Psychology of Sport & Exercise 67 (2023) 102429

Contents lists available at ScienceDirect

Psychology of Sport & Exercise

journal homepage: www.elsevier.com/locate/psychsport

A random forest approach to explore how situational variables affect perceived exertion of elite youth soccer players

Diego Marqués-Jiménez[a,*], Jaime Sampaio[b], Julio Calleja-González[c], Ibon Echeazarra[d,e]

[a] *Valoración del rendimiento deportivo, actividad física y salud y lesiones deportivas (REDAFLED), Department of Didactics of Musical, Plastic and Corporal Expression, Faculty of Education, University of Valladolid, 42004, Soria, Spain*

[b] *Research Centre in Sports Sciences, Health Sciences and Human Development, CIDESD, CreativeLab Research Community, University of Trás-os-Montes e Alto Douro, 5001-801, Vila Real, Portugal*

[c] *Department of Physical Education and Sports, Faculty of Education and Sport, University of the Basque Country, 01007, Vitoria-Gasteiz, Spain*

[d] *Department of Didactics of Musical, Plastic and Corporal Expression, Faculty of Education and Sport, University of the Basque Country, 01007, Vitoria-Gasteiz, Spain*

[e] *Evaluation and Data Department, Real Sociedad, 20014, Donostia – San Sebastián, Spain*

ARTICLE INFO

Keywords:
Situational variables
RPE
Soccer

ABSTRACT

The aim of this study was to explore how situational variables affect youth soccer players' perceived exertion (RPE) after official matches. Thirty-five elite youth male players (14.33 ± 0.86 years; 173.49 ± 6.16 cm; 63.44 ± 5.98 kg) who belonged to two different teams of a professional club participated in this study. Data collection was conducted during two seasons (2016–2017, 2017–2018) and included 60 official matches (30 official matches per team). Ten minutes after each match players rated their RPE and using a modified Borg CR-10 scale. A Random Forest Regression was used to quantify the importance of match-related situational variables in RPE. Afterwards, a linear mixed model analysis was applied to identify the variability in RPE among the situational variables. The game-playing time, the player status (starter or substitute) and the player identity were the strongest predictors of RPE. Moreover, the match outcome and the final scoreline showed significant effects on both starter and substitute players but the main effect of the quality of the opponent was only identified in starter players ($p < 0.05$). These results allow practitioners to know how situational variables interact and modulate RPE after official matches and help them to prescribe and adapt the players' training content and load before and after matches.

1. Introduction

Both internal and external load have been extensively analyzed in training and competitive scenarios (Impellizzeri, Marcora, & Coutts, 2019). Specifically in soccer, the measurement and evaluation of internal match loads, in particular the rating of perceived exertion (RPE), alongside match running performance, help to accurately relate match activities to physical capacity in an attempt to understand the dose-response nature of competitive matches (Weston, 2013). Despite relying on a subjective assessment, RPE-based methods have been widely adopted to quantify internal load in soccer (Rago et al., 2020).

Perceived exertion is the feeling of how heavy and strenuous a physical task is (Borg, 1998). The RPE is generated centrally by the corollary discharge as a duplicate of the signal generated from the motor command, determining local alterations can have an indirect influencing cause determining an increase of motor command to sustain a task (de Morree, Klein, & Marcora, 2012). Moreover, the RPE integrates afferent neural signals from different inputs to the central nervous system (Abbiss, Peiffer, Meeusen, & Skorski, 2015). Afferent feedback from the peripheral organs (e.g., skeletal muscles, heart and lungs) and other interceptors (e.g., knowledge of the exercise work endpoint) might be examples of these inputs. However, physiological and neural determinants do not fully explain the variation of RPE (Morgan, 1973) as other factors influence it as well. For instance, sociological and environmental factors or subjects' characteristics such as gender, age, fitness level, and expertise may affect RPE (Haddad, Stylianides, Djaoui, Dellal, & Chamari, 2017; Morgan, 1973).

It has been suggested that highly trained players use only a proportion of their physical potential due to situational variables (e.g., tactics, opponents, weather and players' expectations) (Waldron & Highton,

* Corresponding author. Faculty of Education, University of Valladolid, Calle Universidad s/n, 42004, Soria, Spain.
E-mail address: diego.marques@uva.es (D. Marqués-Jiménez).

https://doi.org/10.1016/j.psychsport.2023.102429
Received 5 December 2022; Received in revised form 17 March 2023; Accepted 18 March 2023
Available online 21 March 2023

2014). This may explain why RPE after matches could be influenced by situational variables (Barrett, McLaren, Spears, Ward, & Weston, 2018; Brito, Hertzog, & Nassis, 2016). For instance, Barrett et al. (2018) found positional differences during matches across differential RPE (breathlessness, leg muscle exertion and technical exertion) and higher technical RPE when playing against a higher-ranked opponent. However, limited information is available examining how match-related situational variables interact and influence the RPE in soccer matches.

It is worthwhile to deeply understand why soccer players vary their RPE from one match to another. From this perspective, random forests can incorporate non-linear effects and are superior to alternate methods at modelling complex interactions when the interactions are not, or cannot be, pre-specified (Cutler et al., 2007). Moreover, random forests have no distributional assumptions for predictor or response variables and are thus resistant to bias from non-parametric, skewed and even nominal data, and perform exceptionally well even when many predictors are weak (Breiman, 2001; Cutler et al., 2007). The random forest algorithm could also report the variable importance, which may be due to its complex interaction with other variables (Liaw & Wiener, 2002). The algorithm estimates the importance of a variable by looking at how much the prediction error increases when out-of-bag data for that variable is permuted while all others are left unchanged (Liaw & Wiener, 2002). Compared to variables that are not important, permuting the values of an important variable leads to greater changes in prediction performance (Liaw & Wiener, 2002).

To date, the most common approaches using machine learning algorithms in soccer have focused on the relationship between the RPE and different internal and external loads during training sessions or matches for senior players (Geurkink et al., 2019; Jaspers et al., 2018; Rossi, Perri, Pappalardo, Cintia, & Iaia, 2019). Recently, machine learning algorithms have also been applied for predicting RPE based on GPS-derived external measures in youth soccer training (Marynowicz, Lango, Horna, Kikut, & Andrzejewski, 2022). However, match-related situational variables have not been considered in these studies. Therefore, the aim of this study was to explore how situational variables affect youth soccer players' RPE after official matches. This exploratory study represents the first one examining the importance of match-related situational variables with respect to youth soccer players' RPE variation in competitive matches using machine learning algorithms.

2. Materials and methods

2.1. Design

This research was conducted in nonexperimental conditions. The coaching staff and the participants did not receive any input from the research team. It was conducted in accordance with the Declaration of Helsinki (1964), updated in Fortaleza (2013), and was approved by the Ethics Committee of the University of the Basque Country (UPV/EHU): M10_2021_004_CALLEJA GONZÁLEZ. Prior to the start, parental written informed consent was obtained and all participants volunteered to participate in the study. They acknowledged that cannot be identified via the manuscript and were fully anonymized.

2.2. Participants

Thirty-five elite youth male soccer players (14.33 ± 0.86 years; 173.49 ± 6.16 cm; 63.44 ± 5.98 kg) belonging to two different teams of a professional club participated in this study. Players trained four times and played one official match every week. Players were grouped across different playing positions: central defender (CD) (n = 190); full-back (FB) (n = 72); central midfielder (CM) (n = 180); wide midfielder (WM) (n = 133); and attacker (AT) (n = 212). Goalkeepers were not included in the study, because of their different competitive requirements.

2.3. Procedures

Data collection was conducted when each team was competing at the highest level in their U15 category soccer league. This resulted in a two-season data collection period (2016–2017, 2017–2018), including 60 official matches (30 official matches per team). Data were supplied by the coaching staff, which was similar during both seasons.

Ten minutes after each match players rated their RPE and using a modified Borg CR-10 scale (Borg, 1998). This is the "gold standard" scale to collect subjective perception of exertion (Rago et al., 2020). The players answered individually the question "How hard was the match?". This was always asked by the strength and conditioning coach of the team. The RPE verbal anchors associated with the numerical answer were Rest (0) and Max exertion (10). Players were previously familiarized with this scale according to standard procedures (Borg, 1998), they used it at least one year before.

RPE values were included in the statistical analysis only if the participant played in the same position during the entire match. The final analysis included a total of 787 individual observations.

2.4. Situational variables

The following situational variables were considered for statistical analysis: player status, playing position, match location, match outcome, final scoreline and quality of the opponent. The players' game-playing time in each match was also considered as an independent variable. Based on the players' status, they were divided into starters (players who were in the starting line-up) and substitutes (players who played a match but were not included in the starting line-up). Players were categorized into one of five individual playing position: CD, FB, CM, WM and AT. With respect to match location, matches played at home and away were distinguished. With respect to the final outcome, matches were grouped into win, draw or lose. Final scoreline was categorized into one of five different scorelines: two or more goals up (2), one goal up (1), level scores (0), one goal down (−1) and two or more goals down (−2). The quality of the opponent was calculated using their final league position at the end of the season. For each season analyzed, the quality of the opponent was classified into three groups using k-means cluster analysis. Considering both seasons, this resulted in 9 high-level (ranked in the top 4 or 5 league positions), 10 medium-level (ranked from 5th to 9th or 6th to 10th in the league) and 13 low-level teams (ranked in the bottom 7 or 6 league positions). Both teams won a total of 34 matches, drew 14 and loss 12. K-means cluster analysis identified both teams as high-level teams (teams ended seasons in the 3rd and 5th position).

2.5. Statistical analyses

Random forest regression (RFR) was used to quantify the importance of situational variables in RPE. The model included player identity, player status, player game-playing time, playing position, match location, match outcome, final scoreline and quality of the opponent as predictor variables. Mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R-Squared, R^2) were used to evaluate the prediction error rates and model performance in the regression analysis. The final selected model had the highest R^2 and lowest MSE, RMSE and MAE. Next, the variable importance of the final selected model was evaluated using the Mean Decrease in Accuracy (MDA). Variables with a large MDA were strong predictors of the RPE. Default parameters were used for RFR. Considering the output of MDA, two different RFR models were also created depending on the player status. They were created following the abovementioned steps but only considered playing position, match location, match outcome, final scoreline and quality of the opponent. Afterwards, two linear mixed models were created to identify the variability in RPE among the match-related situational variables depending on the player status. The linear mixed models incorporated situational variables as fixed factors and player identity as the random factor. After fitting the models, the Kolmogorov–Smirnov test and Q-Q plot were applied to confirm normality assumptions of the RPE variable. Pairwise comparisons with Bonferroni adjustment on the *p* value were conducted in further analysis when significant main effects were observed. Statistical analysis was conducted in JASP 0.16.3.0 (Amsterdam, Netherlands). Statistical significance was established at $p < 0.05$.

3. Results

The number of RPE values per player ranged from 11 to 48, with a mean of 23.2 ± 6.7. The number of RPE values per match ranged from 10 to 15, with a mean of 13.9 ± 1.1.

The RPE variance explained was 55.2%, and the model showed a MSE of 0.397 arbitrary units (AU), a RMSE of 0.63 AU and a MAE of 0.479 AU (Table 1). The most important variables that affected youth soccer players' RPE were the player game-playing time, the player status and the player identity (Figure 1). However, the model accuracy decreased when it was carried out differentiating between player status and eliminating the abovementioned variables (Table 1). Moreover, for starter players, quality of the opponent, match outcome and final scoreline were the most important variables, but for substitutes only final scoreline was a strong independent predictor of the RPE (Figure 1).

The main effects of situational variables on players' RPE are reported in Table 2. The match outcome and the final scoreline showed significant effects on both starter and substitute players ($p < 0.05$). However, the effect of the quality of the opponent was only identified in starter players ($p < 0.05$).

Mean differences in the RPE of starters and substitutes are reported in Table 3. The RPE was higher in starter players when playing against high-level teams and after a loss with a difference or two or more goals down ($p < 0.05$). However, RPE was higher in substitute players only after a loss compared to a win ($p < 0.05$) and after a loss by two or more goals down compared to a win by two or more goals up ($p < 0.05$).

4. Discussion

The aim of this study was to explore how situational variables affect youth soccer players' RPE after official matches. The main findings indicate that the game-playing time, the player status and the player identity were the strongest predictors of youth soccer players' RPE. Moreover, the match outcome and the final scoreline showed significant effects on both starter and substitute players but the main effect of the quality of the opponent was only identified in starter players' RPE. These results allow practitioners to know how situational variables interact and modulate RPE after official matches and help them to prescribe and adapt the players' training content and load before and after matches.

The RFR reported a higher predictive accuracy than a previous study conducted with a similar population, which showed an RMSE of 1.621 ± 0.001 (Marynowicz et al., 2022). The sample size contributed to the findings. The current analysis included a total of 787 individual observations, similar to previous studies (Geurkink et al., 2019; Marynowicz et al., 2022), but lower than other ones (Jaspers et al., 2018; Rossi et al., 2019). However, the model accuracy decreased when it was conducted with a reduced sample size (differentiating between player status) and spectrum of variables. On the other hand, findings with regard to variable importance imply limited comparability with previous results due to the different machine learning algorithms and variables selected for the prediction of RPE (Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022; Rossi et al., 2019).

The player game-playing time and the status were the most important factors affecting players' RPE. This finding confirm that volume is the main contributor to how soccer players report subjective exertion (Rossi et al., 2019). As may be expected from the difference in accumulated playing time, starter players demonstrate significantly higher average and accumulated physical load compared to the non-starters (Anderson et al., 2016; Dalen & Lorås, 2019; Sydney, Wollin, Chapman, Ball, & Mara, 2022). Considering that some high-intensity external load metrics predict RPE (Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022), the RPE scores depending on game-playing time and player status were reflective of a different time exposed to physical loads. Even so, the MDA indicated that the player identity represents a strong predictor of the RPE. This endorse the importance of personal interpretation of the RPE scale and an individualized interpretation of the perceived exertion (Geurkink et al., 2019; Marynowicz et al., 2022).

The starter players' RPE was significantly different depending on the quality of the opponent, the match outcome and the final scoreline. In fact, the interactive effects of the quality of the opponent and the final scoreline were also significant. However, positional differences were not found, as in a recently research including youth elite players (Sydney et al., 2022), but opposed to findings obtained in adult players (Barrett et al., 2018). Therefore, further research is needed to clarify this finding.

Regarding the quality of the opponent, the current results support those reported by Barrett et al. (2018), who showed an increased exertion when playing against a higher-ranked opponent. The greater game-playing time of the starter players and their associated match-imposed external load likely influenced the higher RPE when played against high-level teams, because the better the quality of the opposing team, the later the substitutions take place (Gomez, Lago-Peñas, & Owen, 2016) and the greater the distance covered in all intensity range categories (Castellano, Blanco-Villaseñor, & Álvarez, 2011). Moreover, playing against high-level opposition is associated with lower ball possession, both in adult (Bradley, Lago-Peñas, Rey, & Sampaio, 2014; Lago, 2009) and in youth players (Varley et al., 2016), because top teams retain more possession to control the match by dictating play (Bradley, Lago-Peñas, & Rey, 2014). Indeed, teams ran more per minute when players did not have the ball (Castellano et al., 2022). This time in possession is related to the difference in very high-speed running distance covered when play against different level teams (Varley et al., 2016), so this may also contributed to the higher RPE reported by starter players when played against high-level teams.

The starter players' RPE was higher after loss matches with a difference of two or more goals down. This could be explained by changes in tactics and style of play according to the scoreline. Losing match-status is associated with an increase in time spent in possession compared with the win or draw situation (Bradley, Lago-Peñas, & Rey, 2014). When ahead, teams often prefer to decrease their possession and engage in counter-attacks or direct play (Lago, 2009). When behind, teams often prefer to control the match by dictating the play, increasing their possession and the number of attacking situations to increase the likelihood for positive game outcomes. Therefore, teams that are heavily defeated have to cover greater high-intensity running distance without the ball in an attempt to close players down and regain possession (Bradley, Lago-Peñas, Rey, & Gomez Diaz, 2013). This explain why both youth and adult players cover greater total distances, perform more high-intensity activity and decrease low-intensity activity when the result was adverse in order to draw or win the match (Algroy et al., 2021; Castellano et al., 2011; Lago, 2009; Lago, Casais, Dominguez, & Sampaio, 2010). Consequently, the higher RPE scores was reflective of a

Table 1
Evaluation metrics of the Random Forest Regression models.

	All players	Starters	Substitutes
MSE	0.397	0.954	0.827
RMSE	0.63	0.977	0.909
MAE	0.479	0.766	0.759
R^2	0.552	0.124	0.116

MSE: mean squared error; RMSE: root mean squared error; MAE: mean absolute error; R^2: coefficient of determination, R squared.

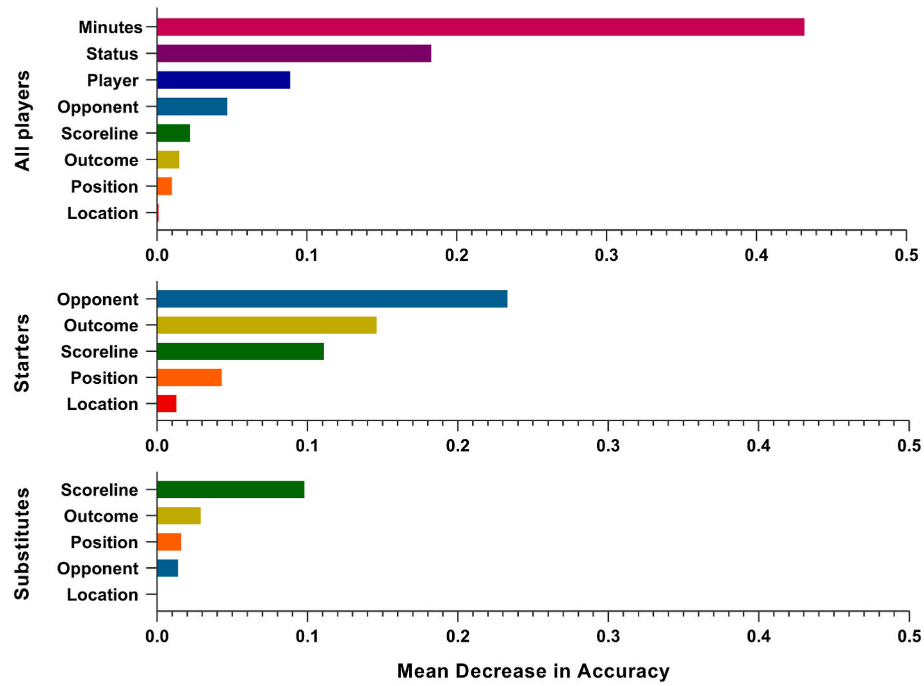

Figure 1. Variables with importance to influence youth soccer players' RPE after official matches, ordered by the Mean Decrease in Accuracy experienced within the model when each metric was permuted.

Table 2
Results of the linear mixed models. Main effects of situational variables and interaction among them.

	Starters				Substitutes			
	Num df	Den df	F	p	Num df	Den df	F	p
Position	4	103.51	1.661	0.165	4	121.10	0.418	0.795
Opponent	2	477.12	10.217	0.000	2	195.05	1.311	0.272
Location	1	458.12	3.114	0.078	1	152.95	0.001	0.987
Outcome	2	483.18	5.047	0.007	2	191.48	4.889	0.008
Scoreline	3	477.87	4.974	0.002	2	187.12	3.641	0.028
Position * Opponent	8	472.08	0.421	0.909	7	187.34	1.446	0.189
Position * Location	4	464.79	0.595	0.666	4	157.27	3.114	0.017
Position * Outcome	8	473.11	1.220	0.285	7	187.93	1.320	0.243
Position * Scoreline	8	469.33	1.090	0.369	7	187.79	1.256	0.275
Opponent * Location	2	470.99	1.839	0.160	2	187.37	0.785	0.457
Opponent * Outcome	4	471.74	0.830	0.506	4	185.35	0.721	0.579
Opponent * Scoreline	2	470.50	4.923	0.008	2	188.78	1.837	0.162
Location * Outcome	2	470.97	2.454	0.087	2	185.47	0.384	0.681
Location * Scoreline	2	470.11	1.117	0.328	2	187.08	1.812	0.166

Den: denominator; df: degrees of freedom; F: F-values; Num: numerator; p: p-values

high-intensity activity performed when losing, as these activities may predict players' RPE (Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022).

The substitute players' RPE was only affected by the match outcome and the final scoreline. Most of the substitutions occur at half-time and during the 60–90 min period (Anderson et al., 2016; Rey, Lago-Ballesteros, & Padrón-Cabo, 2015). This leaves substitute players a short period of time to influence the outcome of the match, at a time when starter players are pre-fatigued or having to adopt a different pacing strategy (Waldron & Highton, 2014). During this period, substitutes are able to achieve relatively higher values in physical performance variables compared to starting and full-match players (Bradley, Lago-Peñas, & Rey, 2014; Bradley & Noakes, 2013; Dijkhuis, Kempe, & Lemmink, 2021; Sydney et al., 2022). Notwithstanding, the most important factor explaining the timing of the first player substitutions on a team is the score as it stands prior to the substitution (Del Corral, Pestana Barros, & Prieto-Rodríguez, 2008). When the teams lost, coaches made the substitutions earlier than when drawing or winning (Del Corral et al., 2008; Gomez et al., 2016; Rey et al., 2015). Additionally, players try to reach their maximal physical capacity when losing (Algroy et al., 2021; Castellano et al., 2011; Lago, 2009; Lago et al., 2010). This explain why RPE was higher in substitute players only after losses compared to wins and after losses by two or more goals down compared to wins by two or more goals up.

Perceptions may reflect how the brain integrates and categorizes the input signals resulting from various different stimuli (Abbiss et al., 2015). In this regard, perceived exertion is likely to be dictated by not only the discrepancy or balance between predicted and actual sensory feedback (Abbiss et al., 2015) but also other complex psychophysiological factors. These include concentrations of hormones, substrate levels, external factors (e.g., environment, spectators), personality (e.g., extraversion, neuroticism), psychological states (e.g., depression, anxiety), motivation, awareness and memory/prior experience of similar exercise (Abbiss et al., 2015; Borresen & Lambert, 2008; Morgan, 1973, 1994; St Clair Gibson et al., 2006). In fact, the low level of cognitive development (Borg, 1998) and the poor education of players result in

Table 3
Summary of Bonferroni post-hoc pairwise comparisons.

	Starters' RPE (au) (M ± SD)	Substitutes' RPE (au) (M ± SD)	Starters mean differences	Substitutes mean differences
Position				
Central Defenders	5.9 ± 1.2	4.0 ± 1.1		
Full-backs	6.6 ± 0.9	3.4 ± 1.4		
Central midfielders	6.1 ± 1.1	4.2 ± 1.1		
Wide midfielders	6.2 ± 1.0	4.4 ± 1.1		
Attackers	6.3 ± 1.1	4.3 ± 1.3		
Opponent				
High-level	6.8 ± 1.1	4.4 ± 1.3	High > Medium *	
Medium-level	6.1 ± 1.1	4.4 ± 1.2	High > Low *	
Low-level	5.9 ± 1.1	4.0 ± 1.1		
Location				
Home	6.1 ± 1.2	4.3 ± 1.1		
Away	6.3 ± 1.1	4.1 ± 1.3		
Outcome				
Win	6.0 ± 1.2	4.0 ± 1.2	Win < Loss *	Win < Loss *
Draw	6.1 ± 1.1	4.4 ± 1.0	Draw < Loss *	
Loss	6.7 ± 0.9	4.7 ± 1.2		
Scoreline				
2 goals	6.0 ± 1.1	4.0 ± 1.2	2 goals < −2 goals *	2 goals < −2 goals *
1 goal	6.0 ± 1.2	4.2 ± 1.4	2 goals < −1 goal *	
0 goal	6.1 ± 1.1	4.4 ± 1.0	1 goal < −2 goals *	
-1 goal	6.6 ± 1.1	4.3 ± 1.1	0 goal < −2 goals *	
-2 goals	6.7 ± 0.8	5.0 ± 1.2		

2: two or more goals up; 1: one goal up; 0: level scores; −1: one goal down; −2: two or more goals down; au: arbitrary units; M: mean; SD: standard deviation.
* Statistically significant difference ($p < 0.05$).

misunderstanding of RPE (Rago et al., 2020). Thus, it is likely that training status and psychophysiological characteristics have been shaped by the demands of players' training and competition context and the results would not be generalizable to adult players. Research aimed at examining the RPE across a continuum of ages and fitness levels would provide much-needed information in this area. Other subjective self-reported measures were not considered in this study. Previous research has shown that RPE may be related to the stress that players support in the previous week (e.g., psychological tension induced by the distance to the matchday) (Rossi et al., 2019). In addition, perceived exertion is likely to be influenced by the most dominant psychophysiological sensation (McLaren, Smith, Spears, & Weston, 2017), yet the response rates of internal biochemical and mechanical stresses are considerably different (Vanrenterghem, Nedergaard, Robinson, & Drust, 2017). Therefore, the inclusion pre-match perceived wellness may further clarify the current findings. Including variables of match-related internal and external loads to explore the role of the different predictive indicators and their interactions offers opportunities for further research. In future works, it could also be interesting to differentiate between earlier substitutes or later substitutes (Bradley & Noakes, 2013), which may provide a deeper understanding of how youth players modulate their perceived exertion.

5. Conclusions

The game-playing time, the player status and the player identity were the strongest predictors of youth soccer players' RPE. Moreover, the match outcome and the final scoreline showed significant effects on both starter and substitute players but the main effect of the quality of the opponent was only identified in starter players' RPE. These results allow practitioners to know how situational variables interact and modulate RPE after official matches and help them to prescribe and adapt the players' training content and load before and after matches, possibly preventing negative training effects. In fact, workload monitoring at the youth level is essential, not only to enable players to reach higher performance levels, but also to preserve athletes' health in the long term and consequently avoiding early retirement (Bourdon et al., 2017). For this reason, the use of different machine learning algorithms that allow predicting internal loads based on match-related situational variables could be a useful tool to improve training periodization of youth soccer players.

Ethical approval

Ethics Committee of the University of the Basque Country: M10_2021_004_CALLEJA GONZÁLEZ.

Authors' contributions

Diego Marqués-Jiménez: research concept and study design, literature review, data collection, data analysis and interpretation, writing of the manuscript, reviewing/editing a draft of the manuscript; Jaime Sampaio: data analysis and interpretation, writing of the manuscript, reviewing/editing a draft of the manuscript; Julio Calleja-González: writing of the manuscript, reviewing/editing a draft of the manuscript; Ibon Echeazarra: research concept and study design, data collection, writing of the manuscript, reviewing/editing a draft of the manuscript.

Declaration of competing interest

The authors report there are no competing interests to declare.

Data availability

The authors do not have permission to share data.

Acknowledgments

The authors thank the athletes, coaching staff and club who made this study possible.

References

Abbiss, C. R., Peiffer, J. J., Meeusen, R., & Skorski, S. (2015). Role of ratings of perceived exertion during self-paced exercise: What are we actually measuring? *Sports Medicine, 45*(9), 1235–1243. https://doi.org/10.1007/s40279-015-0344-5

Algroy, E., Grendstad, H., Riiser, A., Nybakken, T., Saeterbakken, A. H., Andersen, V., & Gundersen, H. S. (2021). Motion analysis of match play in U14 male soccer players and the influence of position, competitive level and contextual variables. *International Journal of Environmental Research and Public Health, 18*(14), 7287. https://doi.org/10.3390/ijerph18147287

Anderson, L., Orme, P., Di Michele, R., Close, G. L., Milsom, J., Morgans, R., … Morton, J. P. (2016). Quantification of seasonal-long physical load in soccer players with different starting status from the English premier league: Implications for maintaining squad physical fitness. *International Journal of Sports Physiology and Performance, 11*(8), 1038–1046. https://doi.org/10.1123/ijspp.2015-0672

Barrett, S., McLaren, S., Spears, I., Ward, P., & Weston, M. (2018). The influence of playing position and contextual factors on soccer players' match differential ratings of perceived exertion: A preliminary investigation. *Sports, 6*(1), 13. https://doi.org/10.3390/sports6010013

Borg, G. (1998). Borg's perceived exertion and pain scales. *Human Kinetics*.

Borresen, J., & Lambert, M. I. (2008). Quantifying training load: A comparison of subjective and objective methods. *International Journal of Sports Physiology and Performance, 3*(1), 16–30. https://doi.org/10.1123/ijspp.3.1.16

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., … Cable, N. T. (2017). Monitoring athlete training loads: Consensus statement. *International Journal of Sports Physiology and Performance, 12*(Suppl. 2), 161–170. https://doi.org/10.1123/ijspp.2017-0208

Bradley, P. S., Lago-Peñas, C., & Rey, E. (2014). Evaluation of the match performances of substitution players in elite soccer. *International Journal of Sports Physiology and Performance, 9*(3), 415–424. https://doi.org/10.1123/ijspp.2013-0304

Bradley, P. S., Lago-Peñas, C., Rey, E., & Gomez Diaz, A. (2013). The effect of high and low percentage ball possession on physical and technical profiles in English FA Premier League soccer matches. *Journal of Sports Sciences, 31*(12), 1261–1270. https://doi.org/10.1080/02640414.2013.786185

Bradley, P. S., Lago-Peñas, C., Rey, E., & Sampaio, J. (2014). The influence of situational variables on ball possession in the English Premier League. *Journal of Sports Sciences, 32*(20), 1867–1873. https://doi.org/10.1080/02640414.2014.887850

Bradley, P. S., & Noakes, T. D. (2013). Match running performance fluctuations in elite soccer: Indicative of fatigue, pacing or situational influences? *Journal of Sports Sciences, 31*(15), 1627–1638. https://doi.org/10.1080/02640414.2013.796062

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Brito, J., Hertzog, M., & Nassis, G. P. (2016). Do match-related contextual variables influence training load in highly trained soccer players? *The Journal of Strength & Conditioning Research, 30*(2), 393–399. https://doi.org/10.1519/jsc.0000000000001113

Castellano, J., Blanco-Villaseñor, A., & Álvarez, D. (2011). Contextual variables and time-motion analysis in soccer. *International Journal of Sports Medicine, 32*(6), 415–421. https://doi.org/10.1055/s-0031-1271771

Castellano, J., Errekagorri, I., Los Arcos, A., Casamichana, D., Martín-Garcia, A., Clemente, F. M., … Echeazarra, I. (2022). Tell me how and where you play football and I'll tell you how much you have to run. *Biology of Sport, 39*(3), 607–614. https://doi.org/10.5114/biolsport.2022.106155

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Dalen, T., & Lorås, H. (2019). Monitoring training and match physical load in junior soccer players: Starters versus substitutes. *Sports, 7*(3), 70. https://doi.org/10.3390/sports7030070

Del Corral, J., Pestana Barros, C., & Prieto-Rodríguez, J. (2008). The determinants of soccer player substitutions A survival analysis of the Spanish soccer league. *Journal of Sports Economics, 9*(2), 160–172. https://doi.org/10.1177/1527002507308309

Dijkhuis, T. B., Kempe, M., & Lemmink, K. A. P. M. (2021). Early prediction of physical performance in elite soccer matches - a machine learning approach to support substitutions. *Entropy, 23*(8), 952. https://doi.org/10.3390/e23080952

Geurkink, Y., Vandewiele, G., Lievens, M., de Turck, F., Ongenae, F., Matthys, S. P. J., … Bourgois, J. G. (2019). Modeling the prediction of the session rating of perceived exertion in soccer: Unraveling the puzzle of predictive indicators. *International Journal of Sports Physiology and Performance, 14*(6), 841–846. https://doi.org/10.1123/ijspp.2018-0698

Gomez, M. A., Lago-Peñas, C., & Owen, A. L. (2016). The influence of substitutions on elite soccer teams' performance. *International Journal of Performance Analysis in Sport, 16*(2), 553–568. https://doi.org/10.1080/24748668.2016.11868908

Haddad, M., Stylianides, G., Djaoui, L., Dellal, A., & Chamari, K. (2017). Session-RPE method for training load monitoring: Validity, ecological usefulness, and influencing factors. *Frontiers in Neuroscience, 11*, 612. https://doi.org/10.3389/fnins.2017.00612

Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and external training load: 15 Years on. *International Journal of Sports Physiology and Performance, 14*(2), 270–273. https://doi.org/10.1123/ijspp.2018-0935

Jaspers, A., De Beéck, T. O., Brink, M. S., Frencken, W. G. P., Staes, F., Davis, J. J., & Helsen, W. F. (2018). Relationships between the external and internal training load in professional soccer: What can we learn from machine learning? *International Journal of Sports Physiology and Performance, 13*(5), 625–630. https://doi.org/10.1123/ijspp.2017-0299

Lago, C. (2009). The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *Journal of Sports Sciences, 27*(13), 1463–1469. https://doi.org/10.1080/02640410903131681

Lago, C., Casais, L., Dominguez, E., & Sampaio, J. (2010). The effects of situational variables on distance covered at various speeds in elite soccer. *European Journal of Sport Science, 10*(2), 103–109. https://doi.org/10.1080/17461390903273994

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News, 2*(3), 18–22.

Marynowicz, J., Lango, M., Horna, D., Kikut, K., & Andrzejewski, M. (2022). Predicting ratings of perceived exertion in youth soccer using decision tree models. *Biology of Sport, 39*(2), 245–252. https://doi.org/10.5114/biolsport.2022.103723

McLaren, S. J., Smith, A., Spears, I. R., & Weston, M. (2017). A detailed quantification of differential ratings of perceived exertion during team-sport training. *Journal of Science and Medicine in Sport, 20*(3), 290–295. https://doi.org/10.1016/j.jsams.2016.06.011

Morgan, W. P. (1973). Psychological factors influencing perceived exertion. *Medicine & Science in Sports, 5*(2), 97–103.

Morgan, W. P. (1994). Psychological components of effort sense. *Medicine & Science in Sports & Exercise, 26*(9), 1071–1077.

de Morree, H. M., Klein, C., & Marcora, S. M. (2012). Perception of effort reflects central motor command during movement execution. *Psychophysiology, 49*(9), 1242–1253. https://doi.org/10.1111/j.1469-8986.2012.01399.x

Rago, V., Brito, J., Figueiredo, P., Costa, J., Krustrup, P., & Rebelo, A. (2020). Internal training load monitoring in professional football: A systematic review of methods using rating of perceived exertion. *The Journal of Sports Medicine and Physical Fitness, 60*(1), 160–171. https://doi.org/10.23736/s0022-4707.19.10000-x

Rey, E., Lago-Ballesteros, J., & Padrón-Cabo, A. (2015). Timing and tactical analysis of player substitutions in the UEFA Champions League. *International Journal of Performance Analysis in Sport, 15*(3), 840–850. https://doi.org/10.1080/24748668.2015.11868835

Rossi, A., Perri, E., Pappalardo, L., Cintia, P., & Iaia, F. M. (2019). Relationship between external and internal workloads in elite soccer players: Comparison between rate of perceived exertion and training load. *Applied Sciences, 9*(23), 5174. https://doi.org/10.3390/app9235174

St Clair Gibson, A., Lambert, E. V., Rauch, L. H. G., Tucker, R., Baden, D. A., Foster, C., & Noakes, T. D. (2006). The role of information processing between the brain and peripheral physiological systems in pacing and perception of effort. *Sports Medicine, 36*(8), 705–722. https://doi.org/10.2165/00007256-200636080-00006

Sydney, M. G., Wollin, M., Chapman, D., Ball, N., & Mara, J. K. (2022). Substitute running outputs in elite youth male soccer players: Less peak but greater relative running outputs. *Biology of Sport, 40*(1), 241–248. https://doi.org/10.5114/biolsport.2023.112969

Vanrenterghem, J., Nedergaard, N. J., Robinson, M. A., & Drust, B. (2017). Training load monitoring in team sports: A novel framework separating physiological and biomechanical load-adaptation pathways. *Sports Medicine, 47*(11), 2135–2142. https://doi.org/10.1007/s40279-017-0714-2

Varley, M. C., Gregson, W., McMillan, K., Bonanno, D., Stafford, K., Modonutti, M., & Di Salvo, V. (2016). Physical and technical performance of elite youth soccer players during international tournaments: Influence of playing position and team success and opponent quality. *Science and Medicine in Football, 1*(1), 18–29. https://doi.org/10.1080/02640414.2016.1230676

Waldron, M., & Highton, J. (2014). Fatigue and pacing in high-intensity intermittent team sport: An update. *Sports Medicine, 44*(12), 1645–1658. https://doi.org/10.1007/s40279-014-0230-6

Weston, M. (2013). Difficulties in determining the dose-response nature of competitive soccer matches. *Journal of Athletic Enhancement, 2*(1), 1–2. https://doi.org/10.4172/2324-9080.1000e107