Validity and Reliability of the PowerCal Device for Estimating Power Output During Cycling Time Trials

Article in The Journal of Strength and Conditioning Research · April 2016
DOI: 10.1519/JSC.0000000000001466

CITATIONS 5
READS 78

3 authors:

Vitor Pereira Costa
Santa Catarina State University (UDESC), Flor...
36 PUBLICATIONS 73 CITATIONS

Luiz Guilherme Antonacci Guglielmo
Federal University of Santa Catarina
142 PUBLICATIONS 687 CITATIONS

Carl D Paton
Eastern Institute of Technology
51 PUBLICATIONS 1,069 CITATIONS

Some of the authors of this publication are also working on these related projects:

Características fisiológicas de remadores e sua relação com o teste de 2000 m em remoergômetro
View project

The effect of prior exercise on the VO2 kinetics View project
Validity and reliability of the PowerCal device for estimating power output during cycling time trials

Running head: PowerCal device for estimating power output in cycling

The research was conducted at Eastern Institute of Technology (EIT) in the Exercise Physiology Laboratory of Faculty of Health and Sport.

Vitor P. Costa¹, Luiz Guilherme A. Guglielmo², Carl D. Paton³

¹Center for Health and Sport Science, Santa Catarina State University, Florianópolis, Santa Catarina State, Brazil.
²Physical Effort Laboratory, Federal University of Santa Catarina, Florianópolis, Brazil.
³Faculty of Health and Sport, Eastern Institute of Technology, Napier, New Zealand.

Corresponding author:
Vitor Pereira Costa
Center for Health and Sport Science, Santa Catarina State University.
Rua Pascoal Simone, 358, Coqueiros, Florianópolis, Santa Catarina State, Brazil.
Phone: +55 48 9649-1169
email: costavp2@yahoo.com.br
Abstract:
This study assessed the validity and reproducibility of the PowerCal device for estimating power output (PO) during cycling hilly time trial (TT). Twenty-one well-trained male cyclists performed an incremental exercise test and three 20-km TT (34.1 ± 10.6 years; 73.2 ± 3.2 kg; 176.8 ± 6.2 cm; maximal power output, 334 ± 31 W; maximal oxygen uptake, 61.0 ± 4.2 mL.kg⁻¹.min⁻¹). The first TT was used as familiarization and the tests were separated by at least 72 hours. Mean PO over the 20-km TT was significantly greater for the Velotron (282 ± 27 W) compared to the PowerCal (242 ± 28 W). The mean power over each kilometer of the trial ranged from 5.8% to 23.4% greater on the Velotron than on the PowerCal. High within-subject variation between the trials was substantially greater for the PowerCal (4.9%) compared to the Velotron (1.8%). High coefficient of variation scores for the Velotron test-retest were found to be concentrated in the beginning and final meters of the TT (~ 6.0%) while the scores were lower in the middle of the trials (~ 3.0%). By contrast, the PowerCal test-retest achieved high coefficient of variation (~ 6.0%) in each km over the TT. Thus, the PowerCal device should be used with caution during cycling activities since it is not reliable and underestimates PO.

Keywords: Reproducibility, power output, cycling.

INTRODUCTION

The frequency and duration of the training sessions in cycling are components of the training program that are relatively easy to control, however, the measurement of exercise intensity is more complex due to the stochastic nature of cycling (13). Cycling is a sport where it is possible to measure power output (PO) directly through the use of mobile power meters, which have become available in the past decade (11). The majority of competitive cyclists are not professionals and they normally use an affordable device to monitor and control training sessions. Heart rate (HR) monitors are affordable and provide useful information about the exercise intensity during aerobic training (2, 8). However, HR scores may not reflect the demands of the activity due to hot environmental conditions, dehydration states, level of the ground, and sprint activities (2, 13).
Recently, a new device called the PowerCal (Cyclops, Madison, USA) was developed which estimates PO by measuring the HR response during exercise (10). Essentially the PowerCal estimates PO using an algorithm developed in order to diminish the variations between PO and HR during exercise. Thus, it is believed that PowerCal predicts accurate PO for aerobic intervals (i.e. ≤ maximal HR). Also, the device is a low-cost method (~ 99 US dollars) compared with the other available power meters (~ 700-3000 US dollars) to detect the magnitude of the power level on a given ride. Furthermore, the PowerCal device is simple to use (i.e. HR strap) and does not require any calibration prior to use. Collectively, these features seem to be an advantage for competitive cyclists to monitor training sessions and racing.

To date there are very few studies assessing the validity and reliability of the PowerCal device. In one recent study examining the PowerCal PO in cycling (10), Costa et al. (10) reported that the PowerCal PO during sprints of 15 s, 30 s, and 45 s showed high within-subject variation (6.7-21.5%) in well-trained cyclists. Also, the authors found high bias ranging from 32 to 129 W and limits of agreement ranging from -236 to 303 W between the PowerCal and the cycle ergometer Velotron (10). In addition the maximal and mean PO during sprints were underestimated by 6.6% to 13.9% and 14.9% to 27.6%, respectively, when compared to the Velotron. It is not clear how the algorithm formula estimates PO for anaerobic stimulus. The HR is not recommended to control such intensity due to a natural slow response in the first seconds of the maximal sprints exercise (2, 8). In this regard, it is not surprising that predictive values for PowerCal during sprints would give inaccurate values compared to the Velotron due to rapid and almost instantaneous PO responses of the cycle ergometer. Thus, from the
first perspective, the estimates of the PO of the PowerCal in short maximal activities is dubious (10); however due to a linear relationship between HR and PO during aerobic exercise (4) it is hypothesized that PowerCal estimates accurate scores of PO compared with the cycle ergometer.

A common measurement of cycling performance is the time trial (TT) races (16). In a TT competition, cyclists race maximally against the clock completing a pre-established distance in the shortest time possible. The majority of TT events cover a distance between 5 to 60 km and are performed individually (3) except in grand tour and velodrome events that include team TT. The distribution of PO can be influenced by a number of factors such as training, nutritional strategies, aerodynamics, ground level of the terrain, environmental conditions and psychological factors (3,1). The mathematical models of cycling performance showed that athletes should vary the distribution of PO in response to changes in environmental resistance, as found in the hilly TT (7). Studies investigating the capability of the PowerCal to estimate PO during hilly TT were not found and could provide important insights showing the level of accuracy of the PO over a task from the beginning to the end point. Therefore, the aim of this study was to determine the validity and reproducibility of the PowerCal device for estimating PO during hilly cycling TT.

METHODS

Experimental approach to the problem

To investigate the validity and reproducibility of the PowerCal device for estimating PO during hilly cycling TT, an incremental exercise test and three TT were carried out
using the reliable and valid cycle ergometer Velotron (Velotron Dynafit Pro, RacerMate Inc, WA, USA) (5,9,15,17,20,21). To this purpose, twenty-one well-trained male cyclists participated in four different sessions. In the first session, the subjects performed an incremental exercise test. The second session, cyclists were familiarized with the self-paced TT, whereas in the third and fourth sessions they also performed a self-paced TT using the PowerCal device.

**Subjects**

Twenty-one well-trained male cyclists (34.1 ± 10.6 years; 73.2 ± 3.2 kg, 176.8 ± 6.2 cm; maximal power output, 334 ± 31 W; maximal oxygen uptake, 61.0 ± 4.2 mL.kg\(^{-1} \text{.min}^{-1}\)) volunteered to participate in this study. The cyclists had a minimum amount of experience from three years of regular competitions. The study was performed in the competitive season following a period of base and pre competition training. Due to the nature of each cyclist’s competition program it was not possible to control their individual training leading up to the study. However immediately prior (2 weeks) to the start of the study cyclists were completing individual self or coach-determined training regimes consisting of a minimum of ten hours (300-km) mixed intensity training per week. At the start of laboratory testing cyclists were required to be in a well-prepared and non fatigued state. All cyclists were informed of the purpose and the risks associated with participation before giving their written informed consent to participate. The study was approved by the institutional research ethics committee in accordance with the declaration of Helsinki and ACSM guidelines.

**Procedures**

All cyclists had previously participated in laboratory cycle ergometer testing and were
familiar with general exercise testing procedures. Cyclists reported to the laboratory on four separate occasions over a period of two weeks. During the initial visit to the laboratory, cyclists completed an incremental exercise test until volitional exhaustion to determine peak power output and maximal oxygen uptake. Following the initial test, cyclists completed three 20-km cycling TT sessions each separated by at least 72 hours. All testing sessions were conducted on an electronically braked cycle ergometer Velotron set up to replicate the cyclists’ individual bicycle position. Furthermore, during the tests cyclists used the PowerCal (Cyclops, Madison, USA), a heart rate strap that calculates PO from the HR response recorded during exercise. All testing was conducted in a laboratory under controlled environmental conditions. Air temperature and relative humidity were 20.4 ± 1.7°C and 53.1 ± 2.9%, respectively.

**Incremental exercise test**

Cyclists completed an incremental exercise test until volitional exhaustion to determine their physiological parameters. The cycle ergometer was adjusted to replicate the participant’s preferred racing position, which was recorded and replicated for the performance sessions. All tests were conducted on an electronically braked cycle ergometer. Cyclists performed a 15 min warm-up at a self-selected intensity followed by 5 mins of rest. Thereafter, the incremental exercise test started at 100 W and PO was increased at a rate of 40W every 4 mins until volitional exhaustion. The participants were instructed to maintain their preferred cadence ~ 90 rpm. If the final stage of the exercise test was not completed, the PPO was calculated using the equation of (14):

\[ \text{PPO} = \text{Pf} + (t/240 \times 40) \]

Pf was the last completed workload, t is the time in seconds of the uncompleted
workload, 240 is the time of each stage in s, and 40 is the workload augments in each stage. The expired respiratory gases were collected into metabolic Metamax 3B system (Cortex, Leipzig, Germany). Prior to each test, the system was calibrated in accordance with the manufacturers instructions using known alpha gas standards. VO$_2$max was defined as the highest oxygen uptake over a 30 s value recorded during the test. The (HR) was registered by using a coded strap and recorded by the metabolic system.

**Time trials**

Cyclists completed three hilly cycling TT on a computer simulated 20-km 3D course using the same cycle ergometer as used previously in the incremental exercise test. The cycle ergometer allowed the cyclists to manually shift the gears attached in the handle bar while pedaling. For each test, the Velotron was connected to a laptop computer interfaced with a projector that displayed the computer-generated image of the 3D course profile in front of the cyclist. The screen in front of the cyclist gives information about the current speed, power, cadence, heart rate, draft, distance, and time. In our study, we blinded the main variables to prevent pacing strategy. Participants were able to view their progress over the course on a computer monitor and were provided with information on distance completed and gear selections. Cyclists initially completed a 20 minute standardised warm up consisting of three repeated increasing intensity bouts. The first two minutes were completed at 2-2.5 W.kg$^{-1}$, followed by two minutes at 3-3.5 W.kg$^{-1}$ and finally one minute at 4-4.5 W.kg$^{-1}$ and repeated consecutively. For the final five minutes cyclists pedalled at a fixed intensity of 100 W. Thereafter, a 20-km self-paced maximal TT was performed. The first TT was used as a familiarization. The TT was completed on a designed course, which replicated a typical racing circuit and
contained numerous changes in gradient represented by both ascents and descents (9). The total elevation gain over the 20-km was 300 meters leading to an average gradient of ~1.5% (9). Furthermore, no verbal encouragement was provided that could possibly interfere with an individual’s pacing effort. Participants were requested to complete each TT as quickly as possible with no restriction on gear selection, cadence or cycling posture (seated or standing). Cyclists used the PowerCal HR strap. Experimental sessions were conducted at the same time of day for each individual in order to allow for diurnal variation and were separated by at least 72 hours. Moreover, cyclists were required to present themselves in a hydrated and non-carbohydrate depleted state. Throughout the experimental sessions, cyclists were cooled with standing floor fans and permitted to consume only water ad libitum.

**Statistical Analysis**

Descriptive statistics are presented as means (± SD). To display the PO pattern response during the TT, PO from the Velotron and the PowerCal were averaged for each interval of 1km. PO during the trials was compared using a two-way analysis of variance with repeated measures, with TT session (1 vs. 2) and distance duration (km). When necessary, subsequent post-hoc comparisons were made using Bonferroni correction. PO of the Velotron vs. PowerCal was compared from the mean scores of the two TT sessions from each device. Two-way analysis of variance with repeated measures was used across the two devices (Velotron and PowerCal) and distance duration (km) with subsequent post-hoc comparisons (Bonferroni) when necessary. The bias and limits of agreement (LoA) of the mean PO differences between PowerCal and Velotron were defined using the method of Bland and Altman (6). Reliability was quantified using the
typical error between successive trials as percentages of coefficient of variation (CV) and intra-class coefficient of correlation (ICC) derived from log-transformed data (12). The confidence intervals (CI) were fixed at 90%. Statistical significance was accepted at \( p < 0.05 \).

RESULTS

The Velotron and PowerCal mean ± SD for the test-retest mean PO, CV, and ICC values are presented in table 1. There were no significant differences between Velotron POTT1 vs. POTT2 (\( p = 0.07 \)). However, significant differences were found between PowerCal POTT1 vs. POTT2 (\( p = 0.02 \)). The mean PO from Velotron was significantly higher than PowerCal (\( p = 0.0001 \)). Low scores were found for the mean PO of the Velotron’s CV (1.8%) and high scores for the mean PO of the PowerCal’s CV (4.9%). High values of ICC were found for the mean PO between the TT under the Velotron (0.96) and PowerCal (0.82), respectively.

XXX Table 1 XXX

High CVs scores for the Velotron test-retest were found to be concentrated in the beginning and final meters of the TT ~ 6.0% while the CV’s were lower in the middle of the trials ~ 3.0% (Figure 1A). By contrast, The PowerCal CVs test-retest achieved high scores (~ 6.0%) in each km over the TT (Figure 1B).

XXX Figure 1 XXX
The PO during each kilometer of the 20-km cycling test for the Velotron vs. PowerCal is shown in Figure 2. Regarding the distribution of PO from the Velotron, cyclists adopted a fast start with the first 4 km above the mean PO (282 ± 27 W), the remaining distance was below mean PO, except in the 6th, 10th, and final 2 km of the test. The PO in each kilometer over the test was significantly (p < 0.01) higher for the Velotron (5.8 to 23.4%) than that produced by using the PowerCal device. The PowerCal distribution of PO showed higher PO above the mean PO (242 ± 28 W) in the first, 11th and final 3 km of the test.

The Bland-Altman plots of mean PO between Velotron vs. PowerCal are presented in Figure 3. The data showed that the mean PO bias was 32 W; and LoA was ranging from -51 to 115 W.

**DISCUSSION**

The aim of this study was to assess the reliability and validity of PO distribution from the PowerCal device during a 20-km hilly cycling TT in competitive cyclists. The finding of this study was that the mean PO of the PowerCal showed high within-subject variation (4.9%). Also, high CV scores were found during each km between the TT test-retest. The mean PO of the PowerCal was underestimated by 15% compared to the
Velotron cycle ergometer. In addition, the PowerCal PO in each km over the TT was underestimated by 5.8% to 23.4%. The mean PO showed higher bias and LoA between the PowerCal and Velotron. Therefore, the PowerCal device displayed poor reliability and the hypothesis of this study was refused because the device is unlikely to be valid during hilly cycling TT.

The time-course of the reliability of mean PO during hilly TT investigated in this study was previously demonstrated in a group of competitive cyclists (9). Clark et al. (9) found that mean PO reliability from Velotron was close to 2% when a short period between the trials was used; however they reported a substantial decrease in reliability of cycling performance when increasing time between trials (~ 3.5%). In our study, the mean PO showed a similar low within-subject variation (1.8%) for the Velotron test-retest. Sporer and McKenzie (17) evaluated the reliability of a 20-km TT using the Velotron and found a CV for the mean PO of 2.1% between Trials 1 and 2 and a CV of 1.9% between Trials 2 and 3. In a further study, Zavorsky et al. (21) found a little higher CV of 3.4% between Trials 1 and 2, and a CV of 3.6% between Trials 2 and 3 for a repeated 20-km TT using the Velotron. However, for the eight top performers the reliability improved (CV of 1.2%). Noreen et al. (15) tested an 8-mile uphill TT using the Velotron and reported a CV for the mean PO of 3.5% between Trials 1 and 2, and a CV of 1.7% between Trials 2 and 3. In contrast to the previous studies using the Velotron, the reliability of the mean PO from the PowerCal device was poor with a high CV score of 4.9%. To the best of our knowledge, we found only one study from our group regarding the reliability of the PowerCal device in cycling (10). Costa et al. (10) found that the mean sprint PO estimation from the PowerCal was also unreliable for
maximal intervals of 15 s, 30 s and 45 s with CV scores of 13.5%, 6.5% and 7.0%; respectively. Taken together, the studies above clearly reported that a short period of cycling TT test-retest using the Velotron in a simulation of a variety of terrain is reliable in well-trained cyclists while the PowerCal device is non-reproducible.

The majority of the studies that have investigated the reproducibility of the performance measures during cycling TT have focused on the mean PO and time to complete the trial without analyzing the pacing strategy during the tests. Thomas et al. (2012) reported a high degree of variability at the start and finish of the 20-km cycling TT using the Velotron. Moreover, the authors found a parabolic “U” shape curve for the PO that means a fast start and high end spurt. According to this, we found high CV values for the Velotron during the start and finish of the trials combined with high PO in the beginning and in the final meters of the trials. Therefore, it was demonstrated in our study that the cyclists displayed a fast start followed by a slow decrease in PO with a high end spurt. In contrast, the CV scores of the PowerCal were higher over the test. Furthermore, the PowerCal distribution of PO over the test was nearly flat with a slight increase in the final meters of the trials.

The mean PO estimated from the PowerCal (242 ± 28 W) was significantly lower compared to the cycle ergometer (282 ± 27 W; p < 0.001). During each km of the performance test, the PowerCal distribution of PO was significantly lower compared to the Velotron (5.8% to 23.4%; p < 0.01). Moreover, the mean PO of the tests showed high bias and LoA between the device and the cycle ergometer (Figure 3). Costa et al. (2015) reported that the estimates of PO from PowerCal during short sprints in cycling
were significantly lower compared to the Velotron ranging from 6.6% to 27.6%. The authors found high bias (32 W to 129 W) and LoA (236 W to W 303) between the peak and mean PO of the PowerCal and Velotron. Taken together, the significant differences found in the PO values combined with high bias and LoA of the PO between the PowerCal and the Velotron showed that PowerCal is unlikely to be valid during a hilly TT as well as during short maximal sprints of 15 s, 30 s, and 45 s in well-trained cyclists (10).

Several factors have influenced the PowerCal reliability and validity. First, the PowerCal device contains a secret embedded algorithm for calculating PO based on cycling-specific HR and time data. Actually, the PowerCal estimation of PO is related with the individual HR score measured while cycling. Individual PowerCal data from our study shows that cyclists with low HR values during the TT displayed low PO (i.e. older and fittest cyclists). In contrast, cyclists with high HR produced high PO (i.e. young and less trained cyclists). For untrained and lower physical fitness subjects, it is expected that with the adaptations of a regular aerobic training program, the HR will decrease at rest and during sub-maximal exercise while sub-maximal and maximal PO will increase. Even though we did not find any training study that detected changes in PO using PowerCal, we hypothesized that the PowerCal estimates of PO will probably produce lower scores after training if the sub-maximal HR decreases.

Second, cycling is a stochastic sport where a variety of factors can influence the relationship between HR and PO, and consequently influence the PO estimates from the PowerCal device (2,13). In a hilly TT, the level of the ground is the major factor that
could influence the relationship between HR and PO (19). The TT in the present study was not flat and was based upon numerous changes in gradient represented by both ascents and descents (9). We found a drop in the Velotron PO characterized by the long descent of the TT regardless of changes in the HR and PO estimated by the PowerCal. The decrease in PO was due to the resistive forces on the flywheel of the cycle ergometer simulating the downhill phase of the course profile. During steep up-hill and downhill climb the dissociation between HR and PO increases (18). In fact, linear relationship between HR and PO was found only in a controlled testing condition (i.e. graded exercise test) (4). Therefore, it is not surprising that the estimates of PO from PowerCal would be different compared to the Velotron.

PRACTICAL APPLICATIONS

PowerCal is a device that estimates PO from HR response during exercise. The results from our study showed that the PowerCal is a non-reproducible device and also under predicts the PO compared to the traditional cycle ergometer Velotron during hilly TT. The PowerCal was designed to be an affordable tool to monitor and control PO during cycling training and racing. Indeed, it has a low price and does not need any calibration prior to training. However, our previous sprint study (10) and the current study demonstrated that the PowerCal device should be used with caution during cycling activities since it is not reliable and under estimates PO.

REFERENCES


13. Jeukendrup, A and Van Diemen, A. Heart rate monitoring during training and


Figure 1. Velotron and PowerCal coefficient of variation during a 20-km hilly time-trial.

Figure 2. Velotron and PowerCal distribution of power output during 20-km hilly time trial.

Figure 3. Bland-Altman plots between the mean power output of the Velotron and PowerCal.
Table 1. Mean power output, coefficient of variation, and intra-class coefficient of correlation from PowerCal and Velotron during the time trials.

<table>
<thead>
<tr>
<th></th>
<th>Velotron</th>
<th>PowerCal</th>
</tr>
</thead>
<tbody>
<tr>
<td>POTT1 (W)</td>
<td>280 ± 26</td>
<td>246 ± 28</td>
</tr>
<tr>
<td>POTT2 (W)</td>
<td>284 ± 29</td>
<td>237 ± 28*</td>
</tr>
<tr>
<td>Mean (W)</td>
<td>282 ± 27</td>
<td>242 ± 28**</td>
</tr>
<tr>
<td>CV (%)</td>
<td>1.8 (1.4 – 2.7)</td>
<td>4.9 (3.1 – 7.5)</td>
</tr>
<tr>
<td>ICC</td>
<td>0.96 (0.91 – 0.99)</td>
<td>0.82 (0.58 – 0.96)</td>
</tr>
</tbody>
</table>

PO: power output; TT: time trial; CV: coefficient of variation; ICC: intra-class correlation. * Significant different (P<0.05); ** Significant different (P<0.001).