# Reliability of the Functional Threshold Power in Competitive Cyclists

### Authors
**Fernando Klitzke Borszcz[1, 2], Artur Ferreira Tramontin[1], Vitor Pereira Costa[1]**

### Affiliations
1   Center for Health and Sport Sciences, University of Santa Catarina State, Florianópolis, Brazil
2   Sports Centre, Federal University of Santa Catarina, Florianópolis, Brazil

### Correspondence
Dr. Vitor Pereira Costa
Center for Health and Sport Sciences
University of Santa Catarina State
Rua Pascoal Simone 350
88080700 Florianópolis
Brazil
Tel : + 55 048 3664-8600, Fax : + 55 048 3664-8600
vitor.costa@udesc.br

## ABSTRACT

Functional threshold power (FTP) is defined as the highest power that a cyclist can maintain in a quasi-steady state without fatigue for approximately 1 hour. To improve practicality, a 20-minute time-trial test was proposed, where FTP is represented by 95 % of the mean power produced. It is preceded by a specific 45-min warm-up, with periods of low intensity, fast accelerations, and a 5-min time-trial. Thus, the aim of this study was to determine the reliability of this protocol, including the reliability of the warm-up, pacing strategy, and FTP determination. For this purpose, 25 trained cyclists performed a familiarization and two other tests separated by seven days. The coefficient of variation (CV [ %]), intraclass correlation coefficient (ICC), and change in the mean between test and retest were calculated. The results show that the 20-min time-trial was reliable (CV = 2.9 %, ICC = 0.97), despite a less reliable warm-up (CV = 5.5 %, ICC = 0.84). The changes in the mean between the test and retest were trivial to small for all measurements, and the pacing strategy was consistent across all trials. These results suggest that FTP determination with a 20-min protocol was reliable in trained cyclists.

## Introduction

Laboratory testing has traditionally been used to determine intensities to prescribe training and to detect changes in cycling performance [1–3]. Nowadays, research literature suggests that field testing carries a high ecological validity [4], and it is preferred by cyclists who use portable power meters on their bicycles [5]. Therefore, using portable power meters, time-trials (TT) can be used as performance tests in the field. In this regard, the functional threshold power (FTP) is defined as the highest power output (PO) that a cyclist can maintain in a quasi-steady-state without fatigue for approximately 60 min (e.g., a 60-min TT [$FTP_{60}$]) [6]. The FTP is a metric to establish training intensity zones and to detect changes in performance. It is also the key metric in a series of calculations of training load and exercise intensity (i.e., training stress score [TSS], intensity factor [IF]) [6]. Moreover, FTP is used as an indirect and practical estimate of the anaerobic threshold (AnT) and maximal lactate steady state (MLSS) [7–9]. Due to the long test duration (i.e., 60 min), it was suggested that the FTP might be determined as 95 % of the mean PO in a 20-min TT ($FTP_{20}$) [6].

According to Allen and Coggan [6], the $FTP_{20}$ test is divided into three different parts (i.e., warm-up, main part, and cool down). The first step is the specific 45-min warm-up, which includes a maximum effort 5-min TT [6]. To the best of our knowledge, Borszcz et al. [7, 8] and Valenzuela et al. [10] are the only studies that have used the warm-up recommended by the authors [6]. Others have not used this section of the protocol [11–13]. Borszcz et al. [7] reported high blood lactate concentrations ([La]) before the start of the test ($6.5 \pm 2.9$ mmol $\cdot$ L$^{-1}$). Therefore, the $FTP_{20}$ warm-up protocol is not in accordance with the recent warm-up guidelines, which suggested a duration of about 20 min and a [La] of < 3 mmol $\cdot$ L$^{-1}$ at the end [14]. Additionally, cyclists adopted a conservative pacing strategy at the start of a TT in a recent research [7], despite the

fast start strategy presented in several TTs with durations that range from 4 to 60 min [7, 12, 13, 15–19].

FTP is one of the most well-known and debated tests in cycling and in the triathlon community. A recent concern in sports sciences is the applicability of the tests developed for athletes in the "real world" [20]. Recent studies [7–9] have investigated the validity of the $FTP_{20}$ test with AnT, MLSS, and $FTP_{60}$. Overall, trivial differences (bias) were found between $FTP_{20}$ (original warm-up) with $FTP_{60}$, AnT, and MLSS. These metrics were moderately to almost perfectly correlated (r = 0.61 to 0.91), despite the individual variability in the prediction (limits of agreement [LoA] of 95 % = 7.8–27.5 %) [7–9]. However, the tests should be reliable to detect the smallest worthwhile physiological or performance changes; reliability is a prerequisite for test validity because a measure that is not reproducible is not considered valid [2, 3, 21]. Thus, the aims of the current study were to verify the reliability of the protocol proposed by Allen and Coggan [6] and the reliability of the pacing strategy during the 20-min TT.

## Materials and Methods

### Subjects

Twenty-five male competitive cyclists (mean ± standard deviation [SD]: age = 34.7 ± 6.7 years, weight = 76.0 ± 9.4 kg, height = 1.77 ± 0.05 m, peak power output [PPO] = 381 ± 38 Watts [W], and $\dot{V}O_2max$ = 57.8 ± 7.3 mL·kg$^{-1}$·min$^{-1}$) voluntarily participated in the study. The cyclists had at least 2 years of experience in regional and national competitions and trained 10 ± 3 h and rode 211 ± 82 km per week. The cyclists were classified as Performance Level 3 (PL-3; trained cyclists) [22]. After verbal and written explanations of the procedures, all participants signed an informed consent form, approved by the institutional ethics committee and conforming to the Declaration of Helsinki. This study was performed in accordance with the ethical standards of the International Journal of Sports Medicine [23].

### Study design

The cyclists completed four laboratory tests during the three visits to the laboratory, each test separated by seven days. In the first session, the cyclists performed an incremental exercise test and a familiarization of the $FTP_{20}$ protocol; each test was separated by a recovery period of 30-min. In the second and third sessions, cyclists performed two identical self-paced 20-min TTs for the determination of $FTP_{20}$ reliability. The cyclists were asked to refrain from strenuous exercise in the 48 h preceding each test. All tests were carried out under standardized laboratory conditions of 20 °C and 40–50 % relative humidity. All tests were performed on an electrically braked bicycle ergometer Velotron (Dynafit Pro, Racer Mate Inc, WA, USA) that was modified with a racing saddle, adjustable stem, and the subject's pedal system.

### Incremental exercise test

The test was preceded by a warm-up of 10 min at 100 W. Immediately after warm-up, the load was increased continuously by 25 W every minute until voluntary exhaustion was reached. During the test, oxygen uptake ($\dot{V}O_2$) and heart rate (HR) were measured continuously with a gas analyzer (Quark PFTergo – Cosmed Srl, Rome, Italy) that was calibrated in accordance with the manufacturer's in-

structions using alpha gas standards. PPO was determined as PO of last completed stage. If the stage was not completed, PPO was calculated according to Kuipers et al. [24]. $\dot{V}O_2max$ was determined as the highest 30-s average in $\dot{V}O_2$.

### Functional threshold power tests

The cyclists performed two laboratory-simulated 20-min TTs to determine the $FTP_{20}$ on the Velotron (RacerMate Inc, WA, USA). The accuracy of the ergometer was previously determined (< 1 % error) [25]. TTs were performed on a flat terrain with no wind as simulated by the software RacerMate Interactive 3D (RacerMate Inc, WA, USA). During the test, participants could view their progress over the course on a computer monitor, and they were provided with information on the time completed and gear selected. All other information was blinded, no verbal encouragement was provided, and water was allowed *ad libitum* [2].

TTs were performed following the procedure described by Allen and Coggan [6]. The duration of the warm-up was 45 min, divided into 5 sections: 1) 20 min at self-selected low-intensity (rating of perceived exertion [RPE] 8–10), 2) 3 × 1-min fast pedaling accelerations (100-105 rpm) with a 1-min recovery period between the efforts, 3) 5 min at self-selected low-intensity (RPE 8–10), 4) 5 min at maximal effort, and 5) 10 min at self-selected low-intensity (RPE 8–10) and 5 min in passive rest. It is important to note that during self-selected low-intensity cycling, the authors [6] recommended that cyclists perform at 65 % of the FTP; therefore, a pretest would be necessary. Thus, we instructed the cyclists to cycle at an RPE 8–10, in accordance with FTP-based training recommendations (i.e., 65 % FTP ~ RPE 8–10). During the fast-pedaling accelerations, no intensity was suggested in the test description; therefore, we instructed the cyclists to select an "easy" gear (i.e., 52 × 21–23) and accelerate between 100–105 rpm. Athletes were instructed to choose similar gears during each part of the warm-up during the retest. The main part of the test consisted of a 20-min TT where the participants were asked to produce the highest average PO possible. Blood samples were obtained from the earlobe before ($[La]_{PRE}$) and immediately after ($[La]_{POST}$) the warm-up and were immediately analyzed (YSI 1500, Yellow Springs, OH, USA) for [La] determination. The RPE was measured at the end of each warm-up step and every 5 min during the 20-min TT. PO, cadence, and HR were monitored throughout the test. An example of the protocol is presented in ▶ **Fig. 1**.

### Data analysis

Descriptive results are reported as mean ± SD. For determination of the systematic error between test and retest, paired Student t tests were performed. Additionally, the magnitude of the test-retest changes was assessed via Cohen's d effect sizes and interpreted using the thresholds of: < 0.2 (trivial), 0.6 (small), 1.2 (moderate), 2.0 (large), 4.0 (very large) [26]. In order to analyze the random error of the measurements, the typical error of measurement (TEM) was calculated in raw units, after logarithmic transformation in percentage units (i.e., coefficient of variation [CV]) and in standardized units (TEMs; calculated as CV divided by the "pure" between-subject SD %) [27]. To interpret the magnitude of TEMs, the Cohen's d effect sizes thresholds were halved [10]. Additionally, the intraclass correlation coefficient (ICC) was calculated and inter-

preted as follows: < 0.10 (trivial), 0.30 (small), 0.50 (moderate), 0.70 (large), 0.90 (very large), 0.99 (nearly perfect), and 1 (perfect) [26]. Confidence intervals (CI) of 95 % were calculated for d, TEM, CV, TEMs, and ICC. The above calculations were performed using Hopkins' spreadsheet [27].

For the pacing strategy analysis, PO data were averaged at every 2-min segment and were normalized in percentages of the mean PO of the 20-min TT. Then, the TEM, in raw units (percentage of mean PO), was calculated for each segment [18, 19]; this method has been well accepted for analyzing the pacing strategy reliability [18, 19, 28]. In addition, a two-way repeated measure ANOVA was used to compare the PO in each part of the warm-up and the pacing strategy. The assumption of normality was verified using Shapiro-Wilk's test. Mauchly's test of sphericity was used to test this assumption, and a Greenhouse–Geisser was used when necessary. When the main effect was significant, the Bonferroni post-hoc test was applied. Effect sizes (partial eta squared, $\eta_p^2$) was calculated for the ANOVA main effects, and the magnitude of $\eta_p^2$ was classified as small ≤ 0.06, moderate 0.07–0.14, and large > 0.14 [29]. Statistical significance for ANOVA was accepted at a $p \leq 0.05$. Analyses were performed using IBM SPSS software version 21.0 (SPSS, Chicago, IL, USA).

We conducted a sensitivity analysis in G * Power (Version 3.1.9.4, Germany) to determine the smallest effect that one could have detected with high probability given n = 25, $p \leq 0.05$, and statistical power = 80 %. In this present study, the effect size for the t-tests was $d = 0.58$, for ANOVA main effect analysis $\eta_p^2 = 0.26$, and ICCs = 0.50.

## Results

For the familiarization trial, the mean PO was 231 ± 41 W during the 20-min TT. Compared with the first test (20-min TT), the familiarization trial results were as follows: d = 0.41 (95 % CI, ± 0.09); p < 0.001; TEM = 7.4 W (95 % CI, × / ÷ 1.3); CV = 3.3 % (95 % CI, × / ÷ 1.3); TEMs = 0.20 ( 95 % CI, × / ÷ 1.3); and nearly perfect ICC (r = 0.96; 95 % CI, 0.91–0.98).
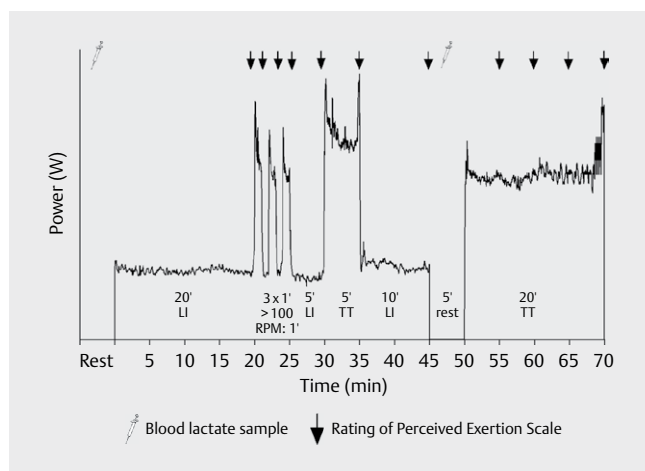


▶ **Fig. 1** FTP$_{20}$ protocol of a representative cyclist. For details, see the Methods section. LI = low-intensity; TT = time-trial.

For test-retest the mean values of PO, HR, RPE, cadence, and [La] for the warm-up and for the 20-min TT are shown in ▶ **Table 1**. All measures presented trivial to small and non-significant changes (d ≤ 0.27; p > 0.05) between test and retest. The average warm-up and 20-min TT measures presented moderate to nearly perfect ICCs (r = 0.55–0.97). The TEMs for all measures during average warm-up were classified as moderate, and during the 20-min TT the TEMs were small for PO and cadence, moderate for HR, and large for RPE (▶ **Table 1**).

For PO measured during warm-up (▶ **Fig. 2c**), the ANOVA showed that there was no main effect between the trials ($F_{(1,24)}$ = 0.699, p = 0.411, $\eta_p^2$ = 0.03) and there was no interaction between the factors trials × time ($F_{(3, 60)}$ = 0.765, p = 0.496, $\eta_p^2$ = 0.03). The low intensity segments presented CV values between 12.7–20.4 % (95 % CI, × / ÷ 1.3) (▶ **Fig. 2b**), classified as moderate to large TEMs, and the 5-min TT presented a CV of 4.8 % (95 % CI, × / ÷ 1.3), classified as moderate TEMs (▶ **Fig. 2a**).

For analysis of the pacing strategy during the 20-min TT (▶ **Fig. 3**), the ANOVA test showed that there was no main effect between the trials ($F_{(1,24)}$ = 0.194, p = 0.664, $\eta_p^2$ = 0.01) and there was no interaction between the factors trials × time ($F_{(5,118)}$ = 2.027, p = 0.081, $\eta_p^2$ = 0.08). The analysis of TEM for each of the 2-min segments (▶ **Fig. 3b**) showed the smallest variability between 4 and 12 min (TEM = 4.4–5.8 %; 95 % CI, × / ÷ 1.3), and the highest variability at the start (TEM = 7.4 %; 95 % CI, × / ÷ 1.3) and between 14 and 20 min of the TT (TEM = 7.9–11.4 %; 95 % CI, × / ÷ 1.3); in all segments, the TEMs were deemed large to very large (▶ **Fig. 3a**).

## Discussion

The main finding of the present study was the nearly perfect test-retest reliability in PO, HR, and cadence in determining FTP using the 20-min TT. In contrast to that, poor reliability was found for the long warm-up. Furthermore, the distribution of PO during the 20-min TT presented large to very large TEMs. Interestingly, despite the high variability in each part of the warm-up and the moderate CV for the overall warm-up (5.5 %), the reliability of the mean PO of the 20-min TT does not seem to be affected (▶ **Table 1**). This is in accordance with the variability found in the TTs of different durations using shorter standardized warm-ups (CV = 2.0–3.4 %) [30–32].

To our knowledge, this is the first study to analyze the reliability of the protocol proposed by Allen and Coggan [6]. The results of the present study show trivial to small changes in the mean (d ≤ 0.27) between test-retest measures (▶ **Table 1**). A likely reason for the high reliability is the familiarization trial [2, 3, 33]. In addition, there was enough time for proper recovery between test and retest (7 days). A meta-analysis showed that between the first two trials the CV is 1.3 × higher than the CV between subsequent trials; there is a reduction in the change in the mean (1.2–0.2 % [3]). Additionally, a study showed an increase in the ICC between familiarization-test and test-retest comparisons [32]. In the present study, the CV between test and familiarization was 1.1 × higher than between test-retest. However, the present study showed a change in the mean of 15 W (6.4 %) between the familiarization and the test. Thus, because familiarization was performed after the incremental test, it is not possible to isolate the learning effects and

Borszcz FK et al. Reliability of the Functional … Int J Sports Med 2020; 41: 175–181

177

► **Table 1** Reliability of the physiological variables, cadence and power output during the FTP$_{20}$ protocol.

| Measures | Test (mean ± SD) * | Retest (mean ± SD) * | Mean change ± SD * | d (95 % CI) | TEM * (95 % CI) | CV % (95 % CI) | TEMs (95 % CI) | ICC (95 % CI) |
|---|---|---|---|---|---|---|---|---|
| **Overall warm-up** | | | | | | | | |
| PO (W) | 128 ± 15 | 129 ± 18 | 1 ± 10 | 0.06[T] (−0.23–0.35) | 7 (5–10) | 5.5 (4.3–7.7) | 0.46[M] (0.36–0.64) | 0.84[VL] (0.67–0.93) |
| HR (bpm) | 114.3 ± 14.0 | 114.1 ± 15.8 | −0.2 ± 6.4 | −0.01[T] (−0.22–0.20) | 4.6 (3.6–6.4) | 3.9 (3.0–5.4) | 0.32[M] (0.25–0.45) | 0.92[NP] (0.82–0.96) |
| RPE (au) | 11.0 ± 1.4 | 11.1 ± 1.8 | 0.1 ± 1.3 | 0.06[T] (−0.33–0.45) | 0.9 (0.7–1.3) | 8.2 (6.4–11.0) | 0.64[M] (0.50–0.89) | 0.72[VL] (0.45–0.87) |
| Cadence (rpm) | 88.3 ± 5.9 | 88.5 ± 6.2 | 0.2 ± 3.1 | 0.03[T] (−0.21–0.27) | 0.4 (0.3–0.6) | 2.4 (1.9–3.3) | 0.37[M] (0.29–0.51) | 0.89[VL] (0.76–0.95) |
| [La]$_{PRE}$ (mmol · L$^{-1}$) | 1.21 ± 0.47 | 1.10 ± 0.32 | −0.1 ± 0.4 | −0.27[S] (−0.75–0.21) | 0.3 (0.2–0.4) | 25.6 (20.0–35.6) | 0.93[M] (0.73–1.30) | 0.55[L] (0.20–0.78) |
| [La]$_{POST}$ (mmol · L$^{-1}$) | 4.90 ± 2.58 | 4.75 ± 2.51 | −0.2 ± 1.8 | −0.06[T] (−0.40–0.28) | 1.3 (1.0–1.8) | 37.0 (28.9–51.5) | 0.64[M] (0.50–0.89) | 0.72[VL] (0.45–0.87) |
| **5-min TT[a]** | | | | | | | | |
| PO (W) | 299 ± 48 | 296 ± 49 | −3 ± 18 | −0.06[T] (−0.24–0.12) | 13 (10–18) | 4.8 (3.7–6.7) | 0.31[M] (0.24–0.43) | 0.93[NP] (0.85–0.97) |
| **20-min TT** | | | | | | | | |
| PO (W) | 246 ± 35 | 245 ± 36 | −1 ± 10 | −0.03[T] (−0.16–0.10) | 7 (5–10) | 2.9 (2.3–4.0) | 0.19[S] (0.15–0.26) | 0.97[NP] (0.93–0.99) |
| HR (bpm) | 159.8 ± 12.2 | 157.6 ± 11.7 | −2.0 ± 5.4 | −0.17[T] (−0.39–0.05) | 3.8 (3.0–5.3) | 2.6 (2.0–3.6) | 0.35[M] (0.27–0.49) | 0.97[NP] (0.93–0.99) |
| RPE (au) | 15.4 ± 1.6 | 15.4 ± 1.7 | 0.0 ± 1.2 | 0.00[T] (−0.35–0.35) | 0.9 (0.7–1.3) | 5.9 (4.6–8.2) | 0.62[L] (0.48–0.86) | 0.74[VL] (0.49–0.88) |
| Cadence (rpm) | 99.6 ± 9.7 | 98.2 ± 9.2 | −1.4 ± 3.5 | −0.15[T] (−0.33–0.03) | 2.5 (2.0–3.5) | 2.7 (2.1–3.8) | 0.28[S] (0.23–0.39) | 0.93[NP] (0.85–0.97) |

 * = the same unit of measure presented in the parameters; a = 5-min TT performed during the warm-up; au = arbitrary units; bpm = beats per minute; CI = confidence interval; CV = coefficient of variation; d = effect size; HR = heart rate; ICC = intraclass correlation coefficient; mmol · L$^{-1}$ = millimole per liter; PO = power output; TEM = typical error of measurement; TEMs = typical error of measurement standardized; RPE = rating of perceived exertion; rpm = revolutions per minute; W = watts; [La] = blood lactate concentration. Magnitude inferences: T = trivial; S = small; M = moderate; L = large; VL = very large; NP = nearly perfect

residual fatigue from the incremental test. The protocol proposed for the FTP$_{20}$ identification was composed of a complex 45-min warm-up. The results demonstrated high variability in the PO (CV = 13.0–14.1 %) during the low-intensity parts of the warm-up. This is in accordance with the findings observed in the lowest intensity (60 % of maximal HR) part of the Lamberts and Lambert Submaximal Cycle Test warm-up protocol (CV = 12.4 %) [34]. Furthermore, we found high variability in PO measured during the high cadence sprints (CV = 14.4–20.4 %). Eston and Williams [35] examined the reliability of $\dot{V}O_2$ measured during the constant load intensity regulated by RPE. The reliability assessed by coefficient of correlation between tests 1–2 and 2–3 for $\dot{V}O_2$ measured at the RPEs 9, 13, and 17 were r = 0.26–0.83, 0.64–0.94, and 0.92–0.96, respectively. Therefore, the regulation of PO is more reliable at high intensities compared to low intensities, in accordance with the results of the present study, where the TTs are more reliable than the low-intensity parts of the warm-up.

A 5-min TT was inserted during the warm-up in the FTP$_{20}$ protocol. According to Allen and Coggan [6], the main goal of the 5-min TT inserted in the warm-up is to "'open' up the legs for the rest of the effort". The FTP$_{20}$ determined using this protocol was not statistically significantly different from FTP$_{60}$, AnT (i.e., IAT and Dmax), and MLSS (p > 0.05), despite the random errors (LoA = 7.8–
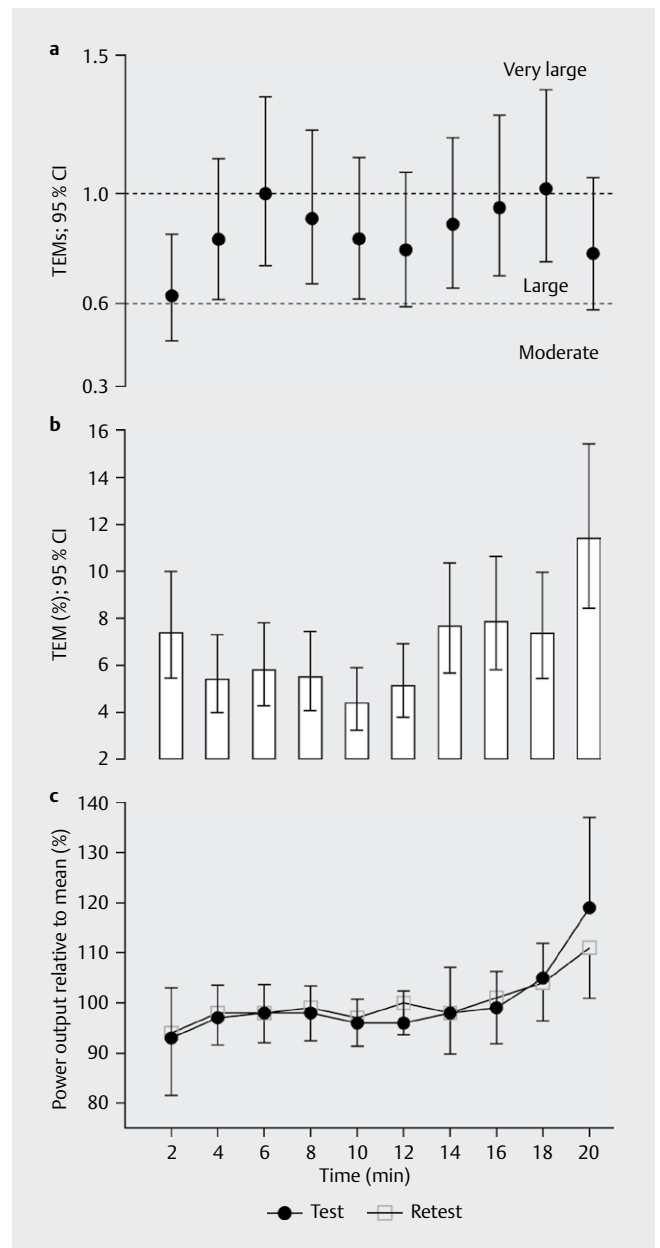
27.5 %) [7, 8,3 7]. However, when the 20-min TT was performed after a warm-up of 15 min at moderate intensity, the FTP$_{60}$ corresponded to 90 % of a 20-min TT PO [12]. This is not in accordance with the suggested 95 % of 20-min TT PO to determine the FTP$_{20}$; thus, the warm-up subsequently influenced the prediction of FTP$_{60}$, changing the correction factor from 5 to 10 % [12]. Previous work demonstrated high reliability in the 4- and 20-min TTs performed in the same session but separated by a 30-min low-intensity exercise in well-trained and elite cyclists (CV = 2 % and 1.6 %, respectively [CVs calculated by us from LoA]) [17], and in the 4- and 20-min TTs performed on separate days (CV = 2.2 and 1.4 %, respectively) [12]. However, we found the CVs of 4.8 and 2.9 % for the TTs of 5- and 20-min, respectively, separated by 15 min. Therefore, the 5-min TT inserted in this warm-up had poor reliability compared to the 5-min TT performed alone (CV = ~1.7 %) [36]; consequently, it had a limited capacity to detect small changes in the performance.

Cycling events include several pacing profiles such as negative, positive, all-out, even, parabolic-shaped, and variable pacing strategies [28]. In the present study during the 20-min TT, the cyclists adopted the "negative" pacing strategy using the full warm-up protocol, in accordance with a previous study [7]. However, when the 20-min TT was performed without the 45 min warm-up, the cyclists

178

Borszcz FK et al. Reliability of the Functional … Int J Sports Med 2020; 41: 175–181
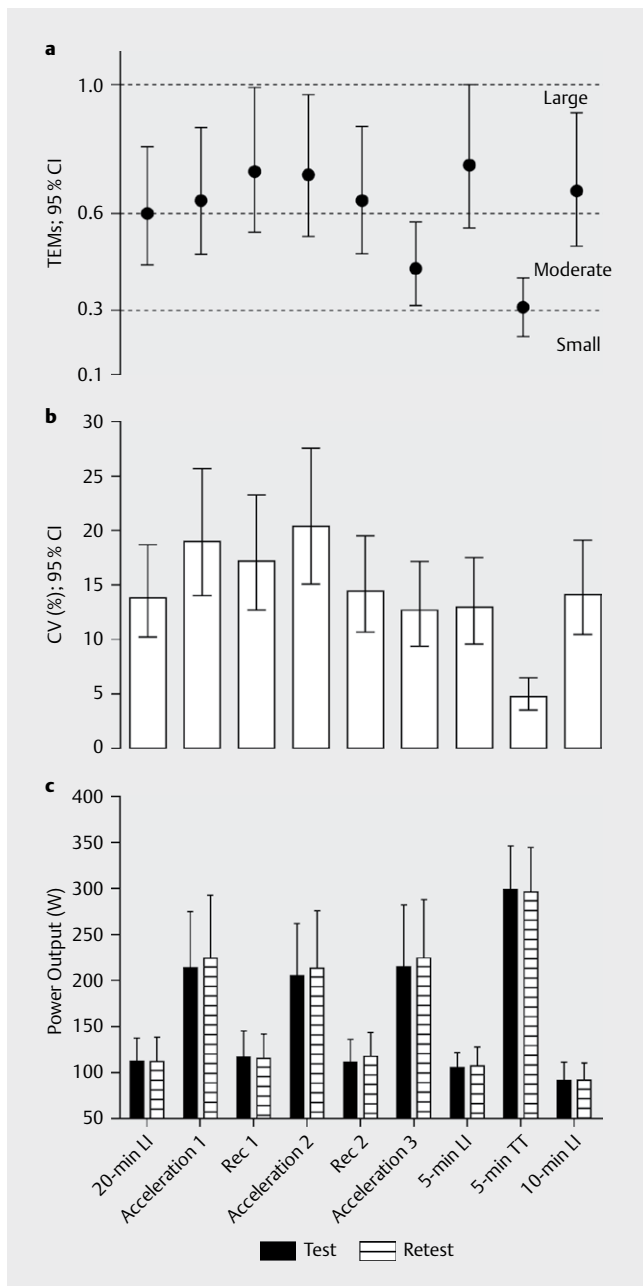
▶ **Fig. 2** T.ypical error of measurement standardized **a**, coefficient of variation for power output measure **b**, and power output **c** at each part of the warm-up. Values are TEMs ± 95 % confidence intervals in A, CV ± 95 % confidence intervals in B, and means ± SD in C. LI = low-intensity; TT = time-trial.



▶ **Fig. 3** Typical error of measurement standardized **a**, typical error of measurement of pacing strategy **b**, and pacing strategy **c** during 20-min TT. Values are TEMs ± 95 % confidence intervals in A, TEM ± 95 % confidence intervals in B, and means ± SD in C.

adopted the "even" [17] or "parabolic-shaped" [12, 13] pacing strategy. Therefore, we believe that the pacing strategy for the 20-min TT was modulated by the long warm-up and by the [La] before the start of the TT ($4.9 \pm 2.6$ and $4.8 \pm 2.5$ mmol L$^{-1}$; test-retest, respectively). The cyclists appear to have adopted a conservative strategy of up to half the time to complete the test. Burnley et al. [37] demonstrated that the majority of cyclists with [La] > 4 mmol L$^{-1}$ before the start of a 5-min TT had decreases in the PO relative to the control warm-up. Moreover, the warm-up with greater [La]

caused the adoption of a more conservative pacing strategy at the beginning [37]. Another factor behind the initial conservative strategy is may because the FTP$_{20}$ is a time-based TT. As showed in a previous study [15], the pacing strategy is more conservative during the initial stages of time-based TTs compared to distance-based TTs of similar duration.

Despite the moderate warm-up reliability (CV = 5.5 %), the pacing strategy presented no systematic error between test and re-test, that is, no statistically significant ANOVA main effect and interaction ($p > 0.05$; $\eta_p^2 < 0.06$, small effect). However, the random error expressed as TEM ( %) in each 2-min interval showed large to

Borszcz FK et al. Reliability of the Functional … Int J Sports Med 2020; 41: 175–181

179

very-large variability in the pacing strategy, especially between the start to 2 min and between 14 to 20 min (▶ **Fig. 3a** and **b**). This variability does not correspond to the variability observed in the 4- and 20-km TT tests [18, 19], wherein the high TEM was observed only during the initial and the final segments. In addition, the TEM of the present study at the initial and final segments (7.4 and 11.4 %, respectively) were higher than the TEM reported during the 20-km TT (~6.7 %) [19]; at the midpoint of the TT, TEM (4.4–7.9 %) was also higher than that reported during the 20-km TT (TEM = 1.8 % to 3.4 %) [16, 19]. The high variability found in the present study was also probably due to the TT type. Previous studies [16, 18, 19] analyzed the reliability of a distance-based TT; a meta-analysis showed that the time-based TTs were 1.2 × less reliable than distance-based TTs [3]. Therefore, it seems that athletes can better control their pacing strategy in TTs similar to official races (i.e., distance-based).

Finally, the smallest worthwhile change (SWC) is defined as $0.3 \times CV$ [26]. In well-trained and professional cyclists (Performance levels 4 and 5 [22]), the SWC for a 20-min TT is 0.4–0.5 % [12, 17]. However, the present study showed an SWC of 0.9 % for $FTP_{20}$, in trained cyclists, using the Allen and Coggan warm-up. Therefore, $FTP_{20}$ should be used to detect performance changes in trained cyclists, but not in well-trained and professional cyclists, because sensitivity is too low for athletes with a higher fitness level. Nevertheless, more data is needed to elucidate about the variability of the 20-min TT using different warm-up protocols at different performance levels.

## Practical applications

These results suggest that the FTP determined from the Allen and Coggan [6] protocol is reliable, which should be considered as valuable information for practitioners implementing this protocol. Therefore, it should be used by coaches and trained cyclists (i.e., Performance levels 2 and 3 [22]) to detect changes in performance.

# Conclusions

In conclusion, despite the low warm-up reliability in trained cyclists, the $FTP_{20}$ determined was reliable based on the protocol proposed by Allen and Coggan [6]. Future studies should verify the influence of the different warm-up strategies on $FTP_{20}$ determination and the reliability of the $FTP_{20}$ in elite professional cyclists.

## Conflict of Interest

The authors declare that they have no conflict of interest.

## References

[1] Borszcz FK, Tramontin AF, de Souza KM et al. Physiological correlations with short, medium, and long cycling time-trial performance. Res Q Exerc Sport 2018; 89: 120–125

[2] Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. Sports Med 2008; 38: 297–316

[3] Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. Sports Med 2001; 31: 211–234

[4] Reilly T, Morris T, Whyte G. The specificity of training prescription and physiological assessment: A review. J Sports Sci 2009; 27: 575–589

[5] Passfield L, Hopker J, Jobson S et al. Knowledge is power: Issues of measuring training and performance in cycling. J Sports Sci 2017; 35: 1426–1434

[6] Allen H, Coggan A. Training and Racing with a Power Meter. 2nd ed. Boulder: Velopress; 2010

[7] Borszcz FK, Tramontin AF, Bossi AH et al. Functional threshold power in cyclists: Validity of the concept and physiological responses. Int J Sports Med 2018; 39: 737–742

[8] Borszcz FK, Tramontin AF, Costa VP. Is the functional threshold power interchangeable with the maximal lactate steady state in trained cyclists? Int J Sports Physiol Perform 2019; 14: 1029–1035

[9] Valenzuela PL, Morales JS, Foster C et al. Is the functional threshold power a valid surrogate of the lactate threshold? Int J Sports Physiol Perform 2018; 13: 1293–1298

[10] Smith TB, Hopkins WG. Variability and predictability of finals times of elite rowers. Med Sci Sports Exerc 2011; 43: 2155–2160

[11] Denham J, Scott-Hamilton J, Hagstrom AD et al. Cycling power outputs predict functional threshold power and maximum oxygen uptake. J Strength Cond Res 2017, doi:10.1519/JSC.0000000000002253

[12] MacInnis MJ, Thomas ACQ, Phillips SM. The reliability of 4-minute and 20-minute time trials and their relationships to functional threshold power in trained cyclists. Int J Sports Physiol Perform 2019; 14: 38–45

[13] Morgan PT, Black MI, Bailey SJ et al. Road cycle TT performance: Relationship to the power-duration model and association with FTP. J Sports Sci 2019; 37: 902–910

[14] McGowan CJ, Pyne DB, Thompson KG et al. Warm-up strategies for sport and exercise: mechanisms and applications. Sports Med 2015; 45: 1523–1546

[15] Abbiss CR, Thompson KG, Lipski M et al. Pacing differs between time- and distance-based time trials in trained cyclists. Int J Sports Physiol Perform 2016; 11: 1018–1023

[16] Costa VP, Guglielmo LGA, Paton CD. The effects of block training on pacing during 20-km cycling time trial. Appl Physiol Nutr Metab 2017; 42: 391–398

[17] Nimmerichter A, Williams C, Bachl N et al. Evaluation of a field test to assess performance in elite cyclists. Int J Sports Med 2010; 31: 160–166

[18] Stone MR, Thomas K, Wilkinson M et al. Consistency of perceptual and metabolic responses to a laboratory-based simulated 4 000-m cycling time trial. Eur J Appl Physiol 2011; 111: 1807–1813

[19] Thomas K, Stone MR, Thompson KG St et al. Reproducibility of pacing strategy during simulated 20-km cycling time trials in well-trained cyclists. Eur J Appl Physiol 2012; 112: 223–229

[20] Buchheit M. Houston, we still have a problem. Int J Sports Physiol Perform 2017; 12: 1111–1114

[21] Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professionals. Phys Ther Sport 2003; 4: 122–128

[22] de Pauw K, Roelands B, Cheung SS et al. Guidelines to classify subject groups in sport-science research. Int J Sports Physiol Perform 2013; 8: 111–122

[23] Harriss DJ, Macsween A, Atkinson G. Standards for ethics in sport and exercise science research: 2020 update. Int J Sports Med 2019; 40: 813–817

[24] Kuipers H, Verstappen F, Keizer H et al. Variability of aerobic performance in the laboratory and its physiologic correlates. Int J Sports Med 1985; 06: 197–201

180

Borszcz FK et al. Reliability of the Functional … Int J Sports Med 2020; 41: 175–181

[25] Abbiss CR, Quod MJ, Levin G et al. Accuracy of the Velotron ergometer and SRM power meter. Int J Sports Med 2009; 30: 107–112

[26] Hopkins WG, Marshall SW, Batterham AM et al. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc 2009; 41: 3–13

[27] Hopkins WG. Spreadsheets for analisys of validity and reliability. Sportscience 2015; 19: 36–42

[28] Abbiss CR, Laursen PB. Describing and understanding pacing strategies during athletic competition. Sports Med 2008; 38: 239–252

[29] Cohen J. Statiscal Power Analysis for the Behavioral Sciences. New Jersey: Lawrance Erlbaum; 1986

[30] Bishop D. Reliability of a 1-h endurance performance test in trained female cyclists. Med Sci Sports Exerc 1997; 29: 554–559

[31] Jeukendrup A, Saris WH, Brouns F et al. A new validated endurance performance test. Med Sci Sports Exerc 1996; 28: 266–270

[32] Triska C, Karsten B, Heidegger B et al. Reliability of the parameters of the power-duration relationship using maximal effort time-trials under laboratory conditions. PLoS One 2017; 12: 1–12

[33] Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000; 30: 1–15

[34] Lamberts RP, Swart J, Noakes TD et al. A novel submaximal cycle test to monitor fatigue and predict cycling performance. Br J Sports Med 2011; 45: 797–804

[35] Eston RG, Williams JG. Reliability of ratings of perceived effort regulation of exercise intensity. Br J Sports Med 1988; 22: 153–155

[36] Paton CD, Hopkins WG. Ergometer error and biological variation in power output in a performance test with three cycle ergometers. Int J Sports Med 2006; 27: 444–447

[37] Burnley M, Doust JH, Jones AM. Effects of prior warm-up regime on severe-intensity cycling performance. Med Sci Sports Exerc 2005; 37: 838–845

Borszcz FK et al. Reliability of the Functional … Int J Sports Med 2020; 41: 175–181

**181**