OPEN ACCESS

Journal of Science & Cycling
Breakthroughs in Cycling & Triathlon Sciences

Original Article

# Reliability of Cyclotronics Smart Trainer Power Output in Trained Cyclists

**Vitor Pereira Costa [1,*], Paula Marline Johann [1], Plinio Martinez Reche Junior [1] and Allan Inoue Rodrigues [2]**

**Received:** 2 August 2025 **Accepted:** 17 March 2026 **Published:** 23 March 2026

[1] Human Performance Research Laboratory, Center for Health and Sports Science, Santa Catarina State University, Florianópolis, Brazil

[2] Research Group on Exercise and Health Science, Rio de Janeiro State University, Rio de Janeiro, Brazil

**Correspondence**

Vitor Pereira Costa

Human Performance Research Laboratory, Center for Health and Sports Science, Santa Catarina State University, Florianópolis, Brazil

vitor.costa@udesc.br

## Abstract

The aim of this study was to assess the test–retest reliability of the Cyclotronics Smart Trainer during a graded exercise test (GXT) and a power profile test (PPT) in trained cyclists. For the GXT reliability analysis, 22 cyclists performed three GXTs, each separated by seven days. For the PPT reliability assessment, a smaller group of 10 trained participants (2 females, including 1 triathlete and 7 cyclists) completed one GXT followed by three PPTs, also separated by seven days. The first test in each protocol was used for familiarization and excluded from the reliability analysis. During the GXT, the coefficient of variation (CV) ranged from 5.5% at the lowest workload (100 W) to 1.6% at the highest workload (300 W), demonstrating reduced variability at higher intensities. A very low CV was also observed for peak power output (PPO), indicating high consistency in maximal performance measures. Regarding the PPT, acceptable test–retest reliability was observed, particularly for efforts lasting 120 seconds or more. As expected, shorter efforts showed slightly greater variability, which is common in high-intensity performance tests with a smaller sample size. In conclusion, the Cyclotronics Smart Trainer demonstrated reliable measurements during both GXT and PPT protocols. These findings support its use in repeated testing scenarios where measurement consistency is essential.

**Keywords**

cycling; graded exercise test; power profile test; reproducibility

## 1 Introduction

Indoor cycling has evolved substantially over the past two decades, transitioning from conventional stationary bikes in gym-based spinning classes to sophisticated smart trainer systems that allow athletes to simulate outdoor riding in immersive virtual environments.

These devices enable cyclists to perform structured workouts, enter online races, and monitor performance through real-time feedback on power output, cadence, and other physiological metrics (Morais, Bragada, Magalhães, & Marinho, 2024). Smart trainers can be broadly categorized into wheel-on or direct-drive systems, with the latter replacing the bicycle's rear wheel to deliver greater accuracy and stability in power measurement. The adoption of direct-drive smart trainers has grown rapidly due to their precision, compatibility with training platforms like Zwift and TrainerRoad, and their ability to replicate realistic resistance profiles (McIlroy, Passfield, Holmberg, & Sperlich, 2021).

Given these technological advances, device reliability has become critical for both athletic monitoring and scientific research. A device's reliability determines whether observed performance changes reflect true physiological adaptations or merely measurement noise (Hopkins, 2000). Early studies with electromagnetically braked systems, such as the wheel-on Computrainer (RacerMate, USA), demonstrated high test-retest reliability, with coefficients of variation (CV) below 2% for both peak and mean power output across standardized cycling protocols (Lamberts, Swart, Noakes, & Lambert, 2009; Sparks, Williams, Massey, Bridge, Marchant, & McNaughton, 2016; Jeker, Gosselin, Drouet, Goulet, 2021). These findings positioned the CompuTrainer as one of the most widely used and historically relevant laboratory ergometers for controlled endurance performance assessments. However, methodological reviews have emphasized that the evaluation of cycling power measurement systems should consider multiple metrological properties and appropriate reference frameworks when interpreting reliability and validity outcomes (Bouillod, Soto-Romero, Grappe, Bertucci, Brunet, & Cassirame, 2022). As technology evolved and the CompuTrainer was discontinued in 2017, the need to evaluate the reliability of newer smart trainer systems became increasingly evident.

Subsequent research has evaluated the reliability of both wheel-on and direct-drive smart trainers. Peiffer and Losco (2011) reported a CV of approximately 2.6% for repeated time trials on the wheel-on Tacx Fortius VR. More recent evaluations of direct-drive systems, such as the Wahoo KICKR, showed CVs as low as 1.5% for power output, alongside similarly low variability in cadence, speed, and time (Zadow, Kitic, Wu, Smith, & Fell, 2016; Hoon, Michael, Patton, Chapman, & Areta, 2016). These and many other studies have encouraged the integration of smart trainers into laboratory testing environments. Nonetheless, the commercial and scientific success of a few globally established brands has overshadowed the validation of emerging models from smaller manufacturers, many of which are entering the market with limited peer-reviewed scrutiny.

The Cyclotronics Smart Trainer (Cyclotronics, Curitiba, Brazil) is a new direct-drive system designed for high-performance indoor cycling and compatible with standard training platforms. Despite offering hardware and software features comparable to leading devices, its reliability under standardized physiological testing conditions has not been formally assessed. This gap in the literature presents a barrier to the scientific use of the device, particularly for longitudinal performance monitoring or intervention studies where measurement precision is critical. Independent investigations are necessary to determine whether Cyclotronics meets the reliability standards expected in sports science research and elite training settings.

Graded exercise tests (GXTs) remain one of the most widely used methods for assessing aerobic capacity and power in endurance athletes (Jamnick, Pettitt, Granata, Pyne, & Bishop, 2020). More recently, the concept of power profiling (PPT) that uses a series of maximal efforts over multiple durations has been integrated into both training and performance diagnostics (Pinot, & Grape, 2011). These methods require accurate and stable power data across a wide range of intensities. With the growing popularity of indoor cycling and the emergence of new smart trainer brands, it is essential to verify whether these tools can be reliably used to assess key performance indicators. Without robust test-retest data, interpretations of fitness improvements or comparisons across time may be compromised.

Therefore, the aim of this study was to assess the test-retest reliability of the Cyclotronics Smart Trainer during a GXT and a PPT in trained cyclists. We hypothesized that the Cyclotronics Smart Trainer would demonstrate high reliability in power output across test sessions, with CVs and intraclass correlation coefficients (ICC) comparable to those reported in studies using other validated smart trainers.

# 2 Material and Methods

## 2.1 Design

To verify the reliability of GXT we first recruited 22 cyclists. Cyclists performed three GXTs separated by seven days. To verify the reliability of the PPT we recruited 10 cyclists. Cyclists performed one GXT and three PPT separated by seven days. The first test of GXT and PPT reliability protocol was used as familiarization. In addition, the first GXT was used to determine maximal oxygen uptake ($VO_{2}max$). 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 cyclist own bicycle attached to the Cyclotronics smart trainer (Cyclotronics, Curitiba, Brazil).

## 2.2 Participants

Thirty-two trained cyclists voluntarily participated in the study (Table 1). The cyclists had at least 2 years of experience in regional competitions and trained approximately 7-8 h and rode 200-250 km per week. After verbal and written explanations of the procedures, all participants signed an informed consent form, approved by the institutional ethics committee number 84837924.1.0000.0118 and conforming to the Declaration of Helsinki.

## 2.3 Methodology

### *2.3.1 Cyclotronics Smart Trainer*

Cyclotronics is a smart trainer developed to communicate power output data with training apps, bike computers, computers, sports watches, and virtual cycling platforms. Cyclotronics uses direct drive technology with a heavy flywheel with electronically controlled magnetic resistance (Figure 1). The exercise protocols were created using cycling power analysis software (Golden Cheetah, V3.5, USA). During the GXTs, the software was in ERG mode while cyclists followed the protocol on a screen computer. The ERG mode allows the smart trainer slightly increases electromagnetic resistance while cyclists were able to produce power output according to each stage of the protocol. During the PPT, the software was in SLOPE mode with a fixed 1.0% grade. The SLOPE mode allows the smart trainer to increase resistance according to the grade, consequently, cyclists augment power with changes in gear during PPT. The Cyclotronics was calibrated in accordance with the manufacturer’s instructions (Cyclotronics, Curitiba, Brazil) prior to warm up before each test.

**Figure 1.** Cyclotronics smart trainer.

### 2.3.2 Graded Exercise Test (GXT)

The test was preceded by a warm-up of 10 min of 100 W. Immediately thereafter, the workload increased by 25 W·min$^{-1}$ until the participant reached voluntary exhaustion. Oxygen uptake ($VO_2$) and heart rate (HR) were measured continuously using a calibrated gas analyzer (Quark PFTergo – Cosmed Srl, Rome, Italy). The $VO_2$ data were plotted as a function of the power in an average of 30 sec, and the highest value was considered the $VO_2max$. Peak power output (PPO) was determined as the highest 15 s mean PO of the test. The GXT was performed to characterize the aerobic physical fitness of the reliability GXT and PPT protocol.

### 2.3.3 Reliability of the GXT

To assess test-retest reliability, participants completed the same incremental protocol on three separate occasions: the first session served as a familiarization trial, followed by two experimental trials separated by a seven-day interval. Power output was analyzed at each stage between 100 W and 300 W. The upper limit of 300 W was chosen because some participants were unable to sustain higher power outputs. Additionally, PPO was evaluated independently. Power output was blinded throughout the test; cyclists received only time feedback, which was displayed on a computer screen connected to the smart trainer.

### 2.3.4 Reliability of the PPT

To assess test-retest reliability, participants completed the power profile test on three separate occasions, each spaced seven days apart. The first session served as a familiarization trial, followed by two experimental sessions. After a standardized 10-minute warm-up at 100 W identical to that used in the GXT cyclists performed a series of seven maximal efforts lasting 5, 15, 30, 60, 120, 300, and 1200 seconds, respectively. Each effort was followed by an active self-selected recovery period of 120, 180, 180, 300, 300, 300, and 1200 seconds, respectively. Participants were allowed to alternate between seated and standing positions to simulate real cycling conditions. They were instructed to perform each effort at the highest sustainable intensity for the given duration. To ensure consistency and experimental control, no verbal encouragement was provided. Additionally, power output was blinded throughout the test; only time feedback was available, displayed on a computer screen connected to the smart trainer.

## 2.4 Statistical Analysis

The Shapiro-Wilk test was applied to ensure a Gaussian distribution of the data. All results are presented as mean ± SD. Independent t-test was used to compare means of the maximal laboratory tests. For the GXT, a two-way repeated-measures ANOVA was performed to assess differences across test and retest conditions at each load (100 to 300 W). Similarly, for the PPT, a, two-way repeated-measures ANOVA was used across test and retest conditions for each effort duration (5 to 1200 s). In case of non-significant interaction, only the main effect of the test was considered. When test-by-loads (GXT) and test-by-efforts (PPT) interactions were significant, a post hoc one-way ANOVA was performed, and Bonferroni-adjusted paired t-test was used as appropriate to identify differences at specific load or effort points. In order to analyze the random error of the measurements during GXT and PPT test-retest, the typical error of measurement (TEM) was calculated in raw units, after logarithmic transformation in percentage units as CV (Hopkins, 2000). Additionally, the ICC was calculated and interpreted 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) (Hopkins, Marshall, Batterham, & Hanin, 2009). Confidence intervals (CI) of 95 % were calculated for TEM, CV, and ICC. The above calculations were performed using Hopkins' spreadsheet (Hopkins, 2015). The bias and limits of agreement (LoA) of the mean PO differences between PPO and PPT (5 to 1200 s) were defined using the method of Bland and Altman (Bland & Altman, 1986). The statistical significance was set at P<0.05.

## 3 Results

Table 1 displays the maximal physiological responses and performance outputs recorded during the GXT and PPT. Absolute maximal oxygen uptake ($VO_{2max}$) was significantly higher for the GXT group (4.1 ± 0.4 $L.min^{-1}$) compared to the PPT group (3.6 ± 0.6 $L.min^{-1}$). Relative $VO_{2max}$, maximal heart rate ($HR_{max}$), absolute and relative PPO were not significantly different between groups (Table 1).

The mean power output in each step during GXT Test 1 and Test 2 are shown in Table 2. Two-way ANOVA with repeated measures across loads and tests revealed no significant intensity-by-time interaction (P=0.76). However, the main effect showed that power output increased significantly overtime from 100 to 300 W. PPO was not significant different between tests (346 ± 33 W vs. 346 ± 31 W, P=0.33).

In addition, Table 2 shows the test–retest reliability for mean power output during the GXT (100 to 300 W). TEM ranged from 0.91 to 2.40 W across the stages, and CV decreased progressively as power output increased (from 5.5% at 100 W to 1.6% at 300 W). These results suggest improved consistency at higher workloads. The ICC values for the GXT ranged from trivial (–0.07 at 275 W) to moderate (0.57 at 225 W) indicating limited relative reliability, particularly in the presence of broad confidence intervals. Despite relatively small typical errors and low CVs at higher intensities (≥200 W), ICCs remained below 0.60, reflecting limited relative reliability under these conditions. PPO displayed a TEM of 0.13 W (0.10-0.19), a very low CV of 1.2 (0.9-1.7) and nearly perfect ICC 0.99 (0.96-0.99).

The mean power output in each effort during PPT Test 1 and Test 2 are shown in Table 3. Two-way ANOVA with repeated measures across efforts and tests revealed no significant intensity-by-time interaction (P=0.37). However, the main effect showed that power output decreased significantly over time from 15 to 1200 s**.** Importantly, there were no significant differences in power output between 5 s and 15 s (P = 0.19).

In addition, test–retest results for the PPT efforts are summarized in Table 3. Across all durations (5 to 1200 s), the PPT showed high absolute and relative reliability. TEM ranged from 0.21 W (5 s) to 0.41 W (15 s), and CVs varied depending on the duration of the effort. As expected, CVs were higher for shorter durations (7.7% at 5 s; 12.2% at 15 s), reflecting greater variability in short, maximal efforts. In contrast, longer efforts demonstrated lower CVs (e.g., 3.8% at 120 s and 3.6% at 1200 s). The ICCs were interpreted as large to nearly perfect across all PPT durations, ranging from 0.89 (15 s) to 0.97 (5 s and 1200 s). These results demonstrate strong relative reliability for repeated measures across sprint and endurance segments of the power profile.

**Table 1.** Maximal laboratory indexes from the participants.

| | GXT (n = 22)** | PPT (n = 10) | P value (equal variance not assumed) |
|---|---|---|---|
| **$VO_{2max}$ (L.min$^{-1}$)** | 4.1 ± 0.4 | 3.6 ± 0.6 | 0.03* |
| **$VO_{2max}$ (mL.kg$^{-1}$.min$^{-1}$)** | 55.7 ± 6.9 | 51.2 ± 6.9 | 1.09 |
| **$HR_{max}$ (bpm)** | 183 ± 6 | 186 ± 9 | 0.49 |
| **PPO (W)** | 346 ± 32 | 329 ± 48 | 0.34 |
| **PPO (W.kg$^{-1}$)** | 4.6 ± 0.5 | 4.8 ± 0.5 | 0.42 |

GXT: graded exercise test; PPT: power profile test; $VO_{2max}$: maximal oxygen uptake; $HR_{max}$: maximal heart rate; PPO: peak power output.
*Significantly different
**Data was averaged from test-retest.

**Table 2.** Test-retest reliability during graded exercise test.

| Power output (W) | Test 1 | Test 2 | TEM (95% CI) | CV (95% CI) | ICC (95% CI) |
|---|---|---|---|---|---|
| 100 | 99 ± 8 | 102 ± 7 | 0.97 (0.78-1.31) | 5.5 (4.4-7.4) | 0.53 (0.22-0.74) |
| 125 | 119 ± 7 | 121 ± 6 | 1.14 (0.88-1.63) | 4.3 (3.3-6.1) | 0.45 (0.05-0.73) |
| 150 | 146 ± 7 | 147 ± 6 | 1.14 (0.91-1.53) | 3.4 (2.7-4.7) | 0.45 (0.12-0.69) |
| 175 | 169 ± 6 | 171 ± 5 | 2.40 (1.92-3.22) | 3.0 (2.4-4.0) | 0.16 (-0.21-0.48) |
| 200 | 198 ± 5 | 199 ± 7 | 1.38 (1.10-1.88) | 2.7 (2.1-3.6) | 0.36 (-0.01-0.64) |
| 225 | 223 ± 7 | 222 ± 6 | 0.91 (0.73-1.24) | 2.0 (1.6-2.8) | 0.57 (0.26-0.77) |
| 250 | 247 ± 7 | 247 ± 5 | 1.20 (0.96-1.63) | 1.9 (1.5-2.6) | 0.43 (0.08-0.68) |
| 275 | 272 ± 6 | 272 ± 6 | 1.03 (0.82-1.40) | 2.2 (1.7-3.0) | -0.07 (-0.42-0.30) |
| 300 | 300 ± 7 | 299 ± 5 | 1.31 (1.04-1.77) | 1.6 (1.3-2.2) | 0.39 (00.3-0.66) |

TEM: typical error of measurement; 95% CI: 95% confidence interval; CV: coefficient of variation; ICC: intraclass correlation coefficient.

**Table 3.** Test-retest reliability during power profile test.

| Effort duration | Test 1 | Test 2 | TEM (95% CI) | CV (95% CI) | ICC (95% CI) |
|---|---|---|---|---|---|
| 5 (s) | 841 ± 311 | 850 ± 309 | 0.21 (0.14-0.28) | 7.7 (5.2-14.5) | 0.97 (0.88-0.99) |
| 15 (s) | 699 ± 229 | 684 ± 169 | 0.41 (0.28-0.74) | 12.2 (8.2-23.3) | 0.89 (0.62-0.97) |
| 30 (s) | 546 ± 129 | 531 ± 100 | 0.27 (0.19-0.49) | 6.0 (4.1-11.1) | 0.95 (0.81-0.99) |
| 60 (s) | 397 ± 80 | 402 ± 71 | 0.33 (0.22-0.59) | 6.2 (4.3-11.7) | 0.93 (0.74-0.98) |
| 120 (s) | 302 ± 46 | 302 ± 48 | 0.24 (0.16-0.43) | 3.8 (2.6-7.0) | 0.96 (0.85-0.99) |
| 300 (s) | 253 ± 42 | 250 ± 37 | 0.25 (0.17-0.46) | 4.0 (2.7-7.4) | 0.96 (0.84-0.99) |
| 1200 (s) | 221 ± 40 | 227 ± 33 | 0.22 (0.15-0.39) | 3.6 (2.4-6.6) | 0.97 (0.87-0.99) |

TEM: typical error of measurement; 95% CI: 95% confidence interval; CV: coefficient of variation; ICC: intraclass correlation coefficient.

Bland–Altman plots (Figures 2 and 3) were used to assess agreement between test and retest for PPO each PPT duration. All efforts showed small biases, and the mean differences ranged from –8.5 W (5 s) to 15.7 W (15 s). The narrowest LoA were found in longer efforts: 300 s (3.3 ± 12 W, LoA: –21 to 27 W) and 1200 s (–4.4 ± 11 W, LoA: –25 to 17 W). Wider LoAs were observed in shorter efforts, such as 15 s (LoA: –157 to 188 W) and 5 s (LoA: –139.1 to 122.1 W), indicating greater absolute variability in short-duration, high-power output efforts.

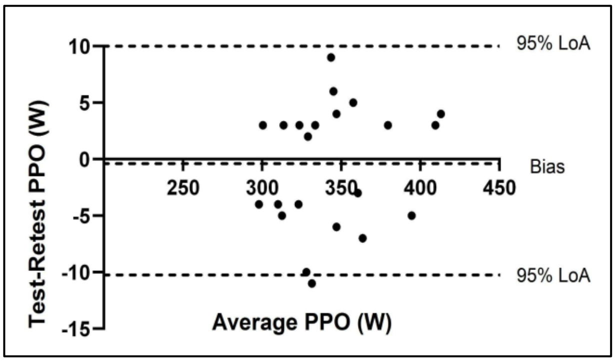


**Figure 2.** Bland-Altmann plots of peak power output during graded exercise test.

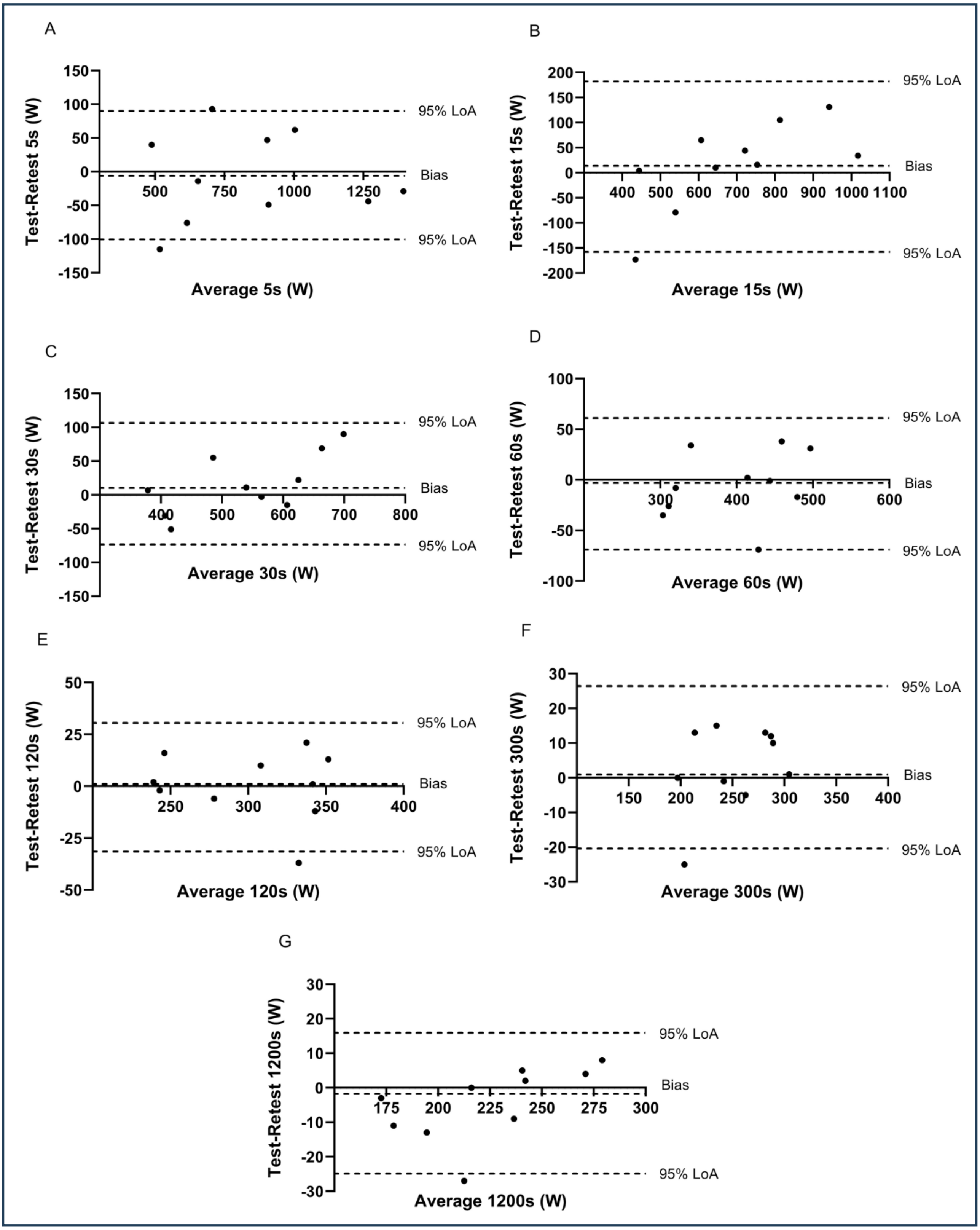


**Figure 3.** Bland-Altmann plots of mean power output during power profile test (5s - A; 15s - B; 30s - C; 60s - D; 120s - E; 300s - F; 1200s – G.

# 4 Discussion

The present study aimed to assess the test–retest reliability of power output from the commercially available Cyclotronics Smart Trainer during a GXT and a PPT in trained cyclists. The main findings demonstrate high reliability for most parameters across both testing protocols, with the PPT showing acceptable test–retest agreement across all durations, particularly for efforts ≥120 seconds. These results support the use of the Cyclotronics Smart Trainer for repeated performance assessments.

Power outputs recorded during the GXT showed ICC values ranging from trivial to moderate, indicating limited relative reliability. However, the CV during the GXT ranged from 5.5% at the lowest workload (100 W) to 1.6% at the highest workload (300 W). Additionally, a very low CV was observed for PPO, indicating reduced variability as exercise intensity increased. This pattern aligns with previous findings showing that electromagnetic resistance-based trainers tend to exhibit greater variability at lower loads, likely due to mechanical limitations and sensor sensitivity at minimal resistance levels (Sparks, Williams, Massey, Bridge, Marchant, & McNaughton, 2016; Zadow, Kitic, Wu, Smith, & Fell, 2016). This trend of improved precision at higher intensities has also been reported for other smart trainers, such as the Wahoo KICKR, which typically present CV values below 3% during incremental protocols and short time trials at moderate to high intensities (Hoon et al., 2016; Zadow et al., 2016). Similar findings have been observed in direct-drive systems such as the Tacx Neo 2T, which demonstrated good accuracy and reliability, as well as strong agreement with external power measurement devices across a range of submaximal intensities (Morais, Bragada, Magalhães, & Marinho, 2024).

In the PPT, the Cyclotronics Smart Trainer showed excellent reliability for efforts lasting longer than 120 seconds, with CV values ranging from 3.6% to 4.0% and ICC values generally exceeding 0.96. Moreover, the narrowest bias and LoA were observed in longer efforts compared to shorter ones. Therefore, for durations between 120 and 1200 seconds, the results demonstrated excellent consistency. These findings reinforce the potential application of the Cyclotronics for assessing aerobic capacity, including indirect estimations of $VO_{2max}$, critical power, and the lactate threshold or Functional Threshold Power (FTP), which are widely used for training prescription and performance monitoring in cycling (Borszcz, Tramontin, Bossi, Carminatti, & Costa, 2018; Borszcz, de Aguiar, Costa, Denadai, & de Lucas, 2024).

Short-duration efforts during the PPT demonstrated high variability, with elevated CV values and wide LoA, despite high ICCs. This observation is consistent with previous literature indicating that power measurements during very short efforts are more susceptible to fluctuations in flywheel acceleration and the response time of electromagnetic resistance systems (Bertucci, Duc, Villerius, & Grappe, 2005; Peiffer & Losco, 2011). These limitations are particularly relevant when compared to laboratory-grade ergometers such as isokinetic cycle ergometers (e.g., Lode Excalibur Sport), which typically report errors below 2% even during maximal sprints (Bertucci, Duc, Villerius, & Grappe, 2005). Although the Cyclotronics does not achieve this level of accuracy for very short sprints, the observed variability is still considered acceptable for performance monitoring, especially when weighing the trade-offs between practicality, cost, and portability (McIlroy, Passfield, Holmberg, & Sperlich, 2021). A possible explanation for the greater variability in the shortest efforts lies in the strong dependency on initial cadence, reaction time, and instantaneous torque application, all of which introduce variability unrelated to the device's measurement precision. Thus, while the Cyclotronics Smart Trainer provides reliable data for moderate to long efforts within acceptable thresholds for performance applications, practitioners should interpret single sprint values with caution or consider averaging multiple trials.

Additionally, the findings of this study corroborate the growing trend of using smart trainers not only for training but also as valid and reliable tools for remote physiological assessments (McIlroy, Passfield, Holmberg, & Sperlich, 2021). This trend was further accelerated by the COVID-19 pandemic, which emphasized the need for home-based physiological testing in diverse contexts. In this regard, the reliability demonstrated by the Cyclotronics supports its use in telemonitoring, remote coaching, and even in scientific research conducted outside traditional laboratory environments broadening possibilities for coaches, practitioners, and researchers alike. A noteworthy strength of the present study is its dual approach combining a GXT and a PPT. This comprehensive assessment provides valuable insights into both endurance and high-intensity performance domains, which are increasingly relevant in modern cycling training models that emphasize polarized and high-intensity interval training strategies (Seiler, 2010; Stöggl & Sperlich, 2015).

## 5 Practical Applications

The present findings have several practical implications for both coaches and athletes. The Cyclotronics Smart Trainer demonstrates sufficient reliability for performance testing and training prescription in trained cyclists, particularly for efforts lasting 120 seconds or longer during PPT. This makes it suitable for assessing endurance capacity, power profiling, and estimations of $VO_{2max}$, critical power, and FTP in both laboratory and home-based settings. Given the increasing use of virtual training platforms and remote coaching, the device provides a practical and accessible tool for tracking training progress without the need for expensive laboratory-grade equipment. This is especially relevant for amateur athletes, development squads, and regions with limited access to sports science resources.

Moreover, the acceptable variability observed in short sprint efforts suggests that the device can be cautiously used to monitor sprint power output, provided that coaches or practitioners apply repeated measures or trend-based analysis rather than relying on single absolute values. This enables the device to support a comprehensive athlete profile encompassing both aerobic and anaerobic components.

However, some limitations must be acknowledged. First, while the GXT included a robust sample size (n = 22), the power-profile analysis involved only 10 participants, including two female cyclists and one triathlete. Although the overall reliability indices were excellent, the limited sample size may restrict the generalizability of the findings to broader athletic populations. Therefore, future studies should examine whether the reliability observed here is consistent across larger and more diverse samples, including female cyclists, junior athletes, and elite competitors.

Second, this study focused on the device's reliability but did not include direct validity comparisons against gold-standard power meters (e.g., SRM, Quarq, PowerTap) or laboratory-grade ergometers. While the device shows excellent consistency across repeated tests, its absolute accuracy relative to external standards remains unconfirmed.

Third, cadence was not standardized during either the GXT or PPT protocols. As demonstrated in previous validation studies of cycling ergometers and smart trainers, cadence can influence both validity and reliability outcomes due to its interaction with torque production and resistance dynamics. However, this limitation was primarily related to the technological constraints of the Cyclotronics trainer, which does not provide integrated cadence monitoring or external cadence control capabilities. Therefore, cadence variability may have contributed to measurement variability and should be considered when interpreting the reliability outcomes.

Another limitation of the present study is the absence of inter-instrument variability assessment. Only a single Cyclotronics smart trainer unit was used throughout the testing procedures. Consequently, it was not possible to determine the extent to which measurement error might vary when switching between different units of the same brand and model. Future investigations should examine between-unit reliability to establish the consistency of power output measurements across multiple devices.

Lastly, the GXT was limited to workloads up to 300 W and PPO, which may not fully reflect the performance demands of highly trained or elite cyclists, especially during efforts at higher absolute intensities. Future research should explore the device's performance at higher workloads and across different test protocols such as ramp-to-exhaustion or long stepwise protocols.

## 6 Conclusions

The Cyclotronics Smart Trainer demonstrates high test–retest reliability for both GXT and PPT in trained cyclists, particularly for efforts lasting 120 seconds or more. Although short sprint efforts show greater variability, the values remain within acceptable limits for practical use, provided they are interpreted with caution. These findings support the use of the Cyclotronics Smart Trainer as a reliable tool for performance assessment and training monitoring in both laboratory and remote settings. Nonetheless, further research is needed to confirm the device's validity across different populations and under higher performance demands.

**Funding:** This research was funded by Foundation for Research and Innovation Support of the State of Santa Catarina (FAPESC).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bertucci, W., Duc, S., Villerius, V., & Grappe, F. (2005). Validity and reliability of the Axiom PowerTrain cycle ergometer when compared with an SRM powermeter. *International Journal of Sports Medicine, 26*(1), 59–65. DOI: 10.1055/s-2004-817855

2. Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307–310. DOI: 10.1016/S0140-6736(86)90837-8

3. Borszcz, F. K., Tramontin, A. F., Bossi, A. H., Carminatti, L. J., & Costa, V. P. (2018). Functional threshold power in cyclists: Validity of the concept and physiological responses. *International Journal of Sports Medicine, 39*(10), 737–742. DOI: 10.1055/s-0044-101546

4. Borszcz, F. K., de Aguiar, R. A., Costa, V. P., Denadai, B. S., & de Lucas, R. D. (2024). Agreement Between Maximal Lactate Steady State and Critical Power in Different Sports: A Systematic Review and Bayesian's Meta-Regression. *Journal of strength and conditioning research*, *38*(6), e320–e339. DOI: 10.1519/JSC.0000000000004772

5. Bouillod, A., Soto-Romero, G., Grappe, F., Bertucci, W., Brunet, E., & Cassirame, J. (2022). Caveats and recommendations to assess the validity and reliability of cycling power meters: A systematic scoping review. *Sensors*, 22(1), 386. DOI: 10.3390/s22010386

6. Hoon, M. W., Michael, S. W., Patton, R. L., Chapman, P. G., & Areta, J. L. (2016). A Comparison of the Accuracy and Reliability of the Wahoo KICKR and SRM Power Meter. *Journal of Science and Cycling*, *5*(3), 11-15. https://jsc-journal.com/index.php/JSC/article/view/240

7. Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine, 30*(1), 1–15. DOI: 10.2165/00007256-200030010-00001

8. Hopkins, W. G. (2015). Spreadsheets for analysis of validity and reliability. *Sportscience, 19*, 36–42. https://sportsci.org/2015/ValidRely.htm

9. Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and science in sports and exercise*, *41*(1), 3–13. DOI: 10.1249/MSS.0b013e31818cb278

10. Jamnick, N. A., Pettitt, R. W., Granata, C., Pyne, D. B., & Bishop, D. J. (2020). An examination and critique of current methods to determine exercise intensity. *Sports Medicine, 50*(10), 1729–1756. DOI: 10.1007/s40279-020-01322-8

11. Jeker, D., Gosselin, J., Drouet, J. M., & Goulet, E. D. B. (2021). Validity and reliability of the Computrainer Lab™ during simulated 40 and 100 km time-trials. *Frontiers in Sports and Active Living*, *3*, 735046. DOI: 10.3389/fspor.2021.735046

12. Lamberts, R. P., Swart, J., Noakes, T. D., & Lambert, M. I. (2009). Changes in heart rate recovery after high-intensity training in well-trained cyclists. *European Journal of Applied Physiology*, *105*(5), 705–713. DOI: 10.1007/s00421-008-0952-y

13. McIlroy, B., Passfield, L., Holmberg, H. C., & Sperlich, B. (2021). Virtual training of endurance cycling: A summary of strengths, weaknesses, opportunities and threats. *Frontiers in Sports and Active Living*, *3*, 631101. DOI: 10.3389/fspor.2021.631101

14. Morais, J. E., Bragada, J. A., Magalhães, P. M., & Marinho, D. A. (2024). The accuracy and reliability of the power measurements of the TACX Neo 2T Smart Trainer and its agreement against the Garmin Vector 3 Pedals. *Journal of Functional Morphology and Kinesiology, 9*(3), 138. DOI: 10.3390/jfmk9030138

15. Peiffer, J. J., & Losco, B. (2011). Reliability and validity of the Fortius trainer. *International Journal of Sports Medicine, 32*(5), 353–356. DOI: 10.1055/s-0031-1271772

16. Pinot, J., & Grappe, F. (2011). The record power profile to assess performance in elite cyclists. *International Journal of Sports Medicine, 32*(11), 839–844. DOI: 10.1055/s-0031-1279773

17. Seiler, S. (2010). What is best practice for training intensity and duration distribution in endurance athletes? *International Journal of Sports Physiology and Performance*, *5*(3), 276–291. DOI: 10.1123/ijspp.5.3.276

18. Sparks, S. A., Williams, E., Jones, H., Bridge, C., 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. https://jsc-journal.com/index.php/JSC/article/view/272

19. Stöggl, T. L., & Sperlich, B. (2015). The training intensity distribution among well-trained and elite endurance athletes. *Frontiers in Physiology*, *6*, 295. DOI: 10.3389/fphys.2015.00295

20. Zadow, E. K., Kitic, C. M., Wu, S. S. X., & Fell, J. W. (2018). Reliability of Power Settings of the Wahoo KICKR Power Trainer After 60 Hours of Use. *International Journal of Sports Physiology and Performance*, *13*(1), 119–121. DOI: 10.1123/ijspp.2016-0732