RESEARCH

Validity and reliability of the physiological and perceptual responses elicited during a novel treadmill-based Soccer Referee Simulation (SRS)

Gary P. McEwan[1] · Viswanath B. Unnithan[1] · Chris Easton[1] · Andrew J. Glover[1] · Rosie Arthur[1,2]

Abstract

Purpose This study explored the validity and reliability of the physiological and perceptual responses elicited during a novel treadmill-based Soccer Referee Simulation (SRS).

Methods Following the collection of baseline measures and habituation procedures, eight sub-elite soccer referees completed a single trial of the SRS whereby measures of heart rate (HR), oxygen uptake ($\dot{V}O_2$), blood lactate concentrations ($[La^-]_b$), and differential ratings of perceived exertion (RPE) were obtained. Referees' HR responses were also monitored during a series of competitive matches (5 match observations per referee). For the reliability aspect of the investigation, eight well-trained males with soccer playing experience and a comparable physiological profile were initially habituated, and thereafter completed three separate trials of the SRS during which the reliability of the selected outcome variables were ascertained. Trials were separated by 3–7 days and performed under standardised conditions.

Results No differences were evidenced between the SRS and match play in relation to measures of mean HR ($P = 0.444$; ES = 0.29), peak HR ($P = 0.074$; ES = 0.74), or HR-based training impulse ($P = 0.498$; ES = 0.25). Additionally, no systematic differences were detected between reliability trials for any of the measured outcome variables ($P \geq 0.293$), whilst good levels of reliability were observed for measures of mean HR (ICC = 0.94; CV = 3.1%), peak HR (ICC = 0.93; CV = 2.2%), HR-based training impulse (ICC = 0.95; CV = 10.0%), mean $\dot{V}O_2$ (ICC = 0.95; CV = 2.6%); $[La^-]_b$ (ICC ≥ 0.89; CV ≤ 11.5%), and differential RPE (ICC ≥ 0.94; CV ≤ 15.1%).

Conclusion The SRS represents a valid and reliable protocol that closely replicates the physiological and decision-making demands of soccer refereeing.

Keywords Team sport · Physiology · Performance · Testing

Introduction

Soccer refereeing represents a highly demanding activity, requiring referees to perform complex decision-making processes under challenging physiological conditions [1]. Growing interest therefore exists with respect to the physical demands of match play [2], and their impact upon decision-making [3]. However, as large levels of inter-match variability manifest within the activity profiles of referees, the detection of real systematic changes in performance outcomes between matches is difficult [4]. Additionally, whilst the continuous monitoring of heart rate (HR) within applied settings is relatively simple, assessing a referee's respiratory, metabolic, and perceptual responses is impractical as collection procedures may interfere with normal refereeing duties [5]. Drawing meaningful conclusions on the in-situ impact of match play on the physical and decision-making performances of soccer referees therefore remains a challenge.

Simulation protocols that mimic match play represent alternative approaches that are used to standardise the internal and external demands imposed on players [6]. Whilst negating the contextual factors that confound match data, simulation protocols enable researchers to assess physiological and perceptual responses in a more controlled

✉ Gary P. McEwan
  Gmcewan@qmu.ac.uk

1 Division of Sport and Exercise, University of the West of Scotland, Glasgow G72 0LH, UK

2 Division of Psychology, Sociology and Education, Queen Margaret University, Edinburgh, UK

environment. To date, several soccer player simulations have been developed utilising treadmill-based [7, 8] and free-running [6] protocols, the validity and reproducibility of which are well established. However, previous simulations mimic the movement patterns of the players and although those of the referees are similar, distinct differences exist [9]. Notably, referees cover greater distances comprising more low-intensity running but less sprinting, than their playing counterparts [10]. Furthermore, as the physical and decision-making demands are imposed concurrently, rationale exists for the inclusion of match-specific decision-making into the physical training and testing protocols of sport officials [11, 12]. Valid and reliable protocols that replicate the unique demands of soccer refereeing are therefore warranted.

Recently, Samuel et al. [13] assessed the efficacy of a soccer referee simulation whereby referees ran at varying paces on a motorised treadmill whilst simultaneously adjudicating on pre-recorded match footage. Whilst acknowledging the need for such protocols, several limitations warrant consideration when interpreting the fidelity of this protocol. Firstly, referees alternated 4-min intervals at 10 km·h^{-1} with 1 min intervals at 13 km·h^{-1}, for a total of 60 min. Although eliciting a comparable total distance to competitive matches [2], the extent to which this protocol replicates the intermittent activity profile of soccer referees is questionable. Indeed, soccer match play necessitates that referees engage in periods of both low-intensity activity such as standing or walking as well as intense activities such as high-speed running [14]. This protocol therefore appears to be of a relatively moderate-intensity as it excludes the frequent bouts of high-intensity activity that are commonplace during matches. The reproducibility of the physiological responses elicited during this protocol also remains unclear. Concerns also exist with regards to the fidelity of the decision-making task as officials were presented with match broadcast footage (filmed from an elevated position in the grandstand) on a small treadmill-mounted monitor. Such approaches have however been criticised as they do not replicate the in-game perspective of the official [15]. Thus, scope exists for the development and refinement of simulation-based protocols that better replicate the demands of soccer refereeing. Accordingly, this study aimed to evaluate: (1) the validity of the physiological and perceptual responses elicited during a novel Soccer Referee Simulation (SRS); and (2) the levels of absolute and relative reliability associated with this protocol.

Methods

Study design

This investigation comprised two parts: (1) determining the validity of the SRS in relation to the internal loads recorded during competitive matches; and (2) determining the test–retest reliability of the physiological and perceptual responses elicited during the SRS. To establish the validity of the SRS, a cohort of sub-elite referees attended the laboratory on two separate occasions. Following preliminary measurements and habituation procedures, a single trial of the SRS was completed whereby referees' physiological and perceptual responses were assessed. The referees' HR responses were also monitored during 40 competitive matches (5 observations per official) and compared against those elicited during the SRS. To assess the test–retest reliability of the SRS, a repeated measures design was employed comprising a habituation session, followed by three separate trials of the SRS. As the referees' competitive schedules restricted their participation in multiple trials within a short period of time, a cohort of well-trained males with soccer experience and a comparable physiological profile participated in the test–retest arm of the investigation. Previous playing experience and a comparable physiological profile was deemed sufficient for the current purpose, as we sought to explore the reliability of the responses elicited during the SRS, and not decision-making accuracy per se. The absolute and relative reliability of the selected outcome variables were subsequently ascertained.

Participants

Eight male soccer referees (age: 30.1 ± 3.8 years; stature: 178.4 ± 8.8 cm; body mass: 77.1 ± 10.7 kg; VO$_{2max}$: 53.2 ± 4.1 mL·kg·min^{-1}) provided informed written consent to participate in the first part of this study. At the time of investigation, referees were enrolled on the Scottish Centre of Refereeing Excellence (SCORE) programme–a 2 year developmental pathway designed to accelerate the development of high-potential soccer officials. Referees possessed 7.0 ± 1.4 years of officiating experience and had officiated nationally for 2.3 ± 1.5 years. On average, referees engaged in physical training for 3.7 ± 0.8 h/week and officiated 1–2 matches per week. For the test–retest reliability element of the investigation, eight well-trained males (age: 25.1 ± 4.2 years; stature: 177.6 ± 9.0 cm; body mass: 79.6 ± 12.0 kg; VO$_{2max}$: 50.6 ± 4.8 mL·kg·min^{-1}) participated. Participants possessed previous soccer playing experience, engaged in regular high-intensity intermittent training (2–3 times per week), and exhibited a maximal oxygen uptake (VO$_{2max}$) commensurate of elite soccer referees [16]. The study received institutional ethical approval and conformed to the Declaration of Helsinki.

Procedures

During the first visit of both parts of the investigation, participants' VO$_{2max}$ and maximal HR (HR$_{max}$) were established,

and they were habituated to the SRS. During the subsequent trials, the SRS was performed under identical environmental conditions (temperature: 19 °C; humidity: 40%) at the same time of day ($\pm$ 1 h). A minimum and maximum of 3 and 7 days separated trials, respectively. Participants were asked to refrain from strenuous exercise, alcohol, and caffeine intake in the 24 h preceding each trial, and to standardise food and fluid intake prior to each visit [17]. All participants provided verbal confirmation of compliance with these instructions.

Preliminary measurements and habituation procedures

Participants' VO_{2max} and HR_{max} were established during a ramp incremental test on a motorised treadmill (Woodway PPS 55sport-I, USA). Following a 5 min warm-up, participants commenced running at 8 km·h^{-1} for 2 min, with the speed increased by 1 km·h^{-1} every minute until 15 km·h^{-1}. Thereafter, speed remained constant with the gradient increased by 1% every minute until volitional exhaustion [18]. Participants were instructed to perform to the best of their ability and received verbal encouragement throughout. Respiratory measurements were obtained throughout using breath-by-breath gas analysis (Jaeger Oxycon Pro, Germany) and HR was monitored via a chest-worn monitor (Polar H10, Finland). VO_{2max} was taken as the highest VO_2 value recorded using 15-breath rolling averages [19] with HR_{max} defined as the highest value recorded. A maximal effort was identified upon achievement of at least two of the following criteria: (1) respiratory exchange ratio ≥ 1.1; (2) plateau in VO_2 (increase of < 2 ml·kg·min^{-1}) despite an increasing speed; and (3) HR within ± 10 beats·min^{-1} of age-predicted HR_{max}) [20]. Upon completion, participants received ~ 15 min recovery before being familiarised to the SRS and main experimental procedures.

Soccer Referee Simulation (SRS)

The SRS was adapted from previous team-sport simulations to replicate the non-uniform movement patterns exhibited by elite referees during match play [7, 21] (Fig. 1). Accordingly, two ~ 16 min blocks interspersed with a 90 s passive recovery period were performed. This recovery period enabled the collection of physiological markers and mimicked the extended breaks in play that manifest during match play (e.g., substitutions, injuries, cooling breaks, and on-field reviews by the Video Assistant Referee) [22, 23]. Whilst future research may wish to extend the SRS to 45 or 90 min, a shortened or condensed protocol was considered more appropriate in the current context given the challenges associated with conducting research with athletes during the in-season. That is, given the busy match schedules of soccer officials during the in-season, it was not possible for officials

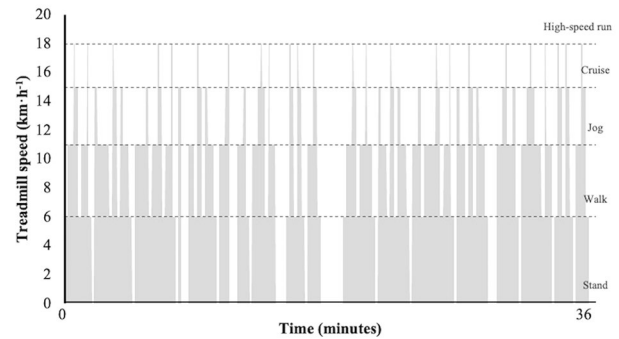| Activity	Treadmill speed (km·h⁻¹)	Number of occurrences (n)	Duration per occurrence (s)	Relative time spent in each activity (%)
Stand	0	20	10	13.9
Walk	6	34	13	22.9
Jog	11	35	23	37.3
Cruise	15	32	12	18.8
High-speed run	18	24	6	7.1

Fig. 1 Schematic representation of the Soccer Referee Simulation (SRS) activity profile

to complete multiple full-duration match simulations within a short period of time. Meanwhile, in anticipating the potential for the SRS to be adopted as a novel training or testing tool, we considered that it would be more practical to implement a shorter protocol into the routine training schedules of soccer officials during the in-season. The protocol incorporated varying periods of standing, walking (6 km·h^{-1}), jogging (11 km·h^{-1}), cruising (15 km·h^{-1}), and high-speed running (18 km·h^{-1}) on a motorised treadmill, with the frequency and duration of each occurrence reflecting previous literature [14, 24]. A threshold of 18 km h^{-1} is commonly used to characterise bouts of high-speed running amongst elite soccer referees [14, 24, 25], and is consistent with the speed at which high-speed runs are performed during the FIFA Interval Test for international and Category 1 referees. Considering the impracticalities of changing speed every few seconds on a motorised treadmill, the frequency and duration of occurrences were manipulated by a factor that resulted in an activity change every 6–23 s [26]. This resulted in 145 activity changes, with the rate of acceleration and deceleration programmed between each activity change set at 2 m·s^{-2}. Match play data suggest that $\sim 36\%$ of the accelerations and decelerations performed by elite referees during match play are performed at rates of 1.5–2.5 m·s^{-2} [25]. To ensure participant safety and mimic anticipation in match play, activity changes were communicated to participants via a visual countdown displayed prior to each activity on a large 40 inch monitor (NEC MultiSync LCD4010, Japan) positioned ~ 2 m in front of the treadmill. Similar approaches have been adopted amongst previous soccer simulations whereby activity changes have been communicated

in the form of either a visual or audio signal [6, 7]. Based upon the modelling of the SRS and the clustering of high-speed running, the SRS elicited a peak running demand of 222 m min^{-1}. In considering the advantages and drawbacks of the various types of soccer simulation [27], a motorised treadmill-based protocol was considered appropriate for several reasons. First and foremost, a key objective of the SRS was to mimic the dual-task nature of soccer refereeing whereby the physical and decision-making demands are imposed concurrently. This would however be impractical using a free-running or non-motorised treadmill-based simulation as it would be difficult to maintain prescribed running speeds whilst attending to a video-based decision-making task. In contrast, the external regulation of running speeds using a motorised treadmill offered the possibility for participants to engage in the physical and decision-making elements of the SRS simultaneously. Motorised treadmill-based protocols indeed offer increased levels of mechanistic rigour and experimental control by eliminating pacing effects and ensuring that a reliable and reproducible physiological stimulus is applied [27].

To replicate the dual-task nature of soccer refereeing, a series of decision-making clips were presented sporadically throughout the SRS. Video clips ($n = 10$) were provided by the Refereeing Department of the Scottish Football Association and portrayed potential foul-play incidents from club and international European matches. Clips were edited to 14–29 s with the elapsed time of the match, the score line, and background sound removed to control for the potential impact of contextual information on participants' decisions [28]. The number of clips presented reflects the frequency of fouls that occur during a match [29]. Upon viewing, participants provided a verbal judgement on whether an offence had been committed (i.e., foul/no foul) and whether to caution the player (i.e., no caution/yellow card/red card). Clips were administered during either a stand ($n = 5$) or jog ($n = 5$) phase, with placement ensuring that a clip succeeded each of the five movement activities on at least two occasions. As we sought to ascertain the validity and reliability of the responses elicited during the SRS, participants' decision-making accuracy was not assessed in the present investigation, with the decision-making component retained to maintain the integrity of the protocol.

Physiological and perceptual outcome variables

Throughout each trial, HR and oxygen uptake (VO_2) were measured as previously described. Intensities of effort were subsequently calculated and expressed in relative terms as percentages of participants' HR_{max} and VO_{2max}. A HR-based training impulse (TRIMP) was also calculated for each trial using Bannister's TRIMP [30].

To assess blood lactate concentrations ([La$^-$]$_b$), capillary samples were obtained pre-, mid-, and immediately post-trial from the fingertip and were analysed within 1 h of collection (Biosen C-Line, Germany). Calibration procedures were performed prior to each trial as per manufacturer guidelines. The inter-assay coefficient of variation (CV) for [La$^-$]$_b$ was 2.0%.

Using the CR100 scale, participants provided differential ratings of perceived exertion (RPE) to differentiate between central (RPE-B), local muscular (RPE-M), and total (RPE-T) exertion [31]. To control for the potential influence of acute fatigue upon perceptual responses, ratings were collected during the jog phase before each clip (epochs; E1–E10) and were obtained in a counterbalanced manner to eliminate order effects [31]. Participants were fully habituated with the correct use of this scale during the preliminary session. To quantify the global intensity associated with the SRS, the average of the 10 ratings was calculated to obtain a single value for each measure.

Match play

To quantify the internal loads imposed during match play, each referee's HR was monitored during 5 competitive matches, resulting in 40 match observations in total. Data were collected during Scottish Championship ($n = 20$) and Scottish League 1 ($n = 20$) matches. Referees officiated at both levels of competition with measures of HR being similar ($80.5 \pm 6.9\%$ and $81.1 \pm 6.3\%$ of HR_{max}, respectively). Data were recorded continuously at a sampling rate of $1H_Z$ using a chest-worn monitor (Polar H10, Finland), with referees wearing the same monitor for the entirety of the study. To account for the condensed duration of the SRS, HR corresponding to the first 36 min of each match were retained for analyses, with measures of mean HR, peak HR, and TRIMP subsequently calculated. Additionally, given the large levels of match-to-match variability evident in performance outcomes [4], the mean HR responses recorded during the 5 matches were retained for comparison to the SRS.

Statistical analyses

The normality of data distribution and homogeneity of variance were verified using Shapiro–Wilk's and Levene's tests, respectively. Paired sample t tests explored potential differences in measures of HR between the SRS and match play. Cohen's effect sizes (ES) were calculated to determine the magnitude of differences and were interpreted as: trivial, < 0.2; small, $0.21–0.6$; moderate, $0.61–1.2$; large, $1.21–1.99$; and very large, > 2.0 [32]. One-way analyses of variance (ANOVA) with repeated measures examined systematic differences in physiological and perceptual responses between reliability trials, with two-way repeated

measures ANOVAs examining variables expressed over multiple time points. Intraclass correlation coefficients (ICC) were calculated to ascertain relative reliability and were interpreted as small (0.10–0.29), moderate (0.30–0.49), large (0.50–0.69), very large (0.70–0.89), and nearly perfect (≥ 0.90) [32]. To establish absolute reliability, the typical error of measurement (TEM) and CV were calculated [33]. All reliability statistics (ICC, TEM, and CV) are reported alongside their respective 95% CI. In addition, the smallest worthwhile change (SWC) was also calculated for each variable as the between-participant SD multiplied by 0.2 [34]. Data are presented as means and standard deviations (mean $\pm$ SD). Statistical procedures were completed using Statistical Package for Social Sciences (SPSS 26.00, IBM, USA) and statistical significance was set at $P < 0.05$.

Results

Validity of the responses during the SRS

The physiological and perceptual responses elicited during the SRS, alongside the HR responses recorded during match play, are summarised in Table 1.

Measures of mean HR ($P = 0.444$; 95% CI -4.2 to 2.0%; ES = 0.29), peak HR ($P = 0.074$; 95% CI -8.1 to 0.5%; ES = 0.74), and TRIMP ($P = 0.498$; 95% CI -14.5 to 7.8 au; ES = 0.25) were similar between the SRS and match play. An illustration of a representative referee's HR profile during

Table 1 Physiological and perceptual responses of the cohort of referees ($n = 8$) during a single trial of the Soccer Referee Simulation (SRS) and during competitive match play

	SRS	Match Play
Heart rate		
Mean HR (% of HR_{max})	79.6 ± 6.6	80.6 ± 6.0
Peak HR (% of HR_{max})	90.0 ± 7.1	93.8 ± 3.5
TRIMP (au)	58.3 ± 20.2	61.6 ± 14.1
Oxygen uptake		
VO_2 (% of VO_{2max})	61.5 ± 6.2	/
Blood lactate concentrations		
Pre-trial (mmol·l^{-1})	0.95 ± 0.17	/
Mid-trial (mmol·l^{-1})	2.04 ± 0.39	/
Post-trial (mmol·l^{-1})	3.18 ± 0.55	/
Differential RPE		
RPE-B (au)	38 ± 14	/
RPE-M (au)	28 ± 12	/
RPE-T (au)	37 ± 14	/

Data presented as mean $\pm$ SD

HR_{max} maximal heart rate, *TRIMP* Bannister's Training Impulse (Bannister, 1993), VO_{2max} maximal oxygen uptake capacity, *RPE-B* breathlessness, *RPE-M* muscular, *RPE-T* total

the SRS and the corresponding time during a competitive match is presented in Fig. 2.

Reliability of the responses during the SRS

The relative and absolute reliability of all outcome variables are outlined in Table 2.

Measures of mean HR ($P = 0.391$), peak HR ($P = 0.836$), TRIMP ($P = 0.660$), and VO_2 ($P = 0.670$) were similar between trials. Progressive increases in $[La^-]_b$ occurred during the SRS ($F_{(2,14)} = 30.317$; $P \leq 0.001$) with increases occurring from pre- (0.97 ± 0.25 mmol·l^{-1}) to mid-trial (3.01 ± 1.23 mmol·l^{-1}; $P = 0.001$), and further increases from mid- to post-trial (3.89 ± 1.47 mmol·l^{-1}; $P = 0.003$). No between-trial differences were detected at any time point ($F_{(2,14)} = 1.342$; $P = 0.293$) whilst the pattern of response remained similar between trials ($F_{(4,28)} = 2.195$; $P = 0.095$).

Progressive increases occurred between E1–E10 in RPE-B ($F_{(9,63)} = 15.316$; $P = 0.000$), RPE-M ($F_{(9,63)} = 3.526$; $P \leq 0.001$), and RPE-T ($F_{(9,63)} = 14.931$; $P \leq 0.001$). No between-trial differences were detected at any time point for any differential RPE (RPE-B: $F_{(2,14)} = 0.854$; $P = 0.447$; RPE-M: $F_{(2,14)} = 0.010$; $P = 0.990$; RPE-T: $F_{(2,14)} = 0.732$; $P = 0.498$). The pattern of response remained similar for all measures between trials (RPE-B: $F_{(18,126)} = 0.547$; $P = 0.930$; RPE-M: $F_{(18,126)} = 0.335$; $P = 0.995$; RPE-T: $F_{(18,126)} = 0.601$; $P = 0.893$).

Measures of mean HR, peak HR, TRIMP, and VO_2 exhibited good levels of absolute reliability (CV $\leq 10.0\%$) and nearly perfect relative reliability (ICC ≥ 0.928). Although good levels of absolute reliability were exhibited for $[La^-]_b$ assessed post-trial (CV = 8.8%), levels were moderate for $[La^-]_b$ assessed pre- and mid-trial (CV $\geq 11.3\%$), with very large to nearly perfect ICC (≥ 0.891) present. Differential RPE exhibited moderate levels of absolute reliability (CV $\geq 13.8\%$) with nearly perfect ICC (≥ 0.937).
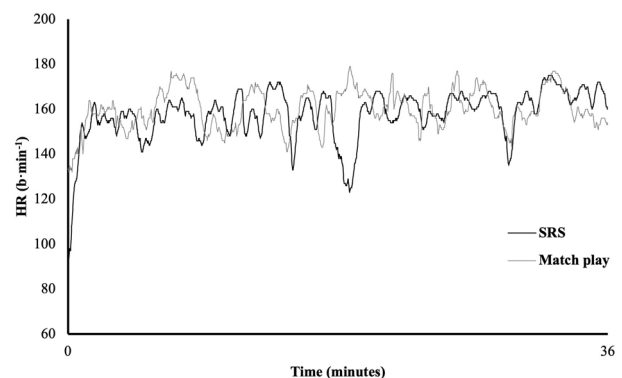
Fig. 2 An illustration of a representative referee's heart rate (HR) profile during the Soccer Referee Simulation (SRS) and the corresponding time during the first half of a competitive match

Table 2 Absolute and relative reliability of physiological and perceptual responses during the Soccer Referee Simulation (SRS) within the cohort of well-trained males (n = 8)

	Trial 1	Trial 2	Trial 3	ICC (95% CI)	CV (95% CI)	TEM	SWC
Heart rate							
Mean HR (% of HR_{max})	80.8 ± 7.0	79.1 ± 5.2	79.4 ± 6.6	0.94 (0.79–0.99)	3.1 (2.2–4.1)	2.6	1.3
Peak HR (% of HR_{max})	91.5 ± 6.5	91.2 ± 4.7	90.8 ± 4.6	0.93 (0.76–0.98)	2.2 (1.2–3.2)	2.2	1.1
TRIMP (au)	63.4 ± 17.2	60.4 ± 14.8	62.0 ± 17.6	0.95 (0.83–0.99)	10.0 (6.0–13.9)	5.0	3.3
Oxygen uptake							
VO_2 (% of VO_{2max})	63.4 ± 4.1	63.4 ± 4.0	64.1 ± 5.1	0.95 (0.83–0.99)	2.6 (1.4–3.7)	1.7	0.9
Blood lactate concentrations							
Pre-trial (mmol·l^{-1})	0.95 ± 0.25	0.98 ± 0.21	0.99 ± 0.34	0.89 (0.63–0.98)	11.5 (6.6–16.3)	0.17	0.05
Mid-trial (mmol·l^{-1})	3.16 ± 1.44	2.99 ± 1.28	2.88 ± 1.08	0.96 (0.87–0.99)	11.3 (6.2–16.4)	0.64	0.25
Post-trial (mmol·l^{-1})	4.14 ± 1.81	3.64 ± 1.33	3.89 ± 1.34	0.97 (0.90–0.99)	8.8 (5.1–12.5)	0.42	0.29
Differential RPE							
RPE-B (au)	36 ± 15	33 ± 14	33 ± 10	0.95 (0.61–0.97)	14.0 (9.2–18.8)	5	3
RPE-M (au)	33 ± 14	32 ± 14	32 ± 9	0.94 (0.79–0.99)	13.8 (7.4–20.1)	5	2
RPE-T (au)	35 ± 14	33 ± 14	32 ± 10	0.94 (0.80–0.99)	15.1 (10.3–19.8)	5	3

Data presented as mean ± SD

ICC intraclass correlation coefficient, *CV* coefficient of variation, *TEM* typical error of measurement, *SWC* smallest worthwhile change, *CI* confidence interval, *HR_{max}* maximal heart rate, *TRIMP* Bannister's Training Impulse (Bannister, 1993), *VO_{2max}* maximal oxygen uptake capacity, *RPE-B* breathlessness, *RPE-M* muscular, *RPE-T* total

Discussion

The current study determined the validity and test–retest reliability of the physiological and perceptual responses elicited during the SRS. Considering first the selected HR measures, no differences were detected between the SRS and match play, whilst the physiological and perceptual responses were aligned with those observed amongst elite soccer referees during match play [35, 36]. The SRS also yielded high levels of reproducibility between trials. Overall, our findings demonstrate the SRS to be a valid and reliable protocol that replicates the demands associated with soccer refereeing.

The SRS elicited mean and peak HRs of 79.6 and 90.0% of HR_{max}, respectively, with no differences evidenced in any HR measure between the SRS and match play. Albeit marginally lower, these values are comparable to published data whereby elite-level referees attain mean and peak HRs of ~ 84% and ~ 97%, respectively [14, 36]. Similarly, the mean VO_2 observed during the SRS was 63.5% of VO_{2max} with slightly higher values (~ 68%) achieved during competitive matches [37]. Although the slightly reduced cardiorespiratory responses could be attributable to the lower competitive level of the studied cohort, it is worth noting that the SRS was developed based upon the activity profiles exhibited by elite-level referees [14]. Further, the peak running demand elicited during the SRS (222 m·min^{-1}) exceeds the ~ 190 m·min^{-1} identified amongst elite English [38] and Italian [39] soccer players during competitive matches.

Previous accounts detailing the [La^{-}]$_b$ of elite referees at half-time have reported values of 3.5 mmol·l^{-1} [14]. In the current investigation, significant increases in [La^{-}]$_b$ were observed from pre- to mid-trial, with [La^{-}]$_b$ peaking post-trial (3.18 ± 0.55 mmol·l^{-1}). Thus, despite the reduced duration, the SRS elicits metabolic responses consistent with official matches. In relation to differential RPE, the grand mean for RPE-T, RPE-B, and RPE-M were 37 ± 14, 38 ± 14, and 28 ± 12, respectively. These ratings correspond to the verbal anchor "somewhat strong" on the CR100 scale. Weston and colleagues [35] previously observed values of 7.8 ("very strong") on the 10-point Borg scale within English Premier League referees, whilst the RPE-B and RPE-M reported by Spanish referees were 6.6 and 7.1 (both "very strong"), respectively [36]. Unfortunately, drawing direct comparisons between these data is difficult, as aside from using different scales, previous studies have obtained one-off ratings ~ 10–30 min following the final whistle [35, 36]. In contrast, we collected differential RPE throughout the SRS, with the grand mean of each measure providing an indication of the global intensity associated with the SRS. Whilst collection procedures often prove impractical within competitive settings, the ability to assess physiological and perceptual responses throughout represents an important benefit of the SRS. Assessing differential RPE throughout the SRS may yield interesting new insights into how perceptions of exertion develop during simulated match play and provide further detail of the internal demands associated with soccer refereeing.

The physical match performances of soccer referees are subject to large levels of match-to-match variation [4], with the volume of high-intensity running being particularly variable (CVs of ~20–54%). Such observations are important when comparing match data as without considerable sample sizes, real systematic changes in outcome variables can prove difficult to detect. The reproducibility of the physiological and perceptual responses during the SRS is therefore promising. Firstly, we failed to detect any systematic bias between trials for any outcome variable ($P \geq 0.288$). Additionally, excellent levels of relative reliability were demonstrated for all physiological and perceptual responses (ICC ≥ 0.891), whilst repeat measures of mean HR, peak HR, and VO_2 exhibited good levels of absolute reliability (CV $\leq 3.1\%$). Similar levels of variation have been reported for measures of HR and VO_2 during existing team-sport simulations [21, 40]. In relation to differential RPE and $[La^-]_b$ collected pre and mid-trial, moderate levels of absolute reliability (CV $\geq 11.3\%$) were present, highlighting $[La^-]_b$ as generally less reliable than other physiological measures such as HR and VO_2 [6, 40]. Interestingly, good levels of absolute reliability were observed for post-trial $[La^-]_b$ (CV $= 8.8\%$), with these levels comparing favourably to the ~20–34% observed previously [6, 40]. Such discrepancies are likely explained by the externally regulated nature of the SRS. Whereas the external loads performed during the SRS were replicated between trials, free-running protocols possess inherent variation in the test itself, which in turn will influence the variability of physiological responses.

Whilst the SRS represents a valid and reliable protocol, readers should remain cognisant of the trade-off that simulation protocols pose between levels of ecological validity and experimental control [27]. Firstly, although replicating the relative time apportioned during matches to each locomotion activity, the impracticalities of changing speed every few seconds on a motorised treadmill make it difficult to replicate the exact duration of each individual running effort [26]. Meanwhile, the absence of changes of direction and unorthodox movements (i.e., sideways and backwards running) may compromise the ecological validity of treadmill-based protocols [7]. This notwithstanding, the modelling of treadmill simulations to incorporate frequent high-intensity activity changes can elicit biomechanical responses representative of match play [8]. The inability to express maximal running velocities represents another potential limitation of motorised treadmill protocols [41]. Sprinting does however account for only ~2% of the total distance covered by elite referees during match play, with maximal velocities rarely achieved during matches [14, 42]. The inability of such protocols to incorporate very high-intensity activities should therefore be considered with respect to their capacity to reliably impose physiological loads consistent with match play. Finally, it must be acknowledged that whilst the SRS

has been modelled on the match activities of elite referees, the physiological and perceptual responses elicited during the protocol have been validated amongst a sub-elite cohort. Referees in the current study were however of a national standard, possessed an aerobic capacity commensurate of elite soccer referees [16], and exhibited physiological and perceptual responses that were consistent with those observed amongst referees at the elite level [35, 36].

Conclusion

The SRS yields highly reproducible physiological and perceptual responses that are consistent with those of match play. Several applications therefore exist for the SRS. Firstly, applied practitioners may wish to use the SRS to profile the physical capacities of referees or to assess the effectiveness of training interventions. The reproducibility of the physiological responses produced during the SRS also lends support to its use within research where factors such as environmental conditions may be manipulated to evaluate their influence on simulated match performances. The inclusion of a referee-specific decision-making stimulus represents another important feature of the SRS. Researchers may therefore use the SRS to investigate the role and impact of dual-tasking on the physiological and perceptual-cognitive demands of soccer referees. The efficacy of the SRS as a training and testing tool also warrants investigation.

Author contributions All authors have participated in the research and/or article preparation. Study concept and design: GM, VU, CE, RA; Data collection: GM, AG; Data analysis: GM, VU, CE, RA; Manuscript development: GM; Contribution to manuscript: GM, VU, CE, AG, RA. All authors read and approved the final version of the manuscript.

Availability of data and materials The data sets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Ethical approval The study received institutional ethical approval and conformed to the Declaration of Helsinki.

Informed consent Participants provided written informed consent.

References

1. Weston M, Castagna C, Impellizzeri FM, Bizzini M, Williams AM, Gregson W (2012) Science and medicine applied to soccer refereeing: an update. Sports Med 42(7):615–631. https://doi.org/10.2165/11632360-000000000-00000

2. Castagna C, Abt G, D'Ottavio S (2007) Physiological aspects of soccer refereeing performance and training. Sports Med 37(7):625–646. https://doi.org/10.2165/00007256-20073 7070-00006

3. Mascarenhas DR, Button C, O'Hara D, Dicks M (2009) Physical performance and decision making in association football referees: a naturalistic study. Open Sports Sci J 2(9):1–9. https://doi.org/10.2174/1875399X00902010001

4. Weston M, Drust B, Atkinson G, Gregson W (2011) Variability of soccer referees' match performances. Int J Sports Med 32(3):190–194. https://doi.org/10.1055/s-0030-1269843

5. Coutts A, Reaburn P, Abt G (2003) Heart rate, blood lactate concentration and estimated energy expenditure in a semi-professional rugby league team during a match: case study. J Sports Sci 21(2):97–103. https://doi.org/10.1080/0264041031000070831

6. Russell M, Rees G, Benton D, Kingsley M (2011) An exercise protocol that replicates soccer match-play. Int J Sports Med 32(7):511–518. https://doi.org/10.1055/s-0031-1273742

7. Aldous JWF, Akubat I, Chrismas BCR, Watkins SL, Mauger AR, Midgley AW, Abt G, Taylor L (2014) The reliability and validity of a soccer-specific nonmotorised treadmill simulation (Intermittent soccer performance test). J Strength Cond Res 28(7):1971–1980. https://doi.org/10.1519/JSC.0000000000000310

8. Page RM, Marrin K, Brogden CM, Greig M (2015) Biomechanical and physiological responses to a contemporary soccer match-play simulation. J Strength Cond Res 29(10):2860–2866. https://doi.org/10.1519/JSC.0000000000000949

9. Weston M, Castagna C, Impellizzeri FM, Rampinini E, Abt G (2007) Analysis of physical match performance in English Premier League soccer referees with particular reference to first half and player work rates. J Sci Med Sport 10(6):390–397. https://doi.org/10.1016/j.jsams.2006.09.001

10. Weston M, Drust B, Gregson W (2011) Intensities of exercise during match-play in FA premier league referees and players. J Sports Sci 29(5):527–532. https://doi.org/10.1080/02640414.2010.543914

11. Kittel A, Elsworthy N, Spittle M (2018) Incorporating perceptual decision-making training into high-intensity interval training for Australian football umpires. J Sports Sci 37(1):29–35. https://doi.org/10.1080/02640414.2018.1480257

12. McEwan GP, Unnithan VB, Easton C, Arthur R (2022) Training practices and perceptions of soccer officials: Insights from the referee training activity questionnaire. Int J Sports Sci Coach In Press. https://doi.org/10.1177/17479541221110707

13. Samuel RD, Galily Y, Guy O, Sharoni E, Tenenbaum G (2019) A decision-making simulator for soccer referees. Int J Sports Sci Coach 14(4):480–489. https://doi.org/10.1177/1747954119858696

14. Krustrup P, Helsen W, Randers MB, Christensen JF, MacDonald C, Rebelo AN, Bangsbo J (2009) Activity profile and physical demands of football referees and assistant referees in international games. J Sports Sci 27(11):1167–1176. https://doi.org/10.1080/02640410903220310

15. Craig C (2013) Understanding perception and action in sport: how can virtual reality technology help. Sports Technol 6(4):161–169. https://doi.org/10.1080/19346182.2013.855224

16. Castagna C, Bizzini M, Araújo Póvoas SC, Schenk K, Büsser G, D'Ottavio S (2019) Aerobic fitness in top-class soccer referees. J Strength Cond Res 33(11):3098–3104. https://doi.org/10.1519/JSC.0000000000002264

17. Jeacocke NA, Burke LM (2010) Methods to standardize dietary intake before performance testing. Int J Sport Nutr Exerc Metab 20(2):87–103. https://doi.org/10.1123/ijsnem.20.2.87

18. Sperlich PF, Holmberg HC, Reed JL, Zinner C, Mester J, Sperlich B (2015) Individual versus standardized running protocols in the determination of VO2max. J Sports Sci Med 14(2):386–393

19. Scheadler CM, Garver MJ, Hanson NJ (2017) The gas sampling interval effect on VO2peak is independent of exercise protocol. Med Sci Sports Exerc 49(9):1911–1916. https://doi.org/10.1249/MSS.0000000000001301

20. Midgley AW, McNaughton LR, Polman R, Marchant D (2007) Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. Sports Med 37(12):1019–1025. https://doi.org/10.2165/00007256-20073 7120-00002

21. Sirotic AC, Coutts AJ (2008) The reliability of physiological and performance measures during simulated team-sport running on a non-motorised treadmill. J Sci Med Sport 11(5):500–509. https://doi.org/10.1016/j.jsams.2007.04.008

22. Spitz J, Wagemans J, Memmert D, Williams AM, Helsen WF (2021) Video assistant referee (VAR): the impact of technology on decision making in association football referees. J Sports Sci 39(2):147–153. https://doi.org/10.1080/02640414.2020.1809163

23. Siegle M, Lames M (2012) Game interruptions in elite soccer. J Sports Sci 30(7):619–624. https://doi.org/10.1080/02640414.2012.667877

24. Barbero-Álvarez J, Boullosa DA, Nakamura FY, Andrín G, Castagna C (2012) Physical and physiological demands of field and assistant soccer referees during America's cup. J Strength Cond Res 26(5):1383–1388. https://doi.org/10.1519/JSC.0b013e31825183c5

25. Castillo D, Castagna C, Cámara J, Iturricastillo A, Yanci J (2018) Influence of team's rank on soccer referees' external and internal match loads during official matches. J Strength Cond Res 32(6):1715–1722. https://doi.org/10.1519/JSC.0000000000002040

26. Thatcher R, Batterham AM (2004) Development and validation of a sport-specific exercise protocol for elite youth soccer players. J Sports Med Phys Fitness 44(1):15–22

27. Field AC, Harper LD, Aldous JWF, Page RM (2022) A commentary on soccer match-play simulations for applied research and practice. Sci Med Footb. https://doi.org/10.1080/24733938.2022.2070268

28. Catteeuw P, Gilis B, Jaspers A, Wagemans J, Helsen W (2010) Training of perceptual-cognitive skills in offside decision making. J Sport Exerc Psychol 32(6):845–861. https://doi.org/10.1123/jsep.32.6.845

29. Mallo J, Frutos PG, Juárez D, Navarro E (2012) Effect of positioning on the accuracy of decision making of association football

top-class referees and assistant referees during competitive match play. J Sports Sci 30(13):1437–1445. https://doi.org/10.1080/02640414.2012.711485

30. Banister EW (1991) Modelling elite athletic performance. In: MacDougall JD, Wenger HA, Green HJ (eds) Physiological testing of elite athletes Champaign. Human Kinetics, Illinois, pp 403–422

31. Weston M, Siegler J, Bahnert A, McBrien J, Lovell R (2015) The application of differential ratings of perceived exertion to Australian Football League matches. J Sci Med Sport 18(6):704–708. https://doi.org/10.1016/j.jsams.2014.09.001

32. Hopkins WG, Marshall SW, Batterham AM, Hanin J (2009) Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc 41(1):3–13. https://doi.org/10.1249/MSS.0b013e31818cb278

33. Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 26(4):217–238. https://doi.org/10.2165/00007256-199826040-00002

34. Buchheit M (2016) The numbers will love you back in return—I promise. Int J Sports Physiol Perform 11(4):551–554. https://doi.org/10.1123/IJSPP.2016-0214

35. Weston M, Bird S, Helsen W, Nevill A, Castagna C (2006) The effect of match standard and referee experience on the objective and subjective match workload of English Premier League referees. J Sci Med Sport 9(3):256–262. https://doi.org/10.1016/j.jsams.2006.03.022

36. Castillo D, Weston M, McLaren SJ, Cámara J, Yanci J (2017) Relationships between internal and external match-load indicators in soccer match officials. Int J Sports Physiol Perform 12(7):922–927. https://doi.org/10.1123/ijspp.2016-0392

37. D'Ottavio S, Castagna C (2001) Analysis of match activities in elite soccer referees during actual match play. J Strength Cond Res 15(2):167–171

38. Fereday K, Hills SP, Russell M, Smith J, Cunningham DJ, Shearer D, McNarry M, Kilduff LP (2020) A comparison of rolling averages versus discrete time epochs for assessing the worst-case scenario locomotor demands of professional soccer match-play. J Sci Med Sport 23(8):764–769. https://doi.org/10.1016/j.jsams.2020.01.002

39. Riboli A, Semeria M, Coratella G, Esposito F (2020) Effect of formation, ball in play and ball possession on peak demands in elite soccer. Biol Sport 38(2):195–205. https://doi.org/10.5114/biolsport.2020.98450

40. Harper LD, Hunter R, Parker P, Goodall S, Thomas K, Howatson G, West DJ, Stevenson E, Russell M (2016) Test-retest reliability of physiological and performance responses to 120 minutes of simulated soccer match play. J Strength Cond Res 30(11):3178–3186. https://doi.org/10.1519/JSC.0000000000001400

41. Williams JD, Abt G, Kilding AE (2010) Ball-sport endurance and sprint test (BEAST90): validity and reliability of a 90-minute soccer performance test. J Strength Cond Res 24(12):3209–3218. https://doi.org/10.1519/JSC.0b013e3181bac356

42. Castillo D, Cámara J, Castellano J, Yanci Y (2016) Football match officials do not attain maximal sprinting speed during matches. Kinesiology 48(2):207–212. https://doi.org/10.26582/k.48.2.10