OPEN ACCESS

Journal of Science & Cycling
Breakthroughs in Cycling & Triathlon Sciences

Original Article

The test-retest reliability of a 16.1 km time trial in trained cyclists using the Wattbike Pro ergometer in hot environmental conditions

Ben Dobson [1,2], Daniel J. Tinnion [3], Lewis A. Gough [4], Lars R. McNaughton [1] and S. Andy Sparks [5,6]

Received: 24 April 2024 **Accepted:** 19 March 2025 **Published:** 11 July 2025

[1] Sports Performance, Exercise and Nutrition Research Group, Department of Sport and Physical Activity, Edge Hill University, Ormskirk, UK

[2] Department of Sport, Hopwood Hall University Centre, Manchester, UK.

[3] Department of Life Sciences, Manchester Metropolitan University, Manchester, UK

[4] Human Performance and Health Research Group, Birmingham City University, Birmingham, UK

[5] Maurten AB, Gothenburg, Sweden.

[6] Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK

Correspondence

Ben Dobson

Sports Performance, Exercise and Nutrition Research Group, Department of Sport and Physical Activity, Edge Hill University, Ormskirk, UK

bendobsonresearch@outlook.com

Abstract

The Wattbike Pro ergometer (Wattbike) is readily available and widely used by athletes, coaches, and researchers as a tool for cycling performance assessment. To-date, no literature has reported the test-reliability of relevant performance criterion using the Wattbike and a 10-mile (16.1 km) TT - which is the most prevalent race distance, often completed in the summer race season. Therefore, the aim of this study was to assess the reliability of 16.1 km TT performance in the heat using the Wattbike Pro ergometer. A cohort of trained cyclists volunteered to take part in this study (n = 16, mean ± SD age 36.4 ± 14.0 y, height, 1.77 ± 0.09 m, body mass 75.2 ± 7.3 kg, PPO 365.1 ± 55.2 W, $\dot{V}O_{2max}$ 55.0 ± 9.5 mL.kg^{-1}.min^{-1}. Participants performed a familiarisation, prior to two 16.1 km TT on the Wattbike Pro ergometer separated by 3-7 days. Differences in mean completion time, power output, and speed were determined using paired samples T-tests, with quartile data assessed using repeated-measures ANOVA. Reproducibility of the performance measures was performed using the coefficient of variation (CV), intraclass correlations, technical error (rTE and sTE) and, Cronbach's α. There were no significant differences between TT1 and TT2 for time, power output and speed (mean difference = 3.25 s, 3.2 W, and 0.15 km·h^{-1}, respectively). All performance data demonstrated excellent reproducibility (CV range = 0.8 – 1.9%) with trivial sTE (0.16 – 0.20). The 16.1 km cycling TT when conducted on a Wattbike Pro ergometer demonstrates a very reliable performance criteria in cohorts of trained cyclists, when exercising in hot conditions. Athletes, coaches, and researchers alike, should be aware of the inter-bike reliability which has been previously reported, and ensure that the same ergometer is used when measuring performance, thereby ensuring the reliability of the 16.1 km TT.

Keywords

Cycling time trial; Coefficient of variation; Endurance exercise; Reproducibility; Temperature

1 Introduction

Competitive cycling performance has been extensively studied in controlled laboratory settings to determine the effectiveness of training, nutritional and/or behavioural interventions (Capostagno, Lambert and Lamberts, 2016). As such, an array of experimental methods and protocols have been used within the literature, such as time-to-exhaustion (TTE) and time trial (TT) and total work done (TWD) protocols. A traditional time-to-exhaustion protocol requires the participant to exercise at a fixed workload (i.e., percentage of VO_{2max}, W_{max} or LT) until they can no longer maintain the necessary power output or speed. In contrast, TT protocols have a clearly defined end point (i.e., distance, duration or energy expenditure) with a self-selected pacing strategy employed to complete the task in the quickest time possible (Laursen et al., 2007). Cycling TT's have been demonstrated to be more reliable than TTE and total work done (TWD) tests, such as the $CCT_{110\%}$ (CV = 4.43% and 4.94%, respectively, and fixed-intensity exercise (CV = 26.6%) (Jeukundrup et al., 1996). Interestingly, the same authors reported a CV of 3.35% when a TT protocol was implemented requiring a preset amount of work to be done, suggesting both a defined end point and ability to modulate effort may be important. Overall, TT's are considered to be more reliable, and ecological validity is greatly enhanced since the workload is dictated by the athlete, allowing for behavioural change in response to an intervention, environmental alterations, and/or fatigue (Tucker, 2009).

Crucially, a performance criterion should be used which best represents actual competitive cycling performance (Sparks et al., 2016). During competition, athletes regularly contend with environmental temperatures ~30°C (Valenzuela et al., 2022). For example, during the 2023 Tour de France the average temperature recorded was 25.7 ± 3.7°C with a high of 32.1°C on stage 7 from Mont-de-Marsan to Bordeaux and a low of 21.1°C on stage 6 Tarbes to Cauterets-Cambasque (Pro Cycling Stats, 2023). The increased temperature during the competitive season represents a greater physiological challenge for an athlete to overcome, compared to a thermoneutral environment (Costa et al., 2019). It is therefore essential that the sensitivity of the test criteria is known, so that a meaningful change can be observed once outside of the expected measurement error range. In the case of well-trained cyclists, it is proposed that a performance change of ~1% is likely to be meaningful, with a technical error of measurement (TEM) during a VO_{2max}, peak power output (PPO) test or 40 km TT using the Computrainer ergometer of 2.2%, 0.9% and 0.9%, respectively (Lamberts et al., 2009).

Originally developed in partnership with British Cycling in 2008, the Wattbike cycle ergometer (Wattbike Pro, Wattbike Ltd, Nottingham, UK) has received increased attention in the cycling literature, predominantly due to its high prevalence in home and commercial gyms, and it's measurement of power by assessing chain tension over a load cell (100 Hz sampling frequency), angular velocity of the crank arms (twice per revolution) (Driller et al., 2014) with a claimed accuracy of ± 2% between 0 – 3760 W (Hopker et al., 2010). Limited literature has reported the reliability of the Wattbike, especially those using an ecologically valid distance. Of these few studies, Hopker et al. (2010) utilised both human participants and a calibration rig to determine the reliability of the Wattbike Pro ergometer against a previously validated crank-based power meter (Science model, SRM, Jülich, Germany). During human constant power trials, a CV of 1.7 % between 50 – 700 W at 90 rev·min^{-1}, and 1.4 % between 100 and 1250 W at 70 rev·min^{-1} was observed. It must be noted that in order to obtain data from an external power source (SRM), the replacement of the manufacturer fitted bottom bracket and crankset is required. Despite performing a 'zero offset', matching crank length (170 mm) and chainring size (48 teeth), the manufacturer (Wattbike) claimed that the altering of the bikes factory calibrated setup invalidated the ability of the Wattbike to accurately measure power (Wainwright et al., 2017). Regardless, both authors agreed that the Wattbike Pro demonstrates high validity and reliability during constant power trials between 50 – 600 W (CV = 1.5 and 1.7 %), across a range of power outputs (CV = 2.6%), but it is unclear if this is the case for a TT.

In respect of an ecologically valid distance, Sparks et al. (2016) investigated the test-retest reliability of a 16.1 km TT using the CompuTrainer wheel-on ergometer in a trained cyclist population (VO_{2max} 56.6 ± 6.6 mL·kg^{-1}·min^{-1}, PPO 365 ± 37 W). Reporting no differences between TT1 and TT2 for any of the performance criteria (Mean difference (MD) = 0.06 min, 0.09 km·h^{-1} and 1.5 W, for time, mean speed and power respectively), with all performance data being highly reproducible (CV = 1.1 – 2.7%) with trivial/small technical error. When split into cohorts of 'slower' and 'faster' athletes, CV ranged from 1.3 – 3.2% and 0.7 – 2.0%, respectively, indicative of a potentially greater ability to regulate and determine effort in the latter group. Whilst this evidence suggests a 16.1 km TT is a reliable protocol, it is unclear whether the environmental temperature and/or a trained cyclist population, or equipment (Computrainer vs Wattbike) would alter the reliability of a 16.1 km TT protocol. Further research is therefore required to investigate the reliability of the Wattbike Pro ergometer during self-paced exercise performance, as this may be subject to greater variability than previously reported using other ergometers. Currently no studies that have reported the test-retest reliability of the 16.1 km TT using the Wattbike Pro ergometer. Despite the high level of ecological validity as a performance criterion, it is unclear how the distance, environmental temperature and/or a trained cyclist population will alter the test-retest reliability of a 16.1 km TT protocol. Therefore, the main aim of this study is to determine the test-retest reliability of the 16.1 km TT on a Wattbike Pro Ergometer in hot conditions, within a cohort of trained cyclists.

2 Material and Methods

2.1 Participants

Sixteen trained cyclists, consisting of 15 males and 1 female, volunteered to participate in this study. The participants were of mean (±SD) age 36.4 ± 14.0 y, height, 1.77 ± 0.09 m, body mass 75.2 ± 7.3 kg, PPO 365.1 ± 55.2 W, Watts per kg BM 5.0 ± 1.0 W·kg^{-1} and relative VO_{2max} 55.0 ± 9.5 mL·kg^{-1}·min^{-1}. All participants were familiar with high intensity cycling TT's on the road and/or in a laboratory and met the inclusion criteria of a weekly training frequency of ≥ 3 times, for a total of ≥ 5 h·week^{-1} with a minimum of two years training experience specifically in cycling (McKay et al., 2021). Participants provided written informed consent prior to undergoing a mandatory health-screening before any experimental trials were conducted. Ethical approval was obtained by the institutional research ethics committee and all procedures were conducted in accordance with the Declaration of Helsinki (2013). Participants were verbally screened to ensure they were not currently ingesting any exogenous buffering agents that may interfere with the outcome of the study in compliance with the pre-experimental controls. Participants were asked to maintain nutritional intake as close to normal as possible throughout the duration of all visits. For the 24 h preceding each visit, participants were instructed to consume the same food and fluids, a process that was verified via the completion of a 24 h food diary. The ingestion of alcohol, additional dietary antioxidants, polyphenols or participating in any form of exercise was prohibited, for the 12 h before attending the laboratory, spicy foods and caffeinated beverages were also prohibited (Westerterp-Plantenga et al., 2006).

2.2 Study Design

Participants attended the laboratory on three occasions, separated by at least 48 h to reduce the risk of fatigue impairing maximal performance. The first visit was used to determine maximal oxygen consumption (VO_{2max}) and body composition, following which participants familiarised themselves with the cycle ergometer, 16.1 km TT protocol and the exercise environment. The remaining two visits required the participants to perform a 16.1 km TT on two separate occasions within a simulated environment chamber (TISS, UK) in hot (28.5 ± 0.6°C) normoxic conditions (20.93%) with a relative humidity of 40 ± 3.2%. The Wattbike Pro cycle ergometer was kept within the simulated environment at all times, and during each trial the environmental chamber was allowed to stabilise at the desired conditions for 1 h prior to any warm-up or TT taking place. Participants were encouraged to ingest at least 1500 ml of water in the 2 h preceding each visit to achieve euhydration. This was subsequently confirmed via the measurement of urine osmolality (UOsm) with ≥700 $mOsmol \cdot kgH_2O^{-1}$ being defined as dehydrated under the guidelines set by The American College of Sports Medicine (Hew-Butler et al., 2018). Each laboratory visit was arranged to occur at the same time of day (± 1 h) to limit the physiological effects of circadian variations (Atkinson and Reilly, 1996).

2.2.1 Maximal Oxygen Consumption Test

Participants completed a graded exercise test (GXT) to volitional exhaustion to determine VO_{2max} on an electromagnetically braked cycle ergometer (Excalibur XL, Lode, Netherlands) as previously described by Gough et al., (2017). Breath-by-breath respiratory gases were continuously analysed using a portable gas analyser (K5, Cosmed, Italy) to measure oxygen uptake (VO_2), carbon dioxide production (VCO_2) and respiratory exchange ratio (RER). Volitional exhaustion was determined to be the inability of the participant to sustain their respective self-selected cadence for more than 5 s, despite strong verbal encouragement. Further criteria for the attainment of VO_{2max} was implemented, specifically, participants reached within 10 $b \cdot min^{-1}$ of age predicted maximal HR (220 $b \cdot min^{-1}$ – age), a blood lactate concentration of ≥8 $mmol \cdot L^{-1}$, a respiratory exchange ratio (RER) ≥1.15, and a rating of perceived exertion (RPE) ≥18 at the end of exercise (Midgley et al., 2007). During analysis, data was averaged into 10 s intervals, with VO_{2max} defined as the highest rolling 30 s average of VO_2 obtained during the test.

2.2.2 Cycling Time Trials

During the first experimental trial, participants were instructed to conduct a self-selected warm-up that was a minimum of 5 min but no longer than 10 min. This was recorded, and subsequently replicated in the following visits. All experimental trials were completed on a Wattbike Pro Cycle Ergometer (Wattbike, Nottingham, UK), despite good (<2%) inter-bike agreement between a sample of 10 Wattbikes (Wainwright et al., 2017) the same ergometer was used for all trials to avoid the possible interaction of inter-bike variation. Handlebar and saddle height was adjusted to replicate each participant's normal racing position, this was also recorded and subsequently replicated for the following visit. A 60 cm fan (CAM5002, Clarke, Essex, UK) was placed on the front right of the ergometer, 2 m away from the participant's side. During the familiarisation participants adjusted the fan speed, this was then subsequently recorded and reproduced during the experimental trials. Prior to the start of each TT, participants were instructed to complete the 16.1 km distance in the quickest time possible. Following this, no verbal encouragement was offered during the trials and all performance data, apart from the current distance completed, was blinded from the participants. Time, speed, and power output data were recorded using the Wattbike HUB software, at a frequency of 100 Hz, then stored, and later downloaded and exported into Excel 2021 (Microsoft Corporation, Redmond, Washington, USA) for further analysis.

2.3 Statistical Analysis

All data was assessed for normality using standard graphical methods prior to analyses (Grafen and Hails., 2002). Thereafter, differences in overall performance data (mean speed, power, and time to complete) were compared between TT1 and TT2 using paired t-tests. When assessing across quartiles repeated measures analysis of variance (ANOVA), followed by post-hoc comparisons with Bonferroni pair-wise comparisons. Effect sizes were calculated using Hedges g to account for the small sample size ($n \leq 20$) and interpreted as ≤0.2 = small effect, 0.5 = medium effect and ≥0.8 = large effect. Reliability was assessed using Cronbach's alpha (α) and was interpreted as < 0.5 = Unacceptable, 0.5 to 0.6 = Poor, 0.6 to 0.7 = Acceptable, 0.7 to 0.9 = Good, and ≥ 0.9 Excellent. Raw technical error (rTE) was calculated using the equation (SDdiff/√2), where SDdiff represents the SD of the differences between TT1 and TT2. Standardised TE (sTE) is reported and calculated using the method of Hopkins (2000) and interpreted using a modified Cohen scale where < 0.2 = trivial, 0.2–0.6 = small, 0.6–1.2 = moderate, 1.2 – 2.0 = large and > 2.0 = very large error (Sparks et al., 2016).

Coefficients of variation (CV) were calculated to allow simple comparison between the present study and existing literature to further interpret the reliability of the ergometer in question. The CV was expressed as a percentage using the equation: CV(%) = (SD/mean)x100. The relationships of variables between TT1 and TT2 were analysed using intraclass correlation coefficients (ICC) presented with CI's as proposed by Atkinson and Nevill (1998). All descriptive data are presented as Mean ± SD, unless stated otherwise.

Statistical significance was set at $p \leq 0.05$, exact p values are given in the text and tables, for p values of '0.000' given by the statistical package values were corrected to '< 0.001'. All statistical data were analysed using Statistical Package for the Social Sciences Version 29.0 (SPSS, IBM, Chicago, IL, USA) and calculations were completed with Excel 2021 (Microsoft Corp., Redmond, Washington, USA).

3 Results

There were no significant differences between trials in respect of total time (Figure 1) to complete the 16.1 km TT distance (MD = 3.25 s (0.22%), t = 0.428, $p = 0.674$, $g = 0.102$), alongside no alterations in mean power output (MD = 3.2 W (1.39%), t = 1.263, $p = 0.226$, $g = 0.300$) and speed (MD = 0.15 km·h^{-1} (0.39%), t = 0.862, $p = 0.402$, $g = 0.205$). Total time to complete demonstrated excellent reproducibility (Table 1) between TT1 and TT2 ($r = 0.97$, $p < 0.001$), with a trivial sTE (0.2). Mean power displayed excellent reproducibility ($r = 0.98$, $p < 0.001$, sTE = 0.16).

During the TT, a significant interaction effect was observed for mean power output (TT*Distance) across quartiles (Figure 2) between TT1 and TT2 ($f(3, 45) = 7.3$, $p < 0.001$, $p\eta^2 = 0.3$). Pairwise comparisons revealed pacing alterations between TT1 and TT2, with increased power output in TT1 at 0 to 4 km (MD = 13.6 W (5.56%), t = 2.841, $p = 0.012$, g = 0.674) and 4 to 8 km (MD = 7.46 W (3.21%), t = 2.589, $p = 0.021$, g = 0.614). No differences between TT1 and TT2 were observed at 8 to 12 km (MD = 0.49 W (0.22%), t = 0.146, $p = 0.886$, g = 0.035) and 12 km to 16.1 km (MD = 7.72 W (3.33%), t = 1.785, $p = 0.094$, g = 0.424). Despite the increased variance, and significant differences identified between 0-4 and 4-8 km, mean power output demonstrated excellent test-retest reliability between TT1 and TT2 (range, CV = 2.2 – 3.6%, $p < 0.001$) with a high level of consistency ($\alpha = 0.96 – 0.98$) (Table 2).

Additionally, mean speed also demonstrated excellent reproducibility (r = 0.97, $p < 0.001$, sTE = 0.19). These data represent excellent reproducibility of 16.1 km TT for all measured performance parameters with a low CV (0.8 – 1.9%) using the Wattbike Pro Ergometer.

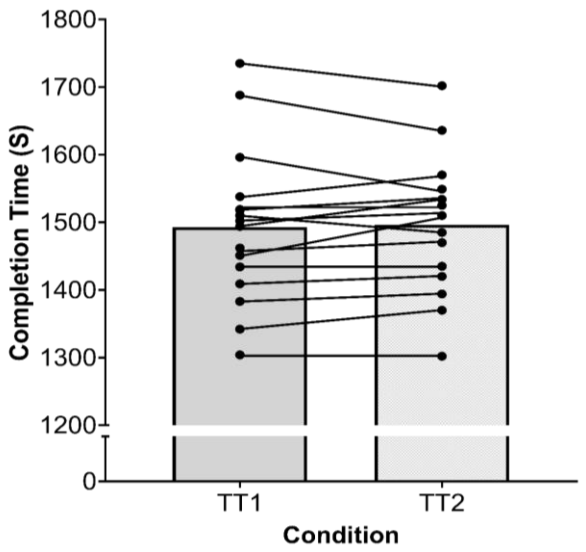

Figure 1. Mean individual performance time to complete the 16.1 km time trial (TT) protocol.

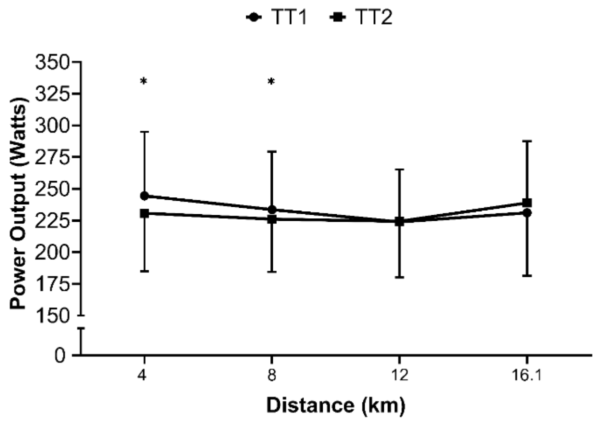

Figure 2. Mean (±SD) power output profile across the 16.1 km time trial (TT) distance. * Indicates a significant difference between TT1 and TT2 at 0 to 4 km ($p = 0.012$) and 4 to 8 km ($p = 0.021$).

Table 1. Reliability of performance variables between time trial 1 (TT1) and time trial 2 *(TT2)*. Data represents raw technical error (rTE), standardised technical error (sTE), coefficient of variation (CV), intraclass correlation coefficient (ICC) and confidence intervals (CI).

	Total Time (s)	**Mean Power (W)**	**Mean Speed (km·h^{-1})**
TT1	1492.8 ± 114.0	233.2 ± 45.9	38.9 ± 2.9
TT2	1496.1 ± 99.3	230.0 ± 43.0	38.8 ± 2.6
TT1-2 Δ	3.3 ± 30.3	3.2 ± 10.1	0.15 ± 0.72
p	0.674	0.226	0.402
g	0.102	0.300	0.205
rTE	21.5	7.17	0.51
sTE	0.20	0.16	0.19
Cronbach's α	0.98	0.99	0.99
CV (%)	1.1	1.9	0.8
CI	0.61 – 1.53	1.19 – 2.68	0.50 – 1.08
ICC			
r	0.97	0.98	0.97
p	< 0.001	< 0.001	< 0.001
CI	0.92 – 0.99	0.95 – 0.99	0.93 – 0.99
Interpretation	Excellent	Excellent	Excellent

Table 2. Reliability of power output between time trial 1 (TT1) and time trial 2 (TT2) across quartiles. Data represents raw technical error (rTE), standardised technical error (sTE), coefficient of variation (CV), intraclass correlation coefficient (ICC) and confidence intervals (CI).

	4 km	**8 km**	**12 km**	**16.1 km**
TT1	244.4 ± 50.6	233.6 ± 45.7	223.9 ± 43.6	231.1 ± 49.7
TT2	230.8 ± 45.8	226.1 ± 41.7	224.4 ± 40.8	238.8 ± 48.8
TT1-2 Δ	13.6 ± 19.1	7.5 ± 11.5	0.5 ± 13.4	7.7 ± 17.3
rTE	13.5	8.2	9.5	12.3
sTE	0.29	0.19	0.23	0.26
Cronbach's α	0.96	0.98	0.98	0.97
CV (%)	3.6	2.2	2.4	3.5
CI	1.75 – 5.41	1.15 – 3.29	1.15 – 3.74	1.84 – 5.23
ICC				
r	0.94	0.98	0.98	0.96
p	< 0.001	< 0.001	< 0.001	< 0.001
CI	0.75 – 0.98	0.90 – 0.99	0.93 – 0.99	0.89 – 0.99
Interpretation	Excellent	Excellent	Excellent	Excellent

4 Discussion

This is the first study to establish that the Wattbike Pro cycle ergometer has excellent time trial test-retest reliability, determined by CV's, ICC, and Cronbach's *α*, for mean performance time, mean speed and mean power output data for 16.1 km TT's completed in a hot environment. The resultant CV of 1.1, 1.9 and 0.8% for completion time, mean power, and speed, respectively, represents an acceptable error to identify whether a true performance effect has been observed. In the context of the trained cyclist cohort assessed in this current study, only trivial mean differences in performance of 3.25 s, 3.2 W and 0.15 km·h^{-1} were observed. These appear trivial in terms of interpreting differences in performance and can be partly attributed to the random and daily biological error of the equipment, protocol (Paton and Hopkins, 2012) and individuals. Biological variation within cyclists (1.2 ± 1.5%) can contribute equally to CV to that of the ergometer's (0.9 – 1.8 %) (Paton and Hopkins, 2005), which is not dissimilar from the CV's observed in the present study. It is important that all efforts to control biological variation are maintained with rigorous pre-experimental test procedures. Most importantly, the reporting of the performance criterion (CV), also allows for the calculation of the smallest worthwhile change for completion time, power output and speed. This value takes into account both biological and equipment error and provides a value range that is necessary to ascertain if a faster or slower performance has been observed.

One critical strength of this study is the accessibility of the ergometer in question, typically, laboratory-grade equipment has been reserved for the research or national governing bodies, however, the Wattbike Pro ergometer is readily accessible as a home and commercial gym piece of equipment, without the large cost typically associated with laboratory equipment. The CV range demonstrated in the present study (0.8 – 1.9%) is well below the stated 5% threshold reported in a review by Currell and Jeukendrup (2008) and most importantly, the Wattbike demonstrates excellent reproducibility across similar performance criterions to that of its competitors such as the SRM power meter (Abbiss et al., 2009), Computrainer (Sparks et al., 2016) and Velotron (Borg et al., 2018; Borszcz, Tramontin and Costa, 2020). Despite the presence of additional heat strain compared to that which has previously been conducted in thermoneutral laboratory conditions, the present study reported an enhanced reliability of mean power output (CV = 1.9%), when using the Wattbike, compared to the findings of Sparks et al. (2016) which observed an overall CV of 2.7% for cyclists completing a 16.1 km TT using the CompuTrainer ergometer. When sub-groups were identified based on performance time, mean power output reliability was improved (CV = 2.0% and 3.2%) for faster and slower riders, respectively. An explanation for the smaller CV in the present study, is the possible role in simulating TT duration on the ergometer itself. Sparks et al. (2016) reported that overall performance times were considerably slower than might be expected for a trained cohort of cyclists (27.05 ± 1.68 min), while overall performance time in the present study was 24.90 ± 1.78 min. This reduction in completion time appears to better represent the athletes own 'expected' performance duration and might have attributed to the enhanced reliability between TT1 and TT2. Additionally, the present study did not investigate the reliability differences between performance in both hot and cool conditions, and it is currently unclear whether reliability may be further enhanced in conditions of reduced thermal strain, but maintaining warm ambient temperatures may contribute to enhanced calibration stability (Davison et al., 2009). Therefore, future research should also look to investigate the reliability of the Wattbike Pro ergometer in cool conditions, to understand the role thermal strain may play in both reliability but also the physics of the Wattbike Pro's fan resistance settings.

The use of reliability and reproducibility are often overlooked within research settings and in particular, during the monitoring/assessment of training programmes. The latter typically requires athletes to complete a self-administered exercise test, for instance the FTP_{20} or a similar length cycling TT (16.1 km). The use of these protocols does not pose a significant issue in appropriately scaled research projects, whereby statistical analysis can determine whether an intervention has had an overall effect (de Oliveira et al., 2017). However, for the individual athlete statistical analysis is not a suitable method of assessing performance changes. It has previously been proposed that the CV of a test is a suitable method to determine an individual's response to training/intervention (Saunders et al., 2014). In the case of this study, if an individual was to complete the 16.1 km TT in 1434 s, a subsequent re-test with a performance criterion reliability of 1.1%, would result in an improved performance if below 1418.23 s (1434-[1434*0.011]), or if slower, above 1449.77 s (1434+[1434*0.011]). It is strongly advised by the present authors that this strategy is adopted by athletes and coaches, where possible, when monitoring changes in overall performance. It is important to mention that despite the reliability of the performance criterion being known, it is plausible that if the ergometer is accessed via a commercial setting, an athlete or individual is unlikely to be able to access the same ergometer repeatedly to complete a performance trial. Previous research by Wainwright et al., (2016) investigated the test-retest reliability on ten individual Wattbike units at power outputs of 100 W – 1000 W using a motorised calibration rig (Lode Calibrator 2000, Groningen, Netherlands). The resultant reproducibility between ergometers was very high, with absolute differences of 0.6 W at 152 W, up to 25.5 W at 983 W, the largest difference represents a 2.6% change which is not dissimilar to the reported variation in performance observed in the present study (CV = 1.9%). Nevertheless, for performance assessment and monitoring purposes the current authors concur that, where possible, the end-users should maintain the same ergometer throughout all trials to account for the individual manufacturing processes which may possibly have resulted in a small, but perhaps meaningful deviation from factory calibration.

During the TT's, significant differences were observed in pacing between TT1 and TT2, via the comparison of 4 km splits. However, this did not result in any overall change in respect of the mean performance variables discussed within this study (p = 0.226 - 0.674). A similar pacing profile was observed by de Oliveira et al., (2017) during the completion of a work-based cycling TT (~420 kJ) (completion time: 1861 ± 142 and 1885 ± 170 s, p = 0.10, d = 0.15, CV = 3.04 ± 2.25%). The authors reported the CV of mean power output at 25, 50 and 75% of the TT ranged from 2.9 to 5.8%, despite the current study using a distance-based protocol rather than work-done. In the present study CV ranged between 2.2 to 3.6% at 4, 8 and 12 km, respectively. The lower degree of variability may be partly attributed to participants being provided live feedback in relation to the distance covered/remaining via the ergometer screen throughout the entire duration of the TT protocol, which was not provided in the work-done protocol completed by de Oliveira et al., (2017). Although not the aim of this study, it is unknown how a further experimental trial would have subsequently influenced pacing, and if the resultant overall performance would have been altered. Previous research by Smith et al. (2001) observed increased reliability following the completion of a second vs third cycling TT, establishing that reliability can be enhanced as athletes become more familiar with the experimental protocol. Nevertheless, trained cyclists have been shown to be able to successfully reproduce reliable cycling TT's across multiple distances (Paton and Hopkins, 2012). Additionally, a similar duration of exercise (20 km) has been shown to be reproducible in cohorts of recreationally active participants who had no prior experience of a TT, with a small effect size, following one familiarisation and two experimental trials (d < 0.49) (Hibbert et al., 2017).

5 Conclusions and Practical Applications

The 16.1 km TT represents a reliable performance criterion for trained cyclists when conducted using the Wattbike Pro cycle ergometer. Practitioners, researchers, and athletes should be aware that differences in test-retest reliability are present in relation to different parameters, with total time to complete demonstrating the lowest variability (0.22%), followed by speed (0.39%) and lastly, power output (1.39%). Further to this, athletes, coaches, and researchers alike, should be aware of the inter-bike reliability, ensuring where possible that the same ergometer is used when measuring performance, thereby maximising the reliability of the 16.1 km TT.

Funding: This research received no external funding.

Acknowledgments: We thank all the participants who gave their time and effort to participate in this study. The authors are extremely grateful for the technical assistance received during data collection by the University's sport and exercise science technician team, namely Scott Mosher, Lorna Caddick, Sam Amirbashiri and Adam Jones.

Conflicts of Interest: The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article. The authors did not receive support from any organisation for the submitted work. No funding was received to assist with the preparation of this manuscript.

References

1. Capostagno, B., Lambert, M. I., & Lamberts, R. P. (2016). A systematic review of submaximal cycle tests to predict, monitor, and optimize cycling performance. *International Journal of Sports Physiology and Performance*, *11*(6), 707-714.

2. Laursen, P. B., Francis, G. T., Abbiss, C. R., Newton, M. J., & Nosaka, K. (2007). Reliability of time-to-exhaustion versus time-trial running tests in runners. *Medicine and Science in Sports and Exercise*, *39*(8), 1374-1379.

3. Jeukendrup, A. E., Saris, W. H., Brouns, F. J. P. H., & Kester, A. D. (1996). A new validated endurance performance test. *Medicine and Science in Sports and Exercise*, *28*(2), 266-270.

4. Tucker, R. (2009). The anticipatory regulation of performance: the physiological basis for pacing strategies and the development of a perception-based model for exercise performance. *British Journal of Sports Medicine*, *43*(6), 392-400.

5. Sparks, A. S., Williams, E. L., Jones, H. J., Bridge, C. A., Marchant, D., & McNaughton, L. (2016). Test-retest reliability of a 16.1 km time trial in trained cyclists using the CompuTrainer ergometer. *Journal of Science and Cycling*, *5*(3), 35-41.

6. Valenzuela, P. L., Mateo-March, M., Zabala, M., Muriel, X., Lucia, A., Barranco-Gil, D., & Pallarés, J. G. (2022). Ambient temperature and field-based cycling performance: insights from male and female professional cyclists. *International Journal of Sports Physiology and Performance*, *17*(7), 1025-1029.

7. *Hottest stages in Tour de France 2023*. (n.d.). https://www.procyclingstats.com/race/tour-de-france/2023/results/hottest-stages

8. Costa, R. J., Gaskell, S. K., McCubbin, A. J., & Snipe, R. M. (2020). Exertional-heat stress-associated gastrointestinal perturbations during Olympic sports: Management strategies for athletes preparing and competing in the 2020 Tokyo Olympic Games. *Temperature*, 7(1), 58-88.

9. Lamberts, R. P., Swart, J., Woolrich, R. W., Noakes, T. D., & Lambert, M. I. (2009). Measurement error associated with performance testing in well-trained cyclists: Application to the precision of monitoring changes in training status. *International SportMed Journal*, *10*(1), 33-44.

10. Driller, M. W., Argus, C. K., Bartram, J. C., Bonaventura, J., Martin, D. T., West, N. P., & Halson, S. L. (2014). Reliability of a 2-bout exercise test on a wattbike cycle ergometer. *International Journal of Sports Physiology and Performance*, *9*(2), 340-345.

11. Hopker, J., Myers, S., Jobson, S.A., Bruce, W. and Passfield, L., 2010. Validity and reliability of the Wattbike cycle ergometer. *International Journal of Sports Medicine*, pp.731-736.

12. Wainwright, B., Cooke, C. B., & O'Hara, J. P. (2017). The validity and reliability of a sample of 10 Wattbike cycle ergometers. *Journal of Sports Sciences*, *35*(14), 1451-1458.

13. McKay, A. K., Stellingwerff, T., Smith, E. S., Martin, D. T., Mujika, I., Goosey-Tolfrey, V. L., ... & Burke, L. M. (2021). Defining training and performance caliber: a participant classification framework. *International Journal of Sports Physiology and Performance*, *17*(2), 317-331.

14. Westerterp-Plantenga, M., Diepvens, K., Joosen, A.M., Bérubé-Parent, S. and Tremblay, A., 2006. Metabolic effects of spices, teas, and caffeine. *Physiology & Behavior*, *89*(1), pp.85-91.

15. Atkinson, G., & Reilly, T. (1996). Circadian variation in sports performance. *Sports Medicine*, *21*, 292-312.

16. Gough, L. A., Deb, S. K., Sparks, A. S., & McNaughton, L. R. (2017). The reproducibility of blood acid base responses in male collegiate athletes following individualised doses of sodium bicarbonate: a randomised controlled crossover study. *Sports Medicine*, *47*, 2117-2127.

17. Midgley, A. W., McNaughton, L. R., Polman, R., & Marchant, D. (2007). Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. *Sports Medicine*, *37*, 1019-1028.

18. Leach, N. K., Hilton, N. P., Tinnion, D., Dobson, B., McNaughton, L. R., & Sparks, S. A. (2023). Sodium Bicarbonate Ingestion in a Fasted State Improves 16.1 km Cycling Time Trial Performance. *Medicine and Science in Sports and Exercise*.

19. Grafen, A., & Hails, R. (2002). *Modern statistics for the life sciences*. Oxford University Press.

20. Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, *30*, 1-15.

21. Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, *26*, 217-238.

22. Paton, C. D., & Hopkins, W. G. (2006). Variation in performance of elite cyclists from race to race. *European Journal of Sport Science*, *6*(01), 25-31.

23. Abbiss, C. R., Quod, M. J., Levin, G., Martin, D. T., & Laursen, P. B. (2009). Accuracy of the Velotron ergometer and SRM power meter. *International Journal of Sports Medicine*, *30*(02), 107-112.

24. Borg, D. N., Osborne, J. O., Stewart, I. B., Costello, J. T., Sims, J. N., & Minett, G. M. (2018). The reproducibility of 10 and 20 km time trial cycling performance in recreational cyclists, runners and team sport athletes. *Journal of Science and Medicine in Sport*, *21*(8), 858-863.

25. Borszcz, F. K., Tramontin, A. F., & Costa, V. P. (2020). Reliability of the functional threshold power in competitive cyclists. *International Journal of Sports Medicine*, *41*(03), 175-181.

26. Currell, K., & Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine*, *38*, 297-316.

27. de Oliveira, L. F., Yamaguchi, G., de Salles Painelli, V., da Silva, R. P., Gonçalves, L. S., Gualano, B., & Saunders, B. (2017). Comprehensive reliability analysis of a 16 km simulated cycling time-trial in well-trained individuals. *Journal of Science and Cycling*, *6*(1), 11-17.

28. Hibbert, A. W., Billaut, F., Varley, M. C., & Polman, R. C. (2017). Familiarization protocol influences reproducibility of 20-km cycling time-trial performance in novice participants. *Frontiers in Physiology*, *8*, 267190.

29. Saunders, B., Sale, C., Harris, R.C. and Sunderland, C., 2014. Sodium bicarbonate and high-intensity-cycling capacity: variability in responses. *International Journal of Sports Physiology and Performance*, *9*(4), pp.627-632.

30. Paton, C.D. and Hopkins, W.G., 2006. Ergometer error and biological variation in power output in a performance test with three cycle ergometers. *International Journal of Sports Medicine*, *27*(06), pp.444-447.

31. Smith, M.F., Davison, R.C.R., Balmer, J. and Bird, S.R., 2001. Reliability of mean power recorded during indoor and outdoor self-paced 40 km cycling time trials. *International Journal of Sports Medicine*, *22*(04), pp.270-274.

32. Davison RR, Corbett J, Ansley L (2009) Influence of temperature and protocol on the calibration of the Computrainer electromagnetically-braked cycling ergometer: original research article. International SportMed Journal 10: 66-76.