Reliability of a Field and Laboratory Test of Repeated Sprint Ability

Jonathan L. Oliver, Craig A. Williams, and Neil Armstrong

The purpose of this study was to assess the reliability of a field and a laboratory test of repeated sprint ability (RSA). Twelve adolescent boys (15.3 ± 0.3 years) completed five trials of both a field RSA test (7 × 30 m sprints) and a laboratory RSA test (7 × 5 s sprints) performed on a nonmotorized treadmill. Mean coefficients of variation (CV) calculated across all trials were < 2.7% for field sprint times, and, in the laboratory, < 2.9% for velocity and < 8.4% for power output. Fatigue indices (FI) calculated from data in both environments exhibited mean CVs > 23%. The inconsistency in the FIs resulted from the mathematical procedures used in the FI calculation methods. Based on the reliability scores, it was concluded that results obtained from measured performance variables in the field and laboratory, and not calculated FIs, should be used to report RSA.

Introduction

Collectively, team sports can be described as multiple-sprint sports because of the pattern of activity that characterizes these sports. The need for young team-game players to be able to repeatedly produce brief maximal or near maximal sprint efforts has been identified through motion analysis studies (3,24). Even though sprint efforts represent a relatively small proportion (< 10%) of activity during a team game, the contribution these efforts make to the outcome of a match is thought to be critical (18). It has been shown that soccer players at the highest level will perform more high-intensity running during a match than players of a lower standard (4,14). Therefore, determining the ability of team-game players to produce repeated sprint efforts provides a worthwhile applied performance measure.

Repeated sprint ability (RSA) is defined as the ability to perform repeated, short-duration sprints over a brief period of time (23). Tests of RSA have been designed to replicate a stressful period of play during a team match and are relatively short in duration (< 3 min; 1,7,28). To test RSA, Reilly (17) recommends that a sprint distance of 30 m should be used and repeated 7 times, with an active recovery period between sprints of 15 to 25 s. This protocol is well supported by motion analyses of such sports as hockey (23). Using such a protocol, previous research has investigated seasonal variations in RSA performance (20), differences
in RSA between players competing in different football codes (25), and differences in RSA between elite and sub-elite adolescent boys (21). It has also been reported that two thirds of Premier League soccer clubs in England use multiple sprint tests to monitor fitness (5). It is a requirement for all Premier League clubs to have a soccer academy for youth players. To date, however, only one study has examined the reliability of RSA tests with adolescent boys (16), and in that study only two trials were performed. Repeated sprint ability is characterized by the ability to maintain sprint performance, as represented by a fatigue index (FI). It is not clear whether FIs are reliable measures of RSA, and limited data suggest that this might not be the case (7,8).

Repeated sprint ability has more commonly been investigated in the laboratory using cycle ergometry (1,2,7). There is, however, an obvious lack of specificity between cycle ergometry and the exercise performed by team-game players. Unsurprisingly, Fitzsimmons et al. (7) observed that the best cycling effort was not significantly correlated with the best running effort during tests of RSA. A nonmotorized treadmill (NMT) is an ergometer which allows rapid changes in running velocity to be monitored in the laboratory (13). Unfortunately, there is a lack of data relating to the reliability of the NMT, with only two published test–retest reliability studies in children (6,26). The purpose of this study was to assess the reliability of a field-based RSA test and the reliability of a new laboratory-based RSA test, each performed by adolescent boys.

**Methods**

Twelve adolescent boys from a local secondary school volunteered and participated in the study; all of the boys were members of the school soccer and/or rugby teams. The boys were a mean age of 15.3 ± 0.3 years, with stature of 1.78 ± 0.07 m, body mass of 64.4 ± 8.0 kg, and sum of triceps and subscapular skinfold thickness of 15.7 ± 5.0 mm. The Institutional Ethics Committee approved the project and written informed assent and consent was obtained from the boys and their parents or guardians, respectively.

Before testing, participants were habituated to both conditions; this included performing a practice attempt of each test. Practice attempts were performed on days separate from the actual test days. A practice attempt of the laboratory protocol was not attempted until each participant was fully familiarized with the NMT and had demonstrated the ability to sprint on the treadmill. The participants were instructed to wear the same clothing during each test and were given diet and activity questionnaires to complete immediately before the first test session. The participants verbally confirmed that they had used the information contained in the questionnaires to ensure a similar preparation before each test. Before each test participants performed a warm-up consisting of 5 min of jogging followed by two brief sprints (< 4 s), and 10 min was then allowed to perform individual stretching exercises. Participants completed a total of five RSA tests in the field and five RSA tests in the laboratory, with seven sprints completed during each test. Field and laboratory RSA tests were conducted concurrently; the order in which testing was completed was largely determined by the weather, with laboratory testing occurring when bad weather prevented field testing. Individual participants completed every
RSA test within the same 1-hr time period, with all testing occurring in the morning. There was a minimum of 24 hrs between tests, and participants completed all 10 test sessions within 17 to 32 days. One participant completed only four laboratory RSA tests and another participant only three.

Field Repeated Sprint Ability Test

The field RSA test was similar to that used previously with male adult and adolescent populations (19,20,25). The test was performed individually outside on an artificial turf pitch (Astro Turf). The field test was only performed in dry conditions with a wind speed of $\leq 8$ km/hr. The course consisted of a 30 m maximal sprint through photoelectric cells (RS Components, Corby, UK) linked to a digital timing system used to record the split times to cover the initial 10-m and 10- to 30-m portion of the sprint (Figure 1). A total of seven maximal sprints commencing every 25 s were completed during each test. After each sprint the participants decelerated and jogged back to the start position, covering a total distance of 51 m. During the active recovery, continuous verbal feedback was provided to ensure the participant had returned to the start position in time to begin the next sprint. Each sprint began with a stationary start from 1 m behind the first timing gate; this was done to ensure the timing system functioned correctly. A verbal countdown was given to begin each sprint, and verbal encouragement was given throughout each sprint. Sprint performance was measured as the fastest 10- (F10) and 30-m (F30) times and the mean 10 (M10) and 30 m (M30) sprint times during each test.

![Diagram of the field RSA test](image-url)
Laboratory Repeated Sprint Ability Test

The laboratory RSA test was performed on a Woodway Tramp nonmotorized treadmill (Woodway GmbH, Germany). The experimental set-up of the nonmotorized treadmill has previously been described by Sutton et al. (26). Briefly, a free-standing custom built safety frame encloses the treadmill allowing the attachment of a safety harness, elastic tether, and strain gauge. The strain gauge (Novatech Measurements, St.-Leonards-on-Sea, UK), with a measurement range of 0 to 40 kg, is mounted to the rear of the safety frame and attached to the participant via an extensible tether system; this allows the measurement of horizontal force. An electronic sensor was attached to the front drum of the treadmill to monitor the treadmill belt velocity. The output voltages from both the velocity and force sensors were connected to a multifunction interface card including a 12-bit analogue-to-digital converter installed in an IBM compatible computer. A custom software application developed using LabVIEW (National Instruments, Newbury, UK) programming software was used to sample the signal voltages from both sensors at 100 Hz. Data were then averaged and reported over 1-s intervals. The product of the horizontal force (N) and treadmill belt velocity (m/s) was used to calculate the horizontal power output (W). A visual display monitor placed in front of the treadmill allowed participants to monitor their velocity.

The laboratory RSA test was designed to replicate the activity pattern of the field RSA test. Each laboratory test consisted of seven 5-s maximal sprints, with each sprint commencing every 25 s. Before the start of each test, participants stood in the middle of the treadmill and were attached to both the tether/strain gauge system and the safety harness. The participant stood upright and the strain gauge position was altered so that the tether was approximately 8° above the horizontal. This was based on the suggestions of Lakomy (13), who found that this positioning minimized the errors in force measurement. Each test commenced with a rolling start, with the participants required to achieve a steady velocity of 8 km/hr. After a verbal countdown the participant performed a maximal 5-s sprint. The experimenter gave a verbal “Stop!” command to signal the end of each sprint, and the participant was then required to decelerate to a velocity of 8 km/hr within 2 s. This was then maintained for the remainder of the recovery bout. The visual display monitor in front of the treadmill also signaled the end of a sprint by a change in color of the display. Verbal encouragement was given throughout each sprint, and verbal feedback was also given during each recovery bout. The peak values attained for both velocity (PV) and power output (PPO) during each test were recorded. The mean values for velocity (MV) and power output (MPO) were calculated from the sprint data recorded throughout each test.

Fatigue Indices

Fatigue indices (FI) were calculated using a variety of methods. A traditional percentage drop-off in performance was calculated between the mean results of the first two and last two sprints for field and laboratory data. For comparison purposes, two further methods were used to calculate the FI of field data. An alternative FI was calculated as the percentage drop-off in performance between F30 and M30 (7). An absolute FI was calculated as the drop-off in time (s) from F30 to M30.
Statistical Analyses

All analyses were performed on the natural logarithms of the performance data. This reduced any effect of nonuniformity of error and allowed the percentage change in the mean and coefficient of variation between consecutive trials to be examined in a pairwise analysis. A pairwise analysis was performed for data collected in consecutive trials for participants who had completed all five test trials. A mean CV was then calculated for each performance variable across all five trials. Data were not included in the mean CV calculation for participants that did not complete all five trials; therefore, the sample size was $n = 10$ for laboratory variables. The mean coefficient of variation was calculated using a two-way ANOVA on the log-transformed data with participants and test trials as effects (28). The antilog of the root mean square error term was substituted into the following equation (22):

$$\text{Mean CV} = 100 (\text{eRMSE} - 1)$$

Confidence intervals (CI) were calculated using an online spreadsheet (12), with the CI calculations being based on the degrees of freedom and chi-square distribution. The combination of sample size and number of trials was chosen to provide a factor of approximately 1.25 for calculating the 95% CI. This provides greater accuracy than previous research examining the reliability of RSA tests in which the factor for calculating 95% CIs has ranged from 1.29 to 1.47 (7,8,16,28).

Results

Pairwise Analysis of Consecutive Trials

Figure 2A displays the group mean results during each trial of the field RSA test ($n = 12$). The group mean sprint times in the field are consistent across all five trials. Pairwise analysis confirmed that the percentage change in the mean between consecutive trials was $\leq 1.1\%$ for the fastest 10-m time, $\leq 0.2\%$ for mean 10-m time, $\leq 1.4\%$ for the fastest 30-m time, and $\leq 1.9\%$ for the mean 30-m time. Figure 2B displays the group mean results for the velocities achieved while sprinting on the treadmill ($n = 10$). Pairwise analysis indicated that the velocity recorded in consecutive trials was consistent, with the percentage change in the mean of $\leq 3.5\%$ for all velocity variables. Figure 2C displays the group mean results for power output measurements during the laboratory RSA tests. For PPO the percentage change in the mean between consecutive trials was 4.9% or below. The percentage change in the mean between consecutive trials for MPO was $\leq 3.3\%$.

Mean Coefficients of Variation

A summary of the mean performance results and the mean CVs (95% CI) are presented in Table 1. Results show that mean sprint times in the field, over both distances (M10 and M30), were less variable than the respective fastest sprint times. Additionally, data recorded over the 30-m distance were more reliable than data recorded over the first 10 m. Similarly analysis of laboratory data revealed that mean CVs were lower for mean compared to peak values, and velocity data were less variable than power outputs.
Figure 2 — Group mean results for (A) sprint times during each trial of the field RSA test ($N = 12$), (B) velocity during each trial of the laboratory RSA test ($N = 10$), and (C) power output during each trial of the laboratory RSA test ($N = 10$).
Fatigue Indices

The calculated traditional FIs for each trial of the field and laboratory RSA tests are shown in Figure 3. Pairwise analysis revealed inconsistent results with the percentage change in the mean exceeding 10% for each variable. The inconsistency was reflected by the mean CVs which were > 46% when FIs for field and laboratory data were calculated using the traditional method. Using the alternative calculation method, based on F30 and M30 sprint times, the mean CV of the FI of field data was reduced to 24.7%. A similar mean CV of 23.0% was found when the FI of field data was expressed in absolute terms.

Table 1  Mean Results and Mean CVs (95 % CI) Across all Five Trials for the Field RSA Test and the Laboratory RSA Test

<table>
<thead>
<tr>
<th>Test</th>
<th>Variable</th>
<th>Mean result</th>
<th>Mean CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field RSA test</td>
<td>F10</td>
<td>1.91 s</td>
<td>2.69 (2.23 – 3.40)</td>
</tr>
<tr>
<td></td>
<td>M10</td>
<td>2.02 s</td>
<td>2.01 (1.67 – 2.54)</td>
</tr>
<tr>
<td></td>
<td>F30</td>
<td>4.76 s</td>
<td>2.22 (1.84 – 2.81)</td>
</tr>
<tr>
<td></td>
<td>M30</td>
<td>5.15 s</td>
<td>1.55 (1.28 – 1.96)</td>
</tr>
<tr>
<td>Laboratory RSA test</td>
<td>PV</td>
<td>6.10 m/s</td>
<td>2.88 (2.34 – 3.74)</td>
</tr>
<tr>
<td></td>
<td>MV</td>
<td>4.93 m/s</td>
<td>2.59 (2.11 – 3.36)</td>
</tr>
<tr>
<td></td>
<td>PPO</td>
<td>619 W</td>
<td>8.32 (6.77 – 10.81)</td>
</tr>
<tr>
<td></td>
<td>MPO</td>
<td>439 W</td>
<td>5.41 (4.40 – 7.03)</td>
</tr>
</tbody>
</table>

Note. F10 = ...
Discussion

The aim of this study was to assess the reliability of an established field RSA test and a new laboratory RSA test with an adolescent population. Sprint times in the field test were found to produce reliable results (CV 1.6–2.7%), as were velocity measurements during the laboratory test (2.6–2.9%). Power output measurements in the laboratory were more variable (5.4–8.3%). A major finding of this study was that three different methods of calculating the FI all produced unacceptable amounts of variability (23–46%). To the best of our knowledge, this is the first study with any population to measure the reliability of the commonly reported RSA test variables of fastest sprint times and FIs using multiple trials.

The sprint times recorded during the field RSA test were slower than those previously reported for both elite and sub-elite male soccer players age 16.4 years (21). The slower sprint times in the present study might be a result of participants being younger (15.3 years), the test surface (Astro Turf), or because of the effects of testing outdoors. Alternatively, seasonal variations might have influenced the results because testing occurred during the off-season; the RSA of Gaelic footballers has been shown to vary throughout the year (20). It is difficult to compare performance data from the laboratory RSA tests with previous research because of the scarcity of data for NMT sprinting, particularly for an adolescent population. Given the differences in protocols, it is only possible to compare peak results across studies. The peak velocities and power outputs achieved on the NMT during the laboratory RSA test were below those reported for adult rugby players (10,27) and adult games players (9), but above those reported for younger children (6). The lower peak values attained by adolescent participants, when compared with adults, is most likely to be a result of their lower body mass when compared with adults, with lighter participants disadvantaged during nonmotorized treadmill sprinting (13).

The use of thorough habituation sessions, including practice attempts of each test, minimized any learning effects during the main study. Pairwise analysis showed that for all measured variables there were only relatively small changes in the group mean between consecutive trials. Although pairwise analysis revealed more variation in laboratory data when compared with field data, the variation was within the expected range. For example, Tong et al. (27) reported a mean CV of 8.2% for PPO during nonmotorized treadmill sprints. When the percentage change in the mean between consecutive trials was converted to a coefficient of variation (CV), the CV for PPO (the most variable measure) ranged from 6.6 to 10.7%. Because the variation between consecutive trials was similar to previously reported data (27), it was deemed acceptable to calculate a single, more precise mean CV across all five trials for each variable in the field and laboratory.

The M30 sprint times during the field RSA test were found to produce the lowest variability, with a mean CV of 1.6%. This compares well with previous adult reliability studies, which have reported the CV for mean sprint times during RSA tests to be 0.8% (7), 1.8% (28), and 2.3% (15). Research investigating RSA has reported results as the fastest sprint times together with a fatigue index (19,20,25), as well as reporting the time taken to sprint 10 m and 30 m during each sprint (20,25). This is the first study to assess the reliability of all the commonly reported RSA measures of split times (10 and 30 m), fastest and mean sprint times, and the
fatigue index. The mean CV of 2.2% for F30 is in good agreement with the value of 2.1% reported by Psotta and Bunc (16) for the reliability of peak velocity during repeated 20-m sprints performed by adolescent boys.

To our knowledge only one previous study has examined the reliability of brief sprints performed on an NMT in adults (27), and only two studies have examined the reliability of children sprinting on an NMT (6,26). Examining the reliability of single 6 s sprints with adults on the NMT, Tong et al. reported that peak speed had a CV of 1.3%. This is below the 2.9% CV reported for PV in the present study. Tong et al., however, examined test–retest data using the highest result from three separate 6-s trials on each day, thereby reducing the amount of variability.

Power output during the laboratory RSA tests was found to be less reliable than velocity. This is not surprising because power output is not directly measured but is calculated from the product of force and velocity. Peak power output was found to be the least reliable measure with a CV of 8.3%. This is very similar to the between-days CV of 8.2% reported for PPO during 6 s NMT sprints (27). Using the same experimental set-up as the present study, Sutton et al. (26) examined the reliability of young children (10.9 years) performing a 30-s sprint on the NMT. Sutton et al. (26) reported coefficients of repeatability of 26.6 W for PPO and 15.3 W for MPO; these can be recalculated to give mean CV values of 6.3 and 5.1%, respectively. The variability of the MPO is similar to that found in the present study. The inconsistency in PPO was the result of variability in both the velocity and force measurements at the time when PPO occurred. The large amount of variability in PPO might have been the result of commencing each sprint from a rolling start, which might have resulted in participants starting each sprint from slightly different velocities and with different amounts of force stored in the tethering system. With data averaged over 5 s for MPO, the affect of any inconsistency in the rolling start was reduced. Consequently MPO provided a more reliable measure.

The reliability CV can be used to determine the likely limits of a “true” value from a single test and to determine the chance that an observed change between two or more test scores is “real,” as well as provide an estimate of future experimental sample sizes. The more reliable test variables allow a greater precision of measurement when calculating the likely limits of a single score or when trying to detect a change between test scores. If observing a change in test score that is equal in magnitude to the CV, then there is a 75% chance that a positive change occurred; there is still a 25% chance, however, that a negative change in performance occurred (12). A coach or sport scientist would then be more confident of detecting a smaller change in performance using the field test compared with the laboratory test. The amount of variability associated with FIs, however, would result in broad likely limits of a single test score and make it more difficult to observe “real” changes between tests. It would be likely that any change in the FI was simply a result of measurement error.

Hopkins (11) provides an equation for estimating the required sample size of a future experimental study using a 95% confidence interval based on the reliability CV and the smallest worthwhile change in performance (d):

\[ N \approx 8 \times \frac{CV^2}{d^2} \]
Although laboratory testing provides the opportunity to measure power output, the variability of power measurements has a large impact on estimated sample sizes for future studies. A smallest worthwhile enhancement of 2.0% would require 10 participants based on F30 field times, but this would increase to 59 participants if MPO was the variable of interest. The equation given by Hopkins (11) is based on a simple crossover study; for a study involving a control group, the sample size would become $4N$. Given the variability associated with power output measurements, it would be difficult to detect a small change in score without testing a large sample. Taking account of the 95% confidence intervals will also have a large effect on the required sample size, which could range from 39 to 99 in the above example for MPO. It might be more practical to study changes in field performance or velocity in the laboratory test rather than power output. Alternatively, the value of $d$ could be increased. When $d$ is increased to the magnitude of the CV, the required sample size is only 8. In the case of the FI, this would give an impractically large $d$ to try and detect. Where changes in the FI (20) or differences in the FI between groups (25,21) have previously been reported in experimental studies, the results must be viewed with caution.

The FI is a routinely reported measure quantifying the decrement in performance during a test. The reliability of FIs, however, has received little attention. A major finding of this study was that regardless of the method of calculation, the FI produced unreliable results. Using the traditional calculation method, the mean CVs were greater than 46%, which is similar to the value of 52% reported from test–retest data of adolescent soccer players performing repeated 20-m sprints (16). The alternative method calculated the FI based on F30 and M30, measures found to be reliable, but still produced variable results (mean CV = 24.7%). This is similar to the variation of 18.5% reported by Fitzsimmons et al. (7) when calculating the FI during repeated field sprints using the same method. Even when fatigue was expressed in absolute terms, the FI still remained unreliable (mean CV = 23%). This finding helped to explain the lack of reliability found with the other calculation methods. The within-participant variability associated with F30 and M30 was low (mean ± mean within-participant SD, 4.76 ± 0.10 and 5.15 ± 0.08 s, respectively), leading to low CVs. Subtracting F30 from M30, however, produces a measure with a much lower absolute value and relatively greater amount of variability (0.38 ± 0.07 s), leading to a large CV. It is this process that was accountable for the large mean CV associated with all FI calculation methods. This problem will always exist where a small decrement in performance is expressed as a drop-off in performance. As an example, consider a player who records F30 sprint times of 4.5- and 4.6-s and M30 sprint times of 5.0 and 5.0 s on two separate test occasions. On the first occasion the absolute FI is 0.5 s and on the second occasion it is reduced to 0.4 s. This is a change in the FI of 20% that was brought about by only a 2.2% change in F30, although M30 remained unchanged. The above scenario might also give the impression that RSA had improved, given a reduced FI. Because fastest sprint time was slower and mean sprint time remained unchanged in the second test occasion, the implication is that RSA had not improved. The inherent lack of reliability in an FI calculated from an RSA test and the difficulty in drawing conclusions from the FI mean that this measure can be misleading. Therefore, a fatigue index should not be used to characterize repeated sprint ability.
Summary

This study was the first to measure the reliability of unidirectional RSA tests using multiple trials performed by adolescent boys. All performance measures obtained during the field RSA test provided reliable results (CV < 2.7%). Results from the laboratory RSA tests showed that velocity (CV < 2.9%) was more reliable than power output (CV < 8.3%). Calculated fatigue indices were shown to be inherently unreliable. It is important to consider the application of reliability studies and how the results impact decisions regarding the precision required, detecting a change between test scores, and estimating sample sizes of future studies. Given the amount of variability associated with the FIs, such measures would not provide adequate precision for monitoring performance, and it would be difficult to detect a change in the FI in an experimental investigation. It is therefore recommended that peak and mean performance measures be used to describe RSA.

Acknowledgments

The authors would like to thank the Principal, PE department, and pupils of Ivybridge Community College for their cooperation and involvement in the study.

References