al., 2005; Dryden and Branch, 2008), suggesting that approval, the first or Ellis' original 12 irrational beliefs (Ellis and Dryden, 1997), is of particular, contextual importance for sports officials.

Although the sample included participants for whom English was not their first language, leading to potential discrepancy in

1 Analysis of missing item data is available upon request.

TABLE 1 Exploratory factor analysis of the 22 retained items.

Items	Communalities	Factor						Eigenvalue	% variance explained	Dimension
		1	2	3	4	5	6			
I83DEP	0.56	0.78								Self-depreciation
I70DEP	0.60	0.76					−0.19			
I48DEP	0.56	0.75								
I22DEP	0.52	0.71		0.14	−0.17					
I14DEP	0.58	0.70		0.17			0.16	7.219	32.812	
I38DEP	0.59	0.68			0.15	0.15				
I53DEP	0.53	0.67				0.11	−0.17			
I58DEP	0.43	0.62								
I17DEP	0.47	0.62								
I9DEP	0.41	0.60			0.15	−0.10				
I79DEM	0.79		0.92							Peer rejection demands
I75DEM	0.56		0.54		0.19	0.11		2.499	11.359	
I78AWF	0.39		0.51			0.17				
I20DEM	0.53			0.74						Emotional control demands
I64DEM	0.57			0.72				1.952	8.871	
I33LFT	0.40			0.55	0.13		0.13			
I15DEM	0.65				0.75		−0.27			Approval
I5DEM	0.56				0.70			1.204	5.473	
I3AWF	0.54				0.60	0.10	0.29			
I32LFT	0.51	0.11			0.57					
I28AWF	0.62	0.23	0.43	0.32	0.25	0.73				Misc.
I26LFT	0.44		0.18			0.54		0.868	3.946	

Extraction method; maximum likelihood; Rotation method; oblimin with Kaiser normalization. Loadings larger than 0.60 are in bold. DEP, Depreciation; DEM, Demands; AWF, Awfulizing; LFT, Low frustration tolerance; Misc., Miscellaneous.

understanding measured constructs (see [Chotpitayasunondh and Turner, 2019](#)), an advantage of such diversity is the potential for generalization regarding the validity and reliability of the measure cross-culturally ([Costello and Osborne, 2005](#)).

Study 2: confirmatory factor analysis and criterion validity

Method

Participants

To avoid cohort effects a new sample was used to test model fit of the measure. This sample consisted of 154 participants ($M_{\text{age}} = 39.92$ years, $SD = 16.26$), 140 of whom were male, 12 were female, and two participants chose not to disclose their gender. Average years' experience as a qualified sports official within the sample was 14.61 years ($SD = 11.96$), with participants officiating for an average of 2.95 h per week ($SD = 1.14$). Participants came from 12 different countries across four continents, and included officials from nine different sports (football, rugby union, rugby league, field hockey, basketball, baseball, cricket, Australian rules football and lacrosse). No missing data was reported in the sample, therefore the sample size

exceeded the recommendation of 150 participants for CFA ([Muthén and Muthén, 2002](#)).

Procedure

As with the EFA stage, data was collected using the online survey platform Qualtrics (Provo, UT). Again, links to surveys were distributed via social media and personal emails to gatekeepers. Participants were instructed to only complete the survey had they not taken part in data collection for the EFA stage of inventory development and were asked to check an option to confirm that they had not completed the original survey. Once informed consent was given, participants completed the surveys by indicating item response on a 3, 4, or 5-point scale depending on the measure. This survey consisted of the 22 retained items of the new measure following the EFA, as well as items included from measures that assessed theoretical correlates. Additional measures were added at this stage to maximize the sample used to assess criterion validity.

Measures

Irrational beliefs scale for sports officials

The irrational beliefs scale for sports officials (IBSSO) consisted of 22 statements, with participants required to identify the extent to which

they agreed with each using a 5-point scale (1 = *strongly disagree* to 5 = *strongly agree*). A score for each factor was determined, along with a composite score for the measure, with higher scores indicating stronger irrational demands and self-depreciation.

Positive mental health scale

The positive mental health scale (PMH-scale) (Lukat et al., 2016) consists of 9 items measuring proximal (e.g., emotional) and distal (e.g., social support) factors that contribute to positive mental health, using a four-point scale (1 = *not true* to 4 = *true*). The PMH-Scale reports high internal consistency (Cronbach's $\alpha = 0.93$), good test-retest reliability ($r = 0.77$), convergent and discriminant validity ranging from 0.26 to 0.81, and is easy to interpret (Lukat et al., 2016). These qualities, along with its assessment of adaptive emotional and behavioral responses, justified the choice to use this as a measure to compare with the IBSSO.

Patient health questionnaire

The patient health questionnaire (PHQ-9) (Kroenke et al., 2001) is a commonly used and validated measure to assess depression in primary care (Cameron et al., 2008). Comprised of 9 items, participants indicate the severity of depression symptoms using a four-point scale (1 = *not at all* to 4 = *every day*). The PHQ-9 reports excellent internal reliability (Cronbach's $\alpha = 0.89$) and excellent test-retest reliability ($r = 0.84$; Kroenke et al., 2001). The assessment of depression severity was chosen to evaluate the criterion validity of the self-depreciation items within the IBSSO.

Perth emotional reactivity scale-short form

The PERS-S (Preece et al., 2018) is a shortened version of the Perth Emotional Reactivity Scale (PERS; Becerra et al., 2017), a measure that distinguishes between positive and negative trait emotional reactivity and reports good to excellent concurrent validity (r 's from 0.80 to 0.98) and internal reliability (Cronbach's α ranging from 0.79 to 0.94 for all factors; Becerra et al., 2017). The shortened version (consisting of 18-items measured using a 5-point scale) was chosen to assist the evaluation of criterion validity to maximize likelihood of survey completion, coupled with its assessment of ease and intensity of emotional activation.

Automatic thoughts questionnaire

The automatic thoughts questionnaire (ATQ-15) (Netemeyer et al., 2002) is a 15-item questionnaire that requires participants to assess the degree of believability across a range of cognitions using a 5-point scale (1 = *not at all* to 5 = *all the time*). The original 30-item inventory (Hollon and Kendall, 1980) reported good to excellent test-retest reliability and good internal stability (Cronbach's $\alpha = 0.97$) with a non-clinical sample. It is also frequently used as a measure of therapeutic progress (see Zettle et al., 2011). The 15-item survey was chosen to maximize response rate and the ATQ-15 also reports a negative correlation with self-esteem ($r = 0.63$), and a positive correlation with social anxiety ($r = 0.56$) and obsessive thoughts ($r = 0.70$; Netemeyer et al., 2002). Its correlation with obsessive thoughts, combined with its assessment of rigid (obsessive) cognitions, informed its selection to compare with the IBSSO.

Affective reactivity index

To assess criterion validity of low frustration tolerance within the IBSSO, the affective reactivity index (ARI) (Stringaris et al., 2012) was

selected. Comprised of six items to assess irritability using a 3-point scale, the ARI reports high self-reporting internal reliability (Cronbach's $\alpha = 0.90$) and has been validated with UK and US samples, reflecting the international sampling regarding the development of the IBSSO.

The social phobia inventory

The social phobia inventory (SPIN) (Connor et al., 2000) is a 17-item self-reported survey that asks participants to indicate the extent to which they have been disturbed by symptoms of social anxiety during the previous 7 days on a 5-point scale (1 = *not at all* to 5 = *extremely*). The measure assesses a range of factors associated with social phobias, with fear of talking to strangers, criticism, authority figures, public speaking, physiological distress and negative social evaluation identified as factors (Connor et al., 2000; Campbell-Sills et al., 2015). The SPIN reports good concurrent validity ($r = 0.57$) with similar measures, such as the Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987) and the Brief Social Phobia Scale (BSPS; Davidson et al., 1997), and excellent internal consistency (Cronbach's $\alpha = 0.95$; Connor et al., 2000). The significance of the factors identified within the survey justified its selection in comparison with the IBSSO.

Data analysis

A confirmatory factor analysis (CFA) was utilized due to five key advantages identified by Brown (2015). First, while the new measure was grounded in theory, there is a requirement for empirical evidence to accept, modify, or reject a measure (see Credé and Harms, 2019). Confirmation of the internal reliability measure of the measure can also decrease the likelihood of Type I and Type II errors, hence the need for statistical analysis (see Credé et al., 2012). Second, as REBT is multifactorial, analysis of both compound and subscale scores is desirable and supported by CFA. Third, CFA estimates scale dimensions allowing for tests of scale reliability, a traditional weakness in reporting practices of new measures (cf. Jackson et al., 2009; Terjesen et al., 2009). Fourth, CFA can be used to produce convergent and discriminant validity, required as irrational beliefs are a multidimensional construct. Finally, CFA allows for the generalization of groups, allowing the measure to identify prevalence of irrational beliefs on specific populations which has been identified as limited and in need of development (Turner, 2014; Turner et al., 2018a).

To overcome difficulties in selecting a goodness of fit index (e.g., contextual influence such as sample size on index choice, cut-off values etc.; see Hu and Bentler, 1999), multiple fit indices were identified for use: absolute fit, parsimony correction and comparative fit. Hence, goodness of fit was analyzed using the χ^2 statistic, standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA) and the comparative fit index (CFI). For transparency, values to establish reasonably good fit between the three, four and five-factor structures of the new measure and obtained data was identified as close to or below 0.08 for the SRMR, close to or below 0.06 for the RMSEA and close to or greater than 0.95 for the CFI. These values are consistent with guidelines outlined by Brown (2015) and those used to validate the original iPBI (Turner et al., 2018a). Internal reliability coefficients (Cronbach's α) for each factor were also calculated, with coefficients exceeding 0.70 indicating good test score reliability (Nunnally and Bernstein, 1994). Data was analyzed using SPSS AMOS v28 (Microsoft, Albuquerque, NM), with three, four and five-factor models being tested for best fit. To develop

model fit, an iterative process whereby the item that reported the lowest standardized factor loading was removed until goodness of fit was acceptable. Analysis began with testing the five-factor model, with an iterative process using the identified fit indices driving further analyses. A five-factor, four-factor, and three-factor model were subsequently tested. To analyze criterion validity, subscales of the new measure were correlated with the additional measures, as well as with participant age, experience and hours per week spent officiating. Positive correlations were expected between the IBSSO and the PERS-S, PMQ-9, ATQ-15, ARI and SPIN measures. Negative correlations were expected between the IBSSO and age, experience and the PMH-Scale.

Results

Construct validity

Initial analysis of the 22-item five-factor model produced an insufficient fit according to the criteria, $\chi^2(199) = 312.68$, $p = 0.001$, SRMR = 0.08, CFI = 0.86, RMSEA = 0.07. The iterative removal of 5 items to improve model fit resulted in the removal of the fifth factor entirely, hence final model fit could not be identified, and the model was rejected.

Initial analysis of the four-factor model produced an unacceptable fit, $\chi^2(203) = 341.27$, $p < 0.001$, SRMR = 0.08, CFI = 0.84, RMSEA = 0.08. An iterative process of removing the lowest loading items, however, led to model fit. In addition to the two items that loaded onto the fifth factor, four items (“It’s awful not to be treated fairly by my peers,” “I cannot stand not being in control of my emotions,” “If I do not act responsibly, it shows I am worthless” and “If I fail in things that matter to me, it means I am a failure”) were removed after reporting factor loadings of 0.409, 0.502, 0.495 and 0.564, respectively. This 16-item model produced an acceptable model fit, $\chi^2(99) = 138.23$, $p = 0.006$, SRMR = 0.06, CFI = 0.95, RMSEA = 0.05, standardized factor loadings were between 0.51 and 1.06 and error variances were between 0.26 and 1.13. Internal consistency (alpha reliability) coefficients were between 0.54 and 0.88, meaning three of the four factors were between acceptable and good, and one (Peer Rejection Demands) was below the acceptable range (0.70; Field, 2009).

Initial analysis of the three-factor model produced an unacceptable fit, $\chi^2(101) = 154.18$, $p = 0.001$, SRMR = 0.08, CFI = 0.90, RMSEA = 0.07. Iteratively, three low factor loading items were removed (“It’s awful not to be treated fairly by my peers,” 0.395; “If I do not act responsibly it shows I am worthless,” 0.497; and “I cannot stand not being in control of my emotions,” 0.527) and a subsequent CFA containing the remaining 13 items produced an acceptable fit to the three-factor structure suggested from the EFA, $\chi^2(52) = 56.72$, $p = 0.303$, SRMR = 0.05, CFI = 0.99, RMSEA = 0.02. For the 13-item measure, standardized factor loadings were between 0.58 and 1.09 and error variances were between 0.33 and 1.18. Internal consistency (alpha reliability) coefficients were between 0.59 and 0.85, meaning three of the four factors reported acceptable to good consistency, and one (Peer Rejection Demands) was below the acceptable range (0.70; Field, 2009).

The identification of an item with a factor loading greater than 1 in both models (item 64: “I want to be, therefore I must be, in control of my emotions”) represents a likely Heywood case (see Farooq, 2022).

The preferred matrix for analysis (Watkins, 2018), the pattern matrix, did not indicate this factor as poorly defined (see Table 1) and therefore equality constraints were imposed on both items in the factor, a common action buttressed by previous literature (Jung and Takane, 2008; Mustaffa et al., 2018). Standardized factor loadings then ranged from 0.54 to 0.81 and 0.58 to 0.80, with error variances between 0.33 to 0.61, and 0.33 to 0.65, for the 4-factor and 3-factor models, respectively (Tables 2, 3).

Concurrent validity

Correlations between the 16-item IBSSO, 13-item IBSSO, and other measures were assessed to test concurrent validity (Tables 4, 5 respectively). Low to medium correlations were mostly reported, with the largest correlations for the self-depreciation factor found between the SPIN ($r = 0.44$ for both models) and the ATQ-15 ($r = 0.42$ and $r = 0.43$ for the 4 and 3-factor models respectively). The composite score of the 4-factor model was significantly and positively correlated with all measures except the ARI, and negatively correlated with the PMH-Scale, supporting concurrent and convergent validity of this model in evaluating mental well-being in officiating populations. Similar correlations were found in the 3-factor model, although the relationship between the compound score of the measure and the PERS-S was not significant. As hypothesized, age was negatively correlated with irrational beliefs in sports officials ($r = -0.18$ and $r = -0.25$ in the 4 and 3-factor models respectively), although age was a more accurate predictor of levels of self-depreciation than other factors ($r = -0.27$ in both models). Experience was also negatively correlated with irrational beliefs in sports officials ($r = -0.04$ and $r = 0.06$ in the 4 and 3-factor models respectively), however this variable was not as significant a predictor of irrational beliefs as age.

Discussion

The CFA produced acceptable model fit for both the 4-factor and 3-factor models of the IBSSO, and the correlation between both models and measures that assess similar constructs support both the concurrent and convergent validity of both models. For example, considering the similar constructs evaluated in the self-depreciation factor extant in both models and the SPIN and ATQ-15 (e.g., depression, fear of evaluation and rigid, obsessive thoughts), a significant and positive correlation is not surprising. A small but negative correlation was found between ARI and the IBSSO. While this was not expected *a priori*, the 4 and 3-factor models only containing one or no LFT items, respectively, supports the convergent validity of the IBSSO (e.g., there are no factors that measure LFT). Additionally, it was hypothesized age would be positively associated with irrational beliefs (Ndika et al., 2012; Turner et al., 2019), a result found in both models supporting the concurrent validity of the measure.

Although the 3-factor model reported better fit, along with a larger and significant negative correlation with age, the level of significance ($p = 0.303$) alongside the elimination of the approval factor that consists of an awfulizing and low frustration tolerance item, means caution should be shown when interpreting the results. While models should be statistically sound, factor analysis alone is not sufficient to justify the removal of an item or factor that has theoretical importance (Flora and Flake, 2017). For example, while Bernard

TABLE 2 Standardized solution and fit statistics for the four-factor 16-item model of the IBSSO (statistics in parentheses indicate results following application of equality constraints).

Items (original item number and original core belief)	Error variances	Standardized factor loadings			
		Self-depreciation	Peer rejection demands	Emotional control demands	Approval
If my competencies did not continually develop and improve it would show what a failure I am (I83:DEP)	0.33	0.58			
If I am not given opportunities, then it shows that I am not a worthwhile person (I70:DEP)	0.50	0.71			
If I am snubbed by people that matter to me, it means I am a worthless person (I48:DEP)	0.51	0.72			
If others think I do not make a valuable contribution, then I am no good (I38:DEP)	0.43	0.66			
If I am dismissed by my peers, then that makes me a loser (I53:DEP)	0.64	0.80			
If I am not treated with consideration, it shows I am unlikeable (I58:DEP)	0.46	0.68			
If people think I cannot control others it shows I am worthless (I17:DEP)	0.40	0.64			
If others are not considerate toward me, then it means I am worthless (I9:DEP)	0.51	0.71			
I do not want to be, therefore I must not be, dismissed by my peers (I79:DEM)	0.35		0.59		
I do not want to be, so absolutely should not be, snubbed by people that matter to me (I75:DEM)	0.40		0.63		
I would like to, and therefore I must, avoid losing control of my emotions (I20:DEM)	0.52			0.72	
I want to be, therefore I must be, in control of my emotions (I64:DEM)	0.57			0.76	
I want to be, therefore I need to be, approved of by others (I15:DEM)	0.65				0.81
I would like, and therefore I need, others to approve of my actions (I5:DEM)	0.61				0.78
It's terrible if others think I am not good at officiating (I3:AWF)	0.29				0.54
I could not stand it if others disapproved of me (I32:LFT)	0.36				0.60

Factor	Mean \pm SD	α (reliability coefficient)	Inter-factor correlations		
Self-depreciation (DEP)	18.63 \pm 5.84	0.88			
Peer rejection demands (PRD)	6.35 \pm 1.60	0.54	0.54**		
Emotional control demands (ECD)	7.94 \pm 1.38	0.70	0.19**	0.39**	
Approval (A)	11.28 \pm 3.33	0.77	0.68**	0.65**	0.06**

$\chi^2(99) = 138.23$, $p = 0.006$, SRMR = 0.06, CFI = 0.95, RMSEA = 0.05. DEM, Demands; DEP, Self-depreciation; LFT, Low frustration tolerance; AWF, Awfulizing.

TABLE 3 Standardized solution and fit statistics for the three-factor 13-item model of the IBSSO (statistics in parentheses indicate results following application of equality constraints).

Items (original item number and original core belief)	Error variances	Standardized factor loadings		
		Self-depreciation	Peer rejection demands	Emotional control demands
If my competencies did not continually develop and improve it would show what a failure I am (I83:DEP)	0.34	0.58		
If I am not given opportunities, then it shows that I am not a worthwhile person (I70:DEP)	0.52	0.72		
If I am snubbed by people that matter to me, it means I am a worthless person (I48:DEP)	0.53	0.73		
If I fail in things that matter to me, it means I am a failure (I14:DEP)	0.32	0.57		
If others think I do not make a valuable contribution, then I am no good (I38:DEP)	0.41	0.64		
If I am dismissed by my peers, then that makes me a loser (I53:DEP)	0.65	0.80		
If I am not treated with consideration, it shows I am unlikeable (I58:DEP)	0.48	0.69		
If people think I cannot control others it shows I am worthless (I17:DEP)	0.40	0.63		
If others are not considerate toward me, then it means I am worthless (I9:DEP)	0.49	0.70		
I do not want to be, therefore I must not be, dismissed by my peers (I79:DEM)	0.33		0.58	
I do not want to be, so absolutely should not be, snubbed by people that matter to me (I75:DEM)	0.42		0.64	
I would like to, and therefore I must, avoid losing control of my emotions (I20:DEM)	0.52			0.72
I want to be, therefore I must be, in control of my emotions (I64:DEM)	0.57			0.76

Factor	Mean \pm SD	α (reliability coefficient)	Inter-factor correlations	
Self-depreciation (DEP)	21.10 \pm 6.01	0.85		
Peer rejection demands (PRD)	6.35 \pm 1.60	0.59	0.53	
Emotional control demands (ECD)	7.94 \pm 1.38	0.71	0.16	0.24

$\chi^2(52) = 56.72$, $p = 0.303$, SRMR = 0.05, CFI = 0.99, RMSEA = 0.02. DEM = Demands; DEP = Self-depreciation.

TABLE 4 Descriptive data of the Positive Mental Health Scale (PMH-Scale), Patient Health Questionnaire (PHQ-9), Perth Emotional Reactivity Scale-Short Form (PERS-S), Automatic Thoughts Questionnaire (ATQ-15), Affective Reactivity Index (ARI), Social Phobia Inventory (SPIN), demographic related variables, and correlations with the 16-item 4-factor model dimensions.

Measures	Mean \pm SD	α (reliability coefficient)	Self-depreciation	Peer rejection demands	Emotional control demands	Approval	IBSSO composite score
PMH-scale	31.11 \pm 4.12	0.87	−0.36**	−0.55	0.04	−0.20*	−0.29**
PHQ-9	13.41 \pm 4.45	0.84	0.30**	0.15	0.04	0.07	0.22*
PERS-S	58.70 \pm 8.67	0.77	0.23*	0.12	−0.08	0.23*	0.20*
ATQ-15	21.82 \pm 7.12	0.90	0.42**	0.12	0.08	0.21*	0.35**
ARI	7.65 \pm 2.29	0.86	−0.04	0.02	−0.18	0.01	−0.08
SPIN	28.66 \pm 10.77	0.92	0.44**	0.06	0.11	0.27**	0.40**
Demographics							
Age	39.54 \pm 16.55	–	−0.27**	−0.05	−0.02	−0.01	−0.18
Experience	12.92 \pm 10.96	–	−0.07	−0.01	0.03	0.04	−0.04
Hours	2.95 \pm 1.14	–	−0.05	−0.11	0.00	−0.03	−0.04

$n = 111$, * $p < 0.05$, ** $p < 0.01$. 16-item 4-factor model composite score: mean = 41.95, SD = 9.56, $\alpha = 0.87$. Contents of scales: PMH-Scale = antecedents to positive mental health; PHQ-9 = Depression; PERS-S = Positive and Negative trait emotional reactivity; ATQ-15 = Self-esteem and obsessive thoughts; ARI = irritability; SPIN = social anxiety. Hours = hours per week officiating.

TABLE 5 Descriptive data of the Positive Mental Health Scale (PMH-Scale), Patient Health Questionnaire (PHQ-9), Perth Emotional Reactivity Scale – Short Form (PERS-S), Automatic Thoughts Questionnaire (ATQ-15), Affective Reactivity Index (ARI), Social Phobia Inventory (SPIN), demographic related variables, and correlations with the 13-item 3-factor model dimensions.

Measures	Mean \pm SD	α (reliability coefficient)	Self-depreciation	Peer rejection demands	Emotional control demands	IBSSO composite score
PMH-scale	31.11 \pm 4.12	0.87	−0.38**	−0.55	0.04	−0.32**
PHQ-9	13.41 \pm 4.45	0.84	0.30**	0.15	0.04	0.30**
PERS-S	58.70 \pm 8.67	0.77	0.21*	0.12	−0.08	0.18
ATQ-15	21.82 \pm 7.12	0.90	0.43**	0.12	0.08	0.38**
ARI	7.65 \pm 2.29	0.86	−0.04	0.02	−0.18	−0.08
SPIN	28.66 \pm 10.77	0.92	0.44**	0.06	0.11	0.40**
Demographics						
Age	39.54 \pm 16.55	–	−0.27**	−0.05	−0.02	−0.25**
Experience	12.92 \pm 10.96	–	−0.07	−0.01	0.03	−0.06
Hours	2.95 \pm 1.14	–	−0.05	−0.11	0.00	−0.04

$n = 111$, * $p < 0.05$, ** $p < 0.01$. 13-item 3-factor model composite score: mean = 37.56, SD = 7.56, $\alpha = 0.85$. Contents of scales: PMH-Scale = antecedents to positive mental health; PHQ-9 = Depression; PERS-S = Positive & Negative trait emotional reactivity; ATQ-15 = Self-esteem & obsessive thoughts; ARI = irritability; SPIN = social anxiety. Hours = hours per week officiating.

(1988) found a 7-factor fit for the original Attitude & Beliefs Scale (and thus was statistically supported), this was seen as having poor “theoretical fit,” justifying further investigation and the eventual development of the Attitudes and Beliefs Scale-2 (DiGiuseppe et al., 2020). As approval is one of the Ellis’ (1962) original irrational beliefs, and low frustration tolerance and awfulizing recognized secondary irrational beliefs (Dryden and Branch, 2008), the 4-factor model was accepted. This model reported acceptable fit and improved level of significance ($p = 0.006$). Additionally, the model preserves more items which is consistent with the recommended policy of conservatism in construct development (DeVellis, 2012) and theoretical alignment (Flora and Flake, 2017).

A limitation of both models is that two factors contain only two items, with a minimum of three items recommended for strong and

stable factors (Costello and Osborne, 2005), potentially indicated by the reliability score of the Peer Rejection Demands ($\alpha = 0.54$ and 0.59 for the 16- and 13-item models respectively) being below the 0.70 threshold recommended by Terjesen et al. (2009). However, Yang and Green (2011) state it is unlikely items contribute equally to a factor, an assumption of Cronbach’s alpha. Hence, Cronbach’s alpha should be treated as a lower-bound estimate of internal consistency, and true reliability could be up to 20% greater (Green and Yang, 2009; McNeish, 2017). Furthermore, underestimation of reliability is strongest when factors contain small number of items, therefore it is likely that true reliability is greater than what is reported (Graham, 2006).

While one item reported a factor loading and error variance of above 1, this was not judged to be a result of a poorly defined factor, as both items within the factor assess the primary irrational belief of

demands with the subscale of emotional control. Theoretically, this is important as contemporary research identifies emotional control as a valuable attribute for sports officials (Carrington et al., 2022; McEwan et al., 2020). Hence, the presence of this Heywood case may be attributable to sample size as, although the sample meets requirements identified by Muthén and Muthén (2002), it is smaller than the minimum of 1,000 recommended by Boateng et al. (2018). Ultimately, the 4-factor model was considered the most suitable model of the IBSSO from a statistical and theoretical perspective.

Study 3: convergent validity analysis

To develop the validity of the IBSSO, the aim of Study 3 was to test convergent validity of both acceptable models of the IBSSO against a similar measure.

Method

Participants

A new sample was used to assess the convergent validity of the 16-item 4-factor model and the 13-item 3-factor model of the IBSSO. The sample consisted of 94 participants ($M_{\text{age}} = 36.74$ years, $SD = 15.03$), 83 of whom were male, 10 were female, and one participant chose not to disclose their gender. Average years' experience as a qualified sports official was 13.89 years ($SD = 10.88$) across eight sports (football, rugby union, rugby league, field hockey, basketball, ice hockey, cricket, and lacrosse). To determine sample size G*Power (version 3.1.9.7; Heinrich-Heine-Universität Düsseldorf, Germany), recommended software to determine sample size to conduct power analysis (Kang, 2021), identified a minimum sample of 85 was required for a medium effect size (0.30), alongside power and α (p value) set at 0.80 and 0.05, respectively. Thus, the sample used for this study exceeds this recommendation.

Procedure

As with Study 1 and Study 2, data was collected using the online survey platform Qualtrics (Provo, UT), with links to surveys distributed via social media and personal emails to gatekeepers. Once informed consent was given by confirming the information sheet and consent form had been read and agreed to, participants completed the surveys by indicating item response on a 5-point Likert scale. The survey consisted of the 22-retained items of IBSSO, as well as items from the iPBI (Turner et al., 2018a) to assess convergent validity.

The iPBI is a 28-item measure that assesses the four core irrational beliefs: demandingness, awfulizing, low-frustration tolerance and depreciation (Dryden, 2014). Participants are asked to report the extent they agree or disagree with a statement using a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*). The iPBI was chosen because of its good construct validity, $\chi^2(344) = 1433.98$, $p < 0.001$, CFI = 0.93, NNFI = 0.92, SRMR = 0.06, RMSEA = 0.07, and internal consistency across factors (between 0.90 and 0.96; Turner et al., 2018a). Additionally, the iPBI reports good concurrent validity with measures of similar constructs. For example, there is a high correlation ($r = 0.86$, $p < 0.001$) between the depreciation factor of the iPBI and self-downing measured by the SGABS (Lindner et al., 1999), and predictive validity through small to medium correlations with trait

subscales of the State-Trait Personality Inventory (STPI; Spielberger, 1979; Turner et al., 2018a). Furthermore, the iPBI evaluates irrational beliefs in a performance context, justifying its use to confirm the convergent validity of the IBSSO.

Upon completion of data collection, scores from the IBSSO and the iPBI were checked for incomplete data, patterned responses, and outliers. There were no incomplete cases nor any patterned responses. A Shapiro Wilks test was conducted with z scores to identify outliers. One score of above 3.29, found within the depreciation subscale of the iPBI, was identified and winsorized (Field, 2009). Data was correlated using SPSS Statistics 30 (Microsoft, Albuquerque, NM). High correlations were expected between the IBSSO and iPBI, along with high correlations between relevant subscales (e.g., depreciation factors of both measures).

Results

As expected, high correlations were reported between the 16-item 4-factor model and the iPBI ($r = 0.84$, $p < 0.01$), and the 13-item 3-factor model and the iPBI ($r = 0.81$, $p < 0.01$; Tables 6, 7 respectively). The largest correlations across subscales was found between the depreciation factor of the iPBI and the self-depreciation factor of the IBSSO ($r = 0.79$, $p < 0.01$). Given the similar conceptualization of these factors, it is not surprising a high correlation was found. Demand-based factors of the IBSSO (peer rejection demands and emotional control demands) and the approval factor reported medium to high correlations with the demandingness factor of the iPBI.

Conclusion

The addition of this study and the high correlation found with a similar measure (the iPBI) supports the convergent validity of the IBSSO.

Study 4: test-retest reliability

Method

Participants

A new independent sample of 29 participants ($M_{\text{age}} = 48.25$ years, $SD = 14.77$), with an average of 14.57 years' experience ($SD = 12.44$) as a sports official, completed the test-retest reliability study. Participants (25 male, 4 female) were from the United Kingdom and Ireland, the United States of America and Canada, and Australia, representing 6 sports (football, cricket, rugby union, baseball, field hockey and Australian Rules Football). Officials reported their level of practice as either recreational ($n = 16$), semi-professional ($n = 8$) or national/international ($n = 5$). While larger sample sizes have been recommended (see Kennedy, 2022), samples of between 25 and 30 are seen as sufficient by others, justified by personal experience and statistical theory, such as a medium effect size and sufficient power (Bonett and Wright, 2014; Clark-Carter, 2010; Cocchetti, 1999). Further, G*Power (version 3.1.9.7; Heinrich-Heine-Universität Düsseldorf, Germany) recommended a minimum sample size of 13 participants for a bivariate correlation analysis with 0.08 power and 0.05 α (p value).

TABLE 6 Descriptive data of the Irrational Performance Beliefs Inventory (iPBI), and its individual factors of Demandingness (DEM), Low-Frustration Tolerance (LFT), Awfulizing (AWF), Self-Depreciation (DEP), demographic related variables, and correlations with the 16-item 4-factor model dimensions.

iPBI and demographic variables	Mean \pm SD	α (reliability coefficient)	Self-depreciation	Peer rejection demands	Emotional control demands	Approval	IBSSO composite score
iPBI	86.57 \pm 18.70	0.95	0.74**	0.58**	0.38**	0.73**	0.84**
DEM	24.11 \pm 4.54	0.78	0.50**	0.61**	0.31**	0.58**	0.65**
LFT	25.51 \pm 5.73	0.89	0.55**	0.50**	0.59**	0.49**	0.67**
AWF	21.77 \pm 5.55	0.86	0.65**	0.51**	0.25*	0.71**	0.75**
DEP	15.19 \pm 5.99	0.92	0.79**	0.38**	0.17	0.71**	0.79**
Demographics							
Age	36.74 \pm 15.03	–	–0.33**	–0.14	–0.36**	–0.15	–0.33**
Experience	13.89 \pm 10.88	–	–0.13	0.02	–0.08	–0.05	–0.10

n = 94, **p* < 0.05, ***p* < 0.01. 16-item 4-factor model composite score: mean = 43.12, SD = 10.39, α = 0.89.

TABLE 7 Descriptive data of the Irrational Performance Beliefs Inventory (iPBI), and its individual factors of Demandingness (DEM), Low-Frustration Tolerance (LFT), Awfulizing (AWF), Self-Depreciation (DEP), demographic related variables, and correlations with the 13-item 3-factor model dimensions.

iPBI and demographic variables	Mean \pm SD	α (reliability coefficient)	Self-depreciation	Peer rejection demands	Emotional control demands	IBSSO composite score
iPBI	86.57 \pm 18.70	0.95	0.74**	0.58**	0.38**	0.81**
DEM	24.11 \pm 4.54	0.78	0.50**	0.61**	0.31**	0.62**
LFT	25.51 \pm 5.73	0.89	0.55**	0.50**	0.59**	0.69**
AWF	21.77 \pm 5.55	0.86	0.65**	0.51**	0.25*	0.70**
DEP	15.19 \pm 5.99	0.92	0.79**	0.38**	0.17	0.76**
Demographics						
Age	36.74 \pm 15.03	–	–0.33**	–0.14	–0.36**	–0.38**
Experience	13.89 \pm 10.88	–	–0.13	0.02	–0.08	–0.12

n = 94, **p* < 0.05, ***p* < 0.01. 13-item 3-factor model composite score: mean = 33.29, SD = 7.46, α = 0.83.

Procedure

The IBSSO, along with demographic information, was completed via Qualtrics (Provo, UT) over three-time points, with the second assessment after 7 days and the third assessment after another 21 days, replicating previous test–retest research (e.g., Turner et al., 2018b). This approach exceeds the two-time points recommended in literature (e.g., Law, 2004) and used previously to develop measures (e.g., SGABS, Lindner et al., 1999). Furthermore, a 28-day interval between the first and third assessments minimized the risk of respondents remembering previous answers, a concern in measuring retest reliability (Polit, 2014). After each survey participants were asked for their email address, with this information handled and stored according to General Data Protection Regulations (e.g., on the first author’s university server requiring a two-factor verification to log in), to correctly identify data and distribute the second and third surveys.

Data analysis

There was no missing data. A Shapiro Wilks test was conducted with *z* scores to identify outliers. Two scores of above 3.29, one from

the first and one in the second-time point of the DEP subscale, were identified and winsorized (Field, 2009).

Data analyses were then conducted in two phases. First, a repeated-measures MANCOVA was performed to evaluate changes in each subscale of the IBSSO: self-depreciation (DEP); peer-rejection demands (PRD); emotional control demands (ECD); and approval (A), across the three timepoints, with no change in direction and non-significant results expected to demonstrate the reliability of the IBSSO. Consistent with literature identifying a negative relationship between age and irrational beliefs (see Ndika et al., 2012; Turner et al., 2019), age was included as a covariate. Second, intra-class coefficients (ICCs) were conducted, with scores above 0.7 in all subscales and compound scores applied to assess and establish consistency of irrational beliefs across three-time points.

Results

The repeated-measures MANCOVA revealed no main effect of time, Wilk’s λ = 0.57, *F* (8, 20) = 1.89, *p* = 0.19, η^2 = 0.43. Direction and non-significance of results was not affected by removing age as a covariate, Wilk’s λ = 0.62, *F* (8, 20) = 1.52, *p* = 0.21, η^2 = 0.39. This

TABLE 8 Repeated-measures ANCOVA, intra-class coefficients, means \pm SD for sports official data across three data collection time points.

Officials' Data Variables	Means \pm SD			Intra-class coefficients	
	Time 1	Time 2 (7 days)	Time 3 (28 days)	ICC	95% CI
Officials' data					
DEP	15.83 \pm 6.94	16.45 \pm 6.79	17.14 \pm 6.11	0.85	0.75–0.92
PRD	5.76 \pm 2.17	5.83 \pm 2.02	6.17 \pm 2.09	0.75	0.60–0.86
ECD	7.03 \pm 2.08	7.31 \pm 2.00	6.87 \pm 2.15	0.71	0.55–0.84
APP	10.52 \pm 4.17	10.41 \pm 3.74	10.76 \pm 4.16	0.83	0.71–0.91
COMP.	39.24 \pm 13.66	40.10 \pm 12.73	40.86 \pm 11.95	0.89	0.82–0.94

Cronbach's α ranges for each subscale: DEP = 0.86–0.96; PRD = 0.64–0.72; ECD = 0.70–0.94; APP = 0.83–0.95.

indicates that there is no effect for time in any of the variables within the IBSSO.

For ICCs, all subscales and compound scores reported results above 0.7, demonstrating strong agreement and consistency across all three measurements points for all variables (Table 8). This indicates that the IBSSO is not influenced by time.

General discussion

The aim of this study was to develop, validate and test the reliability of a specific measure of irrational beliefs in sports officials. An EFA supported analysis of a five, four and three-factor model. Consequently, CFA rejected the five-factor model but supported the four and three-factor models. Low to medium correlations between factors of the IBSSO and existing measures of related emotional dimensions supported criterion validity in the three and four-factor models. Specifically, self-depreciation and the composite score of the IBSSO was negatively correlated with mental well-being, and positively correlated with depression and related social phobias. Predictive validity was established through expected negative correlations with age, consistent with previous research (Ndika et al., 2012), and good test–retest reliability was demonstrated, with stable scores for subscales and compound score reported across three-time points.

The 13-item, three-factor model reported strong model-fit and contained items that measured demands and self-depreciation. While this was not expected, as the inclusion of items that assessed low frustration tolerance and awfulizing was anticipated due to their presence in extant measures and theoretical significance (Dryden and Branch, 2008; Turner et al., 2018a), this finding is less surprising when the nature of irrational beliefs is considered. While Ellis (1962) and others (e.g., Dryden, 1984) consider irrational beliefs to be evaluative cognitions, with demands being the core irrational belief leading to secondary irrational beliefs, a distinction has previously been made between appraising (labeled as “hot” cognitions) and knowing (labeled as “cold” cognitions; Abelson and Rosenberg, 1958). Cold cognitions are subsequently categorized into surface cognitions (descriptions, attributions and inferences) and deep cognitions refer to core beliefs, or schemata, that specify the value that any given event has on a range of appraisals (Anderson, 2000; Szentagotai et al., 2005). This view is endorsed by DiGiuseppe (1996), who posited that irrational beliefs are “evaluative schemas” that represent an individual's concept of reality which, owing to the constructivist nature of this claim, strengthens

the need for bespoke measures for specific populations. Moreover, demandingness and self-depreciation have been classified as evaluative schemas, while low frustration tolerance and awfulizing have been identified as appraisals (see David et al., 2005) and therefore measures that assess demandingness and self-depreciation may be more beneficial in revealing core beliefs. While Szentagotai et al. (2005) suggest DiGiuseppe's (1996) distinction between schemata and appraisals should be treated with caution, as it is “based solely on logical analyses and clinical observations, without any empirical evidence to support these assertions” (p. 142), the same can be said of Ellis' (1962) structuring of primary and secondary irrational beliefs (David et al., 2005). Additionally, DiGiuseppe et al. (1988) found that demandingness, awfulizing and low frustration tolerance loaded onto one factor during a CFA, with self-depreciation loading onto another, providing further support for the findings reported here.

The identification of self-depreciation loading onto one factor, with demandingness, awfulizing and low frustration tolerance loading onto another is also seen in the four-factor model. As seen in Table 2, items measuring self-depreciation load onto factor one (labeled depreciation), with items loading onto the remaining factors (peer rejection demands, emotional control demands and approval) assessing demandingness, awfulizing and low frustration tolerance. The three-factor model's p value (0.054), alongside the good reliability coefficient of the fourth factor (approval) and the policy of conservatism recommended by DeVellis (2012), led to the rejection of the three-factor model in favor of the four-factor alternative.

The identification of the factors within this model is supported in previous literature regarding REBT and sports officials, self-depreciation was an expected factor due to its prominence in REBT literature (e.g., Ellis, 1962, 1988; Dryden and Branch, 2008). Contemporary support for the impact of self-depreciation on maladaptive outcomes can be found in recent investigation into perfectionism in sport. When exploring sport psychology consultants' experiences with athletes who demonstrated perfectionist behaviors, Klockare et al. (2022) reported that such athletes hold the belief that they must be perfect or they are “nothing.” While this perspective is reported as an attitude, this is what Dryden (2016) has previously labeled irrational beliefs and, as Klockare et al. (2022) sought out perspectives from practicing sport psychology consultants, the terms “attitude” and “beliefs” likely hold lexiconic similarities. Additionally, such attitudes were identified as having undesirable effects on performance. For instance, one of the consultants discusses how an athlete holding

this belief feels their time is wasted if they make a mistake in the first minute of competition as they can no longer be perfect. Such a belief would have a significant effect on the performance of sports officials, as another consultant states that a referee in possession of such a belief will fixate on an error, interfering with future performance (Klockare et al., 2022). Additionally, two rugby union officials reported a 31.45% reduction in demandingness, awfulizing and low frustration tolerance after an REBT intervention, while self-depreciation scores only reduced by 8.75% (Maxwell-Keys et al., 2022). Therefore, self-depreciation appears to be a robust irrational belief among sports officials. As elite performers report lower levels of self-depreciation than recreational performers and non-athletes (Turner et al., 2019), and REBT has shown promise in reducing self-depreciation in sporting populations (e.g., Cunningham and Turner, 2016), the measurement and reduction of this secondary irrational beliefs is particularly valuable for enhancing the performance within sporting populations.

Support for the identification of emotional regulation as a factor that officials may hold irrational demands about can be seen in research that acknowledges its importance to performance. For example, calmness (and the avoidance of stress) has been classified as an emotion that contributes to accurate decision-making in sports officials (Anshel and Weinberg, 1995; Castillo-Rodriguez et al., 2021; Neil et al., 2013), with lacrosse and rugby union officials adopting emotion-focussed coping strategies to maintain optimum performance (Friesen et al., 2017; Hill et al., 2016). A possible explanation for the importance of emotional regulation within this population is that it is seen as a source of efficacy specific to officiating (see Guillén and Feltz, 2011). The importance of positive resource appraisal (e.g., being able to control emotions) may contribute to more positive emotional and behavioral outcomes (see Dixon and Turner, 2018), and the promotion of challenge states that has been shown to reduce feelings of anxiety in sports officials (Noguiera et al., 2022). Therefore, the value of effective emotional regulation may promote irrational beliefs (e.g., “I want to, and therefore I must, be in control of my emotions”) in officials.

The final factors of peer rejection demands and approval are well established in officiating research. The construction of the Referees Stress Questionnaire (RSQ; Gomes et al., 2021) identified the receipt of negative evaluations from official observers and the need to be able to control the behavior of others as sources of stress for officials. The importance of significant others (e.g., official observers) was supported during the construction of IBSSO, and evaluation from assessors and colleagues (prior to and during performance) has been identified as an emotion-eliciting event for officials (Friesen et al., 2017). Additionally, the approval of others may also contribute to irrational beliefs held by sports officials. Neil et al. (2013) identified that less experienced officials may sacrifice decision-making accuracy to appease, or avoid criticism from, others. The impact of social pressures has also been found at higher levels of officiating, with social bias contributing to inaccuracy regarding time-added on and disciplinary sanctions (see Dohmen and Sauermann, 2016, for a review). The significant correlation between the SPIN and the IBSSO suggests that fear of negative social evaluation is a factor that influences the beliefs (and, by extension, the performance and well-being) of sports officials.

The development of the IBSSO provides researchers with a bespoke measure of irrational beliefs within officiating populations.

However, it is important to highlight some limitations regarding its development. First, it does not meet *all* recommendations made by Terjesen et al. (2009) regarding inventory construction. For example, there are not an equal number of items for each factor and the assessment of rational beliefs is absent. The unequal number of items may be responsible for the low ($\alpha = 0.54$) internal reliability of the peer rejection demands factor. However, as stated, this may be because Cronbach's α assumes equal contribution of items (Yang and Green, 2011). Further, it may be prudent to assess factors with a low number of items differently to other factors, with Cronbach's α between 0.45 and 0.60 previously reported as “acceptable” with limited items (Berger and Hänze, 2015). Additionally, the measure does not contain the same number of irrational beliefs, with more items for self-depreciation and demandingness than others. However, as stated by Terjesen et al. (2009), this is permissible should an irrational belief be of more clinical use than others, justifying their over-representation. When five REBT practitioners were asked which irrational beliefs interfere with performance the most, all five reported demandingness and self-depreciation as most significant, with these beliefs also impacting an individual's ability to recover from major adversities (e.g., injuries; Turner et al., 2022). Hence, demandingness and self-depreciation are of greater utility from a clinical perspective than other irrational beliefs. Second, the IBSSO is based on the self-reporting of irrational beliefs, identified as a limitation of existing measures with other methods, such as implicit testing, recommended (David et al., 2005; Ellis, 1996; Jordana et al., 2020).

Looking forward, four recommendations are made to develop and use the IBSSO. First, to establish criterion validity, evaluation of sports officials' performance using observers and objective criteria is recommended to overcome the reliance on self-reporting measures (Jordana et al., 2020). Second, to maintain a strong factor with less than two items, future research that utilizes the IBSSO should replicate these studies and recruit large samples (Costello and Osborne, 2005). In particular, the recommendation of 10 participants per new item (Boateng et al., 2018) is advised. This could be achieved by addressing the third recommendation, greater investigation of the IBSSO with female officials (who made up less than 12% of the total sample used across the four studies), particularly as females are more likely to experience negative events (e.g., sexism) when officiating (Tingle et al., 2014). Finally, while the use of scree plot is identified as the optimum method of factor extraction (Costello and Osborne, 2005; Watkins, 2018), justifying its application, additional EFA extracting different factors may assist in further development and validation of the IBSSO (see Fabrigar et al., 1999). While further validation studies of the IBSSO would be welcome, the method of development employed here minimized the impact of cohort effects (e.g., a new sample was recruited for CFA following EFA, as well as for Study 3 and Study 4) and recruited large samples that are essential for effective EFA (Costello and Osborne, 2005). Additionally, requirements for transparent reporting regarding item development and preliminary analysis (see Hughes et al., 2019; Jackson et al., 2009) and requirements for goodness of fit being met in the CFA (Brown, 2015) were fulfilled. Furthermore, the concurrent validity, convergent validity, and test-retest reliability of the IBSSO established in this paper means researchers are provided with a valid and reliable tool for assessing irrational beliefs in sports officials.

Data availability statement

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

Ethics statement

The studies involving humans were approved by this study was approved by the St Mary's University (Twickenham, London) Research Ethics Committee on October 05, 2021. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. MT: Conceptualization, Methodology, Supervision, Writing – review & editing. JN: Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing. AB: Conceptualization, Methodology, Supervision, Writing – review & editing.

References

- Abelson, R. P., and Rosenberg, M. J. (1958). Symbolic psycho-logic: a model of attitudinal cognition. *Behav. Sci.* 3, 1–13.
- Anderson, J. R. (2000). Cognitive psychology and its implications. 5th Edn: Hereford & Worcester: Worth Publishers, Inc.
- Anshel, M. H., and Weinberg, R. S. (1995). Coping with acute stress among American and Australian basketball referees. *J. Sport Behav.* 19, 180–203.
- Becerra, R., Preece, D., Campitelli, G., and Scott-Pillow, G. (2017). The assessment of emotional reactivity across negative and positive emotions: development and validation of the Perth emotional reactivity scale (PERS). *Assessment* 26, 867–879. doi: 10.1177/1073191117694455
- Berger, R., and Hünze, M. (2015). Impact of expert teaching quality on novice academic performance in the jigsaw cooperative learning method. *Int. J. Sci. Educ.* 37, 294–320. doi: 10.1080/09500693.2014.985757
- Bernard, M. E. (1988). Validation of the general attitude and belief scale. *J. Rat. Emot. Cogn. Behav. Ther.* 34, 209–224.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., and Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioural research: a primer. *Front. Public Health* 6:149. doi: 10.3389/fpubh.2018.00149
- Bonett, D. G., and Wright, T. A. (2014). Cronbach's alpha reliability: interval estimation, hypothesis testing, and sample size. *J. Organ. Behav.* 36, 3–15. doi: 10.1002/job.1960
- Bowman, A., and Turner, M. J. (2022). When time is of the essence: the use of rational emotive behaviour therapy (REBT) informed single-session therapy (SST) to alleviate social and golf-specific anxiety, and improve wellbeing and performance, in amateur golfers. *Psychol. Sport Exerc.* 60:102167. doi: 10.1016/j.psychsport.2022.102167
- Brown, T. A. (2015). Confirmatory factor analysis for applied research. 2nd Edn. London: The Guilford Press.
- Bruch, E., and Feinberg, F. (2017). Decision-making processes in social contexts. *Annu. Rev. Sociol.* 43, 207–227. doi: 10.1146/annurev-soc-060116-053622
- Cameron, I. M., Crawford, J. R., Lawton, K., and Reid, I. C. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Br. J. Gen. Pract.* 58, 32–36. doi: 10.3399/bjgp08X263794
- Campbell-Sills, L., Espejo, E., Ayers, C. R., Roy-Byrne, P., and Stein, M. B. (2015). Latent dimensions of social anxiety disorder: a re-evaluation of the social phobia inventory (SPIN). *J. Anxiety Disord.* 36, 84–91. doi: 10.1016/j.janxdis.2015.09.007
- Carrington, S. C., North, J. S., and Brady, A. (2022). Utilising experiential knowledge of elite match officials: recommendations to improve practice design for football referees. *Int. J. Sport Psychol.* 53, 242–266. doi: 10.7352/IJSP.2022.53.242
- Caruso, C., Angelone, L., Abbate, E., Ionni, V., Biondi, C., Di Agostino, C., et al. (2017). Effects of a REBT based training on children and teachers in primary school. *J. Rat. Emot. Cogn. Behav. Ther.* 36, 1–14. doi: 10.1007/s10942-017-0270-6
- Castillo, D. (2016). The influence of soccer match play on physiological and physical performance measures in soccer referees and assistant referees. *J. Sports Sci.* 34, 557–563. doi: 10.1080/02640414.2015.1101646
- Castillo, D. (2017). Effects of the off-season period on field and assistant soccer referees' physical performance. *J. Hum. Kinet.* 56, 159–166. doi: 10.1515/hukin-2017-0033
- Castillo-Rodriguez, A., López-Aguilar, J., and Alonso-Arbiol, I. (2021). Relationship between physical-physiological and psychological responses in amateur soccer referees. *J. Sport Psychol.* 30, 26–37.
- Chadha, N. J., Turner, M. J., and Slater, M. J. (2019). Investigating irrational beliefs, cognitive appraisals, challenge and threat, and affective states in golfers approaching competitive situations. *Front. Psychol.* 10:2295. doi: 10.3389/fpsyg.2019.02295
- Chotpitayasunondh, V., and Turner, M. J. (2019). The development and validation of the Thai-translated irrational performance beliefs inventory (T-iPBI). *J. Rat. Emot. Cogn. Behav. Ther.* 37, 202–221. doi: 10.1007/s10942-018-0306-6
- Clark-Carter, D. (2010). Quantitative psychological research: The complete student's companion. Hove: Psychology Press.
- Cocchetti, D. (1999). Sample size requirements for increasing the precision of reliability estimates: problems and proposed solutions. *J. Clin. Exp. Neuropsychol.* 21, 567–570. doi: 10.1076/jcen.21.4.567.886
- Common, E. A., and Lane, K. L. (2017). "Social validity assessment" in Applied behaviour analysis advanced guidebook. ed. J. K. Luiselli (London: Academic Press), 73–92.
- Connor, K. M., Davidson, J. R. T., Erik, C. L., Sherwood, A., Foa, E., and Weisler, R. H. (2000). Psychometric properties of the social phobia inventory (SPIN): new self-rating scale. *Br. J. Psychiatry* 176, 379–386. doi: 10.1192/bjp.176.4.379
- Costello, A. B., and Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract. Assess. Res. Eval.* 10:7. doi: 10.7275/jjy1-4868
- Credé, M., and Harms, P. D. (2019). Questionable research practices when using confirmatory factor analysis. *J. Manag. Psychol.* 34, 18–30. doi: 10.1108/JMP-06-2018-0272
- Credé, M., Harms, P. D., Niehorster, S., and Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the big five personality traits. *J. Pers. Soc. Psychol.* 102, 874–888. doi: 10.1037/a0027403

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Cunningham, R., and Turner, M. J. (2016). Using rational emotive behaviour therapy (REBT) with mixed martial arts (MMA) athletes to reduce irrational beliefs and increase unconditional self-acceptance. *J. Rat. Emot. Cogn. Behav. Ther.* 34, 289–309. doi: 10.1007/s10942-016-0240-4
- David, D., Szentagotai, A., Eva, K., and Macavei, B. (2005). A synopsis of rational-emotive behaviour therapy (REBT): fundamental and applied research. *J. Rat. Emot. Cogn. Behav. Ther.* 23, 175–221. doi: 10.1007/s10942-005-0011-0
- Davidson, J. R. T., Miner, C. M., DeVeauh-Geiss, J., Tupler, L. A., Colket, J. T., and Potts, N. L. S. (1997). The brief social media scale: a psychometric evaluation. *Psychol. Med.* 27, 161–166. doi: 10.1017/S0033291796004217
- Deen, S., Turner, M. J., and Wong, R. S. K. (2017). The effects of REBT, and the use of credos, on irrational beliefs and resilience qualities in athletes. *Sport Psychol.* 31, 249–263. doi: 10.1123/tsp.2016-0057
- DeVellis, R. F. (2012). Scale development: Theory and applications. London: Sage.
- DiGiuseppe, R. (1996). The nature of irrational and rational beliefs: Progress in rational emotive behaviour theory. *J. Rat. Emot. Cogn. Behav. Ther.* 4, 5–28.
- DiGiuseppe, R., Gorman, B., and Raptis, J. (2020). The factor structure of the attitudes and beliefs scale 2: implications for rational emotive behaviour therapy. *J. Rat. Emot. Cogn. Behav. Ther.* 38, 111–142. doi: 10.1007/s10942-020-00349-0
- DiGiuseppe, R., Leaf, R., Exner, T., and Robin, M. W. (1988). The development of a measure of irrational/rational thinking. Presented at the world congress of behaviour therapy. Scotland: Edinburgh.
- Dixon, M., and Turner, M. J. (2018). Stress appraisals of UK soccer academy coaches: an interpretative phenomenological analysis. *Qual. Res. Sport Exerc. Health* 10, 620–634. doi: 10.1080/2159676X.2018.1464055
- Dohmen, T., and Sauermann, J. (2016). Referee bias. *J. Econ. Surv.* 30, 679–695. doi: 10.1111/joes.12106
- Dryden, W. (1984). Rational-emotive therapy: Fundamentals and innovations. London: Croom Helm.
- Dryden, W. (2014). Rational emotive behaviour therapy: Distinctive features. Oxon: Routledge.
- Dryden, W. (2016). Attitudes in rational emotive behaviour therapy: Components, characteristics and adversity-related consequences. Oxford: Rationality Publications.
- Dryden, W., and Branch, R. (2008). The fundamentals of rational emotive behaviour therapy. Chichester: Wiley.
- Ellis, A. (1957). Rational psychotherapy and individual psychology. *J. Individ. Psychol.* 13, 38–44.
- Ellis, A. (1962). Reason and emotion in psychotherapy. New York: Stuart.
- Ellis, A. (1988). How to stubbornly refuse to make yourself miserable: About anything – Yes, anything! New York: Lyle Stuart.
- Ellis, A. (1996). Responses to criticisms of rational-emotive behavior therapy (REBT) by ray DiGiuseppe, frank bond, windy Dryden, Steve Weinrach, and Richard Wessler. *J. Rat. Emot. Cogn. Behav. Ther.* 14, 97–121. doi: 10.1007/BF02238185
- Ellis, A. (2005). The myth of self-esteem: How rational emotive behaviour therapy can change your life forever. New York: Prometheus Books.
- Ellis, A., and Dryden, W. (1997). The practice of rational emotive behaviour therapy. 2nd Edn: New York: Springer Publishing Co.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strachan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychol. Methods* 4, 272–299. doi: 10.1037/1082-989X.4.3.272
- Farooq, R. (2022). Heywood cases: possible causes and solutions. *Int. J. Data Anal. Tech. Strat.* 14, 79–88. doi: 10.1504/IJDATS.2022.121506
- Field, A. (2009). Discovering statistics using IBM SPSS statistics. London: Sage.
- Fleeson, W., and Noflie, E. E. (2009). In favour of the synthetic resolution to the person-situation debate. *J. Res. Pers.* 43, 150–154. doi: 10.1016/j.jrp.2009.02.008
- Flora, D. B., and Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: decisions for scale development and validation. *Can. J. Behav. Sci.* 49, 78–88. doi: 10.1037/cbs0000069
- Friesen, A. P., Devonport, T. J., and Lane, A. M. (2017). Beyond the technical: the role of emotion regulation in lacrosse officiating. *J. Sports Sci.* 35, 579–586. doi: 10.1080/02640414.2016.1180419
- Funder, D. C. (2006). Towards a resolution of the personality triad: persons, situations, and behaviours. *J. Res. Pers.* 40, 21–34. doi: 10.1016/j.jrp.2005.08.003
- Gomes, A. R., Fontes, L., Rodrigues, M., and Dias, B. (2021). Burnout in referees: relations with stress, cognitive appraisal, and emotions. *Int. J. Psychol. Behav. Anal.* 7, 1–7. doi: 10.15344/2455-3867/2021/177
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability what they are and how to use them. *Educ. Psychol. Meas.* 66, 930–944. doi: 10.1177/0013164406288165
- Grant, J. S., and Davis, L. L. (1997). Selection and use of content experts for instrument development. *Res. Nurs. Health* 20, 269–274. doi: 10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G
- Green, S. B., and Yang, Y. (2009). Commentary cautionary tale. *Psychometrika* 74, 121–135. doi: 10.1007/s11336-008-9098-4
- Guillen, F., and Feltz, D. L. (2011). A conceptual model of referee efficacy. *Front. Psychol.* 2:25. doi: 10.3389/fpsyg.2011.00025
- Hancock, D. J., Bennett, S., Roaten, H., Chapman, K., and Stanley, C. (2021). An analysis of literature on sport officiating research. *Res. Q. Exerc. Sport* 92, 607–617. doi: 10.1080/02701367.2020.1756198
- Helsen, W., and Bultynck, J.-B. (2004). Physical and perceptual-cognitive demands of top-class refereeing in association football. *J. Sports Sci.* 22, 179–189. doi: 10.1080/02640410310001641502
- Hill, D. M., Matthews, N., and Senior, R. (2016). The psychological characteristics of performance under pressure in professional rugby union referees. *Sport Psychol.* 30, 1–32. doi: 10.1123/tsp.2015-0109
- Hollon, S. D., and Kendall, P. C. (1980). Cognitive self-statements in depression: development of an automatic thoughts questionnaire. *Cogn. Ther. Res.* 4, 383–395. doi: 10.1007/BF01178214
- Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equat. Model. Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118
- Hughes, R. A., Heron, J., Sterne, J. A. C., and Tilling, K. (2019). Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int. J. Epidemiol.* 48, 1294–1304. doi: 10.1093/ije/dyz032
- Hyland, P., Shevlin, M., Adamson, G., and Boduszek, D. (2013). Irrational beliefs in posttraumatic stress responses: a rational emotive behaviour therapy approach. *J. Loss Trauma* 20, 171–188. doi: 10.1080/15325024.2013.839772
- Jackson, D. L., Gillaspay, J. A. Jr., and Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychol. Methods* 14, 6–23. doi: 10.1037/a0014694
- Jarrett, T. A. (2013). Warrior resilience and thriving (WRT): rational emotive behaviour therapy (REBT) as a resiliency and thriving foundation to prepare warriors and their families for combat deployment and posttraumatic growth in operation Iraqi freedom, 2005–2009. *J. Rat. Emot. Cogn. Behav. Ther.* 31, 93–107. doi: 10.1007/s10942-013-0163-2
- Jordana, A., Turner, M. J., Ramis, Y., and Torregrossa, M. (2020). A systematic mapping review on the use of rational emotive behaviour therapy (REBT) with athletes. *Int. Rev. Sport Exerc. Psychol.* 16, 231–256. doi: 10.1080/1750984X.2020.1836673
- Jung, S., and Takane, Y. (2008). “Regularized exploratory factor analysis” in New trends in psychometrics. eds. K. Shigemasa, A. Okada, T. Imaizumi and T. Hoshino (Tokyo: University Academic Press), 141–149.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika* 39, 31–36.
- Kang, H. (2021). Sample size determination and power analysis using the G*power software. *J. Educ. Eval. Health Prof.* 18:17. doi: 10.3352/jeehp.2021.18.17
- Kennedy, I. (2022). Sample size determination in test-retest and Cronbach alpha reliability estimates. *Brit. J. Contemp. Educ.* 2, 17–29. doi: 10.52589/BJCE-FY266HK9
- Klockare, E., Olsson, L., Gustafsson, H., Lundqvist, C., and Hill, A. (2022). Sport psychology consultants' views on working with perfectionistic elite athletes. *Sport Psychol.* 6, 219–227. doi: 10.1123/tsp.2021-0055
- Kostrna, J., and Tenenbaum, G. (2022). Developing and testing the expanded sport Official's decision-making model. *Int. J. Sport Exerc. Psychol.* 20, 586–611. doi: 10.1080/1612197X.2021.1891117
- Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 9, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x
- Law, M. (2004). Outcome measures rating form guidelines. CanChild Centre for Childhood Disability Research. Hamilton, Canada: McMaster University.
- Lazarus, R. S. (1991). Emotion and adaptation. Oxford: Oxford University Press.
- Lazarus, R. S. (1999). Stress and emotion. London, New York: A new synthesis.
- Liebowitz, M. R. (1987). “Social phobia” in Anxiety. ed. D. F. Klein (Basel: Karger), 141–173.
- Lindner, H., Kirkby, R., Wertheim, E., and Birch, P. (1999). A brief assessment of irrational thinking: the shortened general attitude and belief scale. *Cogn. Ther. Res.* 23, 651–663. doi: 10.1023/A:1018741009293
- Lloret, S., Ferreres, A., Hernandez, A., and Tomas, I. (2017). The exploratory factor analysis of items: guided analysis based on empirical data and software. *Anales de Psicología* 33, 417–432. doi: 10.6018/analesps.33.2.270211
- Lukat, J., Margraf, J., Lutz, R., van der Veld, W. M., and Becker, E. S. (2016). Psychometric properties of the positive mental health scale (PMH-scale). *BMC Psychol.* 4:8. doi: 10.1186/s40359-016-0111-x
- Maciejewski, D., van Sprang, E., Spinhoven, P., and Penninx, B. (2021). Longitudinal associations between negative life events and depressive symptoms – a 9-year longitudinal study on between-person and within-person effects and the role of family history. *J. Pers. Soc. Psychol.* 121, 707–721. doi: 10.1037/pspp0000381
- MacMahon, C., Mascarenhas, D., Plessner, H., Pizzera, A., Oudejans, R., and Raab, M. (2015). Sports officials and officiating: Science and practice. Oxon: Routledge.
- Madley-Dowd, P., Hughes, R., Tilling, K., and Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *J. Clin. Epidemiol.* 110, 63–73. doi: 10.1016/j.jclinepi.2019.02.016

- Mansell, P. C. (2021). Stress mindset in athletes: investigating the relationship between beliefs, challenge and threat with psychological wellbeing. *Psychol. Sport Exerc.* 57:102020. doi: 10.1016/j.psychsport.2021.102020
- Maxwell-Keys, C., Wood, A. G., and Wood, M. J. (2022). Developing decision making in rugby union match officials using rational emotive behaviour therapy (REBT). *Psychol. Sport Exerc.* 58:102098. doi: 10.1016/j.psychsport.2021.102098
- McEwan, G. P., Unnithan, V., Easton, C., and Arthur, R. (2020). Development and validation of the referee training activity questionnaire (RTAQ): towards a better understanding of the training practices of soccer officials. *J. Sports Sci.* 38, 2782–2793. doi: 10.1080/02640414.2020.1800371
- McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychol. Methods* 3, 412–433. doi: 10.1037/met0000144
- Morris, G., and O'Connor, D. (2017). Key attributes of expert NRL referees. *J. Sports Sci.* 35, 852–857. doi: 10.1080/02640414.2016.1194524
- Mustaffa, W. S. W., Wahid, H. A., Rahman, R. A., and Karim, N. A. H. A. (2018). Overcoming improper solution of second-order measurement model using AMOS. *J. Humanit. Lang. Cult. Bus.* 2, 95–105.
- Muthén, L. K., and Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Struct. Equ. Model.* 9, 599–620. doi: 10.1207/S15328007SEM0904_8
- Ndika, N. A., Olagbaiye, F., and Agiobu-Kemmer, I. (2012). Age differences in irrational beliefs, self-efficacy and self-confidence of adolescents in a Nigerian secondary school. *Psychol. Educ. J.* 46:16.
- Neil, R., Bayston, P., Hanton, S., and Wilson, K. (2013). The influence of stress and emotions on association football referee decision-making. *Sport Exerc. Psychol. Rev.* 9, 22–41. doi: 10.53841/bpssepr.2013.9.2.22
- Netemeyer, R. G., Williamson, D. A., Burton, S., Biswas, D., Jindal, S., Landreth, S., et al. (2002). Psychometric properties of shortened versions of the automatic thoughts questionnaire. *Educ. Psychol. Meas.* 62, 111–129. doi: 10.1177/0013164402062001008
- Nevill, A. M., Hemingway, A., Greaves, R., Dallaway, A., and Devonport, T. J. (2017). Inconsistency of decision-making, the Achilles heel of referees. *J. Sports Sci.* 35, 2257–2261. doi: 10.1080/02640414.2016.1265143
- Noguiera, D., Fontes, L., Gomes, A. R., and Resende, R. (2022). Referees' emotions and performance perception: the importance of stress and cognitive appraisal. *Cuadernos de Psicología del Deporte* 22, 156–170. doi: 10.6018/cpd.468751
- Nunnally, J. C. (1972). *Psychometric theory*. 2nd Edn. New York: McGraw-Hill.
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric theory*. 3rd Edn. New York: McGraw-Hill.
- O'Connor, H. (2018). "Anxiety, unhealthy jealousy, and exercise avoidance: a rational emotive behaviour therapy (REBT) intervention" in Rational emotive behaviour therapy in sport and exercise. eds. M. Turner and R. Bennett (Oxon: Routledge), 38–52.
- Page, K., and Page, L. (2010). Alone against the crowd: individual differences in referees' ability to cope under pressure. *J. Econ. Psychol.* 31, 192–199. doi: 10.1016/j.joep.2009.08.007
- Pargent, F., Hilbert, S., Eichhorn, K., and Bühner, M. (2019). Can't make it better nor worse: an empirical study about the effectiveness of general rules of item construction on psychometric properties. *Eur. J. Psychol. Assess.* 35, 891–899. doi: 10.1027/1015-5759/a000471
- Polit, D. F. (2014). Getting serious about test-retest reliability: a critique of retest research and some recommendations. *Qual. Life Res.* 23, 1713–1720. doi: 10.1007/s11136-014-0632-9
- Praschinger, A., Pomikal, C., and Stieger, S. (2011). May I curse a referee? Swear words and consequences. *J. Sports Sci. Med.* 10, 341–345.
- Preece, D., Becerra, R., and Campitelli, G. (2018). Assessing emotional reactivity: psychometric properties of the Perth emotional reactivity scale and the development of a short form. *J. Pers. Assess.* 101, 589–597. doi: 10.1080/00223891.2018.1465430
- Raab, M., MacMahon, C., Avugos, S., and Bar-Eli, M. (2021). The referee's challenge: a threshold process model for decision making in sport games. *Int. Rev. Sport Exerc. Psychol.* 14, 208–228. doi: 10.1080/1750984X.2020.1783696
- Ramanaiah, N. V., Heerboth, J. R., and Schill, T. R. (1987). "Measurement of irrational beliefs: a critical review" in Advances in personality assessment. eds. J. N. Butcher and C. D. Spielberger (New Jersey: Lawrence Erlbaum Associates, Inc), 49–67.
- Ritchie, J., Basevitch, I., Rodenberg, R., and Tenenbaum, G. (2017). Situation criticality and basketball officials' stress levels. *J. Sports Sci.* 35, 2080–2087. doi: 10.1080/02640414.2016.1255770
- Robb, H. B., and Warren, R. (1990). Irrational belief tests: new insights, new directions. *J. Cogn. Psychother.* 4, 303–311.
- Samuel, R. D., Galily, Y., Guy, O., Sharoni, E., and Tenenbaum, G. (2019). A decision-making simulator for soccer referees. *Int. J. Sports Sci. Coach.* 14, 480–489. doi: 10.1177/1747954119858696
- Samuel, R. D., Tenenbaum, G., and Galily, Y. (2020). An integrated conceptual framework of decision-making in soccer refereeing. *Int. J. Sport Exerc. Psychol.* 19, 738–760. doi: 10.1080/1612197X.2020.1766539
- Schafer, J. L. (1999). Multiple imputation: A primer. *Stat. Methods Med. Res.* 8, 3–15.
- Seery, M. D. (2011). Resilience: a silver-lining to experiencing adverse life events? *Curr. Dir. Psychol. Sci.* 20, 390–394. doi: 10.1177/0963721411424740
- Shorkey, T., and Whiteman, V. L. (1977). Development of the rational behaviour inventory: initial validity and reliability. *Educ. Psychol. Meas.* 37, 527–534. doi: 10.1177/001316447703700232
- Spielberger, C. D. (1979). Preliminary manual for the state-trait personality inventory. University of South Florida, Tampa: Unpublished manual.
- Spitz, J., Put, K., Wagemans, J., Williams, A. M., and Helsen, W. F. (2018). The role of domain-generic and domain-specific perceptual-cognitive skills in association football referees. *Psychol. Sport Exerc.* 34, 47–56. doi: 10.1016/j.psychsport.2017.09.010
- Stringaris, A., Goodman, R., Ferdinando, S., Razdan, V., Muhrer, E., Leibenluft, E., et al. (2012). The affective reactivity index: a concise irritability scale for clinical and research settings. *J. Child Psychiatry* 53, 1109–1117. doi: 10.1111/j.1469-7610.2012.02561.x
- Szentagotai, A., Schnur, J., DiGiuseppe, R., Macavei, B., Kallay, E., and David, D. (2005). The organization and the nature of irrational beliefs: schemas or appraisal? *J. Cogn. Behav. Psychother.* 5, 139–158.
- Terjesen, M. D., Salhany, J., and Scituito, M. J. (2009). A psychometric review of measures of irrational beliefs: implications for psychotherapy. *J. Rat. Emot. Cogn. Behav. Ther.* 27, 83–96. doi: 10.1007/s10942-009-0093-1
- Tingle, J. K., Warner, S., and Sartore-Baldwin, M. L. (2014). The experience of former women officials and the impact on the sporting community. *Sex Roles* 71, 7–20. doi: 10.1007/s11199-014-0366-8
- Turner, M. J. (2014). Smarter thinking in sport. *The Psychologist* 27, 596–599.
- Turner, M. (2022). *The rational practitioner*. Oxon: Routledge.
- Turner, M. J., Allen, M. S., Slater, M. J., Barker, J. B., Woodcock, C., Harwood, C. G., et al. (2018a). The development and initial validation of the irrational performance beliefs inventory (iPBI). *Eur. J. Psychol. Assess.* 34, 174–180. doi: 10.1027/1015-5759/a000314
- Turner, M. J., and Barker, J. B. (2015). Examining the effects of rational emotive behaviour therapy (REBT) on the irrational beliefs of blue-chip professionals. *J. Rat. Emot. Cogn. Behav. Ther.* 33, 17–36. doi: 10.1007/s10942-014-0200-9
- Turner, M. J., and Bennett, R. (Eds.) (2018). *Rational emotive behaviour therapy in sport and exercise*. Oxon: Routledge.
- Turner, M. J., Carrington, S., and Miller, A. (2019). Psychological distress across sport participation groups: the mediating effects of secondary irrational beliefs on the relationship between primary irrational beliefs and symptoms of anxiety, anger, and depression. *J. Clin. Sport Psychol.* 13, 17–40. doi: 10.1123/jcsp.2017-0014
- Turner, M. J., Chadha, N. J., Davis, H., Deen, M. S., Gilmore, H., Jones, J. K., et al. (2022). At the coalface: practitioner perspectives on applying rational emotive behaviour therapy (REBT) in high performance sport. *J. Rat. Emot. Cogn. Behav. Ther.* 41, 251–271. doi: 10.1007/s10942-022-00461-3
- Turner, M. J., Ewan, D., and Barker, J. B. (2020). An idiographic single-case study examining the use of rational emotive behaviour therapy (REBT) with three amateur golfers to alleviate sport performance phobias. *J. Appl. Sport Psychol.* 32, 186–204. doi: 10.1080/10413200.2018.1496186
- Turner, M. J., Slater, M. J., Dixon, J., and Miller, A. (2018b). Test-retest reliability of the irrational performance beliefs inventory. *Eur. J. Sport Sci.* 18, 123–129. doi: 10.1080/17461391.2017.1411527
- Turner, M. J., Wood, A. G., Boatwright, D., Chadha, N., Jones, J. K., and Bennett, R. (2021). Assessing beliefs about emotion generation and change: the conceptualisation, development, and validation of the cognitive mediation beliefs questionnaire (CMBQ). *Psychother. Res.* 31, 932–949. doi: 10.1080/10503307.2020.1871524
- Velicer, W. F., and Jackson, D. N. (1990). Component analysis versus common factor analysis – some further observations. *Multivar. Behav. Res.* 25, 97–114.
- Watkins, M. W. (2018). *Exploratory factor analysis: A guide to best practice*. *J. Black Psychol.* 44, 219–246.
- Webb, T. (2017). *Elite soccer referees: Officiating in the premier league*. La Liga and Serie A. Oxon: Routledge.
- Webb, T., Rayner, M., Cleland, J., and O'Gorman, J. (2021). *Referees, match officials and abuse: Research and implications for policy*. Routledge.
- Werner, C. H., Tang, M., Kruse, J., Kaufman, J. C., and Spörrle, M. (2014). The Chinese version of the revised creativity domain questionnaire (CDQ-R): first evidence for its factorial validity and systematic association with the big five. *J. Creat. Behav.* 48, 254–275. doi: 10.1002/jocb.51
- Weston, M., Drust, B., and Gregson, W. (2011). Intensities of exercise during match-play in FA premier league referees and players. *J. Sports Sci.* 29, 527–532. doi: 10.1080/02640414.2010.543914
- Wood, A. G., Barker, J. B., and Turner, M. J. (2017). "Rational emotive behaviour therapy to help young athletes build resilience and deal with adversity" in Sport psychology for youth athletes. eds. C. J. Knight, C. G. Harwood and D. Gould (Oxon: Routledge) 265–276.
- Yang, Y., and Green, S. B. (2011). Coefficient alpha: a reliability coefficient for the 21st century? *J. Psychoeduc. Assess.* 29, 377–392. doi: 10.1177/0734282911406668
- Zettle, R. D., Rains, J. C., and Hayes, S. C. (2011). Processes of change in acceptance and commitment therapy and cognitive therapy for depression: a mediation reanalysis of Zettle and Rains. *Behav. Modificat.* 35, 265–283. doi: 10.1177/0145445511398344
- Ziegler, M., and Horstmann, K. (2015). Discovering the second side of the coin integrating situational perception into psychological assessment. *Eur. J. Psychol. Assess.* 31, 69–74. doi: 10.1027/1015-5759/a000258