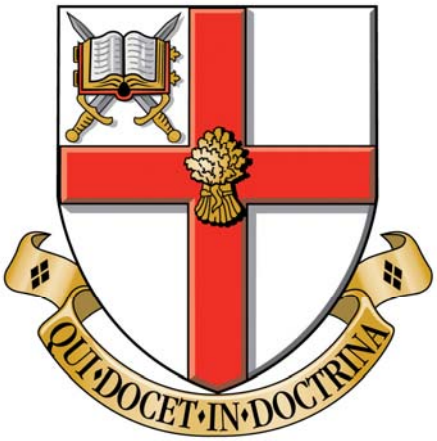

University of Chester

Author(s): James E O Hartshorn; Kevin L Lamb

Title: The reproducibility of perceptually regulated exercise responses during short-term cycle ergometry

Date: 2004

Originally published in: International Journal of Sports Medicine

Example citation: Hartshorn, J. E. O., & Lamb, K. L. (2004). The reproducibility of perceptually regulated exercise responses during short-term cycle ergometry. *International Journal of Sports Medicine*, 25, 362-367

Version of item: Author's post-print

Available at: http://hdl.handle.net/10034/28732

Published in *International Journal of Sports Medicine* (2004)

The reproducibility of perceptually regulated exercise responses during short-term cycle ergometry

J.E.O.Hartshorn[1] and K.L.Lamb[2]

[1]Centre for Exercise and Nutrition Science and [2]Department of Physical Education and Sports Science, Chester College of Higher Education, Parkgate Road, Chester CH1 4BJ, United Kingdom.

Correspondence to:
Dr K.L. Lamb, Department of Physical Education and Sports Science, Chester College of Higher Education, Parkgate Road, Chester CH1 4BJ, United Kingdom.

Tel: (01244) 375444
Fax: (01244) 392889
E-mail: k.lamb@chester.ac.uk

Abstract

The purpose of this study was to assess the reproducibility over four trials of perceptually regulated exercise intensity during short-term cycle ergometry. Recent research has suggested that an improvement in the reproducibility (better agreement) of the exercise output would be observed with a repeated practice of using regulatory tools such as Borg's 6-20 rating of perceived exertion (RPE) scale. Eighteen healthy active volunteers (nine males mean age (± SD) 24.7 ± 3.4 yr, and nine females 27.6 ± 5.4 yr) completed four identical intermittent effort production trials on a cycle ergometer, over a period of two-three weeks, with all trials being between three and five days apart. After warm-up, the volunteers were asked to produce four x three-minute bouts of exercise at RPE levels: 13, 15, 9, and 17 (in this order). Power output (W), percentage maximum heart rate reserve (%MHRR), and oxygen consumption (VO_2; $ml \cdot kg^{-1} \cdot min^{-1}$) were recorded in the final minute of each bout. Analysis revealed that the 95% limits of agreement (LoA) between repeated trials did not decrease for the objective markers of exercise intensity, remaining wide throughout. In the worst case comparisons the LoA represented changes (expressed as a proportion of the mean of two trials) of up to 58.3% in power output (T2 vs. T3 at RPE 9), 65.5% in %MHRR (T1 vs. T2 at RPE 13) and 36.5% in VO_2 (T3 vs. T4 at RPE 17). These findings question the use of ratings of perceived exertion to regulate exercise effort. That the reproducibility of effort is also not seen to improve with practice raises doubts over the validity of using the RPE scale for providing training intensities for this type of exercise.

Key words

Rating of perceived exertion (RPE); production mode; reproducibility; limits of agreement analysis; exercise prescription

Introduction

The employment of an 'effort sense,' or perceived exertion by humans to regulate their exercise output to prescribed levels is common-place in adult fitness and clinical settings, particularly within cardiac rehabilitation (12,20,21,26) and with patients receiving β-blocker therapy (13). This process typically involves the utilisation of the Borg 6-20 rating of perceived exertion (RPE) scale (6,7) in its so-called production mode, instead of or in addition to objective markers of effort, such as heart rate and oxygen uptake. Heart rate as a measure of exercise intensity has a number of drawbacks in these circumstances (13,16,22,23), one being the reduction in heart rate following β-blocker treatment due to competitive blocking of β-adrenoreceptors. During sub-maximal exercise patients receiving β-blocker treatment typically experience heart rates that are between 20 and 30% lower than those of healthy persons (11,13). On the other hand, surgical patients generally exhibit tachycardia, with resting heart rates of 110 to 120 beats·min^{-1} (23). Pollock et al. (23) have argued therefore that it is most likely inappropriate to prescribe fixed low-level heart rate (e.g. 120 beats·min^{-1}) as a training heart rate for the cardiac patient. In situations such as these, ratings of perceived exertion are advocated as being more useful for moderating the intensity of exercise than heart rates.

However, the reproducibility (often referred to as 'test-retest reliability') of the exercise responses generated in such perceptually regulated (PR) trials has been established neither in clinical nor non-clinical situations. This is disconcerting not only from a scientific perspective, but also in terms of the safety of the exercise recipient. Moreover, until very recently a feature common to the limited research on the reproducibility of PR exercise was the lack of regard given to the appropriateness of the statistical techniques used to quantify reproducibility. Typically, this has been inappropriately determined by the use of Pearson product-moment correlation coefficients, which are measures of relation (*relative* reliability) rather than agreement (*absolute*

reliability) and are highly influenced by the range of subjects' measurements. Exercise scientists (2,17,19) have over the past eight years been advocating using Bland and Altman's 95% limits of agreement technique (LoA) (4) as a better means of assessing the within-subject agreement (reproducibility) of their measurement tools.

In recognition of this concern, the LoA technique has been applied in the three most recent studies investigating perceived exertion in its production mode (10,15,27). Thompson and Lamb (27) undertook a test-retest investigation on the effect of a perceptual anchoring technique on the reproducibility of exercise regulation in young children using the 1-10 Children's Effort Rating Table (CERT) (29). The results of this study indicated that the perceptual anchoring technique did not yield better inter-trial reliability than when it was not used. The authors suggested that this might be due partly to the demonstration of the low and high anchors of the CERT scale being incompatible with the demands of the exercise production trials.

Buckley et al.(10) investigated the validity and reproducibility of a Braille version of the RPE scale, whereas Eston et al. (15) investigated a new 10-point scale for use with young children, the Cart and Load Effort Rating (CALER) scale. These studies investigated their scales in production mode with a three and four repeated trial design, respectively, and both showed a marked improvement in the reproducibility of the exercise responses following the additional trials. That is, the additional trials generated narrower (better) limits of agreement, possibly due to the subjects becoming more habituated in the use of the Braille RPE and CALER scales. It is not yet known whether repeated exposure to the RPE scale in sighted adults over three or more trials improves the limits of agreement, and therefore the reproducibility of the exercise responses. Accordingly, the purpose of this study was to assess the reproducibility of perceptually regulated exercise responses during short-term cycle ergometry, using the Borg RPE category scale (7) over four identical exercise trials.

Methods

Subjects

Eighteen healthy active subjects (9 male and 9 female) from Chester College volunteered to take part in this study (males: mean (SD) age 24.7 (3.4) years, height 1.73 (4.1) m and body mass 72.7 (8.3) kg; females: age 27.6 (5.4) years, height 1.66 (5.2) m and body mass 62.4 (8.7) kg). All subjects were recreationally active and reported that they considered themselves to be either moderately fit or fit (trained). Subjects were asked to attend the study's four trials in as near to identical state as possible with exercise, diet and sleep in the 24 hours before testing being similar for all tests. All subjects completed an informed consent form and a health questionnaire prior to the first trial. Ethical approval for the study was granted by the Ethics Committee of the College's Centre for Exercise and Nutrition Science.

Procedures

Over a period of two to three weeks the subjects completed four identical testing sessions, all sessions being between three and five days apart. The subjects were individually tested at the same time of day (as their first test) to control for any within-subject physiological variation due to circadian rhythms (25), although the time of testing varied between subjects. The four sessions took place in the same designated area of the laboratory, using the same cycle ergometer (Monark 824E), heart rate monitor (Polar Electro Oy, Kempele, Finland), telemetric gas analysis system (Metamax, Cosmed, Rome) and RPE scale throughout. The cycle ergometer and telemetry system were calibrated daily and prior to testing. The ambient temperature in the laboratory over the course of the study was 22-27ºC.

The subjects were first familiarised with the cycle ergometer and telemetry system and then introduced to the Borg 6-20 RPE scale (7). Before each session the experimenter

read to each subject a set of RPE instructions adapted from the standardised version (1,8) for use with a production protocol. An anchoring of the perceptual range was also undertaken using the technique described by Noble and Robertson (21, p.78-79), to anchor psychologically the extremes of the scale by memory and definition. The exercise protocol was then explained and the subject's resting heart rate recorded while he/she was sitting on the cycle ergometer.

All trials commenced with a three-minute warm-up at 60W (60 revs min^{-1} x 1kg), with the subject breathing into the Metamax, while pedalling in time to the metronome (Seiko, SQ-44) set at 120 beats·min^{-1}. Each subject was then requested to work at the four specific RPE levels of 13, 15, 9, and 17, presented in this order. A fixed production order was chosen to keep fatigue effects and recovery periods between bouts as constant as possible for each subject. Also a mixed order requires the subject to think up and down in terms of relative effort rather than just progressively upwards. RPE 17 was placed at the end of the production trials where its presumed higher fatigue effect would not have an impact on the rest of the protocol.

To reach the specified RPE level, the subject instructed the experimenter to adjust the cycling resistance by adding or taking away weights while keeping the pedal rate constant (60 revs min^{-1}) in time with the metronome. The weights were always added to or subtracted from the load cradle by adjustments of 0.5kg (30W), until the subject specified that a smaller degree of weight change was required, in which case a 0.25 kg (15W) weight adjustment was used. Whilst two minutes had been allowed for this procedure, in practice all subjects reported to the experimenter that they had reached the desired RPE level within 1.5 minutes of starting the exercise. The load cradle was obscured from the subject's vision by the strategic positioning of the RPE scale. This ensured that the subject could not see the load added or taken away from the cradle, and therefore the effort produced was selected on a "feel-only" basis.

Exercise heart rate was recorded during the final 20 seconds of each bout and subsequently used to determine the percentage maximum heart rate reserve (%MHRR) from the established Karvonen formula (exercise heart rate – resting heart rate) / (predicted maximum heart rate (220 – age) – resting heart rate). Between each of the four bouts of effort the subjects had a minimum rest period of three minutes. If their heart rates had not returned to within 20 beats of resting heart rate during this time, a further rest was allowed (of up to two minutes) until this had occurred. Power output measured in watts was calculated from cycle cadence x external resistance and was recorded for the last minute of each bout. Oxygen uptake (VO_2) was measured by the Metamax system throughout the protocol and the average VO_2 was calculated for the final minute of each bout (expressed in $ml \cdot kg^{-1} \cdot min^{-1}$). After all four bouts of effort had been produced, the subjects then "cooled-down" for five minutes pedalling with no load. No feedback with regards to the exercise intensities produced was given.

Statistical Analysis

The 95% Limits of Agreement procedure (4) was used to examine the reproducibility of the subjects' exercise efforts (based on power output scores, %MHRR and VO_2) between successive trials (T1-T2, T2-T3, and T3-T4). This technique quantifies the amount of systematic error and random error in the repeated measures via the calculation of the mean difference (bias) and 1.96 x standard deviation of the differences (SD_{diff}), respectively, for each comparison of trials. A check on whether the magnitude of these differences was unrelated to the mean of the two trials (homoscedastic) was performed using a Pearson correlation coefficient. The normality of the random errors was assessed with the Shapiro-Wilk statistic. In addition, intraclass correlation coefficients (ICC) for a single measure and their 95% confidence intervals were calculated following the recommendations of Atkinson and Nevill (3) and Rankin and Stokes (24), who argued

that whilst having disadvantages over the limits of agreement (LoA) technique (such as being influenced by the size of the between-subjects variance), ICCs also have the advantage of being relatively simple to interpret. The ICCs were calculated using a two-way (subjects x trials) mixed ANOVA in which the systematic variability due to trials is incorporated (28). Finally, coefficients of variation (CV) were calculated from the ratio of the SD_{diff} and the mean of each comparison, expressed as a percentage All data analyses were performed using SPSS 11.0 for Windows.

Results

Table 1 shows the mean power output (W), MHRR (%) and VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$) values (in each trial) increasing step-by-step as RPE increases. The bias ± 95% limits of agreement (LoA) together with the CVs are displayed for each dependent variable across all RPE levels and trials in Table 2. Similarly, the ICCs and their confidence intervals are displayed in Table 3. It is apparent that there was no clear improvement in absolute reliability (LoA or CV) or relative reliability (ICC) at any RPE level across the four trials. Whilst the normality of the trial differences at each RPE level was confirmed (Shapiro-Wilk statistic; $p>0.05$), half of the paired comparisons revealed heteroscedastic errors. However, log transforming the data in the manner suggested by Nevill and Atkinson (19) did not improve this condition (and in some cases, the degree of heteroscedasticity increased). Accordingly, the 95% LoA were retained in favour of the alternative ratio LoA.

Discussion

The data from this study question the test-retest reproducibility of perceptually regulated exercise responses. Our healthy, active subjects were unable to reproduce successfully exercise outputs over four repeated bouts of cycle ergometry. No marked improvement in the LoA over the repeated trials was evident with the trial-to-trial

agreement remaining poor throughout. This was despite the evidence (increases in power output, %MHRR and VO_2 with increasing RPE level) suggesting that they understood the concept of using the RPE scale in this mode.

Interpreting LoA is a challenge for researchers as the question, "what is an acceptable level of trial-to-trial variability?" needs to be addressed. Atkinson and Nevill (3) have stressed that such a decision requires 'analytical goals' to be set a priori, for example, this might be whether the LoA are narrow enough to allow meaningful changes in physiological variables due to a training or rehabilitation programme to be detected. If 'small' changes are deemed to be important, then the LoA have to be narrow otherwise such error contaminates the 'true' effect. With the present data, it was our concern that the degree of trial-to-trial variability was not so large that responses to a given prescribed RPE level overlapped with those of another RPE level.

The largest disagreement for power output was between T2 and T3 at RPE 15 where the LoA were 5.0 ± 64.6W. However, expressed as a proportion of the average performance over the two trials (percent of the 'grand mean'), the greatest within-subjects variability was at RPE 9, where a change of up to 58.3% was possible between T2 and T3 (compared to 39.7% at RPE 15). On the face of it, this degree of variability seems 'large', but a better interpretation can be made by using a hypothetical subject's data to highlight the "worst case scenario" (19). Applying the relevant 95% LoA (T2-T3) for such a subject producing in trial 2 the mean power output of 66.7W at what he/she perceived RPE 9 to be, he/she might be expected in trial 3 to produce a power output as low as 22W or as high as 103W (values rounded to the nearest watt). At RPE 13, the same 'person' having a trial 2 power output of 120W could possibly have a trial 3 output as low as 91W or as high 156W. Therefore, in trial 3 this subject may have produced RPE 9 ("very light"), with a power output higher than that previously produced in trial 2 for RPE 13 ("somewhat

hard"), a few days earlier. This potential 'overlap' poses a threat to the utility of the RPE scale for exercise prescription.

The ICC values for power output range from being unacceptable (< 0.80) to moderate (0.8 - 0.89) (28) across the four RPE levels. For MHRR and VO_2 the interpretation is no better generally, and as with power output, the correlations are not seen to increase from one trial to the next. Likewise, the CVs for all three dependent variables are unacceptably high (>10%), and like the LoA showed no consistent decrease between trials. This provides additional evidence (irrespective of the observed heteroscedasticity in the data), that the subjects were unable to reproduce successfully their exercise outputs over the four repeated bouts on the basis of their perceptions of exertion.

The 95% LoA for the MHRR and VO_2 were wide throughout the four RPE levels with no reduction apparent across the trials. The widest LoA for MHRR were at RPE 13 for T1-T2 (2.6 ± 27.9%), representing a change of 65.5% relative to the mean of the two trials. Repeating the hypothetical approach described above, a person responding with the mean value (43.9%) in T1 might produce a MHRR as high as 74.4% in T2, or as low as 18.6%. The higher value is greater than the mean response recorded for RPE 17 at any trial, and the lower value is less than that recorded for RPE 9 at any trial. The widest VO_2 responses occurred at RPE 17 for T3-T4 with LoA of -0.4 ± 12.9 $ml \cdot kg^{-1} \cdot min^{-1}$, representing a change of up to 36.5%. Thus, a person utilising the mean value (35.1 $ml \cdot kg^{-1} \cdot min^{-1}$) in T3 might produce a VO_2 as high as 48.0 $ml \cdot kg^{-1} \cdot min^{-1}$ in T4, or as low as 21.8 $ml \cdot kg^{-1} \cdot min^{-1}$. Importantly, this low value is smaller than the mean VO_2 recorded for RPE 13 at any trial.

The above LoA provide evidence that the physiological intensities were not consistently produced at the same RPE levels across the trials. The ICCs for MHRR and VO_2 were lower for T3-T4 than for T1-T2 and T2-T3 in all comparisons indicating a decline in the relative reliability of responses with practice. Additionally, the high CVs for

MHRR and VO_2 provide further evidence that the subjects were unable to use successfully the RPE scale to replicate their levels of physiological strain across the repeated trials.

The present findings are in contrast to the two production studies that have examined the reproducibility of perceptually regulated responses over multiple exercise trials using LoA analysis (10,15). Both of these reported that additional trials appeared to yield narrower (better) limits of agreement. It is important to note, however, that they used different scales (Braille RPE and CALER, respectively) and had different populations (blind students and children) to the present study. Our data, like those of Lamb et al. (18), question the test-retest reproducibility of the RPE scale used in this production mode. To date, no other studies have provided evidence for the reproducibility of the RPE scale amongst sighted adults, in either of its application modes, via the use of statistics other than correlation coefficients. Indeed, it appears that the practical acceptance of the RPE scale by sport and exercise scientists has exceeded the actual scientific support for its reliability. Given the scale's apparent widespread usage in both the assessment and prescription of exercise in healthy subjects (1,9,14) and clinical populations (12,13,20,21) there is a pressing need for further investigations into this key aspect of the scale's application.

The implication of the unacceptable and non-improving trial-to-trial agreement found in this study is that the RPE scale may be unreliable for use in production mode in healthy active individuals, and thereby be of limited practical value for exercise regulation and prescription. Based on the measures taken in the four trials in this study, we suggest that our subjects would be unsuccessful in using the scale to perceptually regulate their exercise intensity for the attainment of physiological training goals (such as training at a specified intensity to improve functional aerobic power). Further research is encouraged amongst other samples, such as exercisers in clinical settings, and/or those exercising for longer bouts and on other forms of apparatus.

References

1. American College of Sports Medicine. Guidelines for Exercise Testing and Prescription (6th ed). Baltimore: Lippincott Williams and Wilkins, 2000.
2. Atkinson G. A comparison of statistical methods for assessing measurement repeatability in ergonomics research. In: Atkinson G, Reilly T (eds). Sport, Leisure and Ergonomics. London: E & FN Spon, 1995: 218-22.
3. Atkinson G, Nevill AM. Statistical methods for addressing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998; 26: 217-38.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; I: 307-310.
5. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med 1990; 20:337-40.
6. Borg GAV. An Introduction to Borg's RPE Scale. New York: Movement Publications, 1985.
7. Borg G. Borg's RPE Scale: a Method for Measuring Perceived Exertion. Stockholm: Borg Perception, 1994.
8. Borg G. Borg's Perceived Exertion and Pain Scales. Champaign, IL: Human Kinetics, 1998.
9. British Association of Sport and Exercise Sciences. Physiological Testing Guidelines (3rd ed). Leeds: British Association of Sport and Exercise Sciences, 1997.
10. Buckley JP, Eston RG, Sim J. Ratings of perceived exertion in braille: validity and reliability in production mode. Br J Sports Med 2000; 38: 297-302.
11. Cohn JN. Clinical implications of the haemodynamic effects of beta-blockade. Am J Cardiol 1985; 5: 125D-8D.

12. Dunbar CC, Glickman-Weiss EL, Bursztyn DA, Kurtich M, Quiroz A., Conley P. A submaximal treadmill test for developing target ratings of perceived exertion for outpatient cardiac rehabilitation. Percept Mot Skills 1998; 87: 755-759.

13. Eston RG, Connolly D. The use of perceived exertion for exercise prescription in patients receiving β-blocker therapy. Sports Med 1996; 21: 176-90.

14. Eston RG, Lamb KL. Effort Perception. In: Pediatric Exercise Science and Medicine, Armstrong N, Van Mechelen W (eds). Oxford: Oxford University Press, 2000: 85-91.

15. Eston RG, Parfitt G, Campbell L, Lamb KL. Reliability of effort perception for regulating exercise intensity in children using the Cart and Load Effort Rating (CALER) Scale. Ped Exerc Sci 2000; 12; 388-397.

16. Gutman MC, Squires RW, Pollock ML. Perceived exertion-heart rate relationship during exercise testing and training in cardiac patients. J Cardiac Rehabil 1981; 1: 52-59.

17. Lamb KL. Test-retest reliability in quantitative physical education research: a commentary. Eur Phys Ed Rev 1998; 4: 145-152.

18. Lamb K, Eston RG, Corns D. The reliability of ratings of perceived exertion during progressive treadmill exercise. Br J Sports Med 1999; 33: 336-339.

19. Nevill AM, Atkinson G. Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. Br J Sports Med 1997; 31: 314-318.

20. Noble, BJ. Clinical applications of perceived exertion. Med Sci Sports Exerc 1982; 14: 406-411

21. Noble BJ, Robertson RJ. Perceived Exertion. Champaign: Human Kinetics, 1996: 280-285.

22. Pollock ML, Wilmore JH, Fox SM. Exercise in Health and Disease: Evaluation and Prescription for Prevention and Rehabilitation. Philadelphia: Saunders, 1984.

23. Pollock ML, Jackson AS, Foster C. The use of the perception scale for exercise prescription. In: Borg G, Ottoson D (eds). Perception of Exertion in Physical Work. London: Macmillan, 1986: 161-78.

24. Rankin G, Stokes, M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil 1998; 12 187-199.

25. Reilly T, Atkinson G, Waterhouse J. Biological Rhythms and Exercise. Oxford: Oxford University Press, 1997: 51-53.

26. Robertson, RJ. Exercise testing and prescription using RPE as a criterion variable. Int J Sport Psychol 2001; 32: 177-188.

27. Thompson J, Lamb KL. The effect of perceptual anchoring on the reliability of exercise regulation in young children. J Sports Sci 2001; 19: 27-28.

28. Vincent J. Statistics in Kinesiology. Champaign: Human Kinetics Books, 1994: 180.

29. Williams JG, Eston RG, Furlong B. CERT: a perceived exertion scale for young children. Percept Mot Skills 1994; 79: 1451-8.

Table 1. Mean power output (W), maximum heart rate reserve (%) and oxygen uptake VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$) at each RPE level for all trials.

Trial	RPE	Power Output Mean (SD)	MHRR Mean (SD)	VO_2 Mean (SD)
1	9	67.5 (13.9)	29.6 (14.6)	15.2 (3.2)
	13	132.5 (34.2)	43.9 (13.4)	22.7 (5.2)
	15	175.8 (39.8)	58.4 (17.4)	29.0 (6.7)
	17	215.8 (50.2)	69.0 (13.5)	34.1 (6.0)
2	9	66.7 (13.8)	26.9 (11.6)	15.3 (2.8)
	13	120.0 (27.2)	41.3 (15.6)	22.4 (6.4)
	15	165.0 (38.2)	56.5 (17.6)	28.3 (6.8)
	17	205.8 (55.1)	67.2 (14.8)	33.9 (7.5)
3	9	70.8 (14.4)	31.2 (11.7)	17.7 (3.9)
	13	116.7 (29.4)	43.8 (12.2)	24.1 (5.0)
	15	160.0 (38.2)	58.7 (12.0)	30.2 (7.0)
	17	193.3 (50.6)	70.4 (11.8)	35.1 (7.7)
4	9	71.7 (16.7)	27.4 (11.4)	17.6 (3.4)
	13	114.2 (27.8)	38.4 (12.0)	23.7 (4.7)
	15	157.5 (43.4)	54.4 (15.0)	29.3 (6.8)
	17	201.7 (67.1)	67.1 (12.9)	35.5 (8.3)

Table 2. The Bias and 95% limits of agreement (LoA), together with the coefficients of variation (CV) for power output (W), MHRR (%) and VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$) for all test-retest paired comparisons.

RPE	Test-retest pairs	Power Output		MHRR		VO_2	
		LoA	CV	LoA	CV	LoA	CV
9	T1 – T2	0.8± 25.7	24.8	2.7 ± 16.2	29.2	-0.1 ± 5.1	16.9
	T2 – T3	-4.2± 40.1	29.8	-4.3 ± 13.6	23.9	-2.4 ± 5.0	15.4
	T3 – T4	-0.8 ± 25.7	18.4	3.8 ± 19.7	34.2	0.1 ± 7.6	21.8
13	T1 – T2	12.5 ± 51.7	20.9	2.6 ± 27.9	33.4	0.3 ± 7.4	16.8
	T2 – T3	3.3 ± 32.8	14.1	-2.5 ± 15.9	19.0	-1.7 ± 8.9	19.6
	T3 – T4	2.5 ± 39.4	17.4	5.5 ± 23.2	28.8	0.4 ± 8.8	18.7
15	T1 – T2	10.8 ± 48.2	14.4	1.9 ± 22.5	19.9	0.7 ± 6.9	12.4
	T2 – T3	5.0 ± 64.6	20.3	-2.2 ± 19.5	17.3	-2.0 ± 10.2	17.9
	T3 – T4	2.5 ± 63.2	20.3	4.3 ± 24.7	22.3	0.9 ± 11.0	18.8
17	T1 – T2	10.0 ± 49.4	11.9	1.9 ± 18.6	13.9	0.1 ± 8.2	12.3
	T2 – T3	12.5 ± 45.4	11.6	-3.2 ± 12.9	9.6	-1.2 ± 9.3	13.8
	T3 – T4	8.3 ± 52.6	13.6	3.3 ± 14.4	10.7	-0.4 ± 12.9	18.6

Table 3. Intraclass correlation coefficients (ICC) and their 95% confidence intervals (CI) for power output (W), MHRR (%) and VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$) for all test-retest paired comparisons.

RPE	Test-retest pairs	Power Output		MHRR		VO_2	
		ICC	CI	ICC	CI	ICC	CI
9	T1 – T2	0.29	-0.22 - 0.66	0.86	0.66 - 0.94	0.63	0.24 - 0.85
	T2 – T3	-0.01	-0.51 - 0.42	0.83	0.53 - 0.94	0.58	0.01 - 0.84
	T3 – T4	0.66	0.28 - 0.86	0.71	0.39 - 0.88	0.45	-0.03 - 0.75
13	T1 – T2	0.60	0.21 - 0.83	0.62	0.23 - 0.84	0.80	0.53 - 0.92
	T2 – T3	0.83	0.60 - 0.93	0.84	0.64 - 0.94	0.66	0.31 - 0.86
	T3 – T4	0.76	0.47 - 0.90	0.59	0.19 - 0.82	0.58	0.17 - 0.82
15	T1 – T2	0.78	0.50 - 0.91	0.81	0.57 - 0.93	0.86	0.68 - 0.95
	T2 – T3	0.64	0.25 - 0.85	0.81	0.57 - 0.92	0.70	0.36 - 0.87
	T3 – T4	0.70	0.35 - 0.88	0.66	0.31 - 0.86	0.58	0.33 - 0.87
17	T1 – T2	0.88	0.70 - 0.95	0.82	0.59 - 0.93	0.82	0.58 - 0.93
	T2 – T3	0.88	0.68 - 0.96	0.87	0.66 - 0.95	0.81	0.56 - 0.92
	T3 – T4	0.89	0.74 - 0.96	0.81	0.56 - 0.93	0.68	0.31 - 0.87