Isokinetics and Exercise Science 23 (2015) 127–132
DOI 10.3233/IES-150573
IOS Press

Reproducibility and validity of the PowerCal device for estimating power output during sprints in well-trained cyclists

Vitor P. Costa[a,*], Luiz Guilherme A. Guglielmo[a] and Carl D. Paton[b]
[a]*Physical Effort Laboratory, Faculty of Physical Education, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil*
[b]*Faculty of Sports Science, Eastern Institute of Technology, Napier, New Zealand*

Received 7 September 2014
Accepted 7 February 2015

Abstract.
BACKGROUND: The PowerCal device purports to effectively estimate power output (PO) using an algorithm computed from the HR response recorded during exercise.
OBJECTIVE: To assess the reproducibility and validity of PO estimation using the PowerCal during sprints.
METHODS: Ten well-trained male cyclists (maximal oxygen uptake -62.7 ± 5.9 ml $\cdot$ kg^{-1} $\cdot$ min^{-1}) performed four separate experimental trials over a two-week period. Cyclists completed an incremental exercise test until volitional exhaustion followed by three maximal sprint sessions on separate days. Each sprint session consisted of 3 repeated maximal-effort sprints (15, 30, and 45 s) formatted with an effort/recovery duration ratio of 1:5.
RESULTS: Peak power output (PPO) and mean power output (MPO) of the PowerCal showed high within-subject variation (6.7 to 21.5%) and low to high intra-class coefficient of correlation (ICC = 0.39 to 0.85). Furthermore, the PPO from PowerCal was underestimated by 6.6 to 13.9% and the MPO by 14.9 to 27.6% when compared to the Velotron. Furthermore, the data exhibited high bias (32 W to 129 W) and large confidence limits (−236 to 315 W) between the PO derived from the Powercal and the Velotron.
CONCLUSION: We do not recommend competitive cyclists or sport scientists to use the PowerCal if reliable and valid PO information is required during maximal sprints.

Keywords: Reproducibility, power output, cycling

1. Introduction

The ability for athletes and coaches to monitor training load is important to cycle training programme development. Heart rate (HR) monitors are devices commonly used to monitor and control exercise intensity during cycling training and races [1]. However, the use of HR as a guide to exercise intensity is of little use when athletes are performing short duration high-intensity activities due to substantial delays in the HR response. Fortunately, the past decade has seen the development of strain-gauge based power measurement devices capable of measuring instantaneous changes in cycling exercise intensity [2,3]. While these devices have proved valuable in monitoring exercise over all intensity domains their complex set-up and high costs often restrict their use to professional or high-level competitive athletes.

*Corresponding author: Vitor Pereira Costa, Physical Effort Laboratory, Federal University of Santa Catarina, Campus Universitário Trindade, Florianópolis, Santa Catarina, Brazil. CEP: 88040-900. Tel.: +55 048 3721 4774; Fax: +55 048 3721 9772; E-mail: costavp2@yahoo.com.br.

ISSN 0959-3020/15/$35.00

Recently, an inexpensive device called the PowerCal (Cyclops, Madison, USA) has been developed which claims to provide accurate estimate of power output (PO) directly from the HR response during exercise. The PowerCal device estimates PO using an algorithm computed from the HR response recorded during exercise. If the PowerCal predicts the PO from HR, the device would give values for aerobic intervals (i.e. ranges from resting HR to maximal HR). Sprint training is performed at maximal effort intensity that involves anaerobic stimulus [4,5]. However, the ability of the PowerCal to estimate power during sprints remains uncertain; no studies have been traced which describe the use of this device to estimate PO during sprints in cyclists.

The reproducibility and validity of various powermeter systems have been tested, and used in several other scientific studies [3,4,6–9]. However, as a new device the error of measurement from the PowerCal still needs to be investigated in different modes of exercise. Therefore, the purpose of this study was to determine the reproducibility and validity of the PowerCal during sprint sessions in competitive cyclists.

2. Methods

2.1. Participants

Ten well-trained male cyclists (mean and standard deviation was 33.2 years $\pm$ 10.9 years; 74.8 kg $\pm$ 6.1 kg, 177.6 cm $\pm$ 6.9 cm; maximal power output, 375.3 W $\pm$ 27.4 W; maximal oxygen uptake, 62.7 ml $\cdot$ kg^{-1} $\cdot$ min^{-1} $\pm$ 5.9 ml $\cdot$ kg^{-1} $\cdot$ min^{-1}) volunteered to participate in this study. The cyclists had a minimum experience of three years in regular competitions and were in a transition period from pre-competitive to competitive phase of the annual cycling racing program during the study period. All cyclists were informed of the purpose and 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.

2.2. Study design

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 their sub-maximal and maximal physiological performance parameters. Following the initial test, cyclists completed three sprints sessions each separated by $\sim$72 hours. All sessions were conducted on an electronically braked cycle ergometer (Velotron Dynafit Pro, RacerMate Inc, WA, USA) set up to replicate the cyclists' individual requirements. Furthermore, during the intervention cyclists used the PowerCal (Cyclops, Madison, USA), a power meter that calculates PO from the HR response recorded during exercise. Cyclists also recorded their PO from PowerCal using a Joule cycling computer (Cyclops, Madison, USA). All testing was conducted in a laboratory under controlled environmental conditions. Air temperature and relative humidity were 20.9°C $\pm$ 1.0°C and 56.2% $\pm$ 5.2%, respectively.

2.3. Maximal incremental exercise test

Initially, the cycle ergometer was adjusted to replicate the participants preferred racing position, which was recorded and replicated for all testing sessions. The cyclists wore their own cycling shoes and cleats in all testing and training. Cyclists performed a 15 min warm-up at a self-selected intensity followed by 5 min of rest. Thereafter, the incremental exercise test started at 100 W and power output was increased by 40 W increments every 4 min until volitional exhaustion. The expired respiratory gases were collected and analysed by a Metamax 3B system (Cortex, Leipzig, Germany). Prior to each test, the system was calibrated in accordance with the manufacturers' instructions. VO_2 max was defined as the highest oxygen uptake (VO_2) over a 30 s period recorded during the test. The HR was recorded using a coded strap (Polar, Kempele, Finland) and recorded by the VO_2 system.

2.4. Maximal sprints

All cyclists completed three sprints sessions conducted in the laboratory using the same cycle ergometer and device previously described. The first sprint session was used as a familiarization. For each sprint session, the Velotron was connected to a computer interfaced with a projector in front of the cyclist to simulate an outdoor cycling condition. The computer was used with interactive 3D software version 1.0 (Racermate Inc.). The software displays performance indexes that were not blinded during the sprints (i.e. gear,

Table 1
Velotron reproducibility scores for peak and mean power output during sprints

	PPO			MPO		
Sprint	15 s	30 s	45 s	15 s	30 s	45 s
1 (W)	581 ± 110	523 ± 77	510 ± 75	545 ± 97	483 ± 57	459 ± 48
2 (W)	580 ± 110	514 ± 72	494 ± 76	540 ± 97	470 ± 54	450 ± 54
CV	5.3	6.1	5.4	4.5	6.2	6.0
(95% CI)	(3.8–8.8)	(4.4–10.3)	(3.9–9.0)	(3.5–7.6)	(4.5–10.4)	(4.4–10.1)
ICC	0.95	0.86	0.90	0.96	0.79	0.77
(95% CI)	(0.85–0.98)	(0.62–0.95)	(0.71–0.97)	(0.87–0.99)	(0.46–0.93)	(0.42–0.92)

CV: coefficient of variation; ICC: intra-class coefficient of correlation.

speed, distance, PO, HR and cadence). In addition, cyclists were also allowed to control the power information from the PowerCal at the Joule device. 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 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.

Initially, cyclists completed a 15 min warm-up at a self-selected intensity. The sprint session was divided in three bouts each. The duration of each bout was 15 s, 30 s, and 45 s performed by the cyclists in this sequence. The bouts were performed at maximal effort and recovery intervals at a self-selected intensity below 50% of maximal aerobic power as a form of active recovery. The exercise-rest ratio was 1:5. Participants did complete each sprints session under the supervision of a researcher to ensure the sprints were adhered to as stringently as possible.

2.5. Statistical analysis

Descriptive statistics are presented as means (± SD). To display the PO pattern response during sprints, the MPO from the Velotron and PowerCal were averaged for each interval of 15 s, 30 s, and 45 s. PPO was calculated as the highest 1 s score achieved in each interval. PO during the trials was compared using a two-way analysis of variance with repeated measures, with sprint session (1 vs. 2) and sprint time duration (15 s, 30 s, and 45 s). 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 sprint sessions of the each device. Two-way analysis of variance with repeated measures was used across devices (Velotron and PowerCal) and sprints (15 s, 30 s, and 45 s) with subsequent post-hoc comparisons (Bonferroni) when necessary. The bias and limits of agreement (LoA) of the PO differences between PowerCal and Velotron for each sprint were defined using the method of Bland and Altman [10]. Reliability was quantified using typical error between successive trials as percentages of coefficient of variation (CV) and intra-class coefficient of correlation (ICC) derived from log-transformed data [11]. The confidence intervals (CI) were fixed at 95%. Statistical significance was accepted at $P < 0.05$.

3. Results

The Velotron mean ± SD for the test-retest PPO and MPO scores of the sprints, the CV and ICC values are presented in Table 1. There were no significant differences between the sprint sessions (1 vs. 2) for each sprint interval. However, the PPO and MPO for each sprint of 15 s were significantly higher than those relating to the 30 s and 45 s. There was no significant difference between the PO scores of 30 s and 45 s. Low to moderate scores were found for the CV ranging from 4.5%–6.2% and high values of ICC ranging from 0.77–0.96 for PPO and MPO between the sprint sessions.

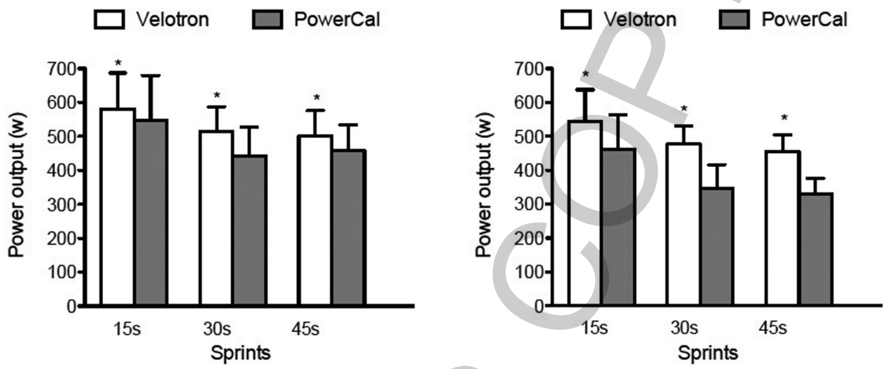
The PowerCal mean ± SD for the test-retest PPO and MPO scores of the sprints, the CV and ICC values are presented in Table 2. There were no significant differences between the training sessions for each sprint interval. However, the PPO and MPO for each sprint of 15 s were significantly higher than 30 s and 45 s. There was no significant difference between the PO scores of 30 s and 45 s. High scores were found for CV ranging from 6.7%–21.5% and wide values of ICC ranging from 0.39–0.92 for PPO and MPO between the sprint sessions.

The PPO was significantly higher for the Velotron compared to the PowerCal during sprints of 15 s (580 W ± 107 W vs. 548 W ± 132 W; respectively), 30 s (514 W ± 73 W vs. 443 W ± 85 W; respectively), and 45 s (502 W ± 74 W vs. 458 W ± 76 W; respectively) (Fig. 1(a)). Furthermore, the MPO was signifi-

Table 2
PowerCal reproducibility scores for peak and mean power output during sprints

	PPO			MPO		
Sprint	15 s	30 s	45 s	15 s	30 s	45 s
1 (W)	553 ± 122	443 ± 89	457 ± 84	474 ± 99	343 ± 64	330 ± 51
2 (W)	542 ± 147	443 ± 85	460 ± 72	450 ± 106	351 ± 78	328 ± 45
CV	21.5	9.8	8.2	13.5	6.7	7.0
(95% CI)	(15.2–37.7)	(9.8–16.6)	(5.9–13.8)	(9.7–23.2)	(4.8–11.3)	(5.0–11.7)
ICC	0.39	0.82	0.85	0.73	0.92	0.84
(95% CI)	(−0.16–0.76)	(0.52–0.94)	(0.58–0.95)	(0.33–0.91)	(0.75–0.97)	(0.56–0.95)

CV: coefficient of variation; ICC: intra-class coefficient of correlation.



(a) (b)

Fig. 1. (a) Peak power output; and (b) mean power output from Velotron (white bar) and PowerCal (black bar) during sprints of 15; 30; 45 seconds. * $p < 0.05$.

cantly higher for the Velotron compared to the PowerCal during sprints of 15 s (543 W ± 95 W vs. 462 W ± 101 W; respectively), 30 s (477 W ± 54 W vs. 347 W ± 69 W; respectively), and 45 s (454 W ± 50 W vs. 329 W ± 47 W; respectively) (Fig. 1(b)).

The Bland-Altman plots of PO between Velotron vs. PowerCal are presented in Fig. 2. The data showed that the PPO bias for the 15 s, 30 s, and 45 s sprints was 32 W, 70 W, and 43 W; and LoA was −236 to 301 W, −142 to 283 W, and −148 W to 235 W; respectively. The MPO bias for the 15 s, 30 s, and 45 s sprints was 80 W, 129 W, and 125 W; and LoA was −155 W to 315 W, −44 W to 303 W, and −0.4 to 250 W.

4. Discussion

The aim of this study was to assess the reliability and validity of PO from the PowerCal during sprints in competitive cyclists. The finding indicated that PPO and MPO of the PowerCal exhibited high within-subject variation (6.7%–21.5%). Furthermore, the PPO of the PowerCal was underestimated by 6.6%–13.9% and MPO by 14.9%–27.6% compared to the Velotron. Also, the PPO and MPO data showed extremely high bias ranging from 32 W to 129 W and limits of agreement ranging 236 to 303 W between the Powercal and Velotron.

The validity of a test should not be considered before its reproducibility has been shown to be acceptable [11–13]. The reproducibility of PO during single bouts of "all out" sprint cycling suggests that the reproducibility of MPO is superior to PPO [13]. In fact, Glaister and colleagues [14] reported a slightly lower CV for the MPO (2.6–3.4%) compared to the PPO (3.3%–3.5%) after two protocols of cycling training session including short sprints of 5 s. According to this, the data from PowerCal showed higher CV for the PPO (8.2%–21.5%) compared to MPO (6.7%–13.5%). Furthermore, the CV associated with the PowerCal was higher compared to Velotron scores ranging from 5.3% to 6.1% and 4.5% to 6.2% for PPO and MPO, respectively.

Scores of about 1.5–2.0 times the typical error could be used as a threshold above which any individual change would be interpreted as a real change following an intervention [12,13]. In the present study, the typical error for the PowerCal MPO expressed as CV in the 15 s sprint, was 13.5% (Table 3). This suggests that any individual change following an intervention needs

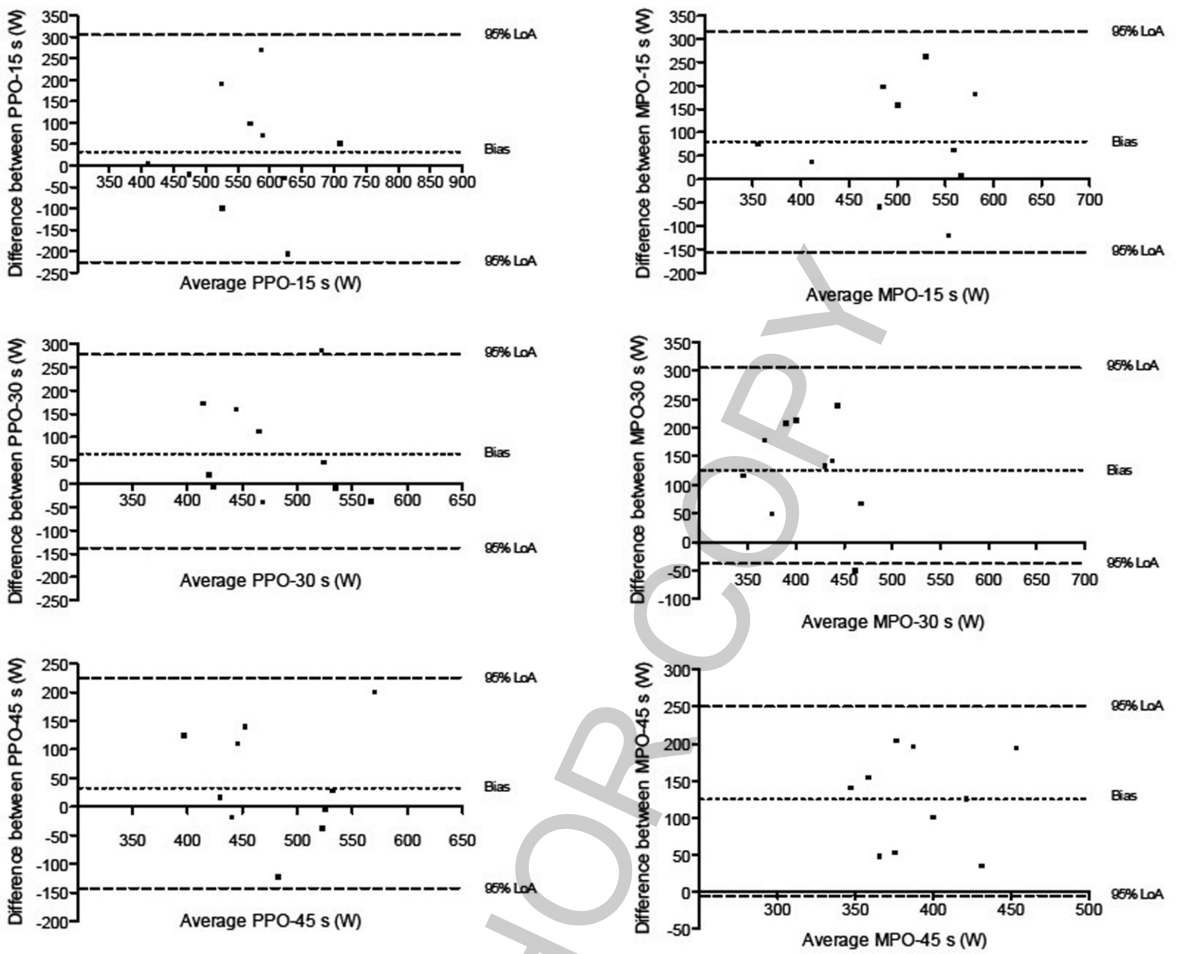


Fig. 2. Bland-Altman plots for the PPO and MPO between Velotron and PowerCal during the sprints of 15, 30, and 45 seconds.

to be greater than 20.3%–27.0% for it to be detectable and the intervention to be considered effective. Considering the fact that enhancements in sub-maximal and maximal power output during graded exercise test and time-trial performance in well-trained cyclists are much lower with scores close to 3%–5% after a short period of high intensity interval training [5,13,15], the efficiency of the PowerCal is inferior indeed.

During the sprints of 15 s, 30 s, and 45 s the PPO was significantly lower for the PowerCal compared to the Velotron. Also, the MPO for the sprints of 15 s, 30 s, and 45 s was significantly lower for PowerCal compared to Velotron. Furthermore, the PPO and MPO between the PowerCal and the Velotron showed high bias and limits of agreement between the devices during all the sprints. The average bias in each sprint for the PPO (32 W to 70 W) was lower compared to the MPO (80 W to 129 W). In addition, the limits of agreement of the PO between both devices were higher in all sprints (−233 W to 315 W). Collectively, the significant differences found in the PO values combined with high bias and limits of agreement of the PO between the PowerCal and the Velotron showed that PowerCal is unlikely to be valid during short maximal sprints of 15 s, 30 s, and 45 s in well-trained cyclists.

There are a number of factors/limitations that could have influenced the differences in PO scores between the PowerCal and the Velotron. First, PowerCal estimates the PO by an algorithm that is sensitive to acute changes in HR that may account for rapid changes in PO. Significantly, in spite of repeated requests the manufacturers refused to release the algorithm so that an independent examination of its structure and contents could not be undertaken. The information given by the manufacturer indicates the algorithm is based on the correlations between HR and PO. In fact strong correlation ($r = 0.97$; $P < 0.001$) can be found between HR and PO when carefully standardized conditions are available [16]. Although HR is closely correlated with PO in a laboratory incremental exercise test, this relationship is not nearly as strong while cycling outdoors. Supposedly the manufactures have analyzed thousands of outdoor cycling data to establish the linear relationship between HR and PO in order to work that into the algorithm. The PO values from the PowerCal are of concern since correlation problems arise with the stochastic nature of cycling. This is due to the wide variety of factors that can influence HR during exercise (i.e. altitude, heat, dehydration, illness, lack of sleep, caffeine, etc.).

Second, high intensity interval training is defined as repeated bouts of short to moderate duration exercise completed at an intensity that is greater than the anaerobic threshold [4,5]. During the first seconds of the sprints the HR responds relatively slowly to changes in PO, whereas the Velotron's response is almost instantaneous. It is no surprise therefore that PO values from the PowerCal were different compared to the cycle ergometer Velotron. A third intervening factor could be the calibration of the PowerCal. The information from the Powercal manufacturers suggests that calibration adds nothing to accuracy. They believe that due to the variability of HR, both between and among individuals, calibration does not statistically increase accuracy of translated energy expenditure. Collectively, these effects discussed above influence the underestimation of the PO values from the PowerCal compared to the Velotron.

5. Conclusions

The PowerCal is a new device designed to calculate PO from HR, but its reliability and validity are questionable. The CV showed poor reproducibility for maximal and mean PO derived from the PowerCal. Furthermore, the maximal and mean PO during the sprints of 15 s, 30 s, and 45 s using PowerCal were underestimated by 6.6% to 13.9% and 14.9% to 27.6%, respectively, when compared to the Velotron. Also, the PO data from the sprints showed high bias (32 W to 129 W) and wide limits of agreement (−236 W to 315 W) between Powercal and Velotron. Thus, we do not recommend competitive cyclists or sport scientists to use the PowerCal if reproducible and valid PO data is required during "all out" maximal sprints.

References

[1] Achten J, Jeukendrup AE. Heart rate monitoring: applications and limitations. Sports Med. 2003;33:517-38.
[2] Duc S, Villerius V, Bertucci W, Grappe F. Validity and reproducibility of the Ergomo® Pro powermeter when compared with the SRM and PowerTap powermeters. Int J Sports Physiol Perform. 2007;2,270-281.
[3] Gardner AS, Stephens S, Martin DT, Lawton W, Lee H, Jenkins D. Accuracy of SRM and Power Tap power monitoring systems for bicycling. Med Sci Sports Exerc. 2004;36,1252-1258.
[4] Laursen PB, Jenkins DG. The scientific basis for high-intensity interval training optimizing training programs and maximizing performance in highly trained endurance athletes. Sports Med. 2002;32,53-73.
[5] Laursen PB, Shing CM, Peake JM, Coombes JS, Jenkins DG. Influence of high-intensity interval training on adaptations in well-trained cyclists. J Strength Cond Res. 2005;19,527-533.
[6] Abbiss CR, Quod MJ, Levin G, Martin DT, Laursen PB. Accuracy of the Velotron Ergometer and SRM Power Meter. Int J Sports Med. 2009;30,107-112. doi: 10.1055/s-0028-1103285.
[7] Gordon RS, Franklin KL, Davies B, Baker JS. Further mechanical considerations between polar and SRM mobile ergometer systems during laboratory-based high-intensity, intermittent cycling activity. Res Sports Med. 2007;15,241-247.
[8] Hurst HT, Atkins S. Agreement between polar and SRM mobile ergometer systems during laboratory-based high-intensity, intermittent cycling activity. J Sports Sci. 2006;24, 863-868.
[9] Millet GP, Tronche C, Fuster N, Bentley DJ, Candau R. Validity and reliability of the Polar® S710 mobile cycling powermeter. Int J Sports Med. 2003;24,156-161.
[10] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1986;1,307-310.
[11] Hopkins WG (2007). Analysis of reliability with a spreadsheet. A new view of statistics: http://sportsci.org/resource/stats/xrely.xls.
[12] Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30,1-15.
[13] Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. Sports Med. 2001;31,211-234.
[14] Glaister M, Stone MH, Stewart AM, Hughes M, Moir GL. Reliability of power output during short-duration maximal-intensity intermittent cycling. J Strength Cond Res. 2003;17, 781-784.
[15] Stepto NK, Hawley JA, Dennis SC, Hopkins WG. Effects of different interval-training programs on cycling time-trial performance. Med Sci Sports Exerc. 1999;31,736-741.
[16] Arts FJ, Kuipers H. The relation between power output, oxygen uptake and heart rate in male athletes. Int J Sports Med. 1994;15,228-31.