"As accepted for publication in International Journal of Sports Physiology and Performance, ©Human Kinetics"
https://doi.org/10.1123/ijspp.2018-0563

The Reliability and Validity of the PowerTap P1 Power Pedals Before and After 100 Hours of Use

Submission Type:

Brief Report

Authors and Affiliations:

James Wright[a]*, Thomas Walker[a], Scott Burnet[a] and Simon A. Jobson[b]

[a]*School of Sport, Health and Social Sciences, Solent University, Southampton, UK;*

[b]*Department of Sport, Exercise & Health, University of Winchester, Winchester, UK*

Mr James Wright (corresponding author*)

Solent University

School of Sport, Health and Social Science

East Park Terrace, Southampton

Hampshire, SO14 0YN, UK

E: james.wright@solent.ac.uk,

T: 023 8031 3955

Preferred Running Head: PowerTap P1 Reliability and Validity

Abstract Word Count: 241 words

Text-Only Word Count: 1714 words

Number of Figures: 2

Number of Tables: 1

"As accepted for publication in International Journal of Sports Physiology and Performance, ©Human Kinetics"
https://doi.org/10.1123/ijspp.2018-0563

Abstract

Purpose: The aims of this study were to 1) evaluate agreement between the PowerTap P1 (P1) pedals and the Lode Excalibur Sport cycle ergometer, 2) investigate the reliability of the P1 pedals between repeated testing sessions, and 3) compare the reliability and validity of the P1 pedals before ($P1_0$) and after ($P1_{100}$) ~100 h of use. ***Methods:*** Ten participants completed four 5-min sub-maximal cycling bouts (100, 150, 200 and 250 W), a 2-min time-trial and two 10-s all-out sprints on two occasions. The above protocol was repeated after fifteen months and ~100 h of use. ***Results:*** Significant differences were seen between the $P1_0$ pedals and the Lode Excalibur Sport at 100 W (P = 0.006), 150 W (P = 0.006), 200 W (P = 0.001) and 250 W (P = 0.006) and during the all-out sprints (P = 0.020). Following ~100 h of use, the $P1_{100}$ pedals did not significantly differ from the Lode Excalibur Sport at 100 W (P = 0.799), 150 W (P = 0.183), 200 W (P = 0.289) and 250 W (P = 0.183), during the 2-min time-trial (P = 0.583) or during the all-out sprints (P = 0.412). The coefficient of variation for the $P1_0$ and $P1_{100}$ ranged from 0.6–1.3% and 0.5–2.0%, respectively, during the sub-maximal cycling bouts. ***Conclusion:*** The P1 pedals provide valid data after ~100 h of laboratory use. Furthermore, the pedals provide reliable data during sub-maximal cycling, even after prolonged use.

Keywords power meter, ergometer, laboratory testing, field testing

"As accepted for publication in International Journal of Sports Physiology and Performance, ©Human Kinetics"
https://doi.org/10.1123/ijspp.2018-0563

Introduction

Physiological testing is frequently performed on a laboratory-based ergometer and is an essential aspect of training for competitive cyclists.[1] The Lode Excalibur Sport is an electromagnetically-braked cycle ergometer commonly used within sports science research and is often regarded as a "gold standard" in testing ergometry.[2,3]

The development of the cycle-mounted power meter has provided athletes, coaches and researchers with the opportunity to monitor power output and cadence using the athlete's own bike, rather than being restricted to a laboratory-based ergometer.[2,4,5] Until recently pedal-based systems have not provided the same measure of reliability when compared to more traditional crank- or hub-based systems with Sparks et al.[6] suggesting that the LOOK Kéo power-pedals were not as reliable as the SRM Powermeter during an incremental testing protocol. Recently, the reliability and validity of the PowerTap P1 pedals have been investigated between 100–500 W at 70, 85 and 100 rev·min^{-1}.[7] These authors reported that the PowerTap P1 pedals slightly underestimated the SRM Powermeter by 2–7 W but suggested that the pedals were reliable and valid, concluding that they were a cost-effective alternative to laboratory-based ergometers.

It has previously been suggested that reliability and validity studies on power measuring devices are limited to using a single test-retest protocol, with suggestions that reliability may be reduced for older systems.[9] To the authors' knowledge, the reliability and validity of pedal-based power meters have not been investigated over an extended period and it is reasonable to suggest that both the reliability and validity of such systems will change over time making monitoring performance changes difficult. Therefore, the aims of the present study were to 1) evaluate agreement between the PowerTap P1 pedals and the Lode Excalibur Sport, 2) evaluate

the reliability of the PowerTap P1 pedals between testing sessions, and 3) compare the reliability and validity of the PowerTap P1 pedals before and after ~100 h of use.

Methods

Participants

Initial testing ($P1_0$) was completed by ten male amateur cyclists using a pair of new PowerTap P1 pedals (mean ± SD: age 34 ± 6 years, body mass 80.8 ± 8.8 kg, stature 1.83 ± 0.05 m). Following a period of 15 months and ~100 h of laboratory use, the testing protocol was repeated ($P1_{100}$) with a further ten cyclists (mean ± SD: age 30 ± 7 years, body mass 80.9 ± 11.9 kg, stature 1.83 ± 0.08 m). During each testing period, the protocol was repeated on two occasions, separated by a minimum of 48 h. All testing was carried out on an electronically-braked cycle ergometer (Excalibur Sport, Lode, The Netherlands) with the pedals installed following the manufacturer's guidelines.

Experimental Procedures

Following a 10-min warm-up, participants completed four 5-min sub-maximal cycling bouts (100, 150, 200 and 250 W) using the ergometer's hyperbolic mode, each separated by a 5-min recovery period at 50 W. The participants were then given a 15-min active recovery period at 100 W before completing a 2-min maximal time-trial effort against a fixed resistance. Following a further 15-min recovery period, participants were required to complete two 10-s maximal sprints, each separated by a 2-min recovery period. Following a period of 15 months and ~100 h of typical laboratory-based testing using the PT1 pedals and Lode Excalibur Sport, the above procedure was repeated. Prior to both testing periods, the Lode Excalibur Sport was calibrated using a dynamic calibration rig (Calibrator 2000, Lode, The Netherlands) at 25–150 W (60 rev·min^{-1}) and 200–500 W (100 rev·min^{-1}).

100

101 **Statistical analyses**

102 Data was exported from the Lode Excalibur Sport and PowerTap P1 pedals with the mean
103 power output for each sub-maximal intensity calculated. For the 10-s sprints, the peak power
104 output from each system was exported for analysis. Comparisons between the Lode Excalibur
105 Sport and the PowerTap P1 pedals were made using a Mann-Whitney-U test with agreement
106 assessed using limits of agreement (LoA). Predicted vs. residual values for power output were
107 plotted to check for heteroscedasticity. Test-retest reliability was measured using CV and
108 typical error of measurement (TEM) and upper and lower 95% confidence limits. Using the
109 equation, $n = 8s^2/d^2$, where CV is used for *s*, and a smallest worthwhile change of 0.2 is used
110 for *d*, the estimated sample size for a test-retest study design was also calculated.[10] Using the
111 example described by Kirkland et al.[11], the smallest worthwhile change was calculated from
112 the data published by Folland et al.[12], where the mean power output during a 16.1 km time-trial
113 was 322 W, with a SD of 15 W (Table 1). Statistical significance was set to $P = 0.05$, with all
114 data reported as mean ± SD.

115

116 **Results**

117 A Mann-Whitney-U test identified significant differences between the Lode Excalibur Sport
118 and the $P1_0$ pedals at 100 W (100.0 W ± 0.0 vs. 100.4 W ± 2.1, $P = 0.006$), 150 W (150.0 W ±
119 0.0 vs. 151.2 W ± 2.1, $P = 0.006$), 200 W (200.0 W ± 0.0 vs. 201.6 W ± 2.5, $P = 0.001$) and
120 250 W (250.0 W ± 0.0 vs. 251.7 W ± 2.1, $P = 0.006$). Significant differences were also seen
121 during the all-out sprints (963.7 ± 111.0 vs. 1026.4 ± 116.2, $P = 0.020$, 95% LoA of -62 ± 195
122 W). No significant differences between the Lode Excalibur Sport and $P1_0$ were observed during
123 the 2-min all-out time-trial (402.7 ± 57.1 W vs. 398.8 ± 54.8 W, $P = 0.718$, 95% LoA of 4 ± 18
124 W) (Figure 2).

Following ~100 h of use, a Mann-Whitney-U test showed no significant differences between the Lode Excalibur Sport and the $P1_{100}$ pedals at 100 W (100.0 W ± 0.0 vs. 100.2 W ± 1.9, P = 0.799), 150 W (150.0 W ± 0.0 vs. 149.0 W ± 2.0, P = 0.183), 200 W (200.0 W ± 0.0 vs. 199.0 W ± 2.6, P = 0.289) and 250 W (250.0 W ± 0.0 vs. 249.2 W ± 3.1, P = 0.289). Furthermore, no significant differences between the Lode Excalibur Sport and the $P1_{100}$ pedals were seen during the 2-min all-out time-trial (379.4 ± 45.0 W vs. 372.7 ± 40.2 W, P = 0.583, 95% LoA of 7 ± 16 W) or during the all-out sprints (979.3 ± 132.6 vs. 936.1 ± 169.5, P = 0.412, 95% LoA of 43 ± 245 W) (Figure 2).

Figure 1 near here

Figure 2 near here

The CV and TEM for the $P1_0$ pedals and $P1_{100}$ during sub-maximal cycling bouts, the 2-min all-out time-trial and all-out sprints can be found in Table 1.

Table 1 near here

Discussion

The results of this study suggest that the PowerTap P1 pedals provide reliable data during sub-maximal cycling and that reliability is maintained after ~100 h of laboratory use. During all-out sprint performance, the P1 pedals appeared to overestimate power output by approximately 60 W when first tested and underestimate power output by approximately 40 W after prolonged use. Figure 2 highlights the heteroscedastic nature of power output data recorded by the P1

pedals, with an increase of error observed at higher power outputs. It is possible that the location of the strain gauges used by each system may help to explain these differences. The strain gauges in the P1 pedals are housed within the pedal body, whereas the Lode Excalibur Sport has strain gauges mounted on the crank and, therefore, some force may dissipate through the pedal before being measured at the crank[7].

The CV of the $P1_0$ (0.6–1.3%) and $P1_{100}$ (0.5–2.0%) pedals during the sub-maximal intervals is comparable, but slightly lower than a recent study by Pallarés and Lillo-Bevia[7] who concluded that the P1 pedals produced a CV of 2.4–3.7% when cycling at 70–100 rev·min^{-1}. The results of the present study are also comparable to alternative systems, with Bertucci et al.[7] reporting the SRM Powermeter to have a CV of 0.7–2.1% at sub-maximal intensities and the PowerTap (hub) a CV of 0.9–2.9%, between testing sessions. According to Hopkins[10], the CV in sports science reliability testing should not exceed 5% and in the present study the new and unused P1 pedals met this criterion for all tested power outputs. However, after a period of ~100 h of use, the CV observed during the all-out sprint performance increased slightly above this recommendation to 6.3%.

The results of the present study would suggest that although not valid when initially purchased, the P1 pedals provide valid data after prolonged use when compared to the Lode Excalibur Sport. During the initial period of testing, a significant difference was seen for all power outputs between 100–250 W; however, no significant differences were seen during repeat testing. Despite the significant differences observed during the initial period of testing, the actual mean percentage difference was less than 1% for all sub-maximal power outputs. Table 1 highlights that some care should be taken if using the P1 pedals during a sprint-based test-retest study design, with a substantially greater sample size required, when compared to sub-maximal

power outputs. This study compared the PowerTap P1 pedals to the Lode Excalibur Sport at a limited selection of power outputs and, although they were typical of those at which amateur cyclists train and race, the fact that a full range of power outputs was not compared is a limitation of this study. It is recommended that future studies investigate the reliability and validity of the P1 pedals between 500–700 W.

Reliability studies are common within sports science when assessing new testing equipment; however, the majority use simple test-retest study designs, separated by several days. For researchers to have confidence in their results, it is essential that the equipment used during data collection demonstrates reliability across the relevant period of assessment, for example, before and after a 12-week training study. Future studies should utilise a more robust study design such as the one presented within this study when assessing the reliability of testing equipment.

Conclusion

The results of this study suggest that PowerTap P1 pedals have acceptable test-retest reliability for amateur cyclists, which is maintained after prolonged use. The P1 pedals were significantly different to the Lode Excalibur Sport during submaximal cycling in early use; however, no significant differences were seen when re-tested and power output was within 1% of the Lode Excalibur Sport before and after ~100 h of use during sub-maximal power outputs.

Acknowledgements

The authors would like to thank Dr Helen Thomas and Dr Michelle Jones for their assistance during this study. The authors would also like to thank all the participants who volunteered their time to take part in this study.

References

1. Paton CD, Hopkins WG. Tests of cycling performance. *Sports Med*. 2001;31(7):489–496.
2. Earnest CP, Wharton RP, Church TS, Lucia A. Reliability of the Lode Excalibur Sport Ergometer and Applicability to Computrainer Electromagnetically Braked Cycling Training Device. *J Strength Cond Res*. 2005;19(2):344–348.
3. Driller MW. The reliability of a 30-minute performance test on a Lode cycle ergometer. *J Sci Cycling*. 2012;1(2):21–27.
4. Bini RR, Carpes FP. *Biomechanics of Cycling*. Switzerland: Springer; 2014.
5. Jones SM, Passfield L. The dynamic calibration of bicycle power measuring cranks. In: Haake SJ (ed.). The Engineering of Sport. Oxford: Blackwell; 1998;265–274.
6. Sparks SA, Dove B, Bridge CA, Midgley AW, McNaughton LR. Validity and Reliability of the Look Keo Power Pedal System for Measuring Power Output During Incremental and Repeated Sprint Cycling. *Int J Sports Physiol Perform*. 2015;10:39–45.
7. Bertucci W, Duc S, Villerius V, Pernin JN, Grappe F. Validity and Reliability of the PowerTap Mobile Cycling Powermeter when Compared with the SRM Device. *Int J Sports Med*. 2005;26:868–873.
8. Pallarés JG, Lillo-Bevia JR. Validity and Reliability of the PowerTap P1 Pedals Power Meter. *J Sports Sci Med*. 2018;17:305–311.
9. Zadow E, Kitic C, Wu SX, Fell JW. Reliability of power setting of the Wahoo KICKR Power Trainer after 60 hours of use. *Int J of Sports Physiol and Perform*. 2018;13(1):119–121.
10. Hopkins WG. A new view on statistics. *Internet Society for Sport Science*. 2000. Available at: http://www.sportsci.org/resource/stats/. Accessed 06 July 2018.
11. Kirkland A, Coleman, J, Wiles, JD, Hopker, J. Validity and Reliability of the Ergomo®pro Powermeter. *Int J Sports Med*. 2008; 913–916.

12. Folland JP, Stern R, Brickley, G. Sodium phosphate loading improves laboratory cycling time-trial performance in trained cyclist. *J Sci Med Sport*. 2008; 464–468.

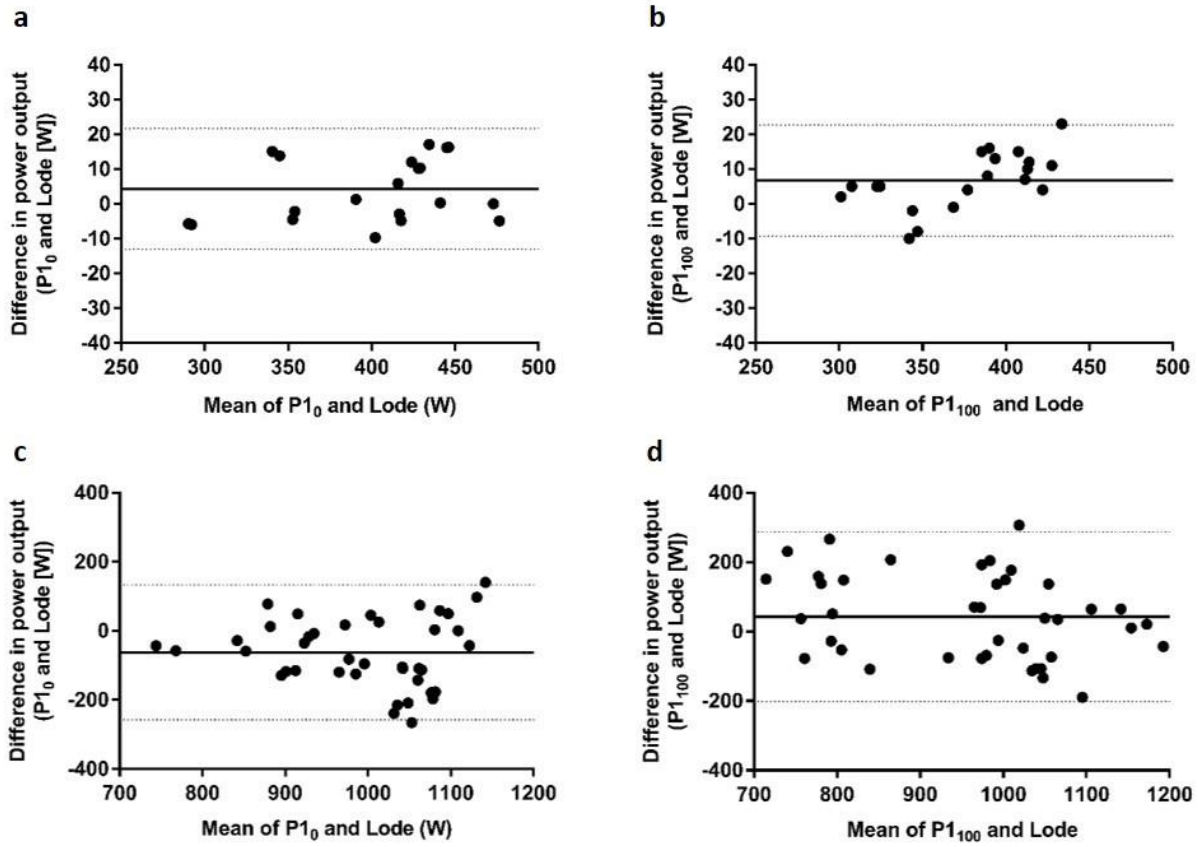

Figure 1 Bland-Altman plots showing the LoA between (a) Lode Excalibur Sport and $P1_0$ pedals during a 2-min time-trial (b) Lode Excalibur Sport and $P1_{100}$ pedals during a 2-min time-trial (c) Lode Excalibur Sport and $P1_0$ pedals during a 10-s all-out sprint, and (d) Lode Excalibur Sport and $P1_{100}$ pedals during a 10-s all-out sprint. The solid line represents the mean difference in power output and the dashed lines represent the 95% LoA.

"As accepted for publication in International Journal of Sports Physiology and Performance, ©Human Kinetics"
https://doi.org/10.1123/ijspp.2018-0563

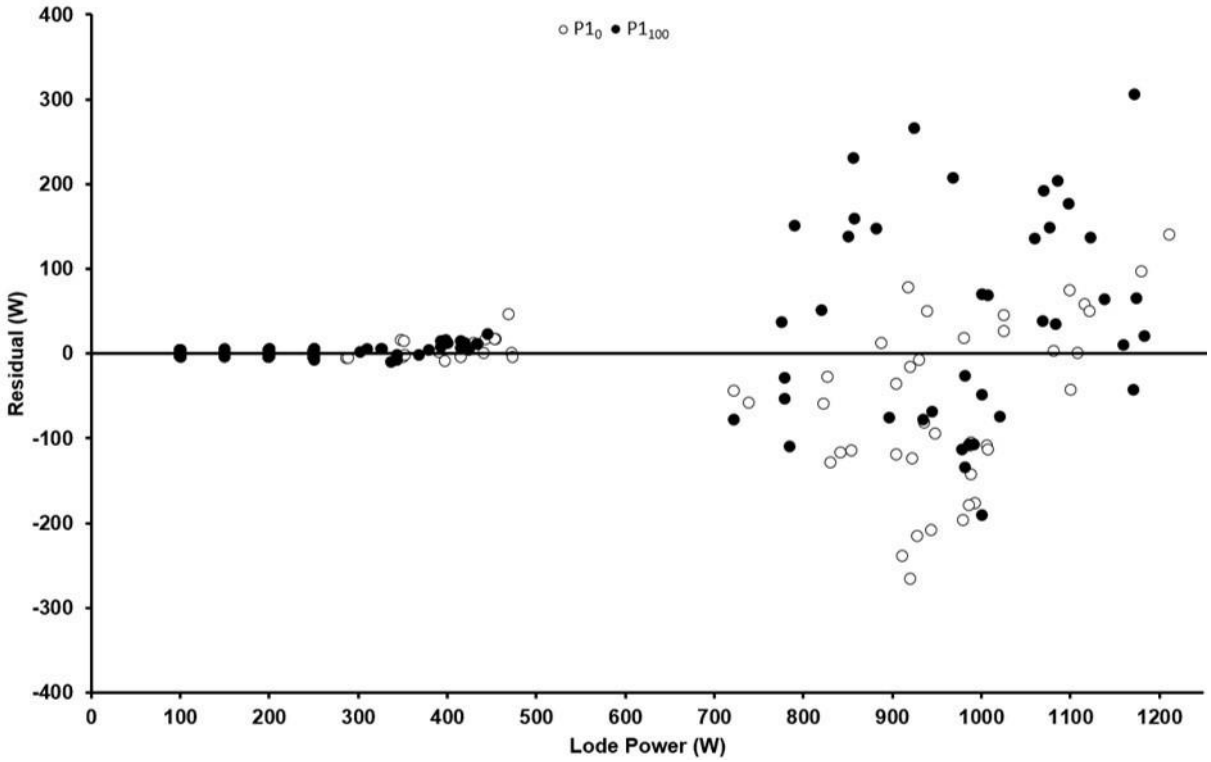

Figure 2 Plot of predicted vs. residual (Lode – P1) values for $P1_0$ pedals (open circles) and $P1_{100}$ pedals (closed circles).

280 **Table 1** Estimated **s**ample sizes required for a test-retest study design, CV and absolute TEM between testing sessions 1 and 2 (including 95% confidence

281 limits).

	PowerTap $P1_0$			PowerTap $P1_{100}$		
	CV (%)	TEM (W)	Sample size required for test-retest study design	CV (%)	TEM (W)	Sample size required for test-retest study design
100 W	0.6 (0.2–1.0)	0.8 (0.4–1.2)	3 (1–10)	1.1 (0.3–1.8)	1.5 (0.8–2.3)	11 (1–29)
150 W	0.7 (0.5–1.0)	1.2 (0.8–1.6)	5 (2–10)	0.5 (0.1–0.8)	1.1 (0.6–1.6)	2 (1–6)
200 W	0.7 (0.3–1.1)	1.9 (1.0–2.7)	5 (1–11)	0.6 (0.4–0.8)	1.3 (0.9–1.7)	3 (1–6)
250 W	0.6 (0.4–1.2)	2.1 (1.1–3.2)	3 (1–13)	1.0 (0.5–1.6)	3.2 (1.9–4.5)	9 (2–24)
2-min TT	1.3 (0.4–2.2)	8.0 (4.1–12.0)	15 (1–44)	2.0 (0.1–3.9)	13.6 (6.2–20.9)	36 (1–140)
All-out sprints	4.2 (1.8–6.7)	50.3 (27.5–73.1)	163 (30–414)	6.3 (4.7–7.9)	75.1 (59.9–90.3)	366 (203–575)

282