The integration of external and internal match load with contextual factors to assess training status in football

Bachelor Thesis to acquire the Bachelor of Science (BSc) in
Human Movement in Sports and Exercise

Submitted to the Institute of Sport and Exercise Sciences

by
Alexander Greif
born 23rd of October 1993

Address: [redacted]

E-Mail: a_grei05@uni-muenster.de
Student Nr.: [redacted]

First Examiner: Ross Julian
Second Examiner: Matthias Hendricks

Gladbeck, March 2021

List of Abbreviationss

ASRM	Athlete self-report measure
EL	External load
HR	Heart rate
HRV	Heart rate variability
HSRD	High speed running distance
IL	Internal load
ML	Match load
RPE	Rating of perceived exertion
TD	Total distance covered
TL	Training load
TRIMP	Training impulse
VO_{2max}	Maximal oxygen consumption

List of Figures

List of Tables

Table of Contents

The evaluation and adjustment of the training process as a mean to reach a certain physical performance goal requires regular monitoring of the training process and outcome (Borresen & Lambert, 2009). While the training process involves the organization, quality and quantity of exercise prescribed to create a specific training stimulus, the training outcome starts with the response to that stimulus and ultimately becomes relevant through its manifestation in changes in the goal-related performance capability, or training status (Impellizzeri et al., 2019) (*Figure 1*). In football, the monitoring of training load (TL) and match load (ML) can help assess training stimulus and training status. In order to do so, professional football clubs monitor various load parameters in an attempt to capture the complex demands of the game (Akenhead & Nassis, 2016). These parameters can be characterized as measures of external load (EL) or internal load (IL) (Impellizzeri et al., 2004). EL is the physical work exerted by an athlete (Impellizzeri et al., 2019). Measures such as total distance covered (TD), high speed running distance (HSRD), accelerations or metabolic power (Osgnach et al., 2010) describe EL. IL refers to the psychophysiological response to EL (Impellizzeri et al., 2019) and commonly includes parameters such as heart rate (HR) derived training impulse (TRIMP), oxygen uptake and rating of perceived exertion (RPE). It is this psychophysiological response that stimulates the adaptations causing a specific training outcome (Impellizzeri et al., 2005).

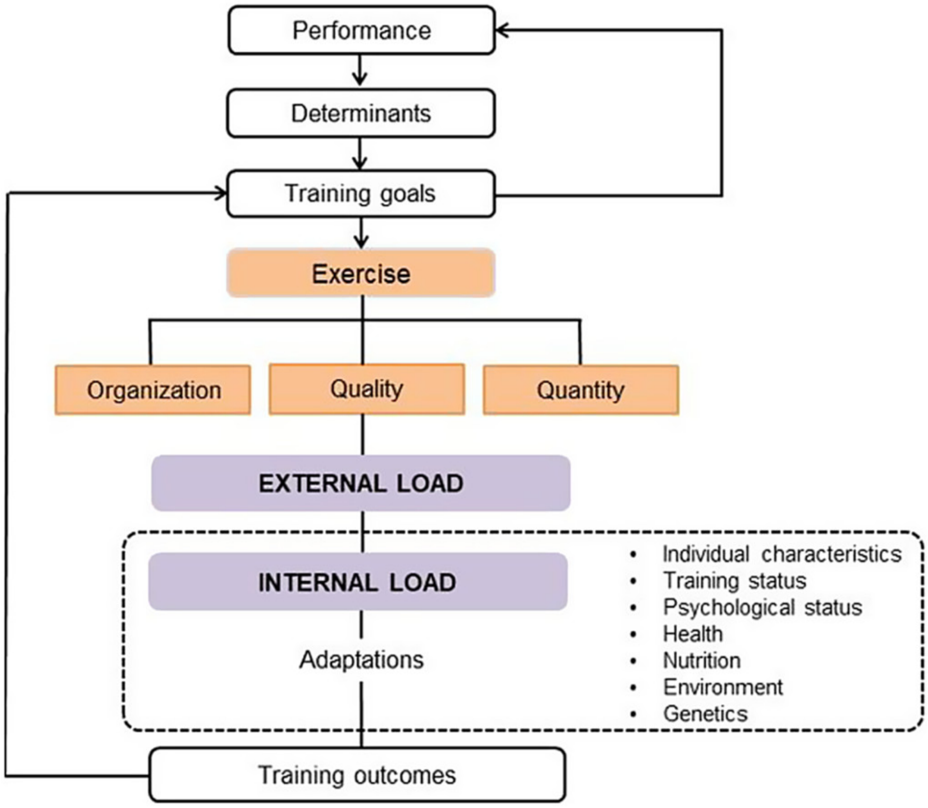
Despite the variety of approaches to load monitoring, the end goal always stays the same – improvement of performance and reduction of injuries (Akenhead & Nassis, 2016). In order to predictably achieve these goals, two requirements need to be fulfilled:

1) The monitoring needs to provide unique and relevant information
2) The information needs to be interpreted correctly to improve decision-making

Information is unique if it adds to the available knowledge. Information is relevant if it helps to systematically influence performance and injury risk. The implementation of load monitoring to

Figure 1

Theoretical framework of the training process (Impellizzeri et al., 2019)



Note. Figure from Impellizzeri et. al – Internal and External Training Load: 15 Years On. (2019)

provide such information relies on the assumption of predictable stimulus-response relationships between training stimuli and training outcomes. The premise is that if the training response is predictable by a measurable stimulus, that stimulus can be deliberately applied to yield the desired result. When the desired training response is performance related, this premise is reasonable, given that various load measures have been directly related to changes in performance indicators (Borresen & Lambert, 2009). The utility of load monitoring in predicting and preventing injury on the other hand is more controversial. Attempts to ascribe injury prediction abilities to measures of TL and ML have been criticized due to conceptual and methodological issues (Impellizzeri et al., 2020a, 2020b; McCall et al., 2017). When testing stimulus-response relationships, load measures are necessary for the quantification of stimuli, while responses can be measured with a variety of available performance tests and physiological assessments. But these methods of response evaluation have certain disadvantages, prompting an interest in the utility of load measures as an assessment tool. These disadvantages will be explained in more detail after elucidating the relationship between performance, fitness and fatigue. Further, the potential of integrated measures of IL and EL to address the issues of traditional training status assessments will be outlined. Afterwards, previous approaches to integrating the two will be summarized along with their advantages and disadvantages before finally, the value of adding context variables in order to increase the sensitivity and specificity of an integrated approach will be assessed.

Performance, fitness and fatigue

The fitness-fatigue model attempts to explain the change in performance that occurs in response to a training stimulus (Banister, 1991). According to it, a stimulus elicits two distinct responses – a fitness and a fatigue response (Banister, 1991). Today it is well established that both fitness and fatigue are complex multifactorial phenomena. A stimulus does not elicit one fitness and one fatigue response. Instead, it triggers multiple physiological processes, that each have distinct time curves and implications for performance (Burton et al., 2004; Hughes et al., 2018; Vanrenterghem et al., 2017). Finally, these processes cause certain anatomical and physiological adaptations with specific effects on performance (Burton et al., 2004; Hughes et al., 2018; Vanrenterghem et al., 2017). Other than being the consequence of physiological processes, changes in performance capability can also arise directly from the effects of mechanical stress on certain tissues, especially those of the musculoskeletal system (Vanrenterghem et al., 2017). Going forward, the terms “fitness-” and “fatigue response” will refer to all chronic performance-enhancing and acute performance-deteriorating effects, respectively, that are triggered by a specific stimulus. As a rule of thumb, changes in fitness can mostly be ascribed to the adaptations, that are the result of the physiological processes caused by both the biochemical and mechanical properties of the stimulus. Fatigue, on the other hand, can be described as a direct effect of those processes, rather than of their result, and an effect of the changes in tissue properties caused by mechanical stress. Both physiological processes and changes in tissue properties can also have acute performance enhancing effects, which are often referred to as postactivation potentiation (Wilson et al., 2013).

The nature and magnitude of a fitness or fatigue response depends on the stimulus. Both are performance-specific, meaning that the manifestation of a given response in performance can be vastly different depending on the type of performance. Different definitions and operationalizations of fitness and fatigue can be useful in different contexts. In this context, it is practical to create a framework in which fitness or fatigue measures are classified as global or partial. Within this framework, global measures do not refer to some sort of general, universally valid type of fitness or fatigue. Instead, global measures are defined in relation to a specific performance. Hence, they encapsulate all components of fitness and fatigue that are relevant to that performance, weighted in a way that reflects their impact on said performance. In football, a global fitness measure is a quantitative indicator of an athlete's preparedness to deal with the physical task requirements of a football game in the absence of fatigue. A global fatigue measure quantifies the acute performance-deteriorating effects on football match physical performance. Partial measures assess specific components of the overall training status. They can directly measure certain sub-performances that are relevant to the overall performance (e.g. maximum sprint speed, maximal aerobic speed or counter-movement-jump height), assess biological performance correlates (e.g. cortisol, resting HR, heart rate variability (HRV), capillary density, muscle cross-sectional area or tendon stiffness) or use subjective methods of training status quantification.

Global measures of fitness and fatigue reflect the multifactorial nature of a performance, thus including a defined number of factors, namely all those that affect a specific overall performance. Partial measures can theoretically assess any number of factors that is smaller than that involved in the overall performance, even possibly isolating single factors. Since, normally, not all factors relevant to a certain performance and their contributions to that performance are perfectly understood and measurable, assessing global fitness or fatigue as a function of their partial measures is prone to error. Hence, a more accurate approach is measuring the actual target performance (Knicker et al., 2011). The complex multifactorial nature of global fitness and fatigue measures makes them infeasible for establishing specific stimulus-response relationships, but their assessment helps determine the relevance of certain partial measures of fitness and fatigue and their relationship with overall performance. These partial measures should, due to the smaller number of factors they pertain, enable more precise evaluation of the exact effects of a stimulus. This will simplify the assessment of relevant stimulus-response relationships, enabling more targeted prescription of stimuli to achieve a desired effect. When aiming to prescribe a stimulus based on a stimulus-response relationship, fitness or fatigue related, the following factors need to be taken into account:

- The dose-response relationship
- The response time curve
- The relationship between the response and overall performance

Assessing training status

The evaluation of physical performance, or performance relevant qualities, is important for two reasons. First, it is necessary in order to establish the relationship between specific stimuli and their associated response. Second, it provides baseline data. Knowledge of the (prescribed) dose and the dose-response relationship alone can only predict change. Only its combination with a known baseline allows to predict outcome. More frequent assessments of performance and/or performance-relevant qualities allow for more precise understanding of the stimulus-response relationship and more frequent opportunities to evaluate and possibly adjust the training process by comparing planned to actual training outcomes. The better the ability to distinguish between different responses (both fitness and fatigue) to different stimuli, the more precise these stimuli can be prescribed in order to reach the desired outcome.

There are numerous ways to assess an athlete's training status (Akenhead & Nassis, 2016, Impellizzeri, 2005). Some are highly specific to certain fitness and fatigue components, others more general. Training status assessments can be performed at the biological, performance or psychological level (Coutts & Cormack, 2014) (Table 1). Typically, measurements at the biological level involve the assessment of certain physiological performance correlates (Djaoui et al., 2017). These can be invasive or non-invasive. Performance correlates can be invasively obtained, for example, through blood sampling or tissue biopsy. Invasive measurements are generally not fit for frequent use (Buchheit et al., 2013). Non-invasive methods assess measures like heart-rate derived indices, oxygen consumption or salivary hormones. In general, they are deemed more appropriate for daily implementation (Thorpe et al., 2017). Some invasive and non-invasive measurements need to be performed at rest (e.g. creatine kinase, muscle glycogen, resting HR, HRV), while others require standardized exercise protocols (e.g. VO_{2max}, lactate, HR recovery). Not all non-invasive methods are suitable for regular implementation however as they can be expensive, time-consuming or require specific expertise and/or equipment (e.g. ultrasound). Furthermore, the assessment of certain markers such as VO_{2max} or lactate thresholds involves highly exhaustive exercise protocols, which can place a substantial physical load on athletes (Pyne et al., 2014). Because these protocols are not designed to provide a specific training stimulus, they can cause a significant fatigue response in the absence of a targeted fitness response. This is especially undesirable during the season (Pyne et al., 2014), when a busy competitive schedule complicates the provision of a sufficient training stimulus while minimizing the performance-compromising fatigue response (Ritchie et al., 2016). Moreover, in order for these tests to validly assess fitness or certain physiological fitness indices, the confounding effects of fatigue need to be controlled for, meaning that accumulated fatigue should be minimal at the start of the test. This further complicates the necessary planning involved in the implementation of such tests and reduces the opportunities for providing fitness-enhancing training stimuli, not only after, but even before a certain test. Especially in-season, resources are scarce and should be used efficiently, mostly through allocation to competition or specific training stimuli, not the evaluation of their effects (Pyne et al., 2014). Accordingly, frequent implementation of these tests is infeasible.

The same applies to some of the tests that aim to assess training status at the performance level. Endurance focused tests like the Yo-Yo Intermittent Recovery Test (Bangsbo et al., 2008) or 15-30 Intermittent Fitness Test (Buchheit, 2008) often involve gradual increments in intensity until the athlete is no longer able to perform sustain the required output. Other endurance tests simply require maximal performance over a specific time (e.g. Cooper Test, Cooper, 1980) or distance (e.g. maximal aerobic speed, Bellenger et al., 2015). Different tests that aim to assess neuromuscular performance do not induce as much fatigue because of the lower loads they are associated with (Halson, 2014). They usually measure kinematic (velocity, time distance, acceleration) or kinetic (e.g. force, rate of force development) aspects of movements like sprints, jumps or certain types of maximal voluntary contractions (Claudino et al., 2017; Halson, 2014). Another way to measure training status at the performance level is the assessment of joint range of motion (Thorpe et al., 2017).

Lastly, training status can be assessed at psychological level through the use of athlete self-report measures (ASRM) (Thorpe et al., 2017). While numerous questionnaires exist that aim to assess well-being and fatigue, their utility in measuring fitness has not been investigated. When it comes to assessing fatigue, some ASRMs show even greater sensitivity to variations in TL and ML than common objective measures (Saw et al., 2015). Their implementation, however, needs to follow certain standards, in order for them to yield reliable results (Saw et al., 2017). These standards can be difficult or to fulfil in practice (Saw et al., 2017). Common single-item ASRMs that attempt to solve some of the difficulties of practical implementation have often not been validated (Jeffries et al., 2020).

The regular objective assessment of some partial measures of fitness and fatigue is possible using neuromuscular performance tests and physiological performance correlates (Halson, 2014). Yet, there is still a need for additional assessments that allow regular implementation in order to gain a better understanding of higher-frequency fluctuations in an athlete's overall training status. Especially endurance related qualities are difficult to monitor on a regular basis, due to the high loads associated with according tests. The use of submaximal performance tests has been suggested in order to address those issues (Buchheit, 2014). They usually involve measurements of internal load while completing a fixed amount of external work at a fixed submaximal intensity, as assessed through EL measures. Running economy is a common example of such a test which measures oxygen consumption at certain running velocities (Barnes & Kilding, 2015). Submaximal performance tests can be described as assessments of efficiency, as they measure the required internal work to produce a fixed external output (cost/output relationship). While these measures have been found to be a valuable tool for acquiring relevant insights into an athlete's training status with relatively little effort, they do have limitations in detecting football relevant changes in fitness and fatigue (Buchheit et al., 2012), thus not fully satisfying the need for additional assessment methods suitable for frequent implementation. The different types of training status assessments are summarized in Table 1.

The discussed assessments of training status either have issues related to practical implementation and/or they only measure specific components of fitness and fatigue. The latter is not an issue per se, since specific partial fitness and fatigue indicators are important to acquire both relevant and unique information about the particular effects of specific stimuli (when assessed in relation to load measures) and their relevance for overall performance (when assessed in relation to global training status measures). In other words, partial measures create a link between a stimulus and its manifestation in the overall performance. Yet, they do not address the need for football-specific global fitness and fatigue indicators which can link stimulus-induced changes in partial performance indicators to overall performance – an insight that the more specific partial measures alone cannot provide.

Table 1

Summary of different training status assessments

	Examples	Suitability for frequent implementation
Biological	Non-invasive: HR_{rest}, HRV, cortisol, muscle architecture, anthropometric measures	Non-invasive: low to very high
	Invasive: lactate, creatine kinase, muscle glycogen	Invasive: Very low to low
Performance	Endurance: Yo-Yo IR, 30-15 IFT, Cooper Test	Endurance: Low
	Neuromuscular: Maximum voluntary contraction, CMJ, maximum sprint speed	Neuromuscular: Medium to high
Psychological	ASRM: wellness questionnaires, DOMS, POMS	Medium to very high
Efficiency	Submaximal tests, IL:EL or EL:IL ratios, Multivariate measures integrating IL and EL	High to very high

Possible benefits of integrated measures of EL and IL

Submaximal physical performance tests aim to assess efficiency by measuring IL while controlling for EL (Buchheit, 2014) – something that is infeasible in training and impossible (or at least strongly recommended against) to do in matches. Integrated measures, instead of controlling for it, simply measure and account for EL. Both assess IL in relation to EL, only that in one, EL is fixed, while in the other it is not. Integrating measures of EL and IL to assess training status through athlete efficiency might be able to add value by

1) providing more football-specific global measures of training status, especially when using ML
2) doing so without requiring any investment of time- or physical resources from athletes.

The theoretical rationale

The EL of a competition is a function of the competition's physical task requirements and the athlete's ability and willingness to meet those requirements. IL is a function of EL, individual characteristics (both fixed and flexible) and environment. Some sports have fixed physical task requirements. In those cases, assuming an athlete's willingness to perform, EL, in time- or repetition-fixed, or completion time, in load-fixed events, are direct indicators of athlete ability. In other words, physical performance is a direct reflection of an athlete's performance-specific training status. In football, on the other hand, the physical task requirements fluctuate between and within games depending on situational factors such as tactics, opponents and behavior of teammates (Buchheit & Simpson., 2017; Carling et al., 2008; Drust et al., 2007; Lago-Peñas, 2012). Hence, despite (semi-) fixed performance duration, EL is a poor indicator of training status (Buchheit & Simpson, 2017; Carling et al., 2008; Drust et al., 2007; Lago-Peñas, 2012). Even assuming perfectly accurate measurements, knowledge of physical match performance allows for only minimal inferences to be made about physical fitness and fatigue. Thus, instead of assessing physical performance, it is more valuable to assess physical performance capability, or training status, as a function of fitness and fatigue.

Figure 1 depicts an updated version of a model first introduced by Impellizzeri et al. in 2005. It illustrates the relationship between EL, IL and training outcome in the training process. It also provides a theoretical rationale for the integration of EL and IL measures to assess training status. According to the model, IL is a direct consequence of EL. Their relationship is further affected by a handful of other variables, including training status. Assuming the influence of other variables in the model such as individual characteristics, genetics or environment to be negligible or stable within individuals, integrating EL and IL allows for inferences to be made regarding within-subject variations in training status.

Current research

Multiple scientific papers have elaborated on the potential benefits of integrating measures of internal and external TL and ML (Carling et al., 2013; Burgess, 2017; Weaving et al., 2017). Akubat et al. (2014) validated two IL:EL ratios against three physiological performance measures (velocity at lactate threshold (vLT), velocity at onset of blood lactate accumulation (vOBLA) and VO_{2max}) in sample of ten amateur football players. TD, HSRD and iTRIMP (Akubat et al., 2012) where assessed during simulated match-play. Significant correlations with large effect sizes were found for iTRIMP:HSRD and vOBLA, as well as iTRIMP:TD and vLT, suggesting the validity of these ratios as measures of fitness. Those findings were followed up on to assess the effects of fatigue on the ratios of HSRD, TD, as well as PlayerLoad and mean metabolic power over iTRIMP (Akubat et al., 2018). In order to do so two match-play simulations were conducted 48 hours apart. The relationships of almost all ratios with vOBLA and vLT (both assessed while rested) were lower for the second simulation than for the first one. Hence, the ratios not only seem to be sensitive to between-subject differences in fitness, but also in fatigue response. Suarez-Arrones et al. (2014) similarly used the ratio of mean running speed and average HR, which they measured during actual professional football match-play, as an efficiency measure. They found this ratio to differ between playing positions with the positions associated with lower efficiency also being assosociated with less TD. In a 2016 follow-up, Arrones et al. compared efficiency between first and second half and, as expected, found it to be lower in the latter. Buchheit et al. (2016) succeeded to reflect a heat acclimatization response over an eight-day training camp by using a ratio of RPE:distance/minute assessed during both training and match-play.

Since EL measures like TD or HSRD alone are insufficient to reflect the overall demands of football match-play (Weaving et al., 2017), Grünbichler et al. (2019) suggested the assessment of workload efficiency as the ratio between equivalent distance (Osgnach et al., 2010) and $TRIMP_{MOD}$ (Stagno et al., 2007). They tracked workload efficiency in 16 competitive matches over a 13-week period and assessed its sensitivity to TL variations in the last 5 days leading up to a competition. A stepwise multiple regression was conducted in which five predictors (each a specific HR measure on a specific day) were identified as significant, that were able to explain 26,6% of the variance in workload efficiency.

An attempt to solve the issue of the multifactorial nature of the physical work performed in rugby was conducted by Delaney et al. (2018a), albeit in a training rather than match context, by comparing the fitness and fatigue sensitivity of multiple combinations of EL and IL parameters. Instead of simply computing ratios between EL and IL, they applied log- and back-transforming to each combination of variables in order to get a more robust estimate of their relationship. The resulting variables were named training efficiency indices. The authors assessed seven EL indicators, three of which, metabolic work (Osgnach et al., 2010), impulse and mechanical work, were compound measures attempting to more holistically quantify EL. The relationship of each of the seven indicators with each, HR-TRIMP (Banister, 1991) and sRPE-TL, was examined. The resulting training efficiency indices were validated by assessing their sensitivity to detect changes in fitness and fatigue in a longitudinal design, in which fitness was assessed by a shuttle-test. Sensitivity to fatigue was tested by correlating the different measures to a four-week exponentially weighted rolling average of external HR, which was assumed to be predictive of a fatigue response. They found comparing measures of EL in relation to HR-TRIMP to be superior to sRPE-TL. Moreover, compound, or global measures of external HR incorporating body mass, speed- and acceleration-based running were deemed to be most meaningful. Combinations between those measures and TRIMP showed a significant correlation with the used indicators of fitness and fatigue. Delaney et al. (2018b) followed up by examining two training efficiency indices (impulse and mechanical work compared to TRIMP) in a sample of female division 1 collegiate soccer athletes. They found that both indices were poorly associated with wellness measures and acute HR (3-day exponentially weighted moving average) and only slightly associated with chronic HR (21-day exponentially weighted moving average).

Lacome et al. (2018a) applied multiple regression analyses to predict the HR response to small-sided games based on measures of EL on an individual level, while accounting for temperature. Afterwards, they compared predicted to actual HR responses, the differences between which (HR_{Δ}) were tested for their ability to assess training status. HR_{Δ} was able to reflect a gradual increase in fitness from July to September associated with pre-season and the start of the season. Furthermore, HR_{Δ} was found to be correlated with HR during standardized submaximal tests. All relationships were assessed on the individual level (within-subject).

Integrated measures of EL and IL can provide information about an athlete's training status without requiring any additional effort from that athlete. Since they assess the relationship between an output (EL) and a cost (IL) they can be described as measures of work efficiency (Lacome et al., 2018a). Only a small number of studies have explored the opportunities of these 'invisible' training status assessments. The existing literature is very heterogenous suggesting various approaches to measuring efficiency. Which type of efficiency they measure (e.g. biomechanical or physiological) depends on the exact parameters they examine. Since none of the studies specified an intention to assess particular partial measures of fitness or fatigue, it can be assumed that that they attempt to provide global measures of efficiency that are highly sensitive to any changes in an athlete's performance-specific training status. Load indicators like mechanical work or equivalent distance aim to assess EL in its entirety. They seem superior to EL measures less representative of the overall demands of football match-play, such as TD, when it comes to assessing training efficiency in conjunction with IL parameters (Delaney et al., 2018a). Yet, they are still likely not able to encompass the whole complexity of football's physical demands. Thus, when attempting to assess global work efficiency, it might be more accurate to incorporate multiple measures of EL into one's model. Lacome et al. (2018a) were the only ones to do so by not isolating the relationship between two single indicators of EL and IL. Instead, they integrated multiple external measures to predict a HR response as a measure of efficiency in an approach that further allows better individualization, thus taking into account the effect of individual differences. Table 2 (Appendix) summarizes all the discussed approaches to integrating EL and IL.

The present study

This study aims to add to the knowledge about integrated measures of EL and IL as a tool for 'invisible' training status assessments Only few studies have explored the opportunities such an approach can provide. Lacome et al. (2018a) proposed a promising multivariate approach to assess athlete readiness. However, while HR measurements provide an excellent opportunity for daily feedback, assessing match-play work efficiency might allow additional insights and enable establishing a relationship between weekly TL and match-day training status, first attempts to which have been made by Grünbichler et al. (2019). Competitive matches are however subject to much more variable contextual factors than team training, including playing surfaces, level of opponents, crowds and travel. In order to adopt an approach like that of Lacome et al. (2018a) to match-play, it might be favorable to take some of these factors into account to increase sensitivity and specificity to actual changes in fitness and fatigue. This study aims to assess the suitability of multivariate models of EL and IL as a measure of match-play efficiency (i.e. training status). Further, confounding effects of certain contextual factors that could compromise sensitivity and specificity of such an approach to changes in training status will be assessed.

First, the benefits of integrating multiple indicators of EL to predict the HR response to football match-play are assessed. Based on the well-established knowledge that relationships between EL and IL vary depending on the exact measures used (McLaren et al., 2018) and the strong conceptual background indicative of different relationships (Impellizzeri et al., 2019, Vanrenterghem et al., 2017), it is believed that a combination of EL measures is able to better explain match-play HR response than each of the measures individually (*hypothesis 1*). Second, the effect of certain contextual variables on HR response will be tested. Various contextual variables have been shown to directly affect HR before (Achten & Jeukendrup, 2003) and thus it is expected that a combination of contextual factors and measures of EL explains more variance in HR response than EL alone (*hypothesis 2*). Lastly, it will be assessed whether taking into account contextual factors will improve the model's ability to detect changes in fatigue, which is believed to be the case (*hypothesis 3)*.

Methods

Research design

The study was conducted in a longitudinal observational design following a male u18 team over the course of one and a half seasons. HR and EL data were not measured particularly for the purpose of this study, but rather as part of the team's monitoring routine. Contextual data were collected retrospectively from various sources (see data collection).

Participants

Study participants were 22 field players that played for the male u18-team of an English professional football club in between July 2018 and December 2019. Players that officially belonged to the u16 and u15 teams of the club's academy but played matches for the u18 team during the measurement period were also included in the analysis. The participants' age range was 14 to 18 years.

Variables

HR-response was measured as average HR. TD, HSRD (>19m*s^{-1}), number of accelerations (>3m*s^{-2}) and number of decelerations (<–3m*s^{-2}) were selected as EL measures. Even though, when using single measures of EL, superior results had been reported using more global measures of EL (Delaney et al., 2018a), the combination of selected parameters was believed to be a practical heuristic for reflecting the overall demands of football match-play without the high efforts associated with computing more advanced metrics from raw data. This assumption was based on the common critique of TD or HSRD for neglecting the important acceleration related demands of football (Osgnach et al., 2010). Thus, combining two measures that are usually more associated with physiological load (TD and HSRD) with two measures more associated with biomechanical load (acceleration and deceleration) (Vanrenterghem et al., 2017) was hoped to provide a pragmatic solution. Furthermore, previous findings suggest that even simple integrations (i.e. EL:IL ratios) of TD or HSRD with HR measures can be a valuable monitoring tool (Akubat et al., 2014, 2018; Buchheit et al., 2016; Delaney et al., 2018a; Suarez-Arrones et al., 2014, 2018). Hence, a combination of the two along with additional consideration of two complementary indicators of EL should provide even better results. All measures of EL were adjusted for time by dividing them by minutes played and then multiplying them by 90 minutes.

The contextual variables assessed were match location (home vs away), temperature and rain (rain vs no rain). Their selection was based on availability but also on previous literature showing relationships of match location (Lago-Peñas, 2012; Trewin et al., 2017) and temperature (Achten & Jeukendrup, 2003; Buchheit et al., 2016) with EL and HR. In football match-play, Mohr et al. (2012) found EL to decrease at high temperatures, while HR did not change significantly compared to normal temperatures, indicating that temperature might contribute to variations in EL-IL relationships. Lacome et al. (2018b) speculated on the potential effects of pitch surface. Despite a lack of research, logical thinking suggests that the lower friction associated with a wet playing surface could have a significant effect on movement efficiency.

HR_{Δ} was computed as the difference between actual and predicted HR (i.e. residuals) and its ability to detect fatigue was tested by looking at congested scheduling. If a player had participated in at least 70 minutes of match-play within the last 72 hours before a match, he was expected to display signs of fatigue in that match. This is in line with prior research investigating match-play induced fatigue and the effects of congested scheduling (Julian et al., 2021; Nédélec et al., 2012).

Inclusion criteria

Only data collected during league matches were included in the analysis. Match recordings were included only if the player had played for 70 minutes or longer. Since locomotor activity (Carling et al., 2013) and efficiency (Suarez-Arrones et al., 2014) decline over the course of the game, 70 minutes were chosen as a compromise to reduce confounding effects of playing time without ending up with too small of a sample size. If HR and/or EL data were not recorded for the whole duration of a player's appearance, they were removed entirely from the analysis. Only players of whom data was available for at least two matches were included in the analysis to emphasize the assessment of within-subject effects, since differences between players are difficult to interpret due to a multitude of confounding factors (Carling et al., 2018).

Data collection

HR and EL were recorded using Polar Team Pro Sensors (Polar Electro), which combine 10Hz GPS, 200Hz accelerometer, gyroscope and magnetometer recordings with HR monitoring (Polar Team Pro | Wearable GPS athlete tracking and performance, n.d.). These and similar devices have previously been validated and found to have acceptable accuracy and reliability, although both seem to decrease at higher intensity activities (Akyildiz et al., 2020; Coutts & Duffield, 2010). Weather conditions were assessed retrospectively using timeanddate (https://www.timeanddate.com/).

Data preparation

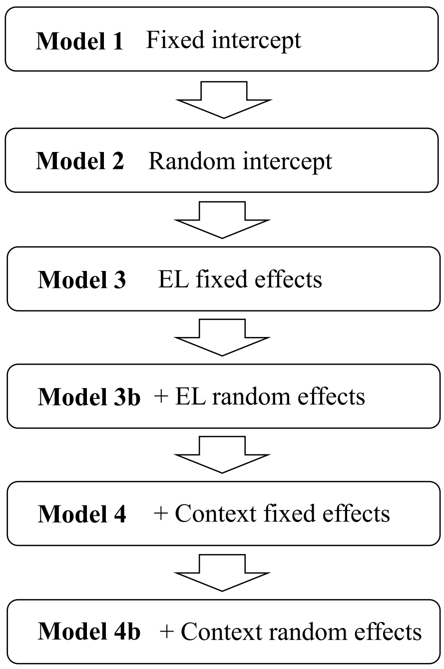
Recordings were uploaded into the Polar Web App (Polar Electro), where they were edited to only include actual playing time (rather than warm-up, halftime break and/or cool-down) and visually inspected for measurement errors. Data were then exported to .csv format. Microsoft Excel for Mac (version 16.13.1) was used to integrate load and contextual data and further check whether inclusion criteria were met.

Data analysis

Hypothesis 1 and 2. A multilevel regression analysis with full maximum likelihood estimation was conducted in RStudio (RStudio Team, 2021) using the *nlme* package. The study's repeated measures design meant that the individual recordings at level 1 were nested within the subjects at level 2. Since the main purpose was the evaluation of within-subject effects, no level 2 predictors were included to explain between-subject variance. In order to test hypothesis 1 and 2, a hierarchical model building approach was applied (*Figure 2)*. Fixed effects were added in blocks (EL and context). Chi-square likelihood ratio tests were used to compare model fit. If random effects did not significantly improve model fit, they were removed before adding the next random effects predictor or proceeding to Model 4. Variables were group mean centered to improve interpretability of level 1 effects. Finally, variances explained by each individual predictor and each block of predictors were computed according to Rights & Sterba (2019) using their *r2mlm* package. The assumptions of heteroscedasticity and normal distribution of residuals were checked by plotting standardized residuals against fitted values and creating a q-q normal plot of the residuals, respectively. Based on visual inspection, it was concluded that there was no heteroscedasticity and residuals were normally distributed.

Figure 2

Hierarchical model building approach



Note. Blockwise entry of EL fixed effects and context fixed effects. Random effects were entered individually.

Hypothesis 3. In order to test hypothesis 3, two more multilevel analyses were conducted comparing residuals (i.e. HR_{Δ}) for observations that a) were affected by congested scheduling and b) were not affected by congested scheduling. This was done twice – Once for HR_{Δ} computed based on the final version of model 3 and once based on the final version of model 4. Only players that were affected by congested scheduling at least twice were included. The dependent variable was HR_{Δ} and the only explanatory variable was congested scheduling at level 1. Models were immediately computed with a random effect of congested scheduling, meaning no hierarchical model building was performed, in order to account for high inter-individual differences in fatigue response (Nédélec et al., 2012). Explained variances of both models were compared in order to test their ability to detect the effects of congested scheduling. This approach was chosen over null-hypothesis significance testing because of the expected small sample size and small number of observations affected by a congested schedule.

Results

Participant flow

397 data points recorded in 40 players over the course of 45 matches were exported to Microsoft Excel. After removing observation that did not meet inclusion criteria, 235 data points (22 players, 41 matches) remained. The number of observations per player ranged from 2 to 28. Regarding hypothesis 3, only seven participants were affected at least twice by congested scheduling, resulting in 116 observations, 17 of which were affected by a congested schedule.

Hypothesis 1 and 2

Model 1, 2, 3 and 4 all exhibited significantly better fit than the previous model (Table 3). The only random effect to significantly improve model fit was that of temperature, which is thus included in the final model (Model 4b). As can be seen in Table 4 and *Figure 3*, that summarize the parameter estimates, not all coefficients were significant. Table 5 breaks down the within-subject variance explained by each predictor for both models. Partitioning of total variance and correlations can be found in Tables 6 and 7 (Appendix), respectively.

Table 3

Hierarchical model building model fit (Hypothesis 1 and 2)

	df	AIC	BIC	logLik	Test	L.Ratio	p-value
Model 1	2	1593.387	1600.306	-794.6937			
Model 2	3	1473.193	1483.572	-733.5967	1 vs 2	122.19404	<.0001***
Model 3	7	1413.791	1438.008	-699.8955	2 vs 3	67.40228	<.0001***
Model 4	10	1389.013	1423.609	-684.5067	3 vs 4	30.77767	<.0001***
Model 4b	12	1386.255	1427.77	-681.1273	4 vs 5	6.75878	0.0341*

Note. Model 4b adds a random effect for temperature, which was the only random effect found to significantly improve model fit.

Table 4

Parameter estimates for full model (Model 4b)

Predictors	*Estimates*	*CI*	*p*	*df*
(Intercept)	166.6	164.3 – 168.9	<0.001***	206
TD	3.37	2.30 – 4.44	<0.001***	206
HSRD	0.00	-0.00 – 0.01	0.453	206
Accelerations	0.09	-0.02 – 0.20	0.126	206
Decelerations	0.13	0.05 – 0.22	0.004**	206
Away	-1.28	-2.30 – -0.25	0.016*	206
Temperature	0.22	0.08 – 0.37	0.003**	206
Rain	-1.15	-2.78 – 0.49	0.175	206
SD intercept	5.19			
SD slope	0.21			
SD residual	3.78			
$N_{\text{Participants}}$	22			
Observations	235			

Note. TD is in km, HSRD in m.

Figure 3

Standardized coefficients of Model 4b

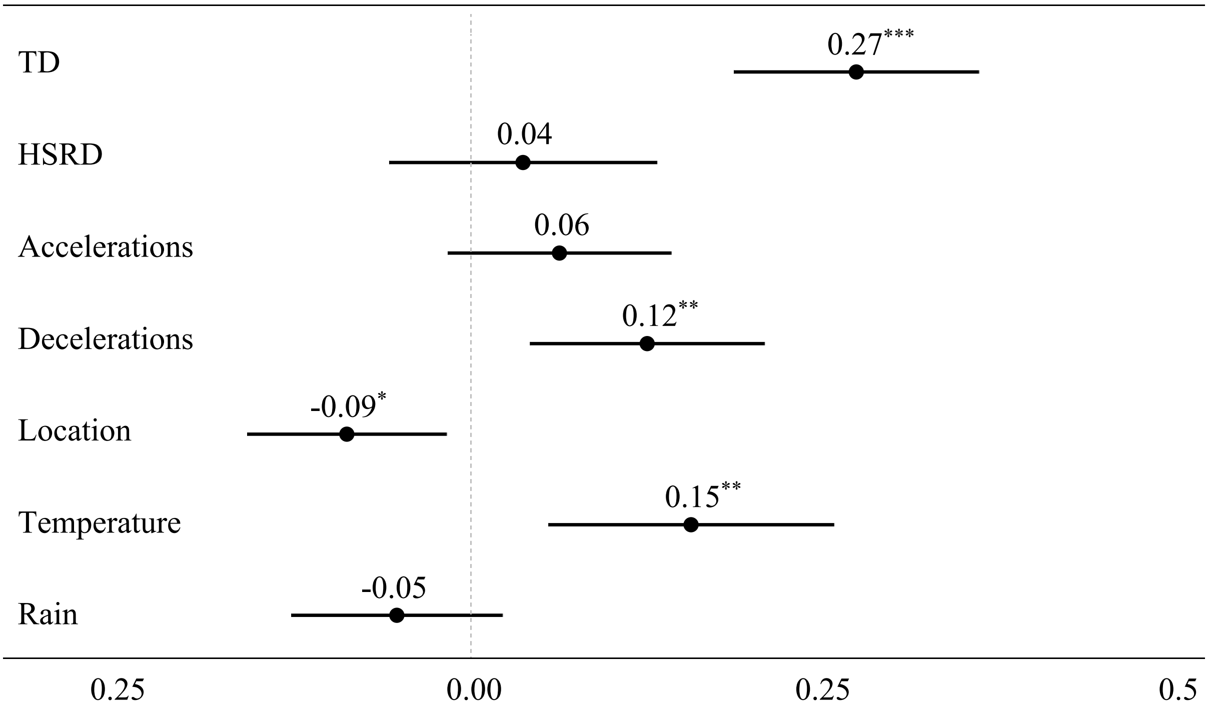


Table 5

Percentage of within-subject variance explained by each predictor and block

	In full model	Compared to null model	Difference
External load	32.05	25.30	-6.75
Context	14.42	7.67	-6.75
TD	11.68	20.95	9.27
HSRD	0.17	12.06	11.89
Accelerations	0.42	5.91	5.48
Decelerations	3.02	9.80	6.78
Away	2.81	0.97	-1.84
Temperature	10.95	6.29	-4.67
Temperature slope	4.97	1.30	-3.67
Rain	0.47	0.31	-0.16
Explained variance	39.72		

Note. Temperature random slope was compared to only temperature fixed slope instead of null model. Temperature entails both fixed and random slope.

Table 6

Total variance explained by EL only (Model 3) and EL + context (Model 4b)

	Model 3	Model 4b	Difference
Between-subject	0.53	0.53	0.00
Within-subject	0.12	0.19	0.07
Combined	0.65	0.72	0.07

Hypothesis 3

The analysis of congested schedule affected matches and others did not reveal any differences between the matches and between the two models. Table 8 contains means and standard deviations of HR_{Δ} for congested schedule affected and others as assessed by both models. Within-subject variance explained by both models can be seen in Table 9.

Table 8

Mean and SD of residual in congested schedule affected (CS) and not affected (Norm)

	EL		EL+Context	
	Norm	CS	Norm	CS
Mean	0.07	-1.08	-0.01	-0.65
SD	3.73	3.82	3.41	2.70

Table 9

Within-subject variance explained by effects of a congested schedule

	EL	EL+Context
Random intercept	0.01	0.00
Random slope	0.01	0.00
Combined	0.02	0.00

Discussion

High performance sports practitioners always seek to tweak their approach in order to maximize their athlete's performance and injury resilience. Regular assessments of athlete status are necessary to evaluate and adjust the training process. Shortcomings in traditional methods have led practitioners and scientists to search for new tools to help navigate the complexities of physical performance and with ever increasing technological opportunities, there is an abundance of data at our hands. A promising avenue is the integration of some of these data, specifically indicators of external and internal training and match load, to assess physical efficiency and allow what has been referred to as 'invisible' training status assessments. This study attempts to shed some light on the use of these methods in football match-play and the role that contextual variables play in this approach. In particular, three questions have been addresses:

1) Can a combination of EL measures better predict HR response than individual measures?
2) Do contextual factors explain variations in HR response that certain measures of EL cannot?
3) Does taking contextual factors into account improve the model's ability to detect fatigue induced through the effects of congested scheduling?

Are combined measures superior to individual ones?

The first hypothesis is confirmed by our data. Even though TD makes by far the greatest contribution, the inclusion of other variables helps further explain variations in HR response, making an argument for the integration of multiple measures of EL in such models to increase accuracy. Lacome et al. (2018a) found acceleration-based parameters like mechanical work and force load to best explain variations in HR response, followed by TD and finally HSRD. In our study, acceleration and deceleration counts did not show the same explanatory power. The arbitrary thresholds used for these, as well as HSRD, neglecting interindividual differences and relevant loads at lower intensities (Abt & Lovell, 2009), might explain their poor association with HR. Interestingly, Jaspers et al. (2018) identified acceleration counts at above $3.5ms^{-2}$ to be the most important predictor of RPE using a machine learning approach, highlighting the dependence of findings on the chosen parameters.

While Lacome et al. (2018a) found large interindividual differences in the relationship between external and internal load, this was not the case in our sample. Taking into account such potential differences for each EL measure did not significantly improve the model. There are three reasons that could possibly explain these differences. First, our sample might be more homogenous than theirs, due to a smaller age range, similar cultural background and similar training history. Second, they could be related to the exact EL measures used and third, we might have simply not had sufficient statistical power due to relatively few observations per participant.

Overall, the integration of EL measures used in this study proved superior to using each measure individually. However, in line with findings by Delaney et al. (2018a), compound measures like metabolic power or mechanical work seem to be preferable over or at least equivalent to a combination of simple EL indicators like those used here.

Do contextual variables add value?

The inclusion of the contextual variables match-location, temperature and rain explain more variation in HR than EL alone. Furthermore, the two combined could explain even more variance than the sum of their individual contributions. Especially the contextual variables showed strong increases in explanatory power when EL was taken into account. This has important practical implications, since it suggests that the given contextual factors affect EL without affecting HR as much. This means that changes in the IL EL relationships can occur without necessarily being related to changes in training status. Not taking these contextual factors into account could thus cause misinterpretations either leading to false positives or false negatives. Especially playing location and the random effect of temperature display this effect, further suggesting relatively large interindividual variability in the way that temperature affects EL. Lacome et al. (2018a) accounted for the effects of temperature using a fixed coefficient, but given our findings and the relatively low additional effort that would be associated with such an approach, taking individually different reactions to temperature changes into account may be a low-hanging fruit to improve model accuracy. The effect of rain, though insignificant, was contrary to what was expected based on changes in playing surface. Instead of a softer playing surface requiring higher internal work for the same external output, rain seemed to be related to slightly decreased heart rate due to its associated decrease in temperature. However, rain also exhibited a positive relationship with acceleration count which seems unrelated to temperature or changes in heart rate. These findings have important practical implications as taking contextual information into account seems to have an important impact on the interpretation of differences in the EL IL relationship.

Is congested schedule induced fatigue detectable?

Despite the inclusion of contextual variables yielding a more accurate model, this did not have any impact on detecting the effects of a congested schedule. There was a slight trend for the EL only model derived HR_{Δ} to be more affected by congested scheduling, but that effect can be explained by the correlation between congested scheduling and temperature and rain. In turn, some of the variance that was actually related to the contextual variables was explained by the congested schedule. Not accounting for weather did in this case pose a risk for type 1 error, even though the effect was very small. The correlation is likely to only be a statistical artifact due to the small number of congested scheduling affected observation, since there is no conceptual relationship between the two, other than congested scheduling perhaps being more common during certain phases of the season.

This insensitivity to acute fatigue is in line with findings from Delaney et al. (2018a, 2018b) who reported their approach to be better suited for recognizing long-term trends. It is also likely that an approach using HR as an indicator of EL is not suitable to detect fatigue due its inconsistent fatigue response (Thorpe et al., 2017) and its poor ability to reflect neuromuscular loads (Borresen

& Lambert, 2009). When attempting to assess global fatigue, RPE might thus be a more appropriate choice.

Important considerations

The complex relationships between different indicators of EL and IL (McLaren et al., 2018) must be taken into account when integrating them in order to assess efficiency. The exact combination of measures has important implications as to which kind of efficiency is assessed. Using biomechanical load measures is probably more specific to neuromuscular efficiency, whereas physiological load likely assesses a combination of neuromuscular and cardiovascular efficiency (Vanrenterghem et al., 2017). The decision of which load measures to combine should be based on the question what exactly one attempts to assess. Conceptual rather than just statistical (i.e. correlational) considerations need to be made before creating a model. The method of validation should then be chosen to reflect that conceptual relationship. A good plan of action would be to consider these four questions:

1) What do I want to measure? Partial vs global? Neuromuscular vs physiological?
2) The relationship between which load measures could reflect that?
3) Which contextual factors might affect their relationship?
4) How can I specifically validate what I attempt to measure?

Limitations

The external validity of this study is limited due to its small number of observations compared to the number of variables tested, especially given the large interindividual differences affecting the relationship between EL and IL. Further, there were only very few observations of certain conditions (rain and congested schedule affected matches) reducing statistical power. The assessed effects of context variables are specific to the load variables included in this model and do not have to be equally relevant when assessing other measures of EL and IL.

Conclusion

Integrated measures of EL and IL are a promising tool to provide training status information without the necessity for explicit testing procedures. However, the research is still limited. Various approaches exist, but none have been thoroughly researched. We found that taking into account the effect of contextual variables can reduce noise, thus potentially improving sensitivity and specificity to actual changes in training status. While these findings are specific to our sample and the load variables included in this model, they do show the relevance that contextual factors can have, emphasizing the need for further research. Going forward, it is important to develop a more streamlined approach in order to better understand the relationship between certain EL and IL and how they relate to performance.

References

Achten, J., & Jeukendrup, A. E. (2003). Heart rate monitoring: Applications and limitations. *Sports Medicine (Auckland, N.Z.)*, *33*(7), 517–538. https://doi.org/10.2165/00007256-200333070-00004

Akenhead, R., & Nassis, G. P. (2016). Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *International Journal of Sports Physiology and Performance*, *11*(5), 587–593. https://doi.org/10.1123/ijspp.2015-0331

Akubat, I., Barrett, S., & Abt, G. (2014). Integrating the Internal and External Training Loads in Soccer. *International Journal of Sports Physiology and Performance*, *9*(3), 457–462. https://doi.org/10.1123/ijspp.2012-0347

Akubat, I., Barrett, S., Sagarra, M., & Abt, G. (2018). The Validity of External:Internal Training Load Ratios in Rested and Fatigued Soccer Players. *Sports*, *6*(2), 44. https://doi.org/10.3390/sports6020044

Akubat, I., Patel, E., Barrett, S., & Abt, G. (2012). Methods of monitoring the training and match load and their relationship to changes in fitness in professional youth soccer players. *Journal of Sports Sciences*, *30*(14), 1473–1480. https://doi.org/10.1080/02640414.2012.712711

Akyildiz, Z., Yildiz, M., & Clemente, F. M. (2020). The reliability and accuracy of Polar Team Pro GPS units. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 175433712097666. https://doi.org/10.1177/1754337120976660

Bangsbo, J., Iaia, F. M., & Krustrup, P. (2008). The Yo-Yo Intermittent Recovery Test: A Useful Tool for Evaluation of Physical Performance in Intermittent Sports. *Sports Medicine*, *38*(1), 37–51. https://doi.org/10.2165/00007256-200838010-00004

Banister, E. W. (n.d.). Modeling Elite Athletic Performance. In J. D. MacDougall, H. A. Wenger, & H. J. Green (Eds.), *Physiological Testing of the High Performance Athletes* (pp. 403–424). Human Kinetics.

Bellenger, C. R., Fuller, J. T., Nelson, M. J., Hartland, M., Buckley, J. D., & Debenedictis, T. A. (2015). Predicting maximal aerobic speed through set distance time-trials. *European Journal of Applied Physiology*, *115*(12), 2593–2598. https://doi.org/10.1007/s00421-015-3233-6

Borresen, J., & Ian Lambert, M. (2009). The Quantification of Training Load, the Training Response and the Effect on Performance: *Sports Medicine*, *39*(9), 779–795. https://doi.org/10.2165/11317780-000000000-00000

Buchheit, M., Racinais, S., Bilsborough, J. C., Bourdon, P. C., Voss, S. C., Hocking, J., Cordy, J., Mendez-Villanueva, A., & Coutts, A. J. (2013). Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. *Journal of Science and Medicine in Sport*, *16*(6), 550–555. https://doi.org/10.1016/j.jsams.2012.12.003

Buchheit, M., Simpson, M. B., Al Haddad, H., Bourdon, P. C., & Mendez-Villanueva, A. (2012). Monitoring changes in physical performance with heart rate measures in young soccer players. *European Journal of Applied Physiology*, *112*(2), 711–723. https://doi.org/10.1007/s00421-011-2014-0

Buchheit, Martin. (2008). The 30-15 Intermittent Fitness Test: Accuracy for Individualizing Interval Training of Young Intermittent Sport Players. *Journal of Strength and Conditioning Research*, *22*(2), 365–374. https://doi.org/10.1519/JSC.0b013e3181635b2e

Buchheit, Martin. (2014). Monitoring training status with HR measures: Do all roads lead to Rome? *Frontiers in Physiology*, *5*. https://doi.org/10.3389/fphys.2014.00073

Buchheit, Martin, Cholley, Y., & Lambert, P. (2016). Psychometric and Physiological Responses to a Preseason Competitive Camp in the Heat With a 6-Hour Time Difference in Elite Soccer Players. *International Journal of Sports Physiology and Performance*, *11*(2), 176–181. https://doi.org/10.1123/ijspp.2015-0135

Buchheit, Martin, & Simpson, B. M. (2017). Player-Tracking Technology: Half-Full or Half-Empty Glass? *International Journal of Sports Physiology and Performance*, *12*(s2), S2-35-S2-41. https://doi.org/10.1123/ijspp.2016-0499

Burgess, D. J. (2017). The Research Doesn't Always Apply: Practical Solutions to Evidence-Based Training-Load Monitoring in Elite Team Sports. *International Journal of Sports Physiology and Performance*, *12*(s2), S2-136-S2-141. https://doi.org/10.1123/ijspp.2016-0608

Burton, D. A., Stokes, K., & Hall, G. M. (2004). Physiological effects of exercise. *Continuing Education in Anaesthesia Critical Care & Pain*, *4*(6), 185–188. https://doi.org/10.1093/bjaceaccp/mkh050

Carling, C. (2013). Interpreting Physical Performance in Professional Soccer Match-Play: Should We be More Pragmatic in Our Approach? *Sports Medicine*, *43*(8), 655–663. https://doi.org/10.1007/s40279-013-0055-8

Carling, C., Bloomfield, J., Nelsen, L., & Reilly, T. (2008). The Role of Motion Analysis in Elite Soccer: Contemporary Performance Measurement Techniques and Work Rate Data. *Sports Medicine*, *38*(10), 839–862. https://doi.org/10.2165/00007256-200838100-00004

Carling, C., Lacome, M., McCall, A., Dupont, G., Le Gall, F., Simpson, B., & Buchheit, M. (2018). Monitoring of Post-match Fatigue in Professional Soccer: Welcome to the Real World. *Sports Medicine*, *48*(12), 2695–2702. https://doi.org/10.1007/s40279-018-0935-z

Claudino, J. G., Cronin, J., Mezêncio, B., McMaster, D. T., McGuigan, M., Tricoli, V., Amadio, A. C., & Serrão, J. C. (2017). The countermovement jump to monitor neuromuscular status: A meta-analysis. *Journal of Science and Medicine in Sport*, *20*(4), 397–402. https://doi.org/10.1016/j.jsams.2016.08.011

Cooper, K. H. (1980). *Aerobics*. Bantam Books.

Coutts, Aaron J., & Duffield, R. (2010). Validity and reliability of GPS devices for measuring movement demands of team sports. *Journal of Science and Medicine in Sport*, *13*(1), 133–135. https://doi.org/10.1016/j.jsams.2008.09.015

Coutts, A.J., & Cormack, S. J. (2014). Monitoring the training response. In D. Joyce & D. Lewindon (Eds.), *High-performance training for sports*. Human Kinetics.

Delaney, J. A., Duthie, G. M., Thornton, H. R., & Pyne, D. B. (2018a). Quantifying the relationship between internal and external work in team sports: Development of a novel training efficiency index. *Science and Medicine in Football*, *2*(2), 149–156. https://doi.org/10.1080/24733938.2018.1432885

Delaney, J., McKay, B., Thornton, H., Murray, A., & Duthie, G. (2018b). Training efficiency and athlete wellness in collegiate female soccer. *Sport Performance & Science Reports*.

Djaoui, L., Haddad, M., Chamari, K., & Dellal, A. (2017). Monitoring training load and fatigue in soccer players with physiological markers. *Physiology & Behavior*, *181*, 86–94. https://doi.org/10.1016/j.physbeh.2017.09.004

Drust, B., Atkinson, G., & Reilly, T. (2007). Future Perspectives in the Evaluation of the Physiological Demands of Soccer: *Sports Medicine*, *37*(9), 783–805. https://doi.org/10.2165/00007256-200737090-00003

Grünbichler, J., Federolf, P., & Gatterer, H. (2020). Workload efficiency as a new tool to describe external and internal competitive match load of a professional soccer team: A descriptive study on the relationship between pre-game training loads and relative match load. *European Journal of Sport Science*, *20*(8), 1034–1041. https://doi.org/10.1080/17461391.2019.1697374

Halson, S. L. (2014). Monitoring Training Load to Understand Fatigue in Athletes. *Sports Medicine*, *44*(S2), 139–147. https://doi.org/10.1007/s40279-014-0253-z

Hughes, D. C., Ellefsen, S., & Baar, K. (2018). Adaptations to Endurance and Strength Training. *Cold Spring Harbor Perspectives in Medicine*, *8*(6). https://doi.org/10.1101/cshperspect.a029769

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

Impellizzeri, F. M., McCall, A., Ward, P., Bornn, L., & Coutts, A. J. (2020). Training Load and Its Role in Injury Prevention, Part 2: Conceptual and Methodologic Pitfalls. *Journal of Athletic Training*, *55*(9), 893–901. https://doi.org/10.4085/1062-6050-501-19

Impellizzeri, F. M., Menaspà, P., Coutts, A. J., Kalkhoven, J., & Menaspà, M. J. (2020). Training Load and Its Role in Injury Prevention, Part I: Back to the Future. *Journal of Athletic Training*, *55*(9), 885–892. https://doi.org/10.4085/1062-6050-500-19

Impellizzeri, F. M., Rampinini, E., Coutts, A. J., Sassi, A., & Marcora, S. M. (2004). Use of RPE-Based Training Load in Soccer: *Medicine & Science in Sports & Exercise*, *36*(6), 1042–1047. https://doi.org/10.1249/01.MSS.0000128199.23901.2F

Impellizzeri, F. M., Rampinini, E., & Marcora, S. M. (2005). Physiological assessment of aerobic training in soccer. *Journal of Sports Sciences*, *23*(6), 583–592. https://doi.org/10.1080/02640410400021278

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

Julian, R., Page, R. M., & Harper, L. D. (2021). The Effect of Fixture Congestion on Performance During Professional Male Soccer Match-Play: A Systematic Critical Review with Meta-Analysis. *Sports Medicine*, *51*(2), 255–273. https://doi.org/10.1007/s40279-020-01359-9

Knicker, A. J., Renshaw, I., Oldham, A. R. H., & Cairns, S. P. (2011). Interactive Processes Link the Multiple Symptoms of Fatigue in Sport Competition: *Sports Medicine*, *41*(4), 307–328. https://doi.org/10.2165/11586070-000000000-00000

Lacome, M, Simpson, B., & Buchheit, M. (2018a). Monitoring training status with player-tracking technology. Part 1: Traditional practices and new concepts. *Aspetar Sports Med J*, *7*, 54–63.

Lacome, Mathieu, Simpson, B., Broad, N., & Buchheit, M. (2018b). Monitoring Players' Readiness Using Predicted Heart-Rate Responses to Soccer Drills. *International Journal of Sports Physiology and Performance*, *13*(10), 1273–1280. https://doi.org/10.1123/ijspp.2018-0026

Lago-Peñas, C. (2012). The role of situational variables in analysing physical performance in soccer. *Journal of Human Kinetics*, *35*, 89–95. https://doi.org/10.2478/v10078-012-0082-9

McCall, A., Fanchini, M., & Coutts, A. J. (2017). Prediction: The Modern-Day Sport-Science and Sports-Medicine "Quest for the Holy Grail." *International Journal of Sports Physiology and Performance*, *12*(5), 704–706. https://doi.org/10.1123/ijspp.2017-0137

McLaren, S. J., Macpherson, T. W., Coutts, A. J., Hurst, C., Spears, I. R., & Weston, M. (2018). The Relationships Between Internal and External Measures of Training Load and Intensity in Team Sports: A Meta-Analysis. *Sports Medicine*, *48*(3), 641–658. https://doi.org/10.1007/s40279-017-0830-z

Mohr, M., Nybo, L., Grantham, J., & Racinais, S. (2012). Physiological Responses and Physical Performance during Football in the Heat. *PLoS ONE*, *7*(6), e39202. https://doi.org/10.1371/journal.pone.0039202

Nédélec, M., McCall, A., Carling, C., Legall, F., Berthoin, S., & Dupont, G. (2012). Recovery in Soccer: Part I – Post-Match Fatigue and Time Course of Recovery. *Sports Medicine*, *42*(12), 997–1015. https://doi.org/10.2165/11635270-000000000-00000

Osgnach, C., Poser, S., Bernardini, R., Rinaldo, R., & di Prampero, P. E. (2010). Energy cost and metabolic power in elite soccer: A new match analysis approach. *Medicine and Science in Sports and Exercise*, *42*(1), 170–178. https://doi.org/10.1249/MSS.0b013e3181ae5cfd

Pyne, D. B., Spencer, M., & Mujika, I. (2014). Improving the Value of Fitness Testing for Football. *International Journal of Sports Physiology and Performance*, *9*(3), 511–514. https://doi.org/10.1123/ijspp.2013-0453

Rights, J. D., & Sterba, S. K. (2020). New Recommendations on the Use of R-Squared Differences in Multilevel Model Comparisons. *Multivariate Behavioral Research*, *55*(4), 568–599. https://doi.org/10.1080/00273171.2019.1660605

Ritchie, D., Hopkins, W. G., Buchheit, M., Cordy, J., & Bartlett, J. D. (2016). Quantification of Training and Competition Load Across a Season in an Elite Australian Football Club. *International Journal of Sports Physiology and Performance*, *11*(4), 474–479. https://doi.org/10.1123/ijspp.2015-0294

Stagno, K. M., Thatcher, R., & van Someren, K. A. (2007). A modified TRIMP to quantify the in-season training load of team sport players. *Journal of Sports Sciences*, *25*(6), 629–634. https://doi.org/10.1080/02640410600811817

Suarez-Arrones, L., Torreño, N., Requena, B., Sáez De Villarreal, E., Casamichana, D., Barbero-Alvarez, J. C., & Munguía-Izquierdo, D. (2015). Match-play activity profile in professional soccer players during official games and the relationship between external and internal load. *The Journal of Sports Medicine and Physical Fitness*, *55*(12), 1417–1422.

Suarez-Arrones, Luis, Núñez, J., Sáez de Villareal, E., Gálvez, J., Suarez-Sanchez, G., & Munguía-Izquierdo, D. (2016). Repeated-High-Intensity-Running Activity and Internal Training Load of Elite Rugby Sevens Players During International Matches: A Comparison Between Halves. *International Journal of Sports Physiology and Performance*, *11*(4), 495–499. https://doi.org/10.1123/ijspp.2014-0523

Thorpe, R. T., Atkinson, G., Drust, B., & Gregson, W. (2017). Monitoring Fatigue Status in Elite Team-Sport Athletes: Implications for Practice. *International Journal of Sports Physiology and Performance*, *12*(s2), S2-27-S2-34. https://doi.org/10.1123/ijspp.2016-0434

Trewin, J., Meylan, C., Varley, M. C., & Cronin, J. (2017). The influence of situational and environmental factors on match-running in soccer: A systematic review. *Science and Medicine in Football*, *1*(2), 183–194. https://doi.org/10.1080/24733938.2017.1329589

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

Weaving, D., Jones, B., Till, K., Abt, G., & Beggs, C. (2017). The Case for Adopting a Multivariate Approach to Optimize Training Load Quantification in Team Sports. *Frontiers in Physiology*, *8*, 1024. https://doi.org/10.3389/fphys.2017.01024

Wilson, J. M., Duncan, N. M., Marin, P. J., Brown, L. E., Loenneke, J. P., Wilson, S. M. C., Jo, E., Lowery, R. P., & Ugrinowitsch, C. (2013). Meta-Analysis of Postactivation Potentiation and Power: Effects of Conditioning Activity, Volume, Gender, Rest Periods, and Training Status. *Journal of Strength and Conditioning Research*, *27*(3), 854–859. https://doi.org/10.1519/JSC.0b013e31825c2bdb

Appendix

Table 2

Summary of integrated measures of EL and IL

Measure	Findings	Training or match	Context	Studies
iTRIMP:TD and iTRIMP:HSRD	Correlated with between-subject differences in vOBLA, vLT and fatigue response	Simulated match-play	-	Akubat 2014; Akubat 2018
TD/min:HRavg	Positional differences Positive relationship with TD Smaller in 2nd half	Match	-	Suarez-Arrones 2014; Suarez-Arrones 2016
RPE:TD/min	Reduction indicative of heat acclimatization	Both	Temperature	Buchheit 2016
EquivalentDistance: $TRIMP_{mod}$	Correlated to training contents within last 4 days before competition	Match	-	Grünbichler et al. 2019
Training efficiency index	Correlated to changes in fitness (Shuttle) Correlated to chronic fatigue (3 and 4-week exponentially weighted rolling average of TL) Not correlated to acute fatigue (3-day exponentially weighted rolling average of TL)	Rugby/football training	-	Delaney et al. 2018a Delaney et al. 2018b
Difference between predicted and actual HR response	Correlated to HR response to standardized submaximal test (within-subject) Sensitive to improvements in fitness associated with pre-season	Small-sided games	Temperature	Lacome et al. (2018)

Table 7

Means, standard deviations, and correlations

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. HRavg	166.27	7.13								
2. TD	9407.01	685.46	.31** [.19, .43]							
3. HSRD	704.15	206.91	.20** [.07, .32]	.43** [.32, .53]						
4. Accelerations	17.89	6.77	.16* [.03, .28]	.05 [-.08, .18]	.41** [.30, .51]					
5. Decelerations	25.91	8.53	.13* [.00, .25]	.15* [.02, .27]	.49** [.38, .58]	.56** [.47, .64]				
6. Location	0.52	0.50	-.10 [-.22, .03]	-.03 [-.16, .09]	.13* [.00, .25]	-.00 [-.13, .12]	.03 [-.10, .16]			
7. Temperature	13.31	5.22	.24** [.12, .36]	-.15* [-.27, -.02]	-.10 [-.22, .03]	.03 [-.10, .16]	.03 [-.10, .16]	-.04 [-.17, .09]		
8. Rain	0.13	0.34	-.05 [-.18, .08]	.15* [.03, .28]	.07 [-.06, .20]	.21** [.08, .33]	.13 [-.00, .25]	-.01 [-.13, .12]	-.27** [-.39, -.15]	
9. Congested Schedule	0.09	0.29	-.06 [-.19, .07]	.09 [-.03, .22]	.05 [-.08, .18]	.05 [-.08, .18]	-.02 [-.15, .11]	-.03 [-.16, .10]	-.18** [-.30, -.05]	.23** [.11, .35]

Eigenständigkeitserklärung

Hiermit versichere ich, dass die vorliegende Arbeit über The integration of external and internal match load with contextual factors to assess training status in football selbstständig von mir und ohne fremde Hilfe verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind. Mir ist bekannt, dass es sich bei einem Plagiat um eine Täuschung handelt, die gemäß der Prüfungsordnung sanktioniert werden kann.

Ich erkläre mich mit einem Abgleich der Arbeit mit anderen Texten zwecks Auffindung von Übereinstimmungen sowie mit einer zu diesem Zweck vorzunehmenden Speicherung der Arbeit in einer Datenbank einverstanden.

Ich versichere, dass ich die vorliegende Arbeit oder Teile daraus nicht anderweitig als Prüfungsarbeit eingereicht habe.

19.03.2021

(Datum, Unterschrift)

Declaration of Academic Integrity

I hereby confirm that this thesis on The integration of external and internal match load with contextual factors to assess training status in football is solely my own work and that I have used no sources or aids other than the ones stated. All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited.

19.03.2021

(date and signature of student)

I agree to have my thesis checked in order to rule out potential similarities with other works and to have my thesis stored in a database for this purpose.

19.03.2021

(date and signature of student)