Validity of the Training-Load Concept

Louis Passfield, Juan M. Murias, Massimo Sacchetti, and Andrea Nicolò

Training load (TL) is a widely used concept in training prescription and monitoring and is also recognized as an important tool for avoiding athlete injury, illness, and overtraining. With the widespread adoption of wearable devices, TL metrics are used increasingly by researchers and practitioners worldwide. Conceptually, TL was proposed as a means to quantify a dose of training and used to predict its resulting training effect. However, TL has never been validated as a measure of training dose, and there is a risk that fundamental problems related to its calculation are preventing advances in training prescription and monitoring. Specifically, we highlight recent studies from our research groups where we compare the acute performance decrement measured following a session with its TL metrics. These studies suggest that most TL metrics are not consistent with their notional training dose and that the exercise duration confounds their calculation. These studies also show that total work done is not an appropriate way to compare training interventions that differ in duration and intensity. We encourage scientists and practitioners to critically evaluate the validity of current TL metrics and suggest that new TL metrics need to be developed.

Keywords: athletic training, endurance training, exercise performance, exercise physiology, exercise training

An athlete’s training can be quantified in terms of its training load (TL), an important concept originally devised by Banister et al that combines training-session intensity and duration in a manner proposed to represent a “dose” of training. Cumulated over multiple sessions, TL is commonly used to monitor and prescribe an athlete’s training.1 The TL concept has received increasing attention in recent years, because its evaluation is considered essential for monitoring whether athletes are adapting to their training programs,3 and because of its proposed relation to risk of injury,4 illness, and overtraining.5 The ability to assess TL has been facilitated by the widespread adoption of wearable and related devices that record training session data with high resolution.6 Consequently, it has been suggested that with the appropriate analysis of TL data obtained from wearables, it may be possible to optimize athletes’ training, thereby maximizing their fitness gains while decreasing their risk of injury, illness, and overtraining.7 However, we are concerned that the validity of the TL concept as a measure of an athlete’s training dose has yet to be established; until this issue is addressed, further progress will be limited. Underlining this concern are recent findings from our research groups that demonstrated circumstances where the calculated TL does not reflect the overall stress of the session and is therefore unlikely to represent its training dose.8–11 Accordingly, the purpose of this brief review is to highlight the issues associated with the validity of the TL concept, and to suggest some directions for future research and practical implementation.

Defining TL

The concept of TL can be ambiguous, but has been defined as the “cumulative amount of stress placed on an individual from a single or multiple training sessions . . . over a period of time.”14 This definition is consistent with the notion that TL metrics should reflect the theoretical construct of a training dose by quantifying the training for a given session in a manner that is proportional to the stress an athlete experiences. Consequently, a higher training dose results in a higher training stress. Specifically, Banister et al proposed that an athlete’s training could be quantified in arbitrary units as training impulses (TRIMPs), and that these TRIMPs could be used to predict the athlete’s resulting changes in performance. Banister et al refer to TRIMPs as a measure of training dose14 that can be related to the consequences of a training session. While Banister et al are widely acknowledged as having first proposed the concept of quantifying training, they did not use the term “training load” in this context. The term seems to have been used separately in the scientific literature of the time13 and only later linked to Banister’s proposed methods of quantification.15 This may explain why there is a lack of clarity in defining the TL concept; indeed, it has led some to argue that it is unscientific and should be abandoned.15 Part of this confusion appears to arise because some researchers consider TL to be related to the amount or volume of an athlete’s training, while others relate it to the dose or stress that arises from a given training session completed by an athlete.15 For the purposes of this review, Banister et al’s original perspective will be adopted where TL metrics are conceived to quantify the training dose (and its resulting stress) as distinct from the amount of training an athlete undertakes.

A further way of differentiating the dose from the amount of training is by categorizing TL as either internal or external to the athlete.16 Impellizzeri et al16 describe the external TL as the physical work prescribed in an athlete’s training program (ie, its amount), and the internal TL as an athlete’s psychophysiological response to this external TL (ie, reflecting the training dose). However, they observe that neither internal nor external TL concepts have a gold standard measure, and that the validity of any TL measure is also likely to be context dependent. As an example, they point out that responses that assume a physiological steady state (eg, heart rate and VO₂) may not be valid when used to quantify TL for short, intermittent, and high-intensity sessions. The absence of a gold standard criterion measure and the context-specific nature of

Passfield is with the School of Sport and Exercise Sciences, University of Kent, Canterbury, Kent, United Kingdom. Passfield and Murias are with the Faculty of Kinesiology, University of Calgary, Calgary, AB, Canada. Sacchetti and Nicolò are with the Dept of Movement, Human and Health Sciences, University of Rome “Foro Italico;” Rome, Italy. Passfield (louis.passfield@ucalgary.ca) is corresponding author.
some TL measures undoubtedly present conceptual and practical challenges for their validation. Indeed, it is possible that these challenges may have obscured the need to examine the validity of the TL concept. Accordingly, we think it imperative that the validity of the TL concept be examined as important conceptual issues may otherwise remain unaddressed. Before exploring the validity of the TL concept, it may be helpful to briefly review its initial development.

The Development of the TL Concept

It is a testament to the farsightedness of Banister et al that almost half a century after their seminal work was published, it still underpins many of the TL measures that are widely used today. In their first paper, they put forward a model that predicted the changes in the performance of a swimmer as a consequence of their 105-day training program and the fitness and fatigue it induced. To calculate the effects of fitness and fatigue on the swimmer’s performance, their swim and weight training were quantified in TRIMPs. The swimmer’s TRIMPs were calculated as arbitrary units by multiplying the distance swam by an intensity factor, and from the number of repetitions performed in the swimmer’s weight training. The resulting performance predictions were presented very simply alongside the actual performances as a proof of concept without any statistical evaluation. In succeeding papers, Banister et al repeated their modeling of training, presenting findings in a similar manner to the original paper. Some 15 years after the original paper, a statistical evaluation and validation of their modeling of training and its predicted effects on performance was published. In this study, 2 participants followed an intensive 28-day training program and completed performance trials twice weekly during the program and for a further 50 days after training. For both participants, the authors reported that their model, derived from TRIMPs, produced a statistically significant prediction of their actual performance trials. However, it should be noted that for this study the authors changed the basis for calculating TRIMPs to use their participants’ training heart-rate data. In addition, they introduced a nonlinear exercise intensity weighting to correct a bias they assumed was caused by sustained low-intensity training sessions.

Several modifications to the way in which TRIMPs are calculated have been published following Banister et al’s papers. Edwards published a coaching book that guided athletes on how to structure their training by using 5 heart rate zones. This 5-zone approach was subsequently adopted by some researchers as an alternative basis for calculating TRIMPs. Using this method, TRIMPs were calculated with a 1 to 5 weighting according to the time spent in each training zone. In contrast, Lucia et al modified the use of their TRIMPs so that they were calculated from the time spent in 3 physiologically specific exercise intensity domains rather than potentially arbitrary heart rate zones. Manzi et al further refined their use of TRIMPs by determining their exercise intensity weighting factors separately for each individual. In this study, they found that training quantified using their individualized TRIMPs correlated with the changes in performance unlike the TRIMPs calculated using Banister’s generalized nonlinear exercise intensity weighting. For the purposes of this review, discussion will refer to the above versions of TRIMPs. Other ways of quantifying TL have also been proposed. The most popular method of these is the session rating of perceived exertion (sRPE) devised by Foster et al. The sRPE is calculated from an athlete’s overall exertion rating for the session using the 0 to 10 category ratio scale. The sRPE was developed and refined over a series of studies and validated by comparison with Edward’s TRIMPs. An indication of the popularity of this measure is that in 2017, this study was reported to have been cited as many as 950 times. Since that time, its citations have increased 3-fold up to the time of this review.

A recent form of TL metric that is becoming increasingly widespread is the prediction of training stress scores and recovery time, typically provided by wearable devices and training data management websites. These TL metrics provide an estimation of internal TL often calculated from external TL measures such as speed and power output. However, most of these TL metrics are proprietary commercial products that have not been developed or validated scientifically; therefore, they will not be reviewed here.

TL and Its Lack of Validity

Validity is the degree to which something reflects what it purports to measure, which, in the case of TL metrics, is the training dose or its resulting stress. From the above historical perspective, it is apparent that the early work of Banister et al when developing TRIMPs did not include any validation exercise or even formal statistical analysis; hence, our concern regarding the validity of TL measures and methods arises for 2 related reasons. First, the construct validity of using TRIMPs as a means of quantifying a training dose was never established and remains to be evaluated fully and appropriately. Second, researchers have subsequently proposed other TL measures and sought to evaluate the criterion validity of these metrics without a recognized gold standard criterion measure of TL. This has resulted in proposed metrics either being validated against previously unvalidated measures or simply not being validated at all. Banister et al’s outstanding contribution was in conceiving and demonstrating that an athlete’s TL, quantified as TRIMPs, could be used to predict changes in performance. This and their subsequent observations appear to have been misinterpreted as a demonstration that TRIMPs constitute a valid measurement of TL. However, the hypothesis that TRIMPs provide a valid means of quantifying a training dose was never tested by Banister et al in their original or subsequent work. Specifically, where criterion validity was considered, Morton et al examined the use of TRIMPs to predict changes in performance, but they did not evaluate whether TRIMPs provided a reasonable estimate of their participants’ TL. Furthermore, as mentioned above, in this study the authors introduced heart rate as a new measure from which to calculate TRIMPs and added an arbitrary nonlinear correction to this data. Without verification, it should not be assumed that such fundamental modifications to TRIMPs enable an accurate representation of the training dose. Surprisingly, however, subsequent modifications to TL metrics were either not subjected to any validation exercise, as is the case with Edwards’ and Lucia’s TRIMPs, or have been validated against Banister’s or others’ unvalidated TL metrics. Foster et al validated the sRPE by comparing it with Edward’s TRIMPs across a range of different types of training sessions. The use of Edward’s TRIMPs as their criterion measure was unfortunate though, as it lacks a scientific basis and has itself never been validated. Despite this history, recent comprehensive and authoritative reviews of TL appear to accept its validity without critical comment, and fueling a concern that this fundamental issue has been overlooked.

One study that did attempt to assess the criterion validity of some measures of TL was performed by Wallace et al. Yet, it too...
suffers from the use of an inappropriate criterion measure for TL because the authors chose to use VO₂ as their proxy for a training dose. Their participants completed a diverse series of training sessions where various metrics of internal and external TL were compared against the total VO₂ for each session. The authors found the strongest association between total VO₂ and total mechanical work done (TWD). They concluded that TWD is therefore the most valid measure of TL, even though this contradicted the recognized premise that measures of internal TL, rather than external TL, are better suited to represent the training stress experienced by the athlete. The justification for the use of VO₂ as a criterion measure was not provided by the authors for this study. Regardless, its use seems indefensible as some training sessions involved short, high-intensity intervals where a VO₂ steady state would not be attainable and cannot therefore provide a valid means for quantifying the session.

Reconsidering the validity of TWD as a measure of TL has important and wide-ranging implications since TWD is frequently used as a key benchmark in both practical and research situations. In practice, training is commonly prescribed and monitored by quantifying its external load, such as TWD. In research settings, TWD is typically used to standardize training interventions that differ in duration and intensity. Moreover, near perfect correlations between TWD and several commonly used internal TL metrics, such as TRIMPs and sRPE, have been reported when analyzing the training and racing of elite cyclists. These findings imply that any limitations affecting the validity of TWD as a TL metric are equally likely to apply to its highly correlated internal TL metrics. Previously, some authors of this review and others, have discussed why the use of TWD is not appropriate as a measure of training dose, or for equating training sessions where exercise duration and intensity vary. Briefly, rather than TWD being held constant between sessions, it must be increased as exercise duration increases in order for an athlete’s effort to be comparable. Moreover, when constant and variable intensity sessions (eg, using intervals) are compared, the constant protocol is found to be much less stressful, or allows greater TWD for the same effort. Additionally, even when the TWD and interval work-to-rest ratio are held constant, longer interval durations increase the metabolic strain (ie, training stress). As Renfree et al point out, the manner in which a session is completed is potentially much more important than its TWD. In summary, neither construct nor criterion measure for TL are concerned that the lack of established validity for TL metrics is equally likely to apply to its highly correlated internal TL metrics. This is the case for the use of VO₂ as a criterion measure for TL. Regardless, its use seems indefensible as some training sessions involved short, high-intensity intervals where a VO₂ steady state would not be attainable and cannot therefore provide a valid means for quantifying the session.

The Acute Performance Decrement

In view of the concerns regarding TL outlined above, the evaluation of a training dose and its related TL metrics is necessary. Yet, as mentioned above, there is no current criterion measure against which putative TL metrics can be tested. Accordingly, we proposed that one way to conceptualize a session’s training dose is by its immediate effect on subsequent performance. The training dose and the consequent stress experienced by an athlete for a session cannot be quantified as a simple product of exercise intensity and duration. However, the session may be quantified by the acute performance decrement (APD) it causes, where the size of a training dose is reflected in the magnitude of the APD. For endurance sports, the APD can be measured immediately after the training session and seen, for example, as a decrease in time to task failure (TTF) or time-trial (TT) performance, although the specific outcome used may result in a different relative change in performance. Quantifying a session in terms of its resulting APD reflects the effect of the training dose and allows it to be compared with the amount of training completed and other TL metrics. This approach is also highly consistent with TL as originally proposed by Banister et al since they conceived TRIMPs to quantify the cumulative detrimental (fatigue) effects of training as well as the beneficial (fitness) effects on subsequent performance. As the etiology of fatigue arising from any training session may be uncertain, we prefer the term “APD” instead, but the intent is the same. Experimentally, measurement of APD is straightforward and has a high and established sensitivity for a variety of environmental, biological, and psychological stressors like heat environment, hypoxia, muscle glycogen depletion, muscle fatigue, and mental fatigue. In terms of evaluating prior training, Passfield and Dour found that a 5-minute TT is sensitive to the effects of a 60-minute session of moderate intensity when compared with a control condition. In a more recent study, APD measured as TTF was sensitive to a change of just 10 W in a 30-minute session. Consequently, we propose APD may be useful for scientific evaluation of new and established TL metrics such as TRIMPs and sRPE. It should be emphasized that we do not recommend that APD be used as a means to evaluate an athlete’s training program, as it is widely recognized that conducting repeated performance trials in training is not practical for this purpose. Instead, APD can be used as a criterion measure for experimental validation of TL metrics by studying the consequences of systematically manipulating TL both in laboratory and field settings. This manipulation can be achieved by varying training session duration and intensity and comparing whether APD and TL metrics change together consistently. Such experimentation forms the basis of our recently published research observations, which are discussed next.

Acute Performance Decrement and TL

In recent studies from our research groups, we have manipulated the duration or intensity of a training session and evaluated how APD measured as TTF or TT performance changes in response. While the sensitivity for both TTF and TT are similar, their percentage changes cannot be equated exactly. Our early laboratory-based study compared TTF after 30 minutes at maximum lactate steady state (MLSS) with TTF after 30 minutes at MLSS + 10 W. Compared to a baseline TTF bout (without prior exercise), 30 minutes at MLSS reduced TTF by 37% while a disproportionate reduction of 65% in TTF was observed after 30 minutes at MLSS + 10 W. This study, though not designed explicitly to examine TL metrics, nonetheless makes evident that TWD does not change subsequent TTF in a proportional manner. An increase in TWD from 0 kJ for the baseline to 415.8 kJ for the MLSS trial reduced TTF by 129 seconds, while a small increase in TWD of only 18 kJ for the MLSS + 10 W trial reduced TTF by a further 95 seconds. In a follow-on study, a systematic examination of the effects on TTF of exercising at 5 different intensities across the moderate to severe exercise domains was conducted. Most laboratory-based training sessions lasted 30 minutes, but participants also completed a 15- and 45-minute session at MLSS so that
the effects of intensity and duration could be evaluated differentially. Consistent with the previous study, the effects of session intensity on subsequent APD, measured as TTF, were nonlinear. When only the 30-minute trials at different exercise intensities are examined, there appears to be a reasonable relationship between TWD and APD. However, when trials where session duration was manipulated are also included in the plot, this relationship breaks down (Figure 1). In Figure 1A, we replot mean responses for all participants from this study to show the relationship between TWD and APD, measured as TTF, with separate trendlines for changes in session intensity and duration shown on the same graph. It is apparent that how the TWD is achieved (i.e., whether by changes in intensity or duration) is more important than its magnitude. Changing the duration of the MLSS ride created the smallest and largest session TWD but these did not result in the smallest and largest APDs. In Figure 1B, we plot estimated session values for Banister et al’s TRIMPs, which are scaled nonlinearly for exercise intensity. These are similarly confounded by the inclusion of sessions with different durations. It seems that the nonlinear intensity correction introduced by Banister et al was miscalculated and that they probably should have introduced a correction for duration. Figure 1C plots the session heart rates from which the TRIMPs were calculated and shows that these are not as markedly affected by different durations, reinforcing the notion that it may be the use of session duration as a multiplier in the calculation of TRIMPs that is the primary underlying issue. The remaining plots for Figure 1 show further measures that indicate a consistent effect of changes in both intensity and duration versus APD.

Figure 1 — Data replotted from Fullerton et al showing correlation for session TWD (A), TRIMPs (B), session HR (C), VE (D), RPE (E), and f_R (F) versus APD measured as time to task failure. Filled circles are for 30-minute sessions at different intensities; open circles are for sessions of 15, 30, and 45 minutes at maximum lactate steady state. See text for further details. APD indicates acute performance decrement; f_R, respiratory frequency; HR, heart rate; RPE, rating of perceived exertion; TRIMPs, training impulses; TWD, total work done; VE, minute ventilation.
The confounding effect of exercise duration is particularly evident in 2 studies of training sessions where we manipulated TL to compare explicitly the agreement of various TL metrics with APD.\textsuperscript{9,10} The first of these studies was conducted in the laboratory using well-trained cyclists. We examined APD measured as a 5-minute TT after maximal session efforts of 5 and 20 minutes, and after submaximal session efforts of 20 and 40 minutes. The TL metrics calculated for these sessions included TWD, 2 different forms of TRIMP, as well as SRPE and a training stress score. The observed APDs were similar after the 2 maximal effort sessions of 5 and 20 minutes, and smaller for the longer, lower intensity exercise bouts, while the TL metrics showed the opposite and contradictory response. This contradiction can be seen in Figures 2B and 2C that show the APD and TWD data replotted and the shorter, higher intensity 5-minute maximal session resulting in a larger APD but lower TL metrics than the longer, lower intensity submaximal sessions. One notable exception in these contradictory responses for APD versus TL metrics was seen for the NASA Task Load Index, which interestingly is not multiplied by exercise duration.\textsuperscript{45} In a subsequent field-based study, participants completed 4 outdoor continuous and intermittent interval running sessions, requiring 10-minute maximal and 25-minute submaximal efforts, before completing a 1500-m TT to measure APD. The APD was broadly similar for all sessions, whereas TL metrics showed large and significant variations between sessions. Again, the TL metrics were biased toward the longer and continuous efforts and contradicted the APD, which indicated the short maximal interval session was the most stressful. Much of the bias in TL metrics toward the longer sessions was removed if they were reported without using duration as a multiplier, as they then more closely resembled the APD response. Our conclusion from these studies is that most TL metrics do not show consistent agreement with the observed APD, and the contradictory responses seem to be caused by using the training session duration as a multiplier when calculating the TL metric. These findings are consistent with those of Weaving et al\textsuperscript{46} who reached a similar conclusion after analyzing the relative contribution of session intensity and duration to daily TL in professional rugby players.

**Insight From Performance Modeling**

Surprisingly, little attention has been paid to the way TL metrics model both session intensity and duration. Yet, we suggest that it may be the way in which TL metrics model the combined effects of exercise intensity and duration that leads to their contradictory

---

**Figure 2** — Data replotted from Kesisoglou et al\textsuperscript{9} for sessions involving 5- and 20-minute maximal and 20- and 40-minute submaximal efforts. (A) The gray line shows estimated maximal power output, and dotted lines indicate how the session duration needs to be extended for it to become maximal (see text for more details). (B) APD, (C) TWD, and (D) \%t\textsubscript{max} for the different sessions. The \%t\textsubscript{max} was calculated from average power-output values of the group. APD indicates acute performance decrement; TWD, total work done; \%t\textsubscript{max}, percentage of the maximum possible duration of the training session.
responses when compared to APD. Useful insight on this fundamental issue may be provided by the field of performance modeling. A very wide range of human performance can be described by the characteristic curvilinear intensity–duration relationship, or alternatively as an almost linear work–duration relationship. Hence, when the maximal effort for increasing session duration is plotted, a progressive increase in TWD is found, as discussed previously. This characteristic relationship enables the duration of any training session to be expressed as a percentage of the maximum possible for its intensity. The %\(t_{\text{max}}\) model locates a training session with respect to the athlete’s maximal intensity–duration curve and expresses the training session’s duration as a percentage of the maximum possible for the intensity sustained. Hence, the %\(t_{\text{max}}\) concept is suggested to provide an indication of the athlete’s relative effort in completing the session. This approach has also been described as the contribution to fatigue concept. Conceptually, this modeling approach combines exercise intensity and duration in a manner that approximates the training dose of a session as opposed to quantifying its amount as with TWD where session intensity and duration are simply multiplied. The implications of applying this concept can be seen in Figure 2 where the previously discussed effects of maximal 5- and 20-minute cycling sessions and submaximal 20- and 40-minute cycling sessions are replotted. Figure 2A shows the participants’ maximal intensity–duration curve with both the maximal 5- and 20-minute sessions sitting on this line. The 20- and 40-minute submaximal sessions are below the line; the 20-minute submaximal dotted arrow indicates the additional session duration required for this session to become maximal. The plot would have to be substantially rescaled to show the same for the 40-minute submaximal session. Figure 2D shows this more clearly as the %\(t_{\text{max}}\) for the submaximal sessions of 20 and 40 minutes can be seen to represent only \(\sim 38\%\) and \(\sim 1\%\) of the calculated maximum duration possible for their intensities. In contrast, the %\(t_{\text{max}}\) for both maximal sessions is the same despite the 4-fold difference in their duration. Importantly, the %\(t_{\text{max}}\) for both maximal sessions is the same despite the 4-fold difference in their duration. It may then be possible for the most effective training stimulus to be modeled and predicted, and for training to be calculated and prescribed on the same basis.

### Practical Applications
Our recent studies highlight that there may be a central issue with the validity of current TL metrics. There is now a need to determine the relative importance of exercise intensity and duration, and how these should be combined to form a TL metric that provides a more accurate estimation of the training dose. Until the validity of current TL metrics has been established, we suggest they are used and interpreted with caution. Indeed, the simplest approach is not to use session duration as a multiplier but to separately report the duration and intensity element(s) of the TL metrics. Given that there are many circumstances where the TWD and related TL metrics do not provide a meaningful interpretation of a session, it may be better to separately present time spent in each distinct exercise domain or training zone. The relationship between exercise intensity and athlete stress is inherently nonlinear and some TL metrics attempt to account for this. However, reasonable agreement with APD was still not found and may be confounded by the influence of the structure of the training session. Accordingly, further research may be required to determine how TL metrics can capture and present information on session structure by, for example, exploring methods such as exposure variation analysis.

Our recent experimental studies also provide some insight into measurements that could be useful in developing novel TL metrics. The NASA Task Load Index has been little used in the context of sport monitoring but showed good agreement with the APD observed following cycling sessions; it performed better than other TL metrics in the variable intensity running study. Although sRPE contradicts APD, we have found end-exercise RPE scores to show good agreement with APD when measured both as TT and TTF. In the latter study, we reported a strong correlation for minute ventilation (VE), recorded at the end of exercise, with subsequent TTF. Further inspection of our data from this study shows a close association for both RPE and respiratory frequency with APD (Figures 1E and 1F). These findings are consistent with the suggestion that effort is a key determinant of APD and supports the potential for respiratory monitoring as an objective marker of this. An additional complication for practitioners is that the present studies have necessarily examined relatively simple structured training sessions. The findings discussed here are mainly from laboratory-based endurance activities. However, the same issues were identified in a field-based study that involved intermittent short maximal efforts. Therefore, it is anticipated that the problems discussed here with the validity of TL metrics are likely to be found when TL metrics are used with a wider diversity of sports and training sessions. Nonetheless, further research is suggested to determine whether the exercise measures discussed can provide useful insights into APD with a wider diversity of training sessions. Ultimately, it is hoped that the present issues with the validity of TL metrics can be fully addressed. It may then be possible for the most effective training stimulus to be modeled and predicted, and for training to be calculated and prescribed on the same basis.

### Conclusion
As the TL concept is used increasingly, it is time to address the long-standing issue of its validity. Various TL metrics are used to quantify a session’s training dose, but current measures of TL have not been evaluated for this purpose against an appropriate criterion. Recent studies from our laboratories suggest that the present use of session duration may be confounding current TL metrics. In addition, new TL metrics and different approaches to their validation are required. Existing TL metrics should be used and interpreted with caution until their validity has been demonstrated. Further research exploring how to model the respective contributions of exercise intensity and duration is warranted. Although not recommended as a means of measuring TL in an athlete’s everyday training, evaluating APD may be useful for validating novel and existing TL measures.

### Acknowledgments
A.N. and L.P. contributed to the conception and design of the work, and drafted the work. All authors revised the work critically for important intellectual content and approved the final version of the manuscript.
References


