Monitoring Training Loads: The Past, the Present, and the Future

Carl Foster, Jose A. Rodriguez-Marroyo, and Jos J. de Koning

Training monitoring is about keeping track of what athletes accomplish in training, for the purpose of improving the interaction between coach and athlete. Over history there have been several basic schemes of training monitoring. In the earliest days training monitoring was about observing the athlete during standard workouts. However, difficulty in standardizing the conditions of training made this process unreliable. With the advent of interval training, monitoring became more systematic. However, imprecision in the measurement of heart rate (HR) evolved interval training toward index workouts, where the main monitored parameter was average time required to complete index workouts. These measures of training load focused on the external training load, what the athlete could actually do. With the advent of interest from the scientific community, the development of the concept of metabolic thresholds and the possibility of trackside measurement of HR, lactate, VO₂, and power output, there was greater interest in the internal training load, allowing better titration of training loads in athletes of differing ability. These methods show much promise but often require laboratory testing for calibration and tend to produce too much information, in too slow a time frame, to be optimally useful to coaches. The advent of the TRIMP concept by Banister suggested a strategy to combine intensity and duration elements of training into a single index concept, training load. Although the original TRIMP concept was mathematically complex, the development of the session RPE and similar low-tech methods has demonstrated a way to evaluate training load, along with derived variables, in a simple, responsive way. Recently, there has been interest in using wearable sensors to provide high-resolution data of the external training load. These methods are promising, but problems relative to information overload and turnaround time to coaches remain to be solved.

Keywords: external training load, internal training load, evaluation

Watching athletes perform well, set personal records or win competitions, are great pleasures for sports scientists. To think that the information that you have collected on the athlete, or synthesized from the literature, has helped the athlete achieve optimal performance is “as good as it gets” for support staff. Conversely, watching a poor performance inspires analysis of what went wrong, with preparation, tactics, or execution of the competitive plan. This provides the basis for the questions that drive sport-science research. Because so much of the preparation of athletes is related to the structure and details of the training program, there is a natural emphasis on how training influences performance. This interest goes into history, to Milo of Crotona, the Italian farm boy who lifted a growing bullock daily until he became the strongest man in the world and legend of the ancient Olympics. This story provides the historical grounding for the quest to understand the training response, most uniquely characterized by the concept of progression of the training load, and to the idea that training loads can be quantitatively expressed and related to performance outcomes. Although it is not known if Milo had a coach, most top athletes throughout history have had one, someone with more knowledge and experience, and the objectivity to evaluate their training and performance. The concept of training monitoring, regardless of historical time frame is in essence about the coach-athlete interface. Although not always appreciated, one can make the argument that the greatest value of sports science is related to optimizing the coach-athlete interface; to give the athlete a smarter, better-informed coach. Accepting the premise that the proper role of sport science is to inform and support the coach-athlete relationship, we need to ask what the coach needs from the sport-science community. A reasonable approximation is provided in Table 1. The reality is that sports scientists are rather good at providing the first 2 of these needs to the coach but less good at the last 2. As addressed previously, index workouts could be performed routinely by groups of athletes as a normal part of the training program, giving the coach high-frequency data useful for predicting progress toward training goals, and decision making regarding when the training program needs to be modified. The laboratory, is hard to schedule, is not well suited to testing large numbers of athletes quickly, and is not available for high-frequency testing. It is also much harder to provide the information which the coach needs to “translate” the results of the training to specifics about the progress and performance of the athlete (Figure 1).

The Past

Limiting our analysis primarily to middle and long distance running, with quantitative performance markers and a history of recording training details, the analysis of the history of training monitoring is straightforward. In the last century, the first evidence of training monitoring was in Scandinavia, by Finnish Olympians Hannes Kolehmainen and Paavo Nurmi. They frequently ran on the track carrying a stopwatch. Whether they were using the stopwatch to learn the feel of race pace, or to track progress is lost to history. Shortly afterward, in Sweden, Coach Gösta Holmér “invented” the concept of “fartlek” training. While often understood as semis-
unstructured training in the forest, fartlek was definitely high intensity training, with some segments matching race pace. There are frequent suggestions that some segments were timed so the idea of a standard, or index, workout is plausible. Of course, the climate in Scandinavia worked against highly structured quantitative monitoring of training, the presence of snow, cold, and inconsistent running surfaces acted to confound quantitative measurement. There are photos of the dominant Swedish runners of the day, Arne Andersson and Gunder Hägg, running on clearly substandard running tracks, or even in waist deep snow. Thus, the idea of quantitative training monitoring was clearly more concept than practice. At the same time, Gunder Hägg reputedly changed his training from very high volume to the much more intense fartlek training, because he felt that the higher volume training was wearing him down and causing frequent illnesses. To be fair, he also had to perform physical labor to earn his living, so the need for a high volume of background training may not have been as necessary as it is today, where athletes earn their living from sport and can spend time recovering when they are not training. Thus, regardless of the technique of monitoring training, athletes have been making decisions based on the perceived value of training programs for a long time.

The contemporary story of training monitoring began in Friburg, Germany in the late 1930s. Coach Wololdemar Gerschler and physician Herbert Reindell, developed interval training as a way to quantitate the training load, based on repetitive runs to a heart rate of 180 beats/min, with a recovery interval to a heart rate of 120. The concept was to run relatively short (100- to 400-m) repetitions, within the heart-rate parameters. If the recovery interval was inadequate to allow the heart rate to return below 120 beats/min, either the pace of the hard segment had to be reduced, the length of the hard segment reduced, or the training session terminated (Figure 2). As the athlete became fitter, it was necessary to run faster, or to run more repetitions, to provide a stress adequate to push the heart rate to 180. The precise reason why HRs of 180 and 120 beats/min evolved as benchmarks is not obvious. However, in young adults, with maximal HRs in the range of 200 beats/min, 180 represents an intensity approximating the second ventilatory or lactate threshold, the critical power or maximal lactate steady state.11,12

Athletes responded very well to interval training, and the method was quickly adopted throughout the world. By the early 1950s virtually every serious athlete in the world was using some version of interval training. One of the attractive elements of interval training is that progress could be easily tracked (Figure 3).13

Within a few years, interval training became understood as a certain number of repetitions, of a certain distance, with a certain recovery interval, with the average time for the completed hard segments as the “monitored” quantity. Since certain workouts could be run with some frequency, the concept of an “index workout” emerged as a way to monitor progress.

Once the concept of interval training was established, a number of well-known coaches and athletes emerged who were closely tied to interval training. The procedure was popular because of the quantitative relationship between the average pace sustainable during “index workouts” and subsequent competition, allowed predictability between performance during training and competition. For example, there is a well-established connection between average pace during 10 × 400 m on 3:00-minute send-off with performance over 1 mile. An athlete who could average 60 seconds per lap, might be expected to approach 4:00 for 1 mile in competition. This training-competition linkage was so important that average times for certain interval sets became the focus for interval training much more than the HR of 180 to 120 popularized by Gerschler and Reindell, perhaps because palpatung HR between interval sets was inaccurate.

The Austrian-English-Australian coach, Franz Stampfl, who was involved in the preparation of Sir Roger Bannister for the first sub-4-minute mile, was very fond of such interval workouts. There are numerous reports that he started his athletes with a relatively easy version of 10 × 400 m on 3:00 minutes and progressed them toward the target workout pace of 60 seconds per lap. Reputedly, when Roger Bannister and his training partners were unable to make progress in that workout, Franzt Stämpfl decided that they were becoming ‘stale’14 (today we would call it nonfunctionally overreached15) and sent them on a hiking holiday. After the holiday they were able to progress with the index workout, and Roger Bannister subsequently broke the 4:00-minute barrier. Similarly, at the

---

**Table 1** What Coaches Need From Sports Science

1. *A way to evaluate athlete potential.* Coaches need to know, beyond early promise and interest, whether a particular athlete is talented enough to perform at a high level or is destined to be an enthusiastic “also ran.” Classically, this is related to the maximal oxygen uptake (VO_{2max}). Although there are average athletes with high values for VO_{2max}, there are very few elite athletes with low values for VO_{2max} (<70 and <65 ml·min^{-1}·kg^{-1}, for males and females, respectively). In the broader context of the Joyner model of running performance, there are few elite athletes without an optimal combination of VO_{2max}, running economy, and %VO_{2max} at the lactate threshold.

2. *A way to evaluate an athlete’s current status.* In particular, it is useful to be able to evaluate performance in a way that is independent of momentary environmental conditions. Historically, athletes have often performed index workouts (eg, 10 × 400 m on a 3:00 send-off) as a surrogate for the average sustainable pace in competition. However, such index workouts are dependent on the stability of external conditions. Factors such as venue surfacing, temperature, humidity, wind, and precipitation can meaningfully affect the results of an index workout. In the laboratory, the sport scientist can often provide more consistent conditions.

3. *A way to evaluate how an athlete is responding to a training program.* In particular, since training programs for groups of athletes are sometimes generic, it is useful to evaluate how one athlete in a group is responding to training compared with another athlete in the same group. Coaches need an answer to the fundamental question of whether the athlete is adequately challenged by the training load, responding to training or being “ground down” by the training load.

4. *A way to measure progress, that is translatable into performance.* It must be remembered that normally measured laboratory outcome measures such as peak power output, maximal treadmill velocity, VO_{2max}, and lactate threshold are rather specific to the laboratory and imperfectly translatable to performance. As such, they lack the real-world linkage that has historically been provided by the index workout. The measurement of progress must also be interpretable in real time, so that the coach can make day-to-day decisions about modifying the training program, something that laboratory studies do not do well. Furthermore, it is a common experience that coaches think that, because of the rest required to provide standard laboratory conditions, athletes are “missing real training” to do laboratory evaluations.
Foster, Rodriguez-Marroyo, and de Koning

University of Oregon, Coach Bill Bowerman developed additional unique concepts, including the value of jogging (low intensity training), the value of hard days and easy days (eg, periodization), and the concept of date pace and goal pace. Like Frantz Stampfl, he might start his athletes with a relatively easy date pace (say 10 × 400 m on 3:00 min in 75 s), and expect progression toward a goal pace of 10 × 400 m on 3:00 minutes in 60 seconds. Like Frantz Stampfl, his program was fundamentally rooted in the concept of the progression of the training load. This period of history might be considered the zenith of the external-training-load period, where the concept of training monitoring meant that if an athlete was able to do a certain training session, they could expect certain competitive results.

With improvements in technology beginning in the early 1980s (radiotelemetric HR monitors, portable blood lactate analyzers, rapidly responding respiratory gas analyzers) the scientific community was able to provide a better marker of the physiological responses during training and competition, and the concept of the internal training load emerged. Beginning with the pioneering studies of Matti Karvonen and the use of the HR reserve, we began a process that has allowed better individualization of training. Likewise, the concept of training intensity zones, related to discrete metabolic thresholds created the ability to monitor training in a much more individual way. Although their work is tied to physiology in a very broad way, the work of several scientists (P.O. Astrand, G. Borg, D.L. Costill, J.T. Daniels, D.B. Dill, C.T.M. Davies, P. de Prampero, B. Ekbloom, E.L. Fox, F. Hagerman, L. Hermansen, W. Hollmann, I. Jacobs, J. Karlsson, W. Kindermann, A. Mader, R. Margaria, L.G.C.E. Pugh, B. Saltin, B. Sjodin, J. Sevedehad, R. Telford, K. Wasserman, B. Whipp) led to a more fundamental understanding of how acute responses and chronic adaptations led to the contemporary concept of training monitoring. Using the combination of threshold concept and simple technology, the idea of an “intensity profile” (Figure 4) emerged, which suggested that training could be organized around distinct zones, which required careful standardization of pretesting conditions. During this period, the concept emerged that the “optimal distribution” of training might be with most of the training load clustered in the “threshold zone,” a concept that is today largely discounted.

Despite the value of being able to fractionate the training intensity with more precision, there was still no overall strategy that linked the training “load” (frequency × intensity × time) to performance changes in any sort of quantitative model. Beginning in the mid-1970s, Eric Banister and his colleagues developed the concept of the “training impulse” or TRIMP. This concept recognized that measuring training intensity as a function of the percent heart rate reserve (%HRR) multiplied by a nonlinear factor (conceptually equivalent to the intensity vs. blood lactate relationship) multiplied by duration yielded a number (TRIMP) that represented both the...
gain in fitness and the gain in fatigue contributed by that training session. If this number is summated over a convenient period (1 wk) one could gain an appreciation of the product of volume and intensity of training. Recognizing that there were distinct “influence curves” related to fitness and fatigue, Banister’s group demonstrated that subsequent performance could be accounted for by the integration of the fitness and fatigue influence curves. In this moment, the concept of training monitoring was moved forward in a fundamental way, such that the details of training could be expressed quantitatively and temporally linked to performance.

The TRIMP concept is fundamentally brilliant. It is also mathematically complex, is grounded in essentially steady state training (eg, mean %HRR), and was developed on subelite athletes. During the next several years, the basic TRIMP concept was modified to allow the concept of HR zones to provide a surrogate TRIMP that solved the complexity of an assumed near-steady-state %HRR and a nonlinear multiplier, while taking advantage of the data-storage capacity of radiotelemetric HR monitors and the metabolic insight provided by the threshold concept. These approaches, however, are limited to situations in which the HR versus intensity zone relationship has been established (requiring laboratory testing). For athletes who are training using multiple modes of activity, this was difficult because “threshold HRs” are often rather mode specific. Taking advantage of the strongly robust nature of the rating of perceived exertion (RPE), which has well established correlations with both %HRR and blood/muscle lactate, we demonstrated that modifying the use of the RPE from a momentary rating to a postexercise rating representing the entire training bout (the Session RPE or sRPE) retained the relationship with %HRR and blood lactate, while allowing multimode, non-steady-state, and even competitive game-type exercise to be included. If the sRPE is multiplied by the duration of activity, then one has a surrogate TRIMP score for training load that is much simpler than the original concept of Banister et al., which is not dependent on accurately measuring HR and can be used with multiple modes of exercise, including resistance exercise. Because of its simplicity, sRPE (which was anticipated by Gastin contemporaneously with first publication of sRPE and used to plan and periodize the training of Australian Football players) has become widely used. There has been interest for many years in whether responses during submaximal exercise could serve as surrogate markers of how

---

**Figure 3** — Schematic of how interval-training repetitions and time per repetition have historically been used to monitor the progression of training. The athlete is assigned the task of progressing a workout (performed every fourth day) by running 200 m with a send-off every 1:30, and the target of running 31 to 33 s/repetition. Over the space of several weeks, the number of repetitions is steadily increased, and the speed of running is either increased or sustained. When an increase in repetition number (day 25) causes slower running, the progression of training volume is halted until the athlete can again run the desired speed. Thus, the volume and intensity of training can be used to monitor progress, with high-frequency evaluations, allowing modification of the training program to maintain progression of training.

---

**Figure 4** — Schematic of running speed versus heart rate (HR) and blood lactate. The athlete runs at progressively faster speeds, usually in stages of 3 to 5 minutes, to allow equilibration of blood and muscle lactate. HR and blood lactate are plotted as a function of speed. Reference markers for the lactate threshold (2.5 mmol/L in this case) and a second threshold thought to be a surrogate of the maximal lactate steady state (OBLA = 4 mmol/L in this case) are plotted, and associated with the HR (which is more practical to measure during routine training). In this case, the athlete would be given a “threshold” HR range of 159 to 179. Low-intensity high-volume training would be accomplished at a speed in the range of 16 km/h with HR just below 159. Sustained (e.g., 10–30 min) hard running would be accomplished at a speed between about 16–18 km/h and a HR within the threshold range. Intensive intervals (eg, 6 × 1600 m/400-m jogging recovery) would be accomplished at a speed just faster than 18 km/h and HR just above 179. In principle, as the athlete adapted to training and became fitter, the athlete could be reevaluated and the velocity and HR defining the reference training zones could be adjusted.
an athlete was responding to training. As early as the mid-1970s, techniques originally used by Astrand and Rhyming to predict VO2max, were adapted to allow computation of the PWC170, a marker that can be responsive to changes in fitness. This has raised the possibility that submaximal exercise performed before a training session (eg, the warm-up) could be used as a monitoring technique, allowing high frequency measurements of athlete status, in a way that would allow the coach to rapidly make day-to-day decisions about whether the planned training needs to be modified. To date, the most attractive model of this concept is the Lamberts and Lambert submaximal cycling test, which appears to be able to track changes in fitness in a way that allows prediction of subsequent cycling performance. Given the importance of providing the coach with a tool suitable for high frequency evaluation and simple enough to make day-to-day decisions, certainly more work in this area is warranted. Given that the HR response can be disrupted, and has technological requirements, it may be that using RPE during standard warm-up activities is an even more simple approach, as it has already been shown to be a useful strategy for predicting VO2max.

Just as RPE has been shown to be a useful method for tracking the intensity of exercise, the process of systematically asking the athlete “how do you feel” holds the potential for monitoring the response to training. Steven Seiler has related the “hair in the cereal” story of observing athlete behavior at the breakfast table. Athletes who are responding poorly to training, are often uncommunicative and have their head down, often to the point where their hair is actually in their cereal at the breakfast table. If the coach, or the support personnel, is observant at the breakfast table, such simple behavioral markers may be as useful for monitoring training as any of the more “physiologic” markers. This concept was demonstrated many years ago, using subscales of the Profile of Mood States (POMS). Since then, a number of validated questionnaires have been developed which address this question. They are simple, easy to administer, and offer the coach the possibility of high-frequency assessment and rapid decision making.

The development of reliable step counters for monitoring the volume of walking in fitness participants gave a new life to the concept of the external training load. Fairly quickly, norms for how many steps were associated with good health emerged. Subsequent development of high-resolution video analysis and digital accelerometers which allow analysis of exercise intensity, and the possibility of 3-dimensional movement analysis, suggest that complex sports activities may be analyzed in a way that provides useful information to the coach. Already of clear value in terms of tactical analysis, these technologies offer the promise of a very insightful analysis of the external training load. Furthermore, the development of highly accurate power meters suitable for cycling and other “cranking” sports has already started to change the landscape of sensor-based training monitoring.

More recently, the very simple concept of the Talk Test has emerged, which allows metabolic zone concepts to be used. This does not require a grounding laboratory evaluation, the technological capacity of HR monitors or blood lactate measurement.

The Future

The future of training monitoring might well be dominated by emerging technologies that allow new possibilities relative to the analysis of the external training load. However, since training monitoring is about individuals, and about letting coaches understand their athletes better as individuals; it seems more likely that elements of the internal training load may be of continuing importance. In particular, the integration of elements already considered as part of the internal training load into single constructs seem likely to be of importance. For example, we know that the total training load, the training intensity distribution pattern and the day-to-day variation in training load (eg, monotony) are independently related to performance. Is there a method of integrating these factors into a single variable that is capable of predicting the response (both good and bad) to training? We know that there is a somewhat predictable pattern of the short-term improvement in performance related to short term reductions in training load. This almost certainly interacts with the recent training load, in that tapering after heavy training is more likely effective than after light training. Can this be expressed quantitatively? If so, coaches would have a rational basis for individually managing the training load. It has been demonstrated by a number of investigators that the training program designed by coaches is not always well executed by athletes. Indeed, even when coaches are present to observe training, their perception does not always match well with the training program as experienced by athletes. Simple techniques allowing visual matching of the designed and executed training programs offer much relative to optimizing the training process. Finally, we need to keep in mind that the information base of how athletes respond to training is actually very thin. For as many well-controlled training studies as have been conducted in healthy nonathletic individuals, there are few controlled studies in competitive athletes, particularly high-level competitive athletes, particularly over the very long time base over which athlete training programs are constructed. Further, because of the relatively short careers of high-level competitive athletes, the likelihood that adequately controlled studies will be conducted is relatively low. Further, because the magnitude of improvement in performance which is meaningful to athletes is fairly small, we struggle with appropriate statistical models to evaluate such training data, even if it were available.

Summary

Training monitoring has advanced significantly during the last century. It has gone from essentially nonexistent to venue-sensitive index-training sessions to a relatively coherent structure for evaluating internal training load, although there is a continuing tension between measuring what was actually done in training (external training load) and the acute experience of training (the internal training load), and how these 2 factors interact to produce the response to training that allows athletes to improve their performance. We have only slowly come to the realization that sports scientists are, at best, support personnel, and of the primacy of the coach-athlete interface. We have a variety of tools (training load, training intensity distribution, training monotony and strain, tapering) which are poorly integrated into a “unified field theory.” We have evidence that we need better tools for matching the coach’s plan with the athlete’s experience of training. We know, but do not have adequate evidence, about the value of frequent, submaximal evaluations as a tool to inform the coach about day-to-day changes in athlete status. We still have serious deficiencies relative to our fundamental understanding of how changes in the training load influences performance in high-level athletes. Lastly, one of the realities of contemporary training monitoring is that the technological capabilities available for our use may provide us with so much information that “seeing the forest for the trees” becomes a very real problem for coaches. Perhaps we need to consider the idea of going “back to the future” in terms of index-training sessions (perhaps with the benefit of contemporary movement monitors), monitored warm-up procedures,
which may allow us to follow the widely used adage in almost any endeavor, “KISS” (keep it simple, stupid), which may be the most important element of training monitoring.

References


