Monitoring load, recovery and performance in young elite soccer players

Brink, Michel S.; Nederhof, Esther; Visscher, Chris; Schmikli, Sandor L.; Lemmink, Koen A. P. M.

Published in: Journal of Strength and Conditioning Research

DOI: 10.1519/JSC.0b013e3181c4d38b

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Early version, also known as pre-print

Publication date: 2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
MONITORING LOAD, RECOVERY, AND PERFORMANCE IN YOUNG ELITE SOCCER PLAYERS

MICHEL S. BRINK,1 ESTHER NEDERHOF,1,2 CHRIS VISSCHER,1 SANDOR L. SCHMIKLI,3 AND KOEN A.P.M. LEMMINK1,4

1Center for Human Movement Sciences, University Center for Sports, Exercise and Health, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 2Department of Human Physiology and Sports Medicine, Free University of Brussels, Brussels, Belgium; 3Rehabilitation and Sport Medicine, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands; and 4School of Sports Studies, Hanze University of Applied Sciences Groningen, Groningen, The Netherlands

ABSTRACT

Brink, MS, Nederhof, E, Visscher, C, Schmikli, SL, and Lemmink, KAPM. Monitoring load, recovery, and performance in young elite soccer players. J Strength Cond Res 24(3): 597–603, 2010—The purpose of this study was to investigate the relation between training load, recovery, and monthly field test performance in young elite soccer players to develop training guidelines to enhance performance. In a prospective, non-experimental cohort design, 18 young elite soccer players registered training and match duration for a full competitive season by means of daily training logs. Furthermore, session rating of perceived exertion (RPE) and total quality of recovery (TQR) scores were recorded. Weekly duration (TLd), load (duration × session RPE = TLrpe), and TQR scores were calculated for 1 and 2 weeks before a monthly submaximal interval shuttle run tests to determine interval endurance capacity. Participants spent on average 394.4 ± 134.9 minutes per week on training and game play with an average session RPE of 14.4 ± 1.2 (somewhat hard) and TQR of 14.7 ± 1.3 (good recovery). Random intercept models showed that every extra hour training or game play resulted in enhanced field test performance (p < 0.05). Session RPE and TQR scores did not contribute to the prediction of performance. The duration of training and game play in the week before field test performance is most strongly related to interval endurance capacity. Therefore, coaches should focus on training duration to improve interval endurance capacity in elite soccer players. To evaluate the group and individual training response, field tests should be frequently executed and be incorporated in the training program.

KEY WORDS training load, interval endurance capacity, hierarchical linear model, football

INTRODUCTION

To reach the top in professional soccer, extensive training is necessary to improve performance. In individual sports, such as distance running, a positive relation between training and performance has been found and extensively described (22). However, little is known about the relation between training and performance in ball team sports.

The duration and intensity of the training are the primary determinants of the training load. It is known that because of individual differences (i.e., training status, school exams, injury), the optimal training load varies between athletes. In ball team sports, the training load prescribed by the coach is often called the external load and is expressed in the duration in minutes and for example high, medium, and low intensity. The internal training load, on the other hand, is the actual physiological stress imposed on the athlete. The internal load accounts for individual differences, for example, starting fitness level and psychosocial aspects (14). It is assumed that the training load in general should be combined with sufficient recovery to enhance performance. This is also known as the supercompensation effect. Therefore, the focus in a program should not only be on training load but also on recovery (15).

To monitor performance, sport-specific field tests have been developed to evaluate the training response (29). These shuttle run tests imitate the physiological profile and are related to quality of play during a match, distance covered, time spent at high intensity, and number of sprints (3,12,24). Gabbett and Domrow incorporated a shuttle run test in their study and investigated the training-performance relation in subelite rugby players training. (10). However, no relationship was observed between training load and field test
performance. Although the players were followed over an entire season, only 4 measurements were taken up to 3 months apart. It is expected that performance changes occur faster, and more frequent testing is needed (4). In addition to that, performance seems also related to different periods of the season, whereas at the beginning of the season, the best aerobic improvement can be expected with the same amount of training (25). This stresses the need for research, which incorporates field test performance on a more regular basis.

In summary, evidence for the training-performance relation in elite soccer is limited, whether this is the external or the internal load. Although recovery is theoretically important to improve performance, evidence in ball team sports is lacking. Therefore, the aim of this study is to monitor training load, recovery, and performance of young soccer players for a full season to develop training guidelines to enhance performance. We hypothesized that a higher training load (external and internal) combined with good recovery would lead to a better performance.

**METHODS**

**Experimental Approach to the Problem**

To determine training load in soccer, usually a more general team training prescription is applied, because it is difficult to control for training intensity on an individual level during group exercises such as small-sided games (6,13,20,26,27). The training load prescribed by the coach is often called the external load and is expressed in the duration in minutes (TLd). The internal training load on the other hand can be monitored by means of heart rate and session rating of perceived exertion (RPE) scores. Although there is a linear relation between heart rate and oxygen consumption (VO2) during aerobic exercise, it is known that heart rate registration in intermittent sports leads to an underestimation of the actual intensity because of frequent anaerobic exercise (1,27). Other disadvantages of heart rate registration for monitoring a team during a full season are that the procedures are time consuming and vulnerable to technical problems.

Using session RPE scores is an alternative to heart rate monitoring and was first described by Foster (8). Impellizzeri et al. showed that session RPE in soccer is related to Banister’s TRaining IMPulse (TRIMP) method (13). In addition, exercise intensity during resistance training can be monitored with use of session RPE (5). For theoretical and practical reasons, the continuous use of the session RPE for measuring the internal load (TLrpe) is in favor. In this study, participants were asked to fill in a daily training log. Approximately 30 minutes after each training session or match, the total amount of playing time in minutes was recorded as well as their session RPE score on a scale from 6 to 20 (Figure 1).

To monitor recovery, Kenttä and Hassmén proposed the use of a recovery score to quantify the “total quality of recovery” (TQR) (15). This method has already been successfully used in the prediction of performance in a sprinter’s case study (28). In this study, participants were asked to record, before each training session and match, their recovery score on a scale from 6 to 20 (Figure 1).

In addition to monitoring the training load and recovery, it is of utmost importance to measure performance as outcome. In the last decades, different soccer-specific field tests have been developed to evaluate the effectiveness of training programs. Most of these are characterized by their interval profile and validated with aerobic VO2 or anaerobic Wingate tests (17,18,29). In this study, the submaximal interval shuttle run test (ISRT) was used to determine interval endurance capacity. The tests were performed every month (Figure 2) on an artificial pitch at the start of the training as a substitute for the warm-up.

A prospective, nonexperimental cohort design was used to monitor load, recovery, and monthly field test performance. Dutch elite young soccer players were monitored during 1...
competitive season from August 2006 until April 2007. The competitive season followed the preseason that started in July and was separated by a winter break from December until January (Figure 1). Training load and recovery were related to changes in field test performance of players.

Subjects
Eighteen young elite soccer players from the same team volunteered to participate in this study for a full season (mean ± SD: age 17 ± 0.5 years, body mass 72.4 ± 7.8 kg, height 180.4 ± 7.3 cm, body fat 9.3 ± 2.7%). Subjects played a cumulative number of years at the highest level. They received a balanced training program by a professional coach with aerobic, speed, agility, technical, and tactical aspects. Once a week, players executed an individualized weight training program. The team competed in the Dutch premier league under 19 years. The study was approved by the Central Committee on Research Involving Human Subjects. Written informed consent was obtained from the subjects and both parents.

Procedures
Training Log. Participants were asked to fill in a daily training log. Before each training session or match, they recorded their recovery score on a scale from 6 to 20 (Figure 1) as proposed by Kenttä and Hassmen (15). Approximately 30 minutes after each soccer session (13) and weight training (5), the total amount of time in minutes was recorded as well as their session RPE score on a scale from 6 to 20 (Figure 2). To control for missing values, the coach was asked to fill in a Web-based training log to register individual training duration or absence. TLd was calculated by adding training and match duration, and TLrpe was calculated by multiplying duration in min with session RPE; TQR = total quality of recovery.

Interval Shuttle Run Test. To determine interval endurance capacity, a submaximal ISRT was used (19). Submaximal intensity was set at 70% of the maximal amount of runs at the start of the season. During the ISRT, players alternately ran for 30 seconds and walked for 15 seconds. Running speed increased from 10 km h⁻¹ every 90 seconds until 14, 14.5, or 15 km h⁻¹ depending on maximal running level. Heart rate was recorded at 5-second intervals (Polar, Kempele, Finland).

A fixed number of runs was used for every individual during an entire season assuming that heart rate decreases with increasing aerobic fitness (30). The submaximal ISRT has sufficient validity and reliability (intraclass correlation coefficient ≥ 0.86) (16,17,19).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>185.4</td>
<td>2.70</td>
</tr>
<tr>
<td>Test 1</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Test 2</td>
<td>-9.6</td>
<td>1.68*</td>
</tr>
<tr>
<td>Test 3</td>
<td>-0.5</td>
<td>1.82</td>
</tr>
<tr>
<td>Test 4</td>
<td>-2.4</td>
<td>2.05</td>
</tr>
<tr>
<td>Test 5</td>
<td>-5.9</td>
<td>1.86*</td>
</tr>
<tr>
<td>Test 6</td>
<td>-7.9</td>
<td>1.61*</td>
</tr>
<tr>
<td>Test 7</td>
<td>-8.7</td>
<td>1.66*</td>
</tr>
<tr>
<td>TLd (h⁻¹)</td>
<td>-0.9</td>
<td>0.33*</td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2 (between tests)</td>
<td>60.1</td>
<td>21.52</td>
</tr>
<tr>
<td>Level 1 (between subjects)</td>
<td>18.2</td>
<td>2.97</td>
</tr>
</tbody>
</table>

Deviance

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty model</td>
<td>-2 × log-likelihood</td>
</tr>
<tr>
<td>Tests</td>
<td>-2 × log-likelihood</td>
</tr>
<tr>
<td>Tests + TLd</td>
<td>-2 × log-likelihood</td>
</tr>
</tbody>
</table>

*Statistical significance, p < .05.
Statistical Analyses
If more than 75% of the scores within 1 week were completed, missing values were replaced with the average session RPE and TQR scores of the week. Weeks with more than 25% missing values were replaced with the mean scores of the weeks before and after.

The data were analyzed using the multilevel modeling program MLwiN. Multilevel analysis is an extension of multiple regression and is developed for analyzing nested data. The advantage of multilevel modeling is that a different number of measurements per subject are allowed, which is inevitable in full-season data collection. Another advantage is that multilevel analyses take the relations within subjects into account. In all models, subjects represented the upper level and measurement occasion the lower level. Performance was modeled using a random intercept model. First, performance over 7 measurements was entered in the model. After that, TLd and TLrpe were added separately to test the hypotheses if higher external and internal load would lead to better performance. Finally, the contribution of the TQR scores for both models were investigated to test if better recovery would lead to improved performance. These models were calculated for 1 and 2 weeks before the performance test. Data of 4 weeks were not incorporated in the analyses because these data were not available before the first test (Figure 1).

A variable significantly contributed to the model if the Z-score reached the critical value for \( p \leq 0.05 \) (11). Z-scores were calculated by dividing the estimate by its standard error. Models as a total were tested against the \( \chi^2 \) distribution, taking into account the additional number of degrees of freedom (i.e., extra parameters). Because a higher training load and good recovery were expected to improve performance, significance was tested 1-tailed.

Results
Data of 1,480 training sessions and matches were collected. On average, subjects completed 6 sessions a week. Table 1 shows the means and SDs of TLd, TLrpe, and TQR scores of 1 and 2 weeks before the ISRT. Average session RPE scores were 14.4 ± 1.2 and 14.3 ± 1.1 for 1 and 2 weeks, respectively, indicating “somewhat hard” intensity. Average TQR scores for both periods were 14.7 ± 1.3 and 14.6 ± 1.3, which corresponds with “good recovery.”

The first model with random intercept represented performance over 7 measurements with TLd 1 week before the ISRT (Table 2). In total, 93 of 144 data points were included. Missing values were caused by absence during performance tests due to injury, illness, school, playing at the national team, or other obligations.

Performance was compared with the first submaximal measurement as reference. Heart rate on the second, fifth, sixth, and seventh tests were significantly lower than the first. The third and fourth tests did not differ from the first. TLd significantly predicted performance outcome. For every hour of training or game play, heart rate decreased 0.9 b-min\(^{-1}\) at the submaximal field test (Figure 3). Adding TLrpe significantly improved the model.

The second model with random intercept represented performance over 7 measurements with TLd 2 weeks before the performance test (Table 3).

Performance was compared with the first measurement as reference. Heart rate on the second, fourth, fifth, sixth, and seventh tests were significantly lower than the first. The third test did not differ from the first. This estimation is not exactly the same compared with the first model and caused by different TLd data points.
Table 3. Multilevel regression model for heart rate at submaximal speed during the Interval Shuttle Run Test (ISRT) and TL_d per hour (h⁻¹) two weeks prior to the performance tests. The deviance of three models is given. First, the empty model, in which submaximal heart rate is estimated according to one fixed factor, the intercept. Second, a model in which heart rate is estimated according to the intercept and the seven test moments (Tests). In the third model, (Tests + TL_d) training load expressed as training duration in minutes (TL_d) was added to the second model (Tests). Estimates and standard errors (SE) are given for the complete model (Tests + TL_d) only (n = 18).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>185.2</td>
<td>3.32</td>
</tr>
<tr>
<td>Test 1</td>
<td>reference</td>
<td></td>
</tr>
<tr>
<td>Test 2</td>
<td>-10.6</td>
<td>1.67*</td>
</tr>
<tr>
<td>Test 3</td>
<td>-2.3</td>
<td>1.81</td>
</tr>
<tr>
<td>Test 4</td>
<td>-6.1</td>
<td>1.53*</td>
</tr>
<tr>
<td>Test 5</td>
<td>-7.8</td>
<td>1.75*</td>
</tr>
<tr>
<td>Test 6</td>
<td>-7.8</td>
<td>1.68*</td>
</tr>
<tr>
<td>Test 7</td>
<td>-9.9</td>
<td>1.84*</td>
</tr>
<tr>
<td>TL_d (h⁻¹)</td>
<td>-0.3</td>
<td>0.18*</td>
</tr>
<tr>
<td>Level 2 (between tests)</td>
<td>63.9</td>
<td>22.62</td>
</tr>
<tr>
<td>Level 1 (between subjects)</td>
<td>19.1</td>
<td>3.13</td>
</tr>
<tr>
<td>Deviance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empty model</td>
<td>-2 × log-likelihood</td>
<td>635.8</td>
</tr>
<tr>
<td>Tests</td>
<td>-2 × log-likelihood</td>
<td>593.5*</td>
</tr>
<tr>
<td>Tests + TL_d</td>
<td>-2 × log-likelihood</td>
<td>590.0</td>
</tr>
</tbody>
</table>

*Statistical significance p < .05.

TL_d 2 weeks before the test significantly predicted performance outcome. For every hour training or game play in 2 weeks, heart rate decreased 0.3 b·min⁻¹ at the submaximal field test. Adding TL_d did not improve the model significantly.

Likewise, for 1 and 2 weeks before the performance tests, models were calculated with TL_api instead of TL_d. TL_api did not significantly contribute to the model in either time frame. Finally, for all models TQR was added to see whether recovery contributed to the prediction of performance outcome. For all cases, TQR did not significantly predict submaximal heart rate, nor did it result in a better model fit.

Discussion

The hypothesis that higher training load (TL_d and TL_api) leads to increased performance was partly confirmed in the models presented in this study. Adding TL_d for 1 week improved the model significantly (Table 2). The 2-weekly TL_d variable significantly contributed to the model but did not improve the model in total. This means that in well-trained soccer players, the amount of training in the week before the ISRT is most strongly related to the outcome of the test.

Although it was hypothesized that the TL_api would lead to a better prediction, the session RPE scores did not contribute to either of the 2 models. Several factors could explain why session RPE scores did not affect submaximal ISRT performance. First of all, in this study, the range of session RPE scores is small compared with the total range of the scale. Foster reported that athletes trained harder than intended on an easy training as prescribed by the coach. The opposite was found when hard training sessions were prescribed (9). This tendency to flatten out differences in RPE scores decreases the difference between external and internal training load. Additionally, the duration of the training in team sport is the same in many sessions. Therefore, the number of training sessions (expressed as duration in minutes) over a week becomes the main determinant of training load (14).

Impellizzeri et al. (13) showed a moderate relation between heart rate and calculated training load in soccer during a 7-week period, but no information about heart rate in combination with duration was provided. Although the age of the players is similar in the current study, the length of the study is different (7 weeks vs. full season). It might be difficult for young elite soccer players, with little experience, to adequately score their perceived exertion for a full season.

Finally, other studies focused on the heart rate–perceived exertion relationship during training and taper in individual sports (21) or internal training load (session RPE × duration) in relation to overtraining or injury (2,8,23). Because ill and injured soccer players were not able to participate in the performance tests, only the fittest players are represented in this study. Therefore, the question arises whether there might be a difference in the additional value of RPE score in relation to overtraining and injury on the one hand and performance increase on the other hand. This could be an interesting research question for future studies. The additional value of TQR could also be higher for prediction of overtraining or injury compared with prediction of performance. Although the TQR score was successfully used in predicting performance in an individual athlete (28), the additional value for the prevention of overtraining and injury should also be further investigated.

Although biological variation in heart rate on an individual level is 2–4 b·min⁻¹ (1), on a group level this variability is smaller (17). To evaluate reliability of the ISRT, Lemmink and Visscher (17) reported heart rate at 2 consecutive measurements with 2 weeks in between. Heart rate changed 0.8, 0.9, and 1.0 b·min⁻¹ at 14, 14.5, and 15 km·h⁻¹, respectively. This indicates that the differences between tests.
in the current study are meaningful. It also indicates that, with almost 400 minutes training in the week before the ISRT, resulting in a 0.9 b min⁻¹ decrease in heart rate, an extra hour training is needed to exceed the normal variation.

In this study, submaximal performance was measured with use of heart rate. There were 2 main reasons to choose submaximal above maximal testing. First, maximal testing on a monthly basis during a full season can interfere with the training schedule. Second, with such heavy repeated measurements, lack of motivation can negatively influence performance outcome of players. However, it is important to be aware of the factors influencing heart rate using a submaximal test procedure.

It is known that physical stress causes an increase in heart rate. Also dehydration caused by high temperature can increase heart rate (1,7). In this study, participants were tested 2 days after the match with rest in between. The submaximal test lasted between 9 and 12 minutes and was performed at the beginning of the training as a substitute for the warm-up. Therefore, the effect of physical stress and dehydration caused by high temperature is expected to be minimal.

This is the first study that monitored training, recovery, and performance during a full competitive season in young elite soccer players. The sample size in this study is limited due to the high training frequency, which results in a large number of data per individual. Although a repeated measures design improves power, a relatively small sample size limits generalizability.

In conclusion, the duration of training and game play in the week before field test performance is most strongly related to interval endurance capacity. Therefore, coaches should focus on training duration to improve interval endurance capacity in elite soccer players. Although there is reported evidence for their contribution to the data collection.

To thank the coaches and players for their participation. They were 2 main reasons to choose daily mean temperature above maximal testing. First, maximal testing on a monthly basis during a full season can interfere with the training schedule. Second, with such heavy repeated measurements, lack of motivation can negatively influence performance outcome of players. However, it is important to be aware of the factors influencing heart rate using a submaximal test procedure.

It is known that physical stress causes an increase in heart rate. Also dehydration caused by high temperature can increase heart rate (1,7). In this study, participants were tested 2 days after the match with rest in between. The submaximal test lasted between 9 and 12 minutes and was performed at the beginning of the training as a substitute for the warm-up. Therefore, the effect of physical stress and dehydration caused by high temperature is expected to be minimal.

This is the first study that monitored training, recovery, and performance during a full competitive season in young elite soccer players. The sample size in this study is limited due to the high training frequency, which results in a large number of data per individual. Although a repeated measures design improves power, a relatively small sample size limits generalizability.

In conclusion, the duration of training and game play in the week before field test performance is most strongly related to interval endurance capacity. Therefore, coaches should focus on training duration to improve interval endurance capacity in elite soccer players. Although there is reported evidence for their contribution to the data collection.

**PRACTICAL APPLICATIONS**

The best aerobic improvement is shown in the beginning of the season. However, the third and fourth tests did not differ from the first, indicating aerobic performance set back in the 2 months before winter break. Usually, trainers start the preseason with aerobic training and shift to technical and tactical training at the start of the competition. Current results show that to prevent soccer players from aerobic performance decrement, attention should be continuously given to aerobic aspects. This finding also supports the notion that frequent testing is needed to detect performance changes over time.

Because the best relation was found in the week before the ISRT, field tests should frequently be executed and be incorporated in the training program to evaluate the group and individual training response. In the week before the field test, subjects spent almost 400 minutes on training and game play. These rough guidelines for the maintenance of aerobic fitness can be used in professional soccer under 19 years. Although the results suggest that more training would lead to better performance, coaches should always keep the danger of overtraining in mind. Individual databases can provide in more detail information about the performance response.

Monitoring training load, recovery, and field test performance might help coaches to warn players with high training load and poor performance.

Based on the results presented in this study, it seems useful to register the duration of every training and match. When both players and coach fill in a daily training log, missing values can be minimized, which will improve reliability.

**ACKNOWLEDGMENTS**

This study was financially supported by the Netherlands Organisation for Health Research and Development (ZonMw), grant number 7502.0006. The authors would like to thank the coaches and players for their participation. They also thank Swen Eeftink, Casper Jorna, and Willem Bossers for their contribution to the data collection.

**REFERENCES**


Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.


