

Test-retest reliability of a maximal graded treadmill test in youth soccer players



Matthew A Mohammadnabi, Andrew D. Fields, Michael R. Esco
The University of Alabama, Department of Kinesiology

Abstract

INTRODUCTION: Maximal graded exercise tests are commonly used to assess aerobic capacity and endurance performance in athletes. However, repeated testing may be influenced by learning effects, potentially affecting the reliability of performance outcomes. The purpose of this study was to determine the test-retest reliability in time-to-exhaustion and maximal work rate during a progressive treadmill running test to failure over two separate days in competitive youth soccer players. **METHODS:** Ten competitive youth soccer players (age = 16.8 ± 0.7 years, height = 177.6 ± 6.2 cm, weight = 69.1 ± 10.0 kg) completed two maximal graded exercise tests on a treadmill, separated by one week. The test protocol began at 6 km·h⁻¹ at 0% grade, with speed increasing by 1 km·h⁻¹ every 3 minutes until the seventh stage (11 km·h⁻¹). Thereafter, the grade of the incline increased by 2% per stage. The test ended when the subjects reached a point of volitional fatigue. Time-to-exhaustion and maximal work rate (speed and fractional grade) were recorded and compared between trials. **RESULTS:** Time-to-exhaustion was significantly longer on Day 2 (23.8 ± 3.6 min) than on Day 1 (21.7 ± 4.3 min; p = 0.005, Cohen's d = 0.52). The intraclass correlation coefficient (ICC) for time-to-exhaustion was 0.94 (p < 0.001). All participants reached at least the seventh stage on both days and hence, maximal speed was maintained at 11 km·h⁻¹ across trials. However, the maximal achieved grade was significantly higher on Day 2 (0.08 ± 0.02) than on Day 1 (0.06 ± 0.03; p = 0.004, Cohen's d = 0.78). The ICC for fractional grade was 0.91 (p = 0.001). **CONCLUSION:** Youth soccer players exhibited significantly greater time-to-exhaustion and maximal work rate on the second testing day, despite the tests being separated by only one week. These findings suggest that the observed improvements were primarily due to increased test familiarity rather than physiological adaptation. **PRACTICAL APPLICATIONS:** Practitioners and researchers assessing maximal running performance in youth soccer players should account for familiarity effects when interpreting test results. Familiarization sessions should be incorporated prior to baseline testing to minimize the influence of learning effects in longitudinal studies or intervention trials.

Introduction

- Maximal graded exercise tests (GXTs) are widely regarded as the gold standard for assessing aerobic fitness.¹
- These tests provide critical information on an athlete's cardiorespiratory fitness and are frequently used to evaluate training adaptations or to predict performance potential.²
- In youth sports, particularly in soccer where endurance and high-intensity running capacity are key performance determinants, GXTs may be employed to monitor development of aerobic fitness over time.^{3,4}
- However, repeated exposure to maximal exercise testing can introduce learning or familiarization effects that may confound the interpretation of performance outcomes.⁵
- Establishing the test-retest reliability of GXT protocols is essential to ensure that changes in performance over time reflect genuine physiological adaptations.⁶
- Although test-retest reliability has been well studied in adults and elite athletes, there is a relative lack of data in adolescent populations, particularly those actively engaged in competitive sports.⁷
- The purpose of this study was to determine the test-retest reliability in time-to-exhaustion and maximal work rate during a progressive treadmill running test to failure over two separate days in competitive youth soccer players.

Methods

- Ten competitive youth soccer players (age = 16.8 ± 0.7 years, height = 177.6 ± 6.2 cm, weight = 69.1 ± 10.0 kg) completed two maximal graded exercise tests on a treadmill, separated by one week.
- The test protocol began at 6 km·h⁻¹ at 0% grade, with speed increasing by 1 km·h⁻¹ every 3 minutes until the seventh stage (11 km·h⁻¹). Thereafter, the grade of the incline increased by 2% per stage. The test ended when the subjects reached a point of volitional fatigue.
- Time-to-exhaustion and maximal work rate (speed and fractional grade) were recorded and compared between trials.
- Paired t-tests were used to compare the mean values in time-to-exhaustion and maximal work rate between testing days 1 and 2.
- Cohen's d effect sizes were calculated to examine the effect sizes of the mean differences.
- Intraclass correlations were used to assess the agreement in the studied variables between testing days.
- Level of significance was set at p < 0.05

Results

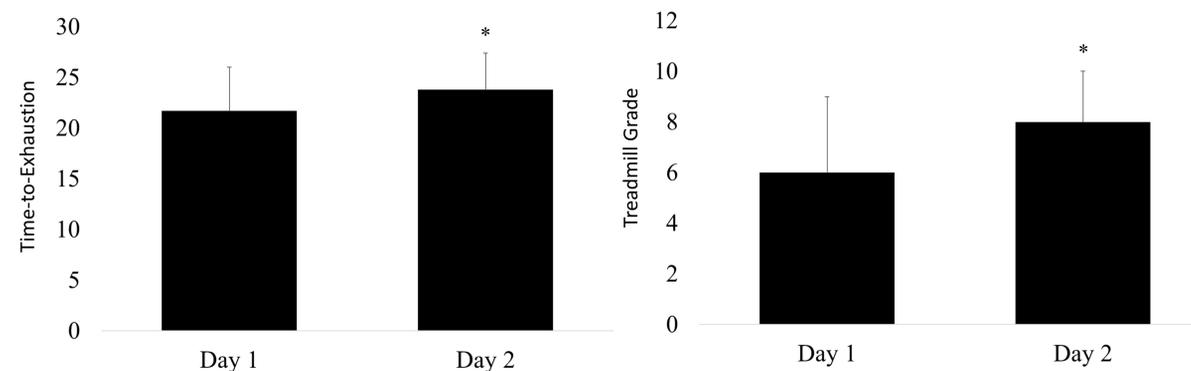


Figure 1. Time-to-exhaustion (minutes) between Day 1 (21.7 ± 4.3 min) and Day 2 (23.8 ± 3.6 min). *indicates Day 2 was significantly higher than Day 1 (p = 0.005, Cohen's d = 0.52).

Figure 2. The grade of the incline (%) at exercise exhaustion between Day 1 and Day 2. *indicates Day 2 was significantly higher than Day 1 (p = 0.004, Cohen's d = 0.78).

Table 1. Intraclass correlations between Day 1 and Day 2 for time-to-completion and the incline achieved at exhaustion

| Variable | ICC | p value |
|--------------------|------|---------|
| Time-to-Exhaustion | 0.94 | <0.001 |
| Incline | 0.91 | 0.001 |

Please note: For the maximal attained speed, all participants reached 11 km·h⁻¹ on Day 1 and Day 2.

Conclusions

- Though the high ICCs observed for both time-to-exhaustion and fractional grade indicate strong reliability of the testing protocol, the statistically significant improvements (after only a 1 week) highlight the importance of accounting for learning effects when interpreting test results in youth soccer players.
- The findings of this study demonstrate that this population of athletes may exhibit significantly improved performance during the second administration of the GXT.
- However, the performance gains occurred in the absence of any formal training intervention between sessions, strongly suggesting that the improvements were driven primarily by increased test familiarity and not underlying physiological changes.
- The findings highlight the importance of accounting for learning effects when interpreting test results, particularly in youth populations who may have limited experience with maximal exertion protocols.
- Therefore, familiarization sessions should be included before baseline testing in research or performance monitoring contexts.

Practical Applications

- Practitioners and researchers working with youth soccer players should incorporate at least one familiarization session prior to baseline testing with maximal GXT protocols.
- Doing so may help minimize the impact of learning effects, ensuring that observed improvements more accurately reflect true physiological adaptations.
- Short-term performance gains, especially in athletes unfamiliar with high-intensity treadmill running, should be interpreted cautiously.
- Even well-trained youth may experience significant learning effects that can influence test outcomes.
- By standardizing protocols and including familiarization trials, practitioners can improve the reliability of performance assessments, leading to more informed training decisions and more accurate tracking of athlete development over time.