Relationships between external and internal training load in elite youth soccer players: gps, heart rate, and subjective questionnaires

Abstract of PhD Thesis

Krisztián Havanecz

Doctoral School of Sport Sciences
Hungarian University of Sports Science

Supervisor: Dr. Gábor Géczi professor, PhD
Consultant: Dr. Sándor Sáfár associate professor, PhD

Official reviewers: Dr. habil. Tamás Szabó private professor, CSc
Dr. habil. Tibor Hortobágyi research professor, DSC

Budapest
2026

Introduction

Training load (TL) can be divided into two main components: external training load (ETL), which represents the physical work performed by the athlete, and internal training load (ITL), which reflects the individual physiological and psychological responses to a given external load (Impellizzeri et al., 2004; Foster et al., 2001). ETL can be further categorized into locomotor training load (LTL) and mechanical training load (MTL). LTL describes displacement-based movements, while MTL refers to changes in movement patterns and micro-movements (Martín-García et al., 2020).

With the widespread use of global positioning system (GPS)-based technologies in modern soccer, objective assessment of ETL has become increasingly accessible. In contrast, ITL monitoring is primarily based on heart rate-derived methods and subjective measures such as rating of perceived exertion (RPE) and session rating of perceived exertion (s-RPE) (Scott et al., 2013; Weaving et al., 2014). Investigating the relationship between these two dimensions of training load is essential, as identical external loads may elicit different internal responses, particularly in youth athletes (Bartlett et al., 2017).

In youth soccer, training load monitoring presents additional challenges due to differences in biological maturation, the development of perceptual systems, and various contextual factors, all of which contribute to variability in ETL–ITL relationships (Teixeira et al., 2022). Previous research suggests that these relationships are generally moderate to strong; however, their magnitude may vary depending on age group, training context (training vs. match), and the specific variables examined (Marynowicz et al., 2020).

Therefore, a comprehensive investigation of the relationship between ETL and ITL variables in youth soccer players is warranted, with particular emphasis on age-related characteristics and different training environments (Connolly et al., 2024).

Objectives

The primary aim of the present research was to investigate the relationship between objective and subjective indicators of TL in elite youth soccer players, with particular emphasis on how these variables represent perceived exertion during both training sessions and match.

The first study aimed to examine the associations between ETL variables—including locomotor load (e.g., total distance covered and distance covered at different intensity zones) and mechanical load (e.g., accelerations, decelerations, inertial movement analysis, and player load)—and subjective internal load measures (rating of perceived exertion (RPE) and session rating of perceived exertion (s-RPE) in both training and match contexts. In addition, the study sought to identify which ETL variables best reflect perceived exertion, and which predictors most strongly influence RPE and s-RPE across different age groups (U15, U17, U19).

The second study aimed to investigate, alongside subjective measures (RPE and s-RPE), the relationship between heart rate-based internal load variables—namely training impulse (TRIMP), Edwards's TRIMP (eTRIMP), and heart rate exertion (HR exertion)—and ETL variables during both training and matches. Furthermore, this study aimed to determine the extent to which heart rate-derived variables can predict ETL variables in a training within an elite U16 soccer team.

Methods

Study 1

In the first study, players from a Hungarian elite youth soccer academy participated (U15, U17, and U19 age groups; N = 50). Data collection was conducted over an 11-week period, including both training sessions and matches, resulting in a total of 1,386 observations, which comprised 145 training sessions and 11 matches per team. ETL was monitored using GPS devices with a sampling frequency of 10 Hz (Catapult S7 Vector, Catapult Sports Ltd., Melbourne, Australia). The GPS units also included additional integrated sensors—namely inertial measurement units (IMUs), such as a triaxial accelerometer, gyroscope, and magnetometer—which were used to assess MTL. The analyzed locomotor variables included total distance covered (TDC, m), moderate-speed running distance (MSR, 14.4–19.79 $km\cdot h^{-1}$, m), high-speed running distance (HSR, 19.8–25.1 $km\cdot h^{-1}$, m), sprint distance (SPR, >25.2 $km\cdot h^{-1}$, m), acceleration (ACC, >1.5 $m \cdot s^{-2}$, m), deceleration (DEC, <−1.5 $m \cdot s^{-2}$, m), inertial movement analysis (IMA, AU), and player load (PL, AU). All variables were also expressed as intensity-based (relative) metrics.

ITL was determined using subjective psychological response variables, including rating of perceived exertion (RPE) and its duration-weighted form, session rating of perceived exertion (s-RPE), which are validated tools for estimating internal load (Foster et al., 2001). RPE data were collected using a standardized protocol: players reported their perceived exertion following training sessions and matches within a predefined time window (typically within 15–30 minutes), using a validated scale (Borg CR10 scale; Borg, 1998).

Regarding statistical analysis, in the first study, training and match data were analyzed both separately and in a combined manner, with additional stratification by age group. Normality of the data was assessed using the Shapiro–Wilk test. Relationships between ETL and ITL variables were examined using Pearson correlation coefficients, with interpretation based on Hopkins et al. (2009). Age-related differences were analyzed using multigroup structural equation modeling (SEM), with model fit evaluated using standard fit indices. In addition, multiple linear regression analyses were conducted to determine the extent to which GPS-derived ETL variables explain the variance in RPE and s-RPE. The level of statistical significance was set at $p < 0.05$.

Study 2

In the second study, elite U16 youth soccer players participated (N = 19). Data collection was conducted in both training and match environments, resulting in a total of 534 observations derived from 50 training sessions and 11 matches.

In addition to GPS-derived ETL variables and subjective measures (RPE and s-RPE), heart rate-based variables were also analyzed to quantify the cardiovascular responses to training load:

- Training impulse (TRIMP; Hopkins, 1991) was calculated as the product of average heart rate and activity duration (AU).
- Edwards TRIMP (eTRIMP; Edwards, 1994) was calculated by multiplying the time spent in each heart

rate zone by a corresponding weighting factor: time spent in zone 1 (50–59% HRmax) was multiplied by 1, in zone 2 (60–69% HRmax) by 2, in zone 3 (70–79% HRmax) by 3, in zone 4 (80–89% HRmax) by 4, and in zone 5 (90–100% HRmax) by 5. These values were then summed and expressed in arbitrary units (AU).

- Heart rate exertion (HR exertion; proprietary variable of Catapult Sports) was calculated by applying weighted coefficients to the time spent in different heart rate zones: zone 1 (≤45% HRmax) ×1, zone 2 (45–55% HRmax) ×1.122, zone 3 (55–65% HRmax) ×1.322, zone 4 (65–75% HRmax) ×1.554, zone 5 (75–85% HRmax) ×2.037, zone 6 (85–95% HRmax) ×3.252, zone 7 (95–105% HRmax) ×5.439, and zone 8 (>105% HRmax) ×9. These values were then summed and expressed in arbitrary units (AU).

Training and match data were analyzed separately. Based on the results of the normality tests, non-parametric Spearman rank correlation coefficients were used to examine relationships between ETL and ITL variables, as well as between objective and subjective ITL variables, with interpretation again based on Hopkins et al. (2009). Multiple linear regression models were applied to identify determinants of internal load, where TRIMP and HR exertion were used as dependent variables and GPS-derived variables as predictors. Model performance was evaluated using the coefficient of determination (R^2) and standardized regression coefficients (β). The level of statistical significance was set at $p < 0.05$.

Results

Study 1

During U15 training sessions, s-RPE showed the strongest relationships with TDC (r = 0.52), PL (r = 0.49), DEC (r = 0.43), and HSR (r = 0.42), whereas in the case of RPE, these relationships were generally weaker (HSR: r = 0.38; MSR: r = 0.29; TDC: r = 0.25). In U15 matches, considerably weaker relationships were observed; s-RPE showed significant associations primarily with TDC (r = 0.53) and PL (r = 0.44).

In the U17 age group, several very strong relationships were identified during training: s-RPE was associated with

TDC (r = 0.77), DEC (r = 0.76), PL (r = 0.73), ACC (r = 0.70), HSR (r = 0.64), and SPR (r = 0.62). In parallel, the strongest relationships for RPE were observed with PL (r = 0.71), IMA (r = 0.63), TDC (r = 0.62), DEC·min^{-1} (r = 0.60), and MSR (r = 0.60). In U17 matches, these relationships became weaker, with s-RPE correlations ranging between r = 0.29 and 0.41.

In the U19 age group, during training, s-RPE showed moderate-to-strong relationships with TDC (r = 0.61), DEC (r = 0.61), PL (r = 0.61), ACC (r = 0.59), and MSR (r = 0.55), while for RPE the strongest associations were found with HSR (r = 0.46), SPR·min^{-1} (r = 0.45), DEC (r = 0.43), and MSR·min^{-1} (r = 0.42). During U19 matches, s-RPE was associated with TDC (r = 0.59), IMA (r = 0.54), and PL (r = 0.45), whereas RPE showed significant relationships with IMA (r = 0.43), PL (r = 0.42), TDC (r = 0.31), HSR (r = 0.29), and MSR (r = 0.29).

The results of the multiple linear regression analyses indicated that the explanatory power of the RPE model varied considerably across age groups (U15: $R^2 = 0.119$; U17: $R^2 = 0.589$; U19: $R^2 = 0.443$), whereas higher values were obtained for s-RPE in all age groups (U15: R^2 =

0.361; U17: $R^2 = 0.689$; U19: $R^2 = 0.617$). Based on standardized regression coefficients (β), TDC emerged as the strongest predictor in the s-RPE models (U15: $\beta = 0.572$; U17: $\beta = 0.511$; U19: $\beta = 0.654$), along with PL (U15: $\beta = 0.572$; U17: $\beta = 0.511$; U19: $\beta = 0.654$). Additionally, SPR·min^{-1} ($\beta = 0.513$) showed a notable contribution in U19, while the roles of ACC, DEC, MSR·min^{-1}, and SPR·min^{-1} varied across age groups. In the RPE model (including TDC, MSR·min^{-1}, SPR·min^{-1}, and DEC·min^{-1} as covariates), a strong predictor was identified only in U19, specifically SPR·min^{-1} ($\beta = 0.508$). The results of the multigroup structural equation modeling (SEM) indicated that the pattern of residual covariances between load variables was similar across age groups. In the constrained model, equality constraints resulted in identical parameter estimates (TDC–s-RPE; PL–s-RPE; DEC–s-RPE), all of which were statistically significant ($p < 0.001$). The comparison between constrained and unconstrained models revealed a significant difference ($\Delta\chi^2 = 69.8$; $\Delta df = 6$; $p < 0.001$); however, the model fit indices indicated poor fit in both cases (e.g., CFI = 0.170; RMSEA = 0.763 for the constrained model).

Study 2

During matches, HR exertion showed very strong correlations with decelerations (DEC), IMA, and PL variables ($r = 0.70–0.85$; $p < 0.001$). For TRIMP, the strongest relationship was observed with TDC ($r = 0.73$; $p < 0.001$), while MSR, IMA, and PL variables showed strong but lower correlations ($r = 0.50–0.64$). Among subjective load measures, both RPE ($r = 0.82$) and s-RPE ($r = 0.80$) showed very strong relationships with TDC.

During training, the strongest relationship among internal load variables was observed between TRIMP and HR exertion ($r = 0.96$; $p < 0.001$). In addition, s-RPE showed a very strong relationship with TRIMP ($r = 0.88$; $p < 0.001$). In match data, the relationship between TRIMP and HR exertion remained strong ($r = 0.81$), while a very strong correlation was observed between s-RPE and TRIMP ($r = 0.74$; $p < 0.001$).

In the U16 age group, regression analysis revealed a significant result for the s-RPE model ($F(1,465) = 596.71$; $p < 0.001$), with high explanatory power ($R^2 = 0.820$). Based on standardized regression coefficients, TDC was identified as an exceptionally strong predictor ($\beta = 0.98$; p

< 0.001). In contrast, HSR, SPR, and ACC contributed only minimally to the model. For the RPE model, a significant relationship was also observed ($F(1,465) = 1156.49$; $p < 0.001$), in which TDC explained 71% of the variance in HR exertion ($R^2 = 0.71$).

Conclusions

The aim of the present research was to explore the relationships between external (ETL) and internal (ITL) training load variables in youth soccer players, with particular emphasis on differences between training and match contexts, as well as the comparability of different measurement methods (GPS-, heart rate-, and subjective-based indicators).

Based on the evaluation of the hypotheses, it can be concluded that the majority of the assumptions formulated in both the first and second studies were confirmed or partially confirmed.

Study 1

H1: LTL variables show stronger relationships with RPE and s-RPE than MTL variables.

- The hypothesis was confirmed. RPE, and especially s-RPE, showed stronger relationships with locomotor, primarily volume-based variables than with mechanical variables, indicating that perceived exertion is mainly determined by training volume.

H2: TDC, HSR, and PL variables show the strongest relationships with RPE and s-RPE.

- The hypothesis was confirmed in training and partially confirmed in matches. TDC and PL consistently showed the strongest relationships, while the role of HSR was dependent on age group and context.

H3: The relationship between HSR and SPR with RPE/s-RPE is stronger in older age groups.

- The hypothesis was confirmed in training and partially confirmed in matches. The relationship between intensity-based variables and perceived exertion increased with age, which can be explained by biological maturation and the development of load perception.

Study 2

H4: Heart rate-based variables show strong relationships with TDC and PL during training, and weaker relationships during matches.

- The hypothesis was confirmed. Very strong relationships were observed during training, while in matches these relationships were more moderate but remained significant.

H5: Heart rate-based variables show stronger relationships with ETL than RPE/s-RPE during matches.

- The hypothesis was partially confirmed. s-RPE did not show weaker relationships, indicating that subjective measures reflect load in a complex and integrative manner.

H6: Strong relationships exist between heart rate-based variables and RPE/s-RPE.

- The hypothesis was confirmed. s-RPE showed very strong relationships with heart rate-based variables, while RPE consistently showed weaker associations.

Based on the findings of the studies, it can be concluded that in youth soccer players, internal load is primarily determined by the volume-related components of external load, while the contribution of intensity and mechanical variables is secondary. The prominent predictive role of TDC was supported by regression models, which demonstrated that it explained 82% of the variance in TRIMP and 71% of the variance in HR exertion. From a practical perspective, this indicates that volume-based metrics should receive particular attention in training load monitoring, especially in youth populations.

A near-perfect relationship was observed between heart rate-based internal load variables, suggesting that these methods consistently describe cardiovascular load. In addition, s-RPE showed very strong relationships with both heart rate-based and external load variables, confirming its practical applicability as a reliable and easily implementable method, particularly in situations where heart rate monitoring is not available.

The comparison of training and match environments revealed that while relationships between external and internal load variables are stronger and more predictable during training, they become weaker in match conditions, reflecting the more complex and less controlled nature of match load. In practical terms, this suggests that multiple indicators should be considered simultaneously when evaluating match load, rather than relying on a single variable.

Furthermore, the relationship between intensity-based variables and perceived exertion increased with age, which can be attributed to biological maturation and the development of perceptual responses. Accordingly, age-specific considerations are essential in training load monitoring, particularly when interpreting high-intensity load variables in youth soccer.

Publications

- Havanecz, K., Sáfár, S., Bartha, C., Kopper, B., Horváth, T., Tóth, P. J., ... & Géczi, G. (2025). Relationship Between Internal and External Load in Under-16 Soccer Players: Heart Rate, Rating of Perceived Exertion, and GPS-Derived Variables. *Sports*, *13*(11), 376.
- Havanecz, K., Tóth, P. J., Kopper, B., Bartha, C., Sáfár, S., Fridvalszki, M., & Géczi, G. (2025). Relationship Between GPS-Derived Variables and

Subjective Questionnaires Among Elite Youth Soccer Players. *Sports*, *13*(8), 246.

- Géczi, G., Sáfár, S., & Havanecz, K. (2022). A kosárlabda-, kézilabda-és labdarúgás-sportteljesítményt meghatározó változók vizsgálata: Fókuszban gps-és lps-rendszerek. *Acta Universitatis: Sectio Sport-Acta Universitatis de Carolo Eszterházy Nominatae*, *52*, 73–81.
- Havanecz, K., Géczi, G., & Sáfár, S. (2020). Áttekintés az eddigi kosárlabda-, kézilabda- és labdarúgás-sportteljesítményt meghatározó változóknak a vizsgálatáról gps- és lps-rendszerekkel. *Acta Universitatis: Sectio Sport-Acta Universitatis de Carolo Eszterházy Nominatae*, *49*(2), 49–57.