International Journal of Sports Physiology and Performance, 2014, 9, 340-345
http://dx.doi.org/10.1123/IJSPP.2013-0103

INTERNATIONAL JOURNAL OF SPORTS PHYSIOLOGY AND PERFORMANCE
www.IJSPP-Journal.com
ORIGINAL INVESTIGATION

Reliability of a 2-Bout Exercise Test on a Wattbike Cycle Ergometer

Matthew W. Driller, Christos K. Argus, Jason C. Bartram, Jacinta Bonaventura, David T. Martin, Nicholas P. West, and Shona L. Halson

Purpose: To determine the intraday and interday reliability of a 2 × 4-min performance test on a cycle ergometer (Wattbike) separated by 30 min of passive recovery (2 × 4MMP). ***Methods:*** Twelve highly trained cyclists (mean ± SD; age = 20 ± 2 y, predicted VO_{2max} = 59.0 ± 3.6 mL · kg^{-1} · min^{-1}) completed six 2 × 4MMP cycling tests on a Wattbike ergometer separated by 7 d. Mean power was measured to determine intraday (test 1 [T1] to test 2 [T2]) and interday reliability (weeks 1–6) over the repeated trials. ***Results:*** The mean intraday reliabilities of the 2 × 4MMP test, as expressed by the typical error of measurement (TEM, W) and coefficient of variation (CV, %) over the 6 wk, were 10.0 W (95% confidence limits [CL] 8.2–11.8), and 2.6% (95%CL 2.1–3.1), respectively. The mean interday reliability TEM and CV for T1 over the 6 wk were 10.4 W (95%CL 8.7–13.3) and 2.7% (95%CL 2.3–3.5), respectively, and 11.7 W (95%CL 9.8–15.1) and 3.0% (95%CL 2.5–3.9) for T2. ***Conclusion:*** The testing protocol performed on a Wattbike cycle ergometer in the current study is reproducible in highly trained cyclists. The high intraday and interday reliability make it a reliable method for monitoring cycling performance and for investigating factors that affect performance in cycling events.

Keywords: power output, cycling, reproducibility, performance testing

Monitoring training loads and evaluating the effects of different training programs in highly trained athletes has become a pivotal aspect of sports performance. The use of regular performance testing may allow coaches and scientists to track athletic performance improvements or even detect athletes who are overreaching or overtraining.[1] Recent research has shown that athletes showing signs of overtraining syndrome have an inability to sustain intense exercise and recover when there is a short duration between repeated exercise bouts.[2] However, little focus has given to the reproducibility of the performance tests used to monitor both improvements and decrements in athletic performance. The tests used must have sufficient precision to detect changes, and any lack of precision makes it very difficult to interpret differences in performance.[3]

The reliability of a performance test refers to its reproducibility when the test is administered over several occasions on the same individual. A common method of evaluating a test's reliability is referred to as the typical error of measurement (TEM). The TEM in performance consists of both random and systematic errors.[4] In cycling, the random error includes the test–retest variation of cyclists who do not always perform each test in an identical fashion. The more highly trained or experienced the cyclist, the lower the chance of random error or test–retest variation.[4,5] Random error can be minimized by using an appropriate type of test. In cycling, the systematic error may relate to the inability of an ergometer to accurately measure power output.[6]

The Wattbike cycle ergometer is currently used in numerous sporting facilities and university laboratories to assess and monitor cycling performance. To date, only a limited number of studies have investigated the accuracy and reproducibility of the ergometer. However, those investigations only examined constant-load or near-maximal-power-output tests.[7,8] The reliability of the Wattbike under constant-load cycling (in the range of 50–300 W) was reported to have a coefficient of variation (CV) of 2.6% in trained cyclists.[8] Furthermore, an investigation in our laboratory examining the interday reliability of a 30-second sprint on the Wattbike ergometer resulted in a CV of 2.4%.[7] However, this test was largely anaerobic in nature; thus, it is relatively unknown whether longer performance tests may result in similar reliability. In addition, the reliability of a repeated-bout (same-day) cycling test on a Wattbike ergometer in highly trained athletes has not yet been investigated.

Having a reliable repeated-bout protocol is highly important to monitor not only fatigue and overreaching

Driller is with the Dept of Sport and Leisure Studies, University of Waikato, Hamilton, New Zealand. Argus, Bartram, Bonaventura, Martin, and Halson are with the Physiology Dept, Australian Institute of Sport, Belconnen, ACT, Australia. West is with the Centre for Medicine and Oral Health, Griffith University, Southport, QLD, Australia.

but also changes in performance from different recovery strategies or training and nutritional interventions. In addition, repeated-bout exercise tests may also simulate real-world sporting performance, such as the individual pursuit (4000 m) at a track cycling competition, whereby cyclists are often required to compete multiple times within a short time frame.

Therefore, the aim of the current study was to determine the interday and intraday reliability of two 4-minute cycling tests, separated by 30 minutes, performed on a Wattbike cycle ergometer in a group of highly trained cyclists.

Methods

Subjects

Twelve highly trained cyclists (mean ± SD; age 20 ± 2 y, mass 69.0 ± 4.2 kg, height 180.9 ± 5.6 cm, maximal aerobic power 354 ± 26 W, predicted VO_{2max} 59.0 ± 3.6 $mL \cdot kg^{-1} \cdot min^{-1}$) completed the current study. All participants were at a nationally competitive level and were in an early phase of their domestic season. Subjects provided informed consent before any testing taking place. The study was approved by the Australian Institute of Sport research ethics committee.

Inclusion Criteria

The data used for analysis in the current study were taken from a larger 6-week training study (unpublished). For the purpose of this article, only the reliability data for 1 of the performance tests is described. The data used for analysis in the current study were taken from the cyclists who performed the 2 × 4-minute cycle test (2 × 4MMP) once each week over the 6 weeks. In cases where there were not full data sets for individual cyclists, the data were excluded (total of 9 subjects excluded). From the subjects included in the current study, 6 were in the experimental group, with the remaining 6 in the control group for the larger training study. We performed an analysis and found no significant difference between groups for the 2 × 4MMP test over the 6 weeks ($P > .05$) and, therefore, pooled the data for both groups (N = 12) in the analysis to determine the reliability of the test.

Design

The subjects attended 8 separate testing sessions at our laboratory over a 6-week period. Initially, they completed an incremental cycling test on an electromagnetically braked cycle ergometer (Lode Excalibur Sport, Groningen, The Netherlands) to establish each individual's peak power output, predicted maximal oxygen uptake (VO_{2max}), and maximum heart rate. A familiarization trial of the 2 × 4MMP was also performed 3 days before the first test. After the familiarization trial, 6 testing sessions were performed, each separated by 7 days. Training during the experimental period was controlled for, with all subjects completing the same prescribed training sessions throughout the 6 weeks. Subjects were asked to refrain from caffeine (<12 h) and to arrive at each session in a hydrated state. Testing was performed at the same time of day (± 1 h, to minimize diurnal variation) and on the same cycle ergometer.

Incremental Exercise Test

The incremental cycling test started at 150 W and increased in power output by 50 W every 3 minutes until volitional exhaustion was reached or the subject could no longer maintain a pedal cadence of >60 rpm. During the incremental exercise test, heart rate was recorded continuously using a RS800 heart-rate monitor (Polar Electro Oy, Kempele, Finland). The maximum heart rate achieved during the exercise test (HR_{max}) was used to set the warm-up for the 2 × 4MMP. While not measured directly, VO_{2max} was predicted using the following regression equation described by Lamberts et al[9]:

$$VO_{2max}\ (mL \cdot kg^{-1} \cdot min^{-1}) = [10.97 \times \text{peak power output (W/kg)}] + 2.598$$

Wattbike Cycle Ergometer

All cycle testing was performed on an air-braked cycle ergometer (Wattbike Ltd, Nottingham, UK). The Wattbike calculates power output by measuring the chain tension over a load cell (sampled at 100 Hz) and the angular velocity of the crank arms (twice per revolution).

Before the start of the study, the Wattbike ergometer was calibrated on a dynamic calibration rig using a first-principles approach by specialists at the Australian Institute of Sport.[10] The reliability of the Wattbike cycle ergometer has been reported previously over a range of power outputs (50–300 W), with a CV of 2.6% (95%CI 0.7–2.0%) in trained cyclists.[8]

The 2 × 4-min Mean Power Test (2 × 4MMP)

The 2 × 4MMP test involved two 4-minute maximal efforts completed 42 minutes apart. A controlled warm-up was performed before each of the 2 efforts. The first warm-up included 3 stages of work set at different percentages of HR_{max} obtained from the incremental step test (6 min at 60%, 6 min at 80%, 3 min at 90%) followed by a 2-minute stationary rest and 2 minutes at 70% of HR_{max}. The second warm-up was shorter and included 6 minutes at 80% and 3 minutes at 90% of HR_{max} before 2 minutes of passive rest. The subjects were seated for 30 minutes between the 2 efforts (passive recovery) in a temperature-controlled room (21°C ± 1°C). The design was based on a repeated-bout protocol previously used to detect overreaching and overtraining syndrome.[2] To ensure that the protocol was applicable to sporting performance, it was also designed to closely mimic an individual pursuit (4000 m) in a track-cycling competition where cyclists are often required to compete in 2 races separated by a time frame similar to that used in the current study.

During the two 4-minute maximal tests (T1 and T2), subjects could view the elapsed time and were required to produce as much work as possible in the time frame, but no other information was provided. During the submaximal stages of the warm-up, preload, and cooldown, subjects were instructed to maintain their target heart rate irrespective of their power output. The gearing and cadence (rpm) were self-selected by subjects on the Wattbike ergometer throughout each testing session. The complete protocol of the 2 × 4MMP test is as follows:

Warm-up 1: 6 minutes at 60% HR_{max}, 6 minutes at 80% HR_{max}, 3 minutes at 90% HR_{max}, 2- minute stationary recovery, 2 minutes at 70% peak power output, 1-minute setup for test

T1: 4-minute maximal effort

Recovery: 30-minute seated recovery

Warm-up 2: 6 minutes at 80% HR_{max}, 3 minutes at 90% HR_{max}, 2-minute stationary recovery, 1-minute setup for test

T2: 4-minute maximal effort

Statistical Analysis

All data were log-transformed and analyzed using an Excel spreadsheet for reliability.[11] TEM was expressed in both absolute terms and as a CV% along with upper and lower 95% confidence interval (CI). An individual's CV was calculated as the SD of an individual's repeated measurement expressed as a percentage of his individual mean test score.[12] The intraclass correlation between trials was also calculated in combination with the 95% CI.[12]

Results

The mean power output (± SD) for T1 over the 6 weeks was 396 ± 35 W, compared with 393 ± 36 W for T2 (Table 1), resulting in a mean drop-off in power output of 3 W in the repeated bout (T2). The combined mean power output (T1 and T2) increased by 13.4% from week 1 to week 6 (373 ± 32 to 423 ± 38 W), indicating a clear training and/or learning effect.

Intraday Reliability

The mean TEM of T1 to T2 over the 6 weeks was 10.0 W, which equated to 2.6% when expressed as a CV% (Table 2). The lowest TEM and CV occurred during week 6, where the reliability between T1 and T2 was just 5.9 W and 1.5%, respectively. In contrast, the highest TEM and CV occurred in week 4, where the reliability between T1 and T2 was 12.2 W and 3.2%, respectively.

Interday Reliability

The mean TEM (W) and CV% for T1 over the 6 testing days were 10.4 W and 2.7%, respectively, and 11.7 W and 3.0% for T2. When evaluating the best performance (either T1 or T2) over the 6 testing days, the mean TEM was 9.8 W and CV was 2.5%. The combined mean (T1 and T2) over the 6 testing days resulted in a TEM of 8.8 W and a CV of 2.3% (Table 3).

Discussion

The current study is the first to determine the reproducibility of a repeated-bout performance test in highly trained cyclists on the Wattbike cycle ergometer. The primary findings from this investigation suggest that using a protocol consisting of two 4-minute exercise bouts on the Wattbike ergometer separated by ~30 minutes recovery results in reproducible mean power output in highly trained cyclists. Both the intraday and interday reliability of this test in highly trained cyclists were associated with a low TEM (when expressed as absolute watts and CV%) and a high within-subject intraclass correlation All measures of reliability (both intraday and interday) displayed a CV <10% and ICC >.8, which are commonly reported reliability criteria in sports-science research.[13] Even though these numbers are commonly used as minimum reliability criteria in athletic testing, they are not necessarily appropriate. Indeed, CVs as small as 0.9% have been previously reported when measuring peak power output in trained cyclists,[4] and changes of this magnitude may be required to assess changes in athletes' training status.

The interday TEM for T1 was 10.4 W (CV 2.7%), 11.7 W (CV 3.0%) for T2, and 8.8 W (CV 2.3%) for the average of T1 and T2. These findings are in line with

Table 1 Power Output (W) During the 4-min Performance Tests Over 6 weeks of Testing, Mean ± SD

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Average
T1	371 ± 34	383 ± 30	386 ± 33	400 ± 41	407 ± 31	428 ± 37	396 ± 35
T2	376 ± 31	385 ± 34	386 ± 28	393 ± 40	402 ± 38	417 ± 39	393 ± 36
Best (T1 and T2)	380 ± 31	390 ± 34	391 ± 33	403 ± 40	409 ± 32	428 ± 37	400 ± 35
Average (T1 and T2)	373 ± 32	384 ± 31	386 ± 30	396 ± 40	404 ± 34	423 ± 38	394 ± 34
T2 – T1	6 ± 15	2 ± 6	0 ± 16	–6 ± 17	–5 ± 13	–11 ± 8	–2 ± 14

Abbreviations: T1, first 4-min bout; T2, second 4-min bout.

Table 2 Intraday Reliability for T1 to T2 Over the 6 Weeks, Mean (95% Confidence Interval)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Average
ICC (T1 to T2)	.91 (.72–.97)	.90 (.68–.97)	.91 (.71–.97)	.92 (.76–.98)	.93 (.78–.98)	.98 (.93–.99)	.93 (.90–.95)
TEM, W (T1 to T2)	10.3 (7.3–17.5)	11.3 (8.0–19.2)	11.0 (7.8–18.6)	12.2 (8.6–20.7)	9.4 (6.7–15.9)	5.9 (4.2–10.0)	10.0 (8.2–11.8)
CV, % (T1 to T2)	2.9 (2.0–4.9)	2.9 (2.0–5.0)	2.7 (1.9–4.6)	3.2 (2.2–5.5)	2.5 (1.8–4.3)	1.5 (1.0–2.5)	2.6 (2.1–3.1)

Abbreviations: ICC, within-subject intraclass correlation , T1, first 4-min bout; T2, second 4-min bout; TEM, absolute typical error of measurement; CV, typical error as a coefficient of variation.

Table 3 Interday Reliability Over the 6 Weeks, Mean (95% Confidence Interval)

	T1	T2	Best of T1 and T2	Average of T1 and T2
ICC				
2–1	.95 (.85–.99)	.92 (.76–.98)	.97 (.91–.99)	.98 (.92–.99)
3–2	.91 (.72–.97)	.87 (.60–.96)	.92 (.74–.98)	.93 (.78–.98)
4–3	.91 (.73–.97)	.84 (.55–.95)	.91 (.71–.97)	.91 (.71–.97)
5–4	.90 (.69–.97)	.97 (.89–.99)	.92 (.75–.98)	.95 (.84–.99)
6–5	.92 (.75–.98)	.92 (.75–.98)	.94 (.81–.98)	.95 (.84–.99)
mean	.92 (.83–.97)	.91 (.81–.97)	.93 (.85–.98)	.94 (.88–.98)
TEM				
2–1	7.3 (5.2–12.4)	9.6 (6.8–16.3)	6.3 (4.4–10.6)	5.2 (3.7–8.8)
3–2	10.5 (7.4–17.8)	13.1 (9.3–22.3)	10.5 (7.4–17.8)	8.8 (6.2–14.9)
4–3	11.5 (8.2–19.6)	15.2 (10.8–25.9)	12.0 (8.5–20.3)	11.7 (8.3–19.9)
5–4	11.6 (8.2–19.7)	7.9 (5.6–13.5)	10.3 (7.3–17.4)	8.3 (5.9–14.1)
6–5	10.4 (7.3–17.6)	11.2 (8.0–19.1)	9.1 (6.4–15.4)	8.9 (6.3–15.0)
mean	10.4 (8.7–13.3)	11.7 (9.8–15.1)	9.8 (8.2–12.6)	8.8 (7.3–11.3)
CV				
2–1	2.0 (1.4–3.4)	2.6 (1.8–4.4)	1.5 (1.1–2.6)	1.4 (1.0–2.4)
3–2	2.6 (1.9–4.5)	3.2 (2.3–5.6)	2.7 (1.9–4.6)	2.3 (1.6–3.9)
4–3	3.1 (2.2–5.3)	3.9 (2.7–6.7)	3.1 (2.2–5.3)	3.0 (2.1–5.2)
5–4	3.2 (2.2–5.4)	2.0 (1.4–3.5)	2.8 (2.0–4.8)	2.3 (1.6–3.9)
6–5	2.5 (1.8–4.3)	2.9 (2.0–4.9)	2.2 (1.6–3.8)	2.1 (1.5–3.7)
mean	2.7 (2.3–3.5)	3.0 (2.5–3.9)	2.5 (2.1–3.3)	2.3 (1.9–2.9)

Abbreviations: T1, first 4-min bout; T2, second 4-min bout; ICC, within-subject intraclass correlation; TEM, absolute typical error of measurement; CV, typical error as a coefficient of variation.

previous research in which a meta-analysis reported CVs in constant-duration cycling tests (1–60 min) of 2.4% to 5.4%.[14] Zavorsky et al[5] reported a slightly higher average CV of 3.6% over 3 repeated 20-km time trials performed on a Velotron cycle ergometer in 16 recreational to trained cyclists. In that study, the researchers retrospectively divided their results into the top 8 and bottom 10 performers (based on power output) and reported CVs of 2.5% and 4.5%, respectively. These results would suggest that the more highly trained the cyclist, the lower the chance of random error or test–retest variation. Given the highly trained status of the athletes used in the current study and the relatively low CV, our findings would support this suggestion.

Meeusen et al[2] reported a decrement in performance (in a similar 2-bout exercise protocol) of 6% in an overreached athletic population and 11% in clinically diagnosed overtrained athletes. Therefore, as the intraday precision (ie, the difference between T1 and T2) of the protocol used in the current study was ~10 W with a CV of ~2.6%, it may be an appropriate method for indicating signs of overreaching and/or overtraining in an athletic population. Furthermore, in addition to detecting signs of chronic fatigue, this test protocol may have the potential to assess acute fatigue and/or recovery interventions, although future research is required.

A limitation of using stand-alone cycle ergometers is that they cannot be set up exactly the same as the subject's own bicycle, (eg, same components, gearing, angles, and dimensions). Indeed, this has been suggested as a critical factor for producing reliable results,[6] and ergometers that allow attachment of an individual's own bicycle have been shown to produce more reliable results than stand-alone ergometers.[4] Geometry and lower-limb kinematics that most closely replicate a cyclist's position on his or her own bike are associated with improved economy,[15] and for that reason an ergometer that closely reflects the feel of cycling may provide a superior means of assessing exercise performance when compared with ergometers that allow minimal adjustments. However, the reliability of the Wattbike ergometer in the current study is highly comparable with studies performed on a subject's own bicycle attached to an ergometer.[4,16] This is likely due to the high level of adjustability of the Wattbike ergometer when compared with other stand-alone ergometers.[7]

Practical Applications

A 2 × 4MMP test on the Wattbike cycle ergometer is highly reproducible in trained cyclists. The typical-error (and other) data presented enable testers to determine if longitudinal changes across time using the 2 × 4MMP test on the Wattbike are "true" improvements or reductions in performance variables For use in future investigations, a minimum of 15 highly trained cyclists would be required to detect the smallest worthwhile or harmful change in the intraday difference between T1 and T2. Alternatively, only 12 highly trained cyclists would be required to detect a small change in the interday average of T1 and T2. These calculations to determine subject numbers are based on 0.2 of the within-subject change and the between-subjects standard deviation.[17,18] Given the typically low error of measurement across all trials, it appears that only a single familiarization trial is required for this test protocol when using highly trained cyclists.

Conclusion

The current study is the first to show that by using a well-controlled, practical testing protocol that includes two 4-minute time trials on a Wattbike cycle ergometer, separated by ~30 minutes, it is possible to detect small meaningful changes in performance in highly trained cyclists. Although performing the test on the subjects' own bicycles might further improve the reliability and lower the TEM, the Wattbike cycle ergometer appears to be highly reliable when it comes to stand-alone cycle ergometers and may provide an appropriate and more readily available alternative. The low TEMs found in the current protocol suggest that this test can help scientists and coaches better understand factors that may influence cycling performance.

Acknowledgments

We would like to thank the cyclists for volunteering their time to complete this study. The study was funded by the Australian Institute of Sport. There were no conflicts of interest relevant to this manuscript.

References

1. Lamberts RP. *The Development of an Evidenced-Based Submaximal Cycle Test Designed to Monitor and Predict Cycling Performance. The Lamberts and Lambert Submaximal Cycle Test (LSCT)*. Cape Town, South Africa: University of Cape Town; 2009.
2. Meeusen R, Nederhof E, Buyse L, et al. Diagnosing overtraining in athletes using the two-bout exercise protocol. *Br J Sports Med*. 2010;44:642–648. PubMed doi:10.1136/bjsm.2008.049981
3. Driller M. The reliability of a 30-minute performance test on a Lode cycle ergometer. *J Sci Cycling*. 2012;1:21–27.
4. Lamberts RP, Swart J, Woolrich RW, et al. Measurement error associated with performance testing in well-trained cyclists; application to the precision of monitoring changes in training status. *Int J Sports Med*. 2009;10:33–44.
5. Zavorsky GS, Murias JM, Gow J, et al. Laboratory 20-km cycle time trial reproducibility. *Int J Sports Med*. 2007;28:743–748. PubMed doi:10.1055/s-2007-964969
6. Paton CD, Hopkins WG. Tests of cycling performance. *Sports Med*. 2001;31:489–496. PubMed doi:10.2165/00007256-200131070-00004
7. Driller MW, Argus CK, Shing CM. The reliability of a 30-s sprint test on the Wattbike cycle ergometer. *Int J Sports Physiol Perform*. 2013;8(4):379–383.
8. Hopker J, Myers S, Jobson SA, Bruce W, Passfield L. Validity and reliability of the Wattbike cycle ergometer. *Int J Sports Med*. 2010;31(10):731–736. PubMed doi:10.1055/s-0030-1261968
9. Lamberts RP, Lambert MI, Swart J, Noakes TD. Allometric scaling of peak power output accurately predicts time trial performance and maximal oxygen consumption in trained cyclists. *Br J Sports Med*. 2012;46(1):36–41. PubMed doi:10.1136/bjsm.2010.083071
10. Gardner AS, Stephens S, Martin DT, et al. Accuracy of SRM and power tap power monitoring systems for bicycling. *Med Sci Sports Exerc*. 2004;36:1252–1258. PubMed doi:10.1249/01.MSS.0000132380.21785.03
11. Hopkins WG. A new view of statistics. *Sportscience*. 1997.
12. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30:1–15. PubMed doi:10.2165/00007256-200030010-00001
13. Atkinson G, Nevill AM, Edwards B. What is an acceptable amount of measurement error?: the application of meaningful "analytical goals" to the reliability of sports science measurements made on a ratio scale. *J Sports Sci*. 1999;17:18.
14. Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. *Sports Med*. 2001;31:211–234. PubMed doi:10.2165/00007256-200131030-00005
15. Heil DP, Derrick TR, Whittlesey S. The relationship between preferred and optimal positioning during submaximal cycle ergometry. *Eur J Appl Physiol Occup Physiol*. 1997;75:160–165. PubMed doi:10.1007/s004210050141
16. Palmer GS, Dennis SC, Noakes TD, et al. Assessment of the reproducibility of performance testing on an air-braked cycle ergometer. *Int J Sports Med*. 1996;17:293–298. PubMed doi:10.1055/s-2007-972849
17. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform*. 2006;1:50–57. PubMed
18. Hopkins WG. Estimating sample size for magnitude-based inferences. *Sportscience*. 2006;10:63–70.