Kent Academic Repository

**O'Malley, Callum A., Fullerton, Christopher L. and Mauger, Alexis R. (2023) *Test-retest reliability of a 30-min fixed perceived effort cycling exercise.* European Journal of Applied Physiology, 123 (4). pp. 721-735. ISSN 1439-6319.**

## Downloaded from

https://kar.kent.ac.uk/114858/ The University of Kent's Academic Repository KAR

## The version of record is available from

https://doi.org/10.1007/s00421-022-05094-z

## This document version

Publisher pdf

## DOI for this version

## Licence for this version

CC BY (Attribution)

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

European Journal of Applied Physiology (2023) 123:721–735
https://doi.org/10.1007/s00421-022-05094-z

ORIGINAL ARTICLE

# Test–retest reliability of a 30-min fixed perceived effort cycling exercise

Callum A. O'Malley[1,2] · Christopher L. Fullerton[1] · Alexis R. Mauger[1]

Received: 9 September 2022 / Accepted: 9 November 2022 / Published online: 27 November 2022

## Abstract

**Purpose** Using exercise protocols at a fixed rating of perceived effort (RPE) is a useful method for exploring the psychophysical influences on exercise performance. However, studies that have employed this protocol have arbitrarily selected RPE values without considering how these values correspond to exercise intensity thresholds and domains. Therefore, aligning RPE intensities with established physiological thresholds seems more appropriate, although the reliability of this method has not been assessed.

**Methods** Eight recreationally active cyclists completed two identical ramped incremental trials on a cycle ergometer to identify gas exchange threshold (GET). A linear regression model plotted RPE responses during this test alongside gas parameters to establish an RPE corresponding to GET ($RPE_{GET}$) and 15% above GET ($RPE_{+15\%GET}$). Participants then completed three trials at each intensity, in which performance, physiological, and psychological measures were averaged into 5-min time zone (TZ) intervals and 30-min 'overall' averages. Data were assessed for reliability using intraclass correlation coefficients (ICC) and accompanying standard error measurements (SEM), 95% confidence intervals, and coefficient of variations (CoV).

**Results** All performance and gas parameters showed excellent levels of test–retest reliability (ICCs = > .900) across both intensities. Performance, gas-related measures, and heart rate averaged over the entire 30-min exercise demonstrated good intra-individual reliability (CoV = < 5%).

**Conclusion** Recreationally active cyclists can reliably replicate fixed perceived effort exercise across multiple visits when RPE is aligned to physiological thresholds. Some evidence suggests that exercise at $RPE_{+15\%GET}$ is more reliable than $RPE_{GET}$.

**Keywords** Perceived effort · Test–retest reliability · Psychophysiology · Cardiorespiratory · Affect

## Abbreviations

| | |
|---|---|
| $[La^-]_b$ | Blood lactate |
| ANOVA | Analysis of variance |
| BF | Breathing frequency |
| CoV | Coefficient of variation |
| CI | Confidence interval |
| GET | Gas exchange threshold |
| HR | Heart rate |
| ICC | Intraclass correlation coefficient |
| RCP | Respiratory compensation point |
| RPE | Ratings of perceived effort |
| $RPE_{+15\%GET}$ | Ratings of perceived effort at 15% above gas exchange threshold |
| $RPE_{GET}$ | Ratings of perceived effort at gas exchange threshold |
| SEM | Standard error measurement |
| TZ | Time zone |
| $\dot{V}CO_2$ | Carbon dioxide production (absolute) |
| $\dot{V}_E$ | Minute ventilation |
| $\dot{V}O_2.kg^{-1}$ | Oxygen uptake (relative) |
| $\dot{V}O_2max$ | Maximum oxygen uptake |
| W | Power output |

## Introduction

Perceived effort is a crucial determinant in the regulation of exercise intensity (Marcora 2008; Tucker 2009). In short, perceived effort is characterised as a

---

Communicated by Susan Hopkins.

✉ Callum A. O'Malley
C.O'Malley@kent.ac.uk

[1] School of Sport and Exercise Sciences, University of Kent, Canterbury, Kent, United Kingdom

[2] School of Kinesiology and Physical Activity Sciences, Université de Montréal, Montréal, Quebec, Canada

psychophysiological phenomenon (Borg 1982) involving a complex interaction between physical stimuli (e.g., power/velocity) and perceptual responses (Gescheider 1997). Crucially, interpretations of perceived effort consider both subfactors. For instance, a lower perception of effort is denoted by an individual achieving a higher power/velocity for a given rating of perceived effort (RPE) value *or* a lower rating of effort for a given velocity/power.

Marcora (2009) highlights that perceived effort has two components, locomotor effort (Marcora et al. 2008) and respiratory effort (Dempsey et al. 2008). Locomotor effort encapsulates how hard, heavy, and strenuous the exercise task feels to drive the working muscles (Marcora 2010). Although it is still contested (see Pageaux 2016), effort perceptions surrounding locomotor effort are likely derived from the accumulation of central motor command by-products (e.g., corollary discharge) that are sent to working muscles (de Morree et al. 2012; Pageaux 2016). The accumulation of corollary discharge is believed to accumulate within cerebral centres such as the prefrontal cortex (de Morree et al. 2012) and anterior cingulate cortex (Pageaux et al. 2014; Meeusen and Roelands 2018) wherein perceptions of effort are generated.

Alternatively, respiratory effort is one of the perceptions associated with the multidimensional sensation of dyspnea (O'Donnell et al. 2009). Specifically, respiratory effort concerns the perception of how hard one is breathing (Laviolette and Laveneziana 2014). It is believed that respiratory effort originates within the brain's anterior cingulate cortex where the efferent copies of motor command from respiratory muscles are centrally processed (Gigliotti 2010). Notably, the changes in the partial pressure of oxygen/carbon dioxide, and neuromuscular work of respiratory muscles may contribute towards the perceived difficulty to breathe (Amann et al. 2010; O'Donnell et al. 2020). Therefore, a combined model which acknowledges the combination of afferent feedback (e.g., chemical changes, breathing discomfort, and chest tightness) and perceptual/affective responses (e.g., inspiratory effort, unsatisfied inspiration) can help to explain the role of respiratory effort within the wider sensation of dyspnea (O'Donnell et al. 2020).

Borg's 15-point RPE scale (Borg 1982) is widely accepted as the most convenient measure of assessing perceived effort. Initially conceived as a surrogate measure of exercise intensity/load (Borg 1982; Gescheider 1997), the use of the RPE scale has adapted to also allow contemporary researchers to obtain a singular gestalt value that simultaneously considers physical stimuli (i.e., velocity/power output), perceptual integration, and the individual inferences gleaned from the present context (Halperin and Emanuel 2020). In addition, the RPE scale (Borg 1982) and its derivatives (e.g., category-ratio 10 and 100, [Borg and Borg 2002]) have also been used to prescribe exercise intensity (Faulkner et al. 2007), quantify training load (Seiler and Kjerland 2006), and assess cardiorespiratory fitness (Faulkner et al. 2007; Mauger et al. 2013).

A novel method that has recently been employed is the use of fixed perceived effort exercise, during which individuals are required to exercise in accordance with their perceptions of effort (Cochrane et al. 2015a, b; Cochrane-Snyman et al. 2016, 2019; Astokorki and Mauger 2017a). Such a task is a unique opportunity for individuals to self-regulate their exercise whilst maintaining a fixed perceived intensity. Furthermore, recent studies (Cochrane et al. 2015a, b) have aligned RPE intensities with established physiological boundaries such as gas exchange threshold (GET) and respiratory compensation point (RCP). In doing so, researchers can begin to characterise the common psychophysiological response patterns that occur during fixed RPE exercise. Therefore, the procedure also allows researchers to examine the influence of additional psychophysiological phenomena (other than perceived effort) on exercise regulation within known intensity domains (Halperin and Emanuel 2020).

However, before implementing a specific protocol in practice, it is important for researchers to compare measures over repeated instances to determine whether they are reliable and that measures are precise. Across numerous laboratories, researchers, and studies, measured values should be accurately reproduced when the same procedure and measurements are repeated (Hopkins 2000). This concept is known as test–retest reliability and must apply to both inter (between individuals) and intra (within individual) levels with intraclass correlation coefficient (ICC) calculations determining whether a test is sufficiently reliable. Additionally, measures such as the standard error measurement (SEM) allow researchers to calculate the precision of these measurements and ascertain whether a substantial difference has occurred within subsequent studies that use the same methodology (Weir 2005).

Several studies have identified that fixed perceived effort activity is reliable. For instance, O'Grady et al. (2021) discerned that exercise at three separate RPE intensities was considered reliable at both the intra- and inter-individual level. Notably, the more intense the fixed effort exercise was, the more reproducible the findings were (i.e., RPE 17 demonstrated better reliability than RPE 9). Likewise, Cochrane-Snyman et al. (2016)—who utilised the more novel method of appropriating RPE intensities to known physiological boundaries—found that performance and electromyographic responses were consistent during 60-min fixed effort exercises. However, this study did not measure the cardiorespiratory markers despite the methodological aim to tailor RPE intensity to a known physiological boundary. Although a later study by the same group (Cochrane-Snyman et al. 2019) did investigate cardiorespiratory responses during fixed perceived effort exercise using this model, no results were presented to determine whether the cardiorespiratory responses were reliable.

Therefore, the purpose of the current study was to examine the test–retest reliability of three separate 30-min cycling trials whereby fixed perceived effort intensities were paired with exercising *at* ($RPE_{GET}$) and *above* ($RPE_{+15\%GET}$) GET. This study tested two main hypotheses. First, both fixed perceived effort intensities would be consistently reproduced. Second, based on findings by previous studies (Eston and Williams 1988; Cochrane-Snyman et al. 2016; O'Grady et al. 2021), performance (e.g., power output [W]), physiological (e.g., heart rate [HR], relative oxygen uptake [$\dot{V}O_2.kg^{-1}$], minute ventilation [$\dot{V}_E$], breathing frequency [BF]), and psychological (e.g., affect, self-efficacy) variables during a higher intensity fixed effort exercise would indicate higher reliability values compared to lower intensity fixed effort exercise.

## Methods

### Participants

Eight healthy (seven male; one female) recreationally active cyclists ([M ± SD] age: 24 ± 2.6 years; stature: 1.75 ± 0.1 m; mass: 72 ± 11.5 kg and maximum oxygen uptake [$\dot{V}O_2max$]: 54 ± 5.8 $ml.kg^{-1}.min^{-1}$) participated in the present study. All participants had at least 2 years of cycling experience (9 ± 3.4 years) and met nationally recognised guidelines for weekly physical activity (659 ± 386 $min·wk^{-1}$). This met the level 3 classification from de Pauw et al. (2013). In addition, all participants were free from underlying cardiorespiratory or other pre-existing medical conditions and injuries that may have inhibited physical performance. None of the participants were currently taking any medication. Prior to providing written informed consent, participants were informed of the procedures, benefits, and risks of the study. The study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the School of Sport and Exercise Sciences Research Ethics Advisory Group (Prop 31_2019_20).

### Perceptual scales

In accordance with recent recommendations by Halperin and Emanuel (2020), the following steps were taken to ensure that the selection, use, and analysis of the RPE scale was adherent to maximising measurement validity. To reduce the ambiguity in the semantic representation of perceived effort, researchers provided a precise and consistent definition of perceived effort as "How hard, heavy and strenuous the exercise consciously feels to drive the working muscles and for your breathing" (Pageaux 2014). Throughout the study, the RPE scale was outlined with the same definition, instructions, and anchors on the 15-point Borg scale (1982) which participants rated their perceptions on. Alongside RPE, the 11-point Feeling Scale (Hardy and Rejeski 1989), measuring in-task affect, was incorporated to acknowledge similar phenomena such as discomfort and tiredness that may not be fully captured by the RPE scale alone. This use of the RPE scale was in accordance with the researchers' collective ontological views.

The Feeling scale considered "How are you feeling at the present moment of the exercise?" on a scale from + 5 'I feel very good' to − 5 'I feel very bad'. Finally, a single-item 11-point Likert scale questioned "How confident are you that you can tolerate the physical and mental effort associated with the cycling task", with responses ranging from 0 'Not Confident at All' to 10 'Extremely Confident' with a mid-point of 5 'Moderately Confident'. This scale was adapted in line with Bandura's (1997) framework. All scales were first explained during the recruitment process to participants.

### Experimental design

This study employed a within-participants randomised crossover design, wherein participants were required to visit the laboratory on eight separate occasions. All experimental sessions were conducted a minimum of 2 days and maximum of 7 days apart. Each participant's visits were scheduled at the same time of day (± 2 h). *Visits 1 and 2* involved identical ramped incremental $\dot{V}O_2max$ tests on a cycle ergometer with an ensuing fixed effort familiarisation cycle. *Visits 3–8* consisted of 30-min fixed effort cycling bouts that matched to one of two intensities corresponding to $RPE_{GET}$ and $RPE_{+15\%GET}$. Each condition was completed three times in a randomised fashion to prevent any order effects. Female participants completed each condition/intensity through one stage of menses (Luteal phase) to reduce any added confounding effects. After completion of all trials, participants were debriefed before being cleared to leave. All procedures took place in the same laboratory setting which had a constant temperate environment ([M ± SD] temperature, 19.3 ± 0.6 °C; humidity, 40.2 ± 4.3%; barometric pressure, 751.5 ± 3.2 mmHg). Participants were instructed to refrain from alcohol and intense exercise in the 48 h preceding testing and to abstain from caffeine consumption in the 4 h pre-testing. All testing took place at least 2 h after the last meal and participants were asked to replicate their eating habits before each session.

## Procedures

### Visits 1 and 2: ramped incremental $\dot{V}O_2$max tests and familiarisations

Upon arrival to the laboratory, anthropometric data were obtained along with a 20 µl resting $[La^-]_b$ sample from the right-hand index finger which was lysed and assessed using an automated analyser (Biosen: C-Line, EKF Diagnostics, GmbH, Barleben, Germany). After this, participants were briefed on the protocols of the ramped incremental test, the scales used during the test, and subsequent familiarisation whilst being fitted with an HR monitor (Cyclus 2: ANT+, Leipzig, Germany) for measurements on a beat-by-beat basis. Participants were then asked to perform a short self-selected five-minute warm-up on the cycle ergometer (Cyclus 2, Leipzig, Germany) which allowed participants to mount their own bike frame for familiarity. Each participant used the same bike frame throughout all visits.

During the completion of the warm-up, the researcher re-explained the use and protocols concerning the RPE scale which would be administered throughout the test. After a completing the warm-up, participants were fitted with a mask that covered the nose and mouth and connected to a flowmeter that was attached to a metabolic cart system (Cortex Metalyser: Model 3B, Leipzig, Germany) which measured gas exchange parameters and pulmonary ventilation (inspired and expired flow rates) on a breath-by-breath basis. The gas analyser was pre-calibrated using a fixed 3-L syringe (Hans Rudolph, Kansas, USA) and known gas concentrations. After participants were fitted to the equipment, confirmed an understanding of the perceptual scales, and provided a resting value for the RPE scale, the ramped incremental test began. The affect and self-efficacy scales were used exclusively during the familiarisation and experimental trials.

For the ramped incremental tests, males were required to cycle at 80 W for 3 min to allow gas parameters to stabilise before commencing the test. Once elapsed, the incremental ramped test began at 100 W and increased incrementally by 25 W·min$^{-1}$. In contrast, females were required to cycle at 40 W for three minutes to allow gas parameters to stabilise before the commencement of the $\dot{V}O_2$max test at 50 W with identical 25 W·min$^{-1}$ ramped increments. These intensities were selected as pilot testing showed that these starting intensities and progressions resulted in all participants reaching volitional exhaustion within the recommended 8–10-min period (Yoon et al. 2007). All participants were informed to maintain a cadence above 80 revolutions·min$^{-1}$ which should gradually increase as cycling intensity became harder until they could no longer sustain the exercise. Each minute (including at 50 [females] or 100 [males] W), RPE was recorded. Cardiorespiratory and power output were monitored continuously (each second) throughout the test. Participants were expected to perform to their maximum perceived ability. Whereupon the participant a) believed they had reached volitional exhaustion or b) cadence dropped below 60 revolutions·min$^{-1}$ for more than 5 s despite strong verbal encouragement, the test was stopped. Additional RPE measures were taken at exhaustion alongside a final $[La^-]_b$ sample.

After the cessation of the ramped incremental test, participants received 15-min passive recovery and then conducted a 10-min familiarisation (5 min at RPE 13 and 15 each) to the fixed perceived effort cycling trials. During these familiarisation trials, participants maintained a cadence between 80 and 90 revolutions·min$^{-1}$ which was then used as reference for the experimental visits. Intensities of RPE 13 and 15 were selected based on previous studies findings as to what $RPE_{GET}$ and $RPE_{+15\%GET}$ correspond to (Cochrane et al. 2015b; Cochrane-Snyman et al. 2016).

### Determination of $RPE_{GET}$ and $RPE_{+15\%GET}$

Individual GETs were determined by utilising a $\dot{V}$-slope method (Beaver et al. 1986) whereby GET corresponded to the point at which $\dot{V}O_2$ values above and below the breakpoint with $\dot{V}CO_2$ diverged from the intersection of the two linear regression lines. For validation, $\dot{V}$-slope was used in conjunction with secondary criteria including: ventilatory equivalents; end-tidal volumes and respiratory exchange ratio. A secondary researcher was used to confirm that GET was assigned at the same place. Once GET was determined, $\dot{V}O_2$ values that were 15% above GET were also calculated. Using these values, the W that was exerted over the course of the ramped incremental test was plotted against the $\dot{V}O_2$ and a linear regression equation ($y = mx + c$) derived the W that corresponded to GET and 15% above GET. Finally, the ramped incremental power output data were plotted against the obtained RPE values in which an identical linear regression equation was used to identify $RPE_{GET}$ and $RPE_{+15\%GET}$. These RPE values were rounded to the nearest whole number. An average of the two values from *Visits 1 and 2* was used as reference RPE points for *Visits 3–8*, experimental visits.

### Fixed perceived effort cycling (experimental sessions)

After participants completed an identical warm-up and baseline measures to *Visits 1 and 2*, participants mounted the ergometer and were asked to cycle at RPE 10 (between "very light" and "light") for 2 min. Once 2 min had elapsed, approximately 30–60 s was afforded for participants to ramp up to the required RPE intensity based on average times to reach the required RPE in pilot testing.

The researcher(s) stressed that the task was a fixed effort trial, meaning RPE must remain constant throughout. As a result, power output changes were expected; therefore, participants could change their power output by increasing/decreasing the virtual gears on the ergometer to ensure that the appropriate RPE was maintained throughout the entirety of the fixed effort cycles. It was advised that participants maintained a cadence between 80 and 90 revolutions min$^{-1}$ throughout and that this cadence was replicated (± 2 revolutions·min−1) in all subsequent experimental visits.

Throughout the fixed effort trials, all exercise-related data except cadence were screened from the participants to ensure that performance was appropriated according to a fixed perceived effort. Every 2 min, the researcher would reaffirm with the participant that exercise intensity was being tailored to the appropriate perceived effort rating. During fixed effort cycling, power output and cardiorespiratory markers were extracted continuously (each second) throughout the 30-min exercise. Every 5 min, including baseline (Minute 0), [La$^-$]$_b$, affective valence and self-efficacy were recorded. Fig. 1 depicts all testing procedures.

After the completion of all visits, participants were fully debriefed before being permitted to leave.

## Analysis

Continuous data (e.g., HR, gas parameters) from experimental session data were averaged into six discrete 5-min time zones (TZ) (e.g., TZ1 = average from Minute 00:00–Minute 04:59). Other data (e.g., [La$^-$]$_b$, perceptual measures) were grouped based on when they were extracted (e.g., minute 0, 5, etc.). Finally, all data were also averaged over the entirety of the exercise as 'overall' (average from Minute 0–Minute 30 or TZ1–TZ6).

All data were exported to SPSS (IBM: v.26, New York, USA) where data were assessed for normality and symmetry. Normality was assessed using the Shapiro–Wilk test and visual inspection of Q–Q plots before any subsequent analysis.

Power output, cardiorespiratory (e.g., HR, $\dot{V}O_2$.kg$^{-1}$), and RPE responses from the ramped incremental tests were analysed according to 30-s averaged values. For *Visits 1 and 2*, a mean across both visits was calculated for values at peak, GET and 15% above GET. A single-measures, two-way random ICC (2,1) was calculated between both ramped incremental tests for peak, GET, and 15% above GET values with accompanying standard error measurements (SEM) to assess the test–retest reliability of *Visits 1 and 2*. ICC values were interpreted as $> 0.9$ excellent reliability, $> 0.8$ good reliability, $> 0.6$ questionable reliability and $< 0.6$ poor reliability A Pearson ($r$) correlation coefficient was also conducted to assess the relationship of performance (W), physiological (HR, $\dot{V}O_2$.kg$^{-1}$), and psychometric (RPE) values between each ramped incremental test with values $\geq 0.9$

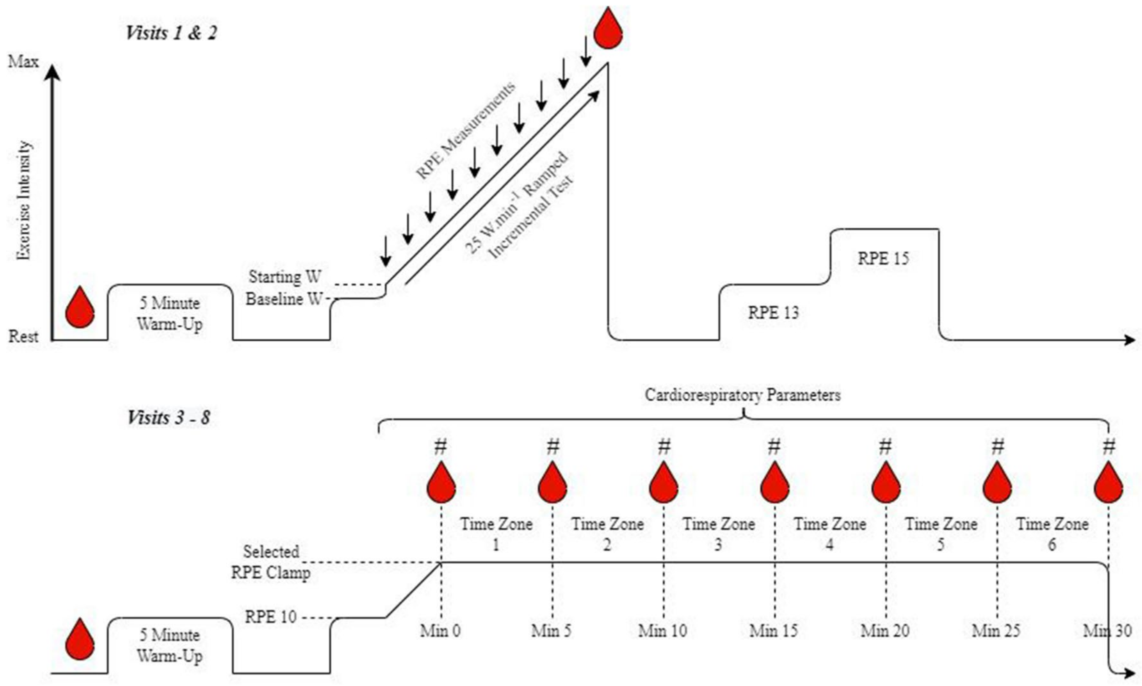


**Fig. 1** Representation of study protocols. Legend: # denotes affect and self-efficacy measurements

indicating very strong, ≥ 0.8 strong, ≥ 0.6 moderate, ≥ 0.4 weak, and < 0.4 no association.

Test–retest (inter-individual) reliability for data within *Visits 3–8* (experimental sessions) was assessed across TZ averaged and ‘overall’ (30-min averaged) data for power output, HR, $[\mathrm{La}^-]_\mathrm{b}$ gas parameters ($\dot{V}\mathrm{O}_2.\mathrm{kg}^{-1}$,$\dot{V}_\mathrm{E}$, BF), and psychometric (affect and self-efficacy) data. When calculating reliability using a single-measures, two-way random ICC (2,1) and accompanying SEM, data from each visit within each condition were used. The SEM was used to calculate a minimal difference (see Eq. 1). Subsequent 95% confidence intervals (95% CI) for each of these variables were calculated by subtracting and adding the minimal difference to the group mean. A coefficient of variation (CoV) was also used to identify intra-individual variation for ‘overall’ 30-min averaged W, $\dot{V}\mathrm{O}_2.\mathrm{kg}^{-1}$, HR, $\dot{V}_\mathrm{E}$, BF, and $[\mathrm{La}^-]_\mathrm{b}$ with measurement errors of ≤ 5% indicative of reliability (Hopkins 2000; Tate and Klett 1959). As coefficients of variations were presented as percentages, the Tate and Klett (1959) method was used to calculate 95% CI for measures of intra-individual reliability (Weir 2005)

$$\text{Minimal Difference} = \text{SEM} \times 1.96 \times \sqrt{2}. \tag{1}$$

A series of 2 × 6 repeated-measures ANOVAs were used to examine the condition and condition × time effects at every 5 min (TZ) for performance (W) and physiological (HR, $\dot{V}\mathrm{O}_2.\mathrm{kg}^{-1}$, $\dot{V}_\mathrm{E}$, and BF) variables between conditions. Similar 2 × 7 repeated-measures ANOVAs were used for $[\mathrm{La}^-]_\mathrm{b}$ and psychological (affect, self-efficacy) variables between conditions that were taken at every 5-min interval (min 0, 5, 10, etc.). Values for each TZ were taken as an average across all three visits. Averages of the three visits for 30-min ‘overall’ values were assessed for differences between conditions using a paired samples *t* test or non-parametric equivalent. Repeated-measures ANOVA tests used a Mauchley’s test wherein if sphericity was violated, a Greenhouse–Geisser adjustment was employed to the appropriate degrees of freedom to counter the increased risk of type one error. For all repeated-measures ANOVAs, significant main effects across condition and time were followed up with a one-way repeated-measures ANOVA and a subsequent Bonferroni post hoc test for specific TZ pairwise comparisons. Non-parametric equivalents (Friedman’s test, Wilcoxon signed-rank test) were used when data violated normality. An alpha level of $P \leq 0.05$ was employed to assess statistical significance, whilst partial eta squared ($\eta_\mathrm{p}^2$) provided an estimate of effect size of the ANOVAs (small = 0.01, medium = 0.10, large = 0.25). Any follow-up pairwise comparisons and *t* tests used a Cohen’s *d* calculation to determine effect size (≥ 0.2 = small, ≥ 0.5 = moderate, ≥ 0.8 = large).

# Results

## Visits 1 and 2 (ramped incremental tests)

Correlation coefficient between visits: mean group data demonstrated a Peak W of 349 ± 36 W which showed a strong correlation between ramped incremental visits (ICC = 0.962, SEM = 6.97, $r = 0.962$). Mean peak $\dot{V}\mathrm{O}_2.\mathrm{kg}^{-1}$ was 52 ± 7 $\mathrm{mL.kg}^{-1}.\mathrm{min}^{-1}$ and demonstrated a questionable correlation between ramped incremental trials (ICC = 0.792, SEM = 3.05, $r = 0.925$). Finally, mean peak HR was 194 ± 6 $\mathrm{b.min}^{-1}$ and demonstrated a strong correlation between ramped incremental trials (ICC = 0.916, SEM = 1.62, $r = 0.945$).

Mean W corresponding to GET was 201 ± 29 W and demonstrated a strong correlation between ramped incremental tests (ICC = 0.957, SEM = 6.01, $r = 0.968$). Mean $\dot{V}\mathrm{O}_2.\mathrm{kg}^{-1}$ at GET was 33 ± 4 $\mathrm{mL.kg}^{-1}.\mathrm{min}^{-1}$ and demonstrated a strong correlation (ICC = 0.929, SEM = 1.12, $r = 0.960$). Finally, mean HR at GET was 158 ± 7 $\mathrm{b.min}^{-1}$ and demonstrated a questionable correlation between ramped incremental visits (ICC = 0.668, SEM = 4.14, $r = 0.629$).

Mean W corresponding to 15% above GET was 236 ± 34 W and demonstrated a strong correlation between ramped incremental trials (ICC = 0.955, SEM = 7.31, $r = 0.963$). Mean $\dot{V}\mathrm{O}_2.\mathrm{kg}^{-1}$ at 15% above GET was 38 ± 5 $\mathrm{mL.kg}^{-1}.\mathrm{min}^{-1}$ and demonstrated a strong correlation between ramped incremental trials (ICC = 0.910, SEM = 1.49, $r = 0.962$). Finally, mean HR at 15% above GET was 168 ± 8 $\mathrm{b.min}^{-1}$ and demonstrated a questionable reliability between ramped incremental trials (ICC = 0.664, SEM = 4.36, $r = 0.677$).

Mean RPE at GET was 13.0 (13–somewhat hard). Mean RPE at 15% above GET was 14.7 (15–hard). Participant RPE values at GET ranged from 12 to 14, whilst RPE values at 15% above GET ranged from 14 to 16.

## Visits 3–8 (experimental sessions)

Test–retest reliability: Single measure test–retest reliability measures indicated that overall (30-min averaged) measures of W and $\dot{V}\mathrm{O}_2.\mathrm{kg}^{-1}$ demonstrated an excellent degree of reliability within the $\mathrm{RPE}_\mathrm{GET}$ condition (Table 1). Overall HR, $[\mathrm{La}^-]_\mathrm{b}$ (Table 1), $\dot{V}_\mathrm{E}$ (ICC = 0.839, SEM = 5.08), and self-efficacy (ICC = 0.807, SEM = 0.45) measures showed a good degree of reliability, whilst overall BF (ICC = 0.728, SEM = 1.66) and affect (ICC = 0.749, SEM = 0.48) showed a questionable reliability within the $\mathrm{RPE}_\mathrm{GET}$ condition Within the $\mathrm{RPE}_{+15\%\mathrm{GET}}$ condition, overall measures of W, $\dot{V}\mathrm{O}_2.\mathrm{kg}^{-1}$, $[\mathrm{La}^-]_\mathrm{b}$ (Table 2), $\dot{V}_\mathrm{E}$ (ICC = 0.963, SEM = 3.26), and BF (ICC = 0.969,

**Table 1** Group mean $\text{RPE}_{\text{GET}}$ inter- and intra-individual results for each time zone and overall

| Variable | TZ | Mean | SD | ICC (2,1) | SEM | 95% CI | *CoV* |
|---|---|---|---|---|---|---|---|
| W | 1 | 184 | 8.1 | 0.903 | 2.5 | 177–192 | 4.4 |
| | 2 | 182 | 8.0 | 0.919 | 2.3 | 176–188 | |
| | 3 | 179 | 7.3 | 0.924 | 2.0 | 174–185 | |
| | 4 | 176 | 8.4 | 0.906 | 2.6 | 169–184 | |
| | 5 | 176 | 9.7 | 0.884 | 3.3 | 166–184 | |
| | 6 | 175 | 9.8 | 0.887 | 3.3 | 166–184 | |
| | Overall | 179 | 8.0 | 0.915 | 2.3 | 172–185 | |
| HR | 1 | 144 | 8.8 | 0.566 | 5.8 | 128–160 | 3.1 |
| | 2 | 153 | 12.4 | 0.882 | 4.2 | 142–165 | |
| | 3 | 155 | 13.2 | 0.884 | 4.5 | 143–168 | |
| | 4 | 156 | 12.6 | 0.806 | 5.5 | 141–171 | |
| | 5 | 157 | 12.7 | 0.778 | 6.0 | 141–174 | |
| | 6 | 158 | 13.0 | 0.805 | 5.8 | 142–174 | |
| | Overall | 154 | 11.9 | 0.825 | 5.0 | 140–168 | |
| $\dot{V}O_2$.kg$^{-1}$ | 1 | 33 | 5.5 | 0.915 | 1.6 | 29–38 | 4.2 |
| | 2 | 35 | 6.7 | 0.950 | 1.5 | 31–39 | |
| | 3 | 35 | 6.9 | 0.943 | 1.7 | 30–40 | |
| | 4 | 35 | 7.1 | 0.921 | 2.0 | 29–40 | |
| | 5 | 35 | 7.3 | 0.928 | 2.0 | 29–40 | |
| | 6 | 35 | 7.6 | 0.910 | 2.3 | 29–41 | |
| | Overall | 35 | 6.8 | 0.932 | 1.8 | 30–40 | |
| $[\text{La}^-]_b$ | Min 0 | 2.46 | 0.6 | 0.735 | 0.3 | 1.55–3.37 | 12.7 |
| | Min 5 | 3.63 | 1.3 | 0.837 | 0.5 | 2.21–5.04 | |
| | Min 10 | 4.04 | 1.9 | 0.820 | 0.8 | 1.85–6.23 | |
| | Min 15 | 4.24 | 2.2 | 0.881 | 0.8 | 2.10–6.37 | |
| | Min 20 | 4.10 | 2.1 | 0.823 | 0.9 | 1.61–6.60 | |
| | Min 25 | 4.05 | 2.3 | 0.835 | 0.9 | 1.51–6.59 | |
| | Min 30 | 4.20 | 2.6 | 0.831 | 1.1 | 1.26–7.14 | |
| | Overall | 3.34 | 1.6 | 0.849 | 0.6 | 1.67–5.01 | |

SEM = 0.96) demonstrated an excellent degree of reliability, whilst HR showed a good degree of reliability (Table 2), and affect (ICC = 0.770, SEM = 0.65), and self-efficacy (ICC = 0.711, SEM = 0.65) demonstrated questionable reliability. Main group mean overall and TZ results can be seen in Tables 1 and 2. Additional tables concerning $\dot{V}_E$, BF, affect, and self-efficacy can be found in supplementary materials.

When assessing 5-min TZ data, W reliability within the $\text{RPE}_{\text{GET}}$ condition was excellent from TZ1–4, whilst TZ5–6 were considered good. Within the $\text{RPE}_{+15\%\text{GET}}$ condition, all time zones except TZ1 indexed an excellent degree of reliability.

During the $\text{RPE}_{\text{GET}}$ and $\text{RPE}_{+15\%\text{GET}}$ condition, all $\dot{V}O_2$.kg$^{-1}$ values demonstrated an excellent degree of reliability across all time zones. During the $\text{RPE}_{\text{GET}}$ condition, HR values showed a good degree of reliability within TZ2, 3, 4, and 6, whilst TZ5 showed questionable reliability and TZ1 showed poor reliability. Alternately, within the $\text{RPE}_{+15\%\text{GET}}$ condition, all HR TZ data showed a good degree of reliability.

During the $\text{RPE}_{\text{GET}}$ condition, $\dot{V}_E$ showed good reliability across all time zones (ICC = 0.801–0.871, SEM = 3.54–6.92) except TZ5 which showed questionable reliability (ICC = 0.778, SEM = 6.78). During the $\text{RPE}_{+15\%\text{GET}}$ condition, excellent reliability across all time zones (ICC = 0.933–0.951, SEM = 4.03–5.27) was observed except at TZ1 which showed good reliability (ICC = 0.827, SEM = 4.76). During the $\text{RPE}_{\text{GET}}$ condition, BF showed questionable validity across all time zones (ICC = 0.640–0.776, SEM = 1.37–2.15), whereas the $\text{RPE}_{+15\%\text{GET}}$ condition showed excellent reliability across all time zones (ICC = 0.903–0.961, SEM = 1.21–1.85) except TZ1 which showed good reliability (ICC = 0.889, SEM = 1.31).

During the $\text{RPE}_{\text{GET}}$ condition, $[\text{La}^-]_b$ demonstrated good reliability at every timepoint except minute 0 (questionable) (Table 1), whereas the $\text{RPE}_{+15\%\text{GET}}$ condition demonstrated excellent reliability of measures taken at minute 15–30

**Table 2** Group mean $RPE_{+15\%GET}$ inter- and intra-individual results for each time zone and overall

| Variable | TZ | Mean | SD | ICC (2,1) | SEM | 95% CI | *CoV* |
|---|---|---|---|---|---|---|---|
| W | 1 | 219 | 10.9 | 0.896 | 3.52 | 209–229 | 2.2 |
| | 2 | 208 | 5.0 | 0.941 | 1.22 | 205–212 | |
| | 3 | 201 | 7.0 | 0.928 | 1.89 | 195–206 | |
| | 4 | 199 | 4.7 | 0.945 | 1.11 | 196–202 | |
| | 5 | 195 | 4.8 | 0.960 | 0.95 | 193–198 | |
| | 6 | 193 | 5.5 | 0.943 | 1.32 | 190–197 | |
| | Overall | 203 | 4.3 | 0.962 | 0.84 | 201–206 | |
| HR | 1 | 159 | 9.0 | 0.807 | 3.97 | 148–170 | 1.6 |
| | 2 | 167 | 10.5 | 0.849 | 4.10 | 156–179 | |
| | 3 | 168 | 11.1 | 0.853 | 4.24 | 156–180 | |
| | 4 | 169 | 10.4 | 0.874 | 3.70 | 159–179 | |
| | 5 | 170 | 11.0 | 0.853 | 4.22 | 158–182 | |
| | 6 | 171 | 11.9 | 0.868 | 4.31 | 159–183 | |
| | Overall | 167 | 10.5 | 0.876 | 3.69 | 157–178 | |
| $\dot{V}O_2.kg^{-1}$ | 1 | 39 | 5.5 | 0.902 | 1.73 | 34–44 | 2.7 |
| | 2 | 40 | 6.1 | 0.947 | 1.40 | 37–44 | |
| | 3 | 39 | 6.1 | 0.931 | 1.59 | 35–44 | |
| | 4 | 39 | 6.0 | 0.939 | 1.47 | 35–43 | |
| | 5 | 39 | 6.4 | 0.937 | 1.62 | 35–43 | |
| | 6 | 39 | 6.5 | 0.936 | 1.64 | 34–43 | |
| | Overall | 39 | 6.0 | 0.951 | 1.34 | 36–43 | |
| $[La^-]_b$ | Min 0 | 3.36 | 0.9 | 0.813 | 0.4 | 2.28–4.44 | 9.2 |
| | Min 5 | 6.25 | 2.2 | 0.819 | 0.9 | 3.68–8.82 | |
| | Min 10 | 6.95 | 2.9 | 0.871 | 1.0 | 4.07–9.84 | |
| | Min 15 | 6.76 | 3.2 | 0.948 | 0.7 | 4.74–8.79 | |
| | Min 20 | 6.86 | 3.5 | 0.941 | 0.8 | 4.51–9.20 | |
| | Min 25 | 6.85 | 3.8 | 0.953 | 0.8 | 4.58–9.11 | |
| | Min 30 | 6.70 | 3.8 | 0.917 | 1.1 | 3.69–9.72 | |
| | Overall | 5.47 | 2.4 | 0.939 | 0.6 | 3.80–7.13 | |

and good reliability at measures taken from minute 0–10 (Table 2).

During the $RPE_{GET}$ condition, affect demonstrated good reliability at minute 0–5 (ICC = 0.831 and 0.826, SEM = 0.53 and 0.45), questionable reliability at minute 10, 15, and 25 (ICC = 0.686–0.786, SEM = 0.41–0.68), and poor reliability at minute 20 and 30 (ICC = 0.597 and 0.488, SEM = 0.69 and 0.81). During the $RPE_{+15\%GET}$ condition, affect demonstrated questionable reliability from minute 0–15 and minute 30 (ICCs = 0.621–0.720, SEM = 0.80–0.95), and poor reliability at minute 20–25 (ICCs = 0.552–0.592, SEM = 0.79–0.95).

Self-efficacy data during the $RPE_{GET}$ condition demonstrated good reliability at minute 0, 5, and 30 (ICCs = 0.812–0.883, SEM = 0.43–0.63), questionable reliability at minute 10–20, (ICCs = 0.636–0.765, SEM = 0.59–0.63), and poor reliability at minute 25 (ICC = 0.505, SEM = 0.57). Self-efficacy data during the $RPE_{+15\%GET}$ condition demonstrated a good reliability at minute 0 and 5 (ICCs = 0.850 and 0.815, SEM = 0.75 and 0.77), questionable reliability at minute 10 (ICC = 0.607, SEM = 0.99), and poor reliability at minute 15–30 (ICCs = 0.427–0.524, SEM = 0.84–0.99).

Intra-individual reliability: Measures of intra-individual reliability demonstrated that overall W varied by a mean ± SD of 4.4 ± 1.5% (95% CI 2.9–8.9%) within the $RPE_{GET}$ condition, whereas the $RPE_{+15\%GET}$ condition varied by 2.2 ± 1.1% (95% CI 1.5–4.5%) on average.

Overall $\dot{V}O_2.kg^{-1}$ was 4.2 ± 1.5% (95% CI 2.8–8.5%) during the $RPE_{GET}$ condition and 2.7 ± 1.3% (95% CI 1.8–5.5%) during the $RPE_{+15\%GET}$ condition. Variability in Overall HR was 3.1 ± 1.1% (95% CI 2.0–6.2%) in the $RPE_{GET}$ condition and 1.6 ± 1.2% (95%CI 1.1–3.3%) in the $RPE_{+15\%GET}$ condition.

Mean ± SD overall $\dot{V}_E$ variability was 6.2 ± 1.2% (95% CI 3.2–9.3) during the $RPE_{GET}$ condition and 2.8 ± 1.1% (95% CI 1.0–4.6) during the $RPE_{+15\%GET}$ condition. Overall BF variability was 4.0 ± 2.0% (95% CI 3.1–5.0) during the $RPE_{GET}$ condition and 2.6 ± 1.1% (95% CI 1.9–3.3) during the $RPE_{+15\%GET}$ condition. Mean ± SD overall $[La^-]_b$

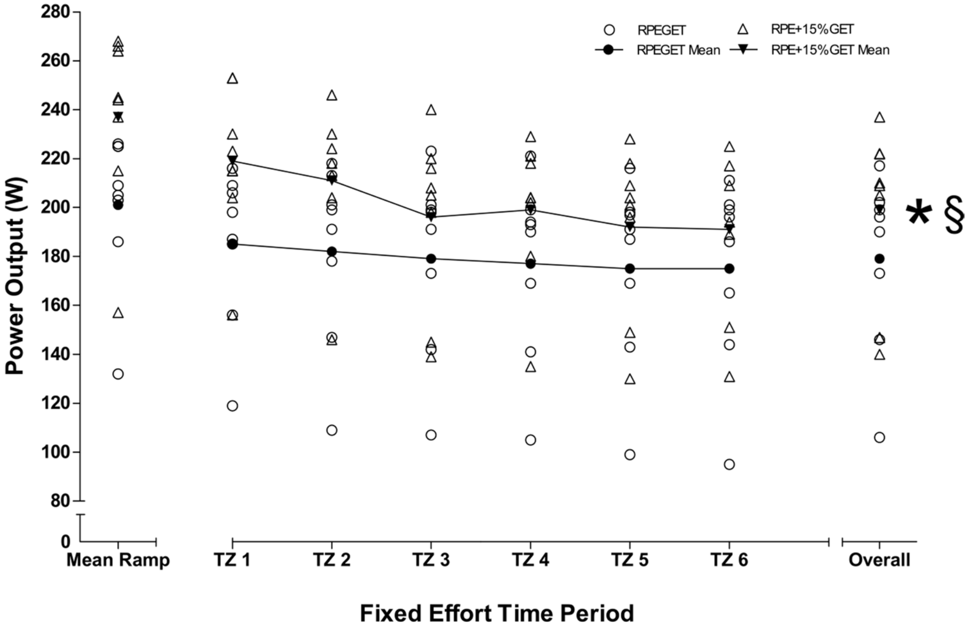


**Fig. 2** Mean ± SD across all three condition experimental visits in time-lapsed changes in W at each 5-min TZ and overall, during the 30-min fixed effort cycling exercise. Legend: * denotes a significant difference in overall values between conditions ($P < .05$), § denotes a moderate effect size

variability was 12.7 ± 9.6% (95% CI 12.4–13.0) during the RPE$_{GET}$ condition and 9.2 ± 7.3% (95% CI 8.9–9.4) during the RPE$_{+15\%GET}$ condition.

Differences between RPE$_{GET}$ and RPE$_{+15\%GET}$ conditions and time zones: A series of 2 × 6 repeated-measures ANOVAs determined significantly large condition effects for W, HR, $\dot{V}O_2$.kg$^{-1}$, $\dot{V}_E$, and BF measures ($F = 43.377–69.336$, $P = 0.001–0.002$, $\eta_\rho^2 = 0.861–0.908$). Significantly large condition × time effects were observed for W, $\dot{V}O_2$.kg$^{-1}$, and BF ($F = 4.950–6.609$, $P = 0.002–0.007$, $\eta_\rho^2 = 0.366–0.486$).

A series of 2 × 7 repeated-measures ANOVAs determined significantly large condition effects for [La$^-$]$_b$, affect, and self-efficacy measures ($F = 19.505–59.163$, $P = 0.001–0.003$, $\eta_\rho^2 = 0.736–0.894$). Significantly large condition × time effects were observed for [La$^-$]$_b$ and affect ($F = 6.811–10.241$, $P = 0.001–0.017$, $\eta_\rho^2 = 0.493–0.594$).

Additional one-way repeated-measures ANOVAs determined significant changes over time in W, HR, and BF during the RPE$_{GET}$ condition ($F = 5.530–20.494$, $P = 0.001–0.017$). Significant changes over time were observed for W, HR, BF, [La-]b, and affect during the RPE$_{+15\%GET}$ condition ($F = 6.485–28.295$, $P = 0.001–0.031$).

During the RPE$_{GET}$ condition, follow-up Bonferroni corrected post hoc analyses revealed significant differences in HR at TZ1 and 4–6 ($P = 0.019–0.023$) and TZ2 and 3 ($P = 0.018$), and BF at TZ1–2 and 4 ($P = 0.029–0.042$). During the RPE$_{+15\%GET}$ condition, Bonferroni post hoc analyses determined significant differences in: W at TZ1 and 3–6 ($P = 0.006–0.024$) and TZ2 and 3–6 ($P = 0.003–0.025$); HR at TZ1 and 2–6 ($P = 0.010–0.025$); $\dot{V}O_2$.kg$^{-1}$ at TZ2 and 3–4 ($P = 0.001–0.018$); BF at TZ2 and 5 ($P = 0.024$); and affect at minute 0–20 and minute 30 ($P = 0.036–0.050$). Overall W, HR, $\dot{V}O_2$.kg$^{-1}$, BF, [La$^-$]$_b$, and self-efficacy were significantly different between conditions ($t = 4.362–8.497$, $P = 0.001–0.003$). Overall $\dot{V}_E$ and affect were significantly different between conditions ($Z = 2.524–2.527$, $P = 0.012$). Large effect sizes were observed for HR, $\dot{V}_E$, BF, [La$^-$]$_b$, affect, and self-efficacy (d = 1.00–1.58). Moderate effect sizes were observed for W and $\dot{V}O_2$.kg$^{-1}$ ($d = 0.58–0.75$). Figures 2, 3, 4, 5 depict the changes of three visit averages in performance, physiological, and psychological during the fixed perceived effort trials.

## Discussion

The present study aimed to assess the test–retest reliability of 30-min fixed perceived effort cycling trials which used a linear regression model to fix RPE intensity according to physiological thresholds. Foremostly, results showed that 30-min fixed effort cycling demonstrated good test–retest and intra-individual reliability amongst a cohort of recreationally active cyclists. This was supported by ICC values which evidenced that overall performance measures (e.g., W) demonstrated an excellent degree of reliability (> 0.900) between visits in both conditions. In addition, overall physiological variables, such as $\dot{V}O_2$.kg$^{-1}$, $\dot{V}_E$, BF, and [La$^-$]$_b$, also demonstrated an excellent degree of reliability (> 0.900) in the RPE$_{+15\%GET}$ condition. Test–retest reliability for HR demonstrated good reliability (> 0.800) across both conditions.

Other research has also exhibited that perception of effort remains consistent over different exercise tasks such

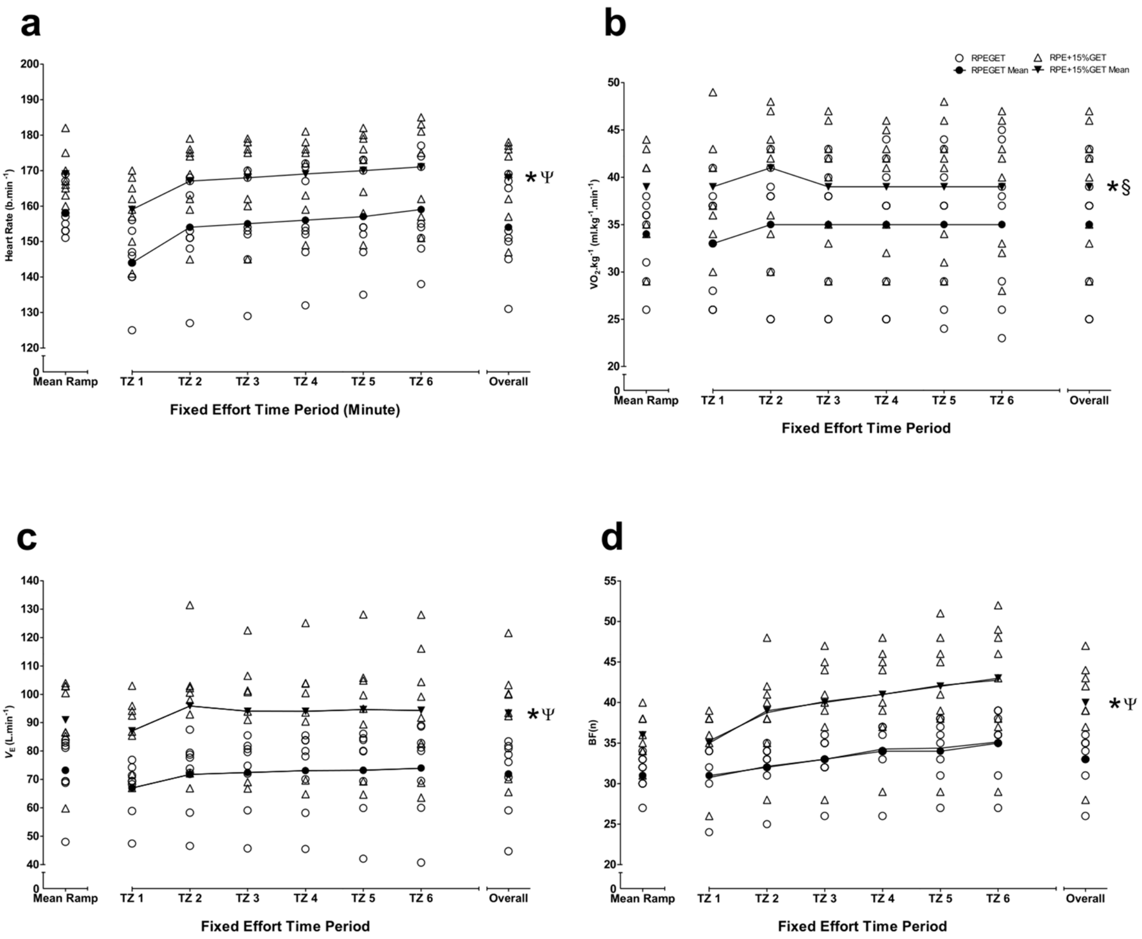
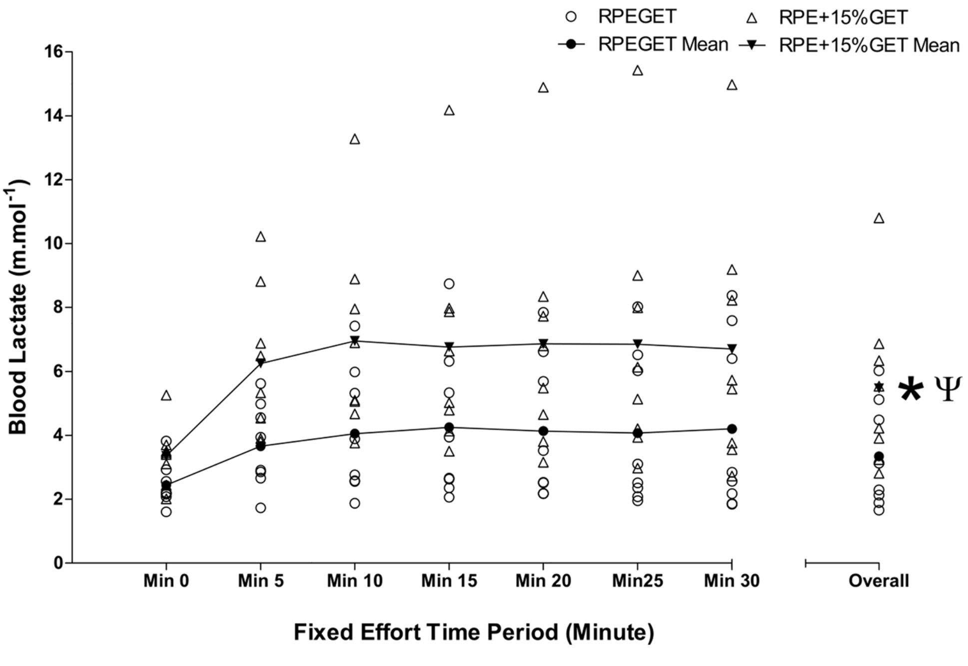


**Fig. 3** Mean ± SD across all three condition experimental visits in time-lapsed changes in cardiorespiratory parameters (**a** = HR, **b** = $\dot{V}O_2.kg^{-1}$, **c** = $V_E$, **d** = BF) at each five-minute TZ and overall, during the 30-min fixed effort cycling exercise. Legend: * denotes a significant difference in overall values between conditions ($P < .05$), § denotes a moderate effect size, and Ψ denotes a large effect size

**Fig. 4** Mean ± SD across all three condition experimental visits in time-lapsed changes in $[La^-]_b$ at each 5-min timepoint and overall, during the 30-min fixed effort cycling exercise. Legend: * denotes a significant difference in overall values between conditions ($P < .05$), § denotes a moderate effect size, and Ψ denotes a large effect size



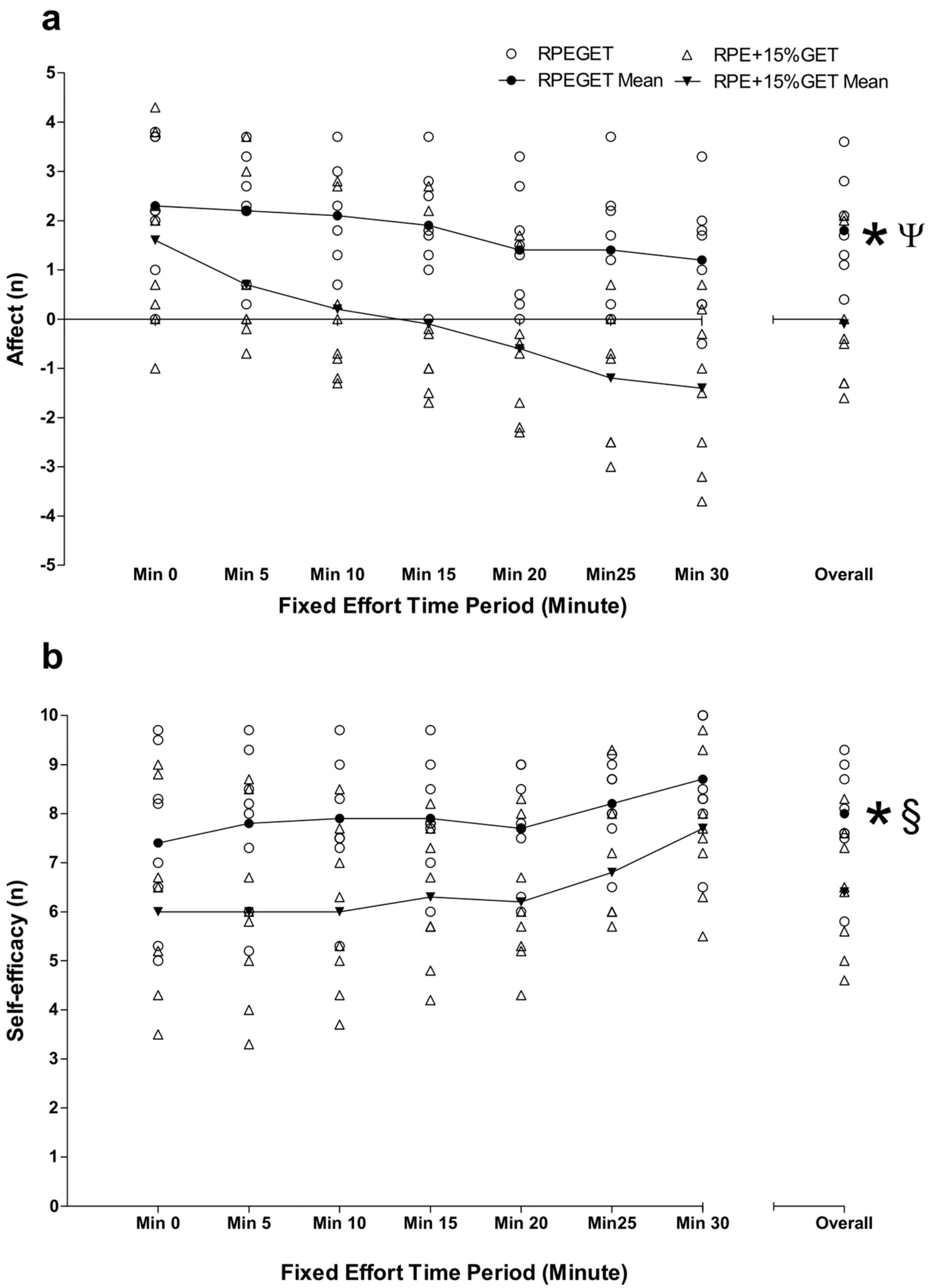


**Fig. 5** Mean ± SD across all three condition experimental visits in time-lapsed changes in psychological parameters: **a** = affective valence, **b** = self-efficacy at each 5-min timepoint and overall, during the 30-min fixed effort cycling exercise. Legend: * denotes a significant difference in overall values between conditions ($P < .05$), § denotes a moderate effect size, and Ψ denotes a large effect size

time-to-exhaustion trials (Okuno et al. 2015) and time-trials (Borg et al. 2018). Furthermore, irrespective of exercise modality, previous studies (Cochrane et al. 2015a, b; Eston and Williams 1988) have identified that fixed perceived effort exercise can be reliably replicated across visits. Such findings are consistent with those observed in this study as measures of performance (W) and physiological response ($\dot{V}O_2$.kg$^{-1}$,$\dot{V}_E$, BF, and [La$^-$]$_b$) showed excellent measures of test–retest reliability (ICC = > 0.900 with small < 6% SEM from the group mean) (Weir 2005). Therefore, it appears that recreationally active athletes can consistently reproduce physical efforts that are regulated by perceptions alone. This may be beneficial for practitioners and coaches alike in the future who lack the resources to measure intricate psychophysical markers that relate to specific workloads and physiological thresholds. Instead, RPE can be used as a surrogate measure during physical activity.

In addition, the present study also assessed intra-individual reliability measures, in which participants demonstrated low CoV values (≤ 5%) and narrow 95% CI for overall performance (W) and physiological ($\dot{V}O_2$.kg$^{-1}$, HR, $\dot{V}_E$, and BF) variables. However, it was notable that [La$^-$]$_b$ varied significantly (12.7% in RPE$_{GET}$ and 9.2% in RPE$_{+15\%GET}$). This finding may discredit the use of lactate as a reliable indicator of exercise intensity if variations between individuals exist so prominently. For instance, the use of maximal lactate steady state has come under increased scrutiny in recent years as opposed to other mathematical models to determine maximal aerobic capacity (Jones et al. 2019). As such, these arguments may be further validated by the findings of the current study.

As noted, only one study to date (Cochrane-Snyman et al. 2016) has explored the reliability of performance and physiological parameters during a fixed effort exercise in which RPE has been tailored to known physiological thresholds/domains. However, this study only utilised correlation coefficients and ICCs to assess the reliability of repeated fixed effort performance, despite research advocating that 95% confidence intervals are a more robust alternative (Hopkins 2000). At the intra-individual level, participants of the present study were able to replicate their efforts consistently between visits in both the RPE$_{GET}$ and RPE$_{+15\%GET}$ condition. Moreover, the 95% CI for most participants remained below 5% to further substantiate this conviction. Paton and Hopkins (2001) identified that self-paced cycling trials usually produce variances of 2–3%. The findings of the current study—particularly data in the RPE$_{+15\%GET}$ condition—remain close to this range of variances as PO, $\dot{V}O_2$.kg$^{-1}$, and HR demonstrated CoVs between 3.1 and 4.4% in the RPE$_{GET}$ condition, and 1.6–2.7% in the RPE$_{+15\%GET}$ condition.

Many have ascribed this consistency in performance to the athlete's familiarity (i.e., experience level, practice) to the exercise tasks. With this is mind, several factors can help rationalise why this study showed the degree of reliability it did, and subsequently inform future research studies to obtain similarly reliable and comparable data. First, the participants that were recruited within this study were all healthy, active, and experienced cyclists. In doing so, this likely led to a more homogenous sample which has consequences for the reliability measures that are calculated (Hopkins 2000). All participants demonstrated very good-to-excellent physiological measures (e.g., $\dot{V}O_2$max, %$\dot{V}O_2$max at GET) during the ramped incremental trials (de Pauw et al. 2013). Therefore, having a collection of participants with a narrower distribution of physiological capabilities compared to other studies (Cochrane et al. 2015a; Bergstrom et al. 2015) could explain the low CoV values and confidence intervals observed in this study.

In addition, as all participants were trained, albeit recreationally, it may be assumed that participants in this study were more attuned to the underlying physiological signals (Elferink-Gemser and Hettinga 2017) during the fixed effort trials compared to previous studies that have used less-trained cohorts (e.g., Cochrane et al. 2015a). Notably, this study involved fixed effort exercise which was aligned to known physiological thresholds, such as GET. Thus, a cohort of currently active individuals who are aware of the typical physiological sensations and perceptions associated with such thresholds could mean that it became substantially easier to taper their efforts according to the RPE value itself as well as the physiological sensations associated with that RPE (Lamb et al. 1999).

Moreover, another critical factor to the reliability of this study could have been the employment of multiple familiarisation trials. Conducting exercise at a fixed RPE is a relatively artificial exercise task; therefore, the opportunity for participants to familiarise themselves twice before the experimental trials could be a key factor. Extant literature has evidenced that the inclusion of familiarisation trials significantly improves the validity and reproducibility of performance indices during self-regulated RPE-based exercise (Lim et al. 2016). Furthermore, Mauger et al. (2014) determined that a cohort active males could replicate fixed effort exercises even without reference to the scale, relying solely on internal psychophysical sensations due to previous experience.

Another notable finding of this study was that RPE$_{+15\%GET}$ results demonstrated much lower variability at both the inter- and intra-individual levels compared to the RPE$_{GET}$ condition. A previous study by O'Grady et al. (2021) determined that fixed effort exercise at higher RPE values rendered lower between and within individual variances in power output and cardiorespiratory parameters compared to fixed effort exercise at lower RPE values. In addition, other studies appear to share similar conclusions based on their results (Eston and Williams 1988; Cochrane-Snyman et al. 2016).

However, it was not explained why harder intensity fixed effort exercise appears to be better replicated than lower intensity fixed effort exercise.

One possible suggestion is that during harder intensity exercise, participants may employ different methods of decision-making according to the different physiological sensations associated with harder intensity compared to lower intensity exercise (Renfree et al. 2014). To illustrate, when exercising at $\mathrm{RPE}_{+15\%\mathrm{GET}}$, participants usually begin exercising within the heavy intensity domain (Gaesser and Poole 1996). Whilst in this domain, athletes experience growing levels of metabolites (e.g., H + ions), nociceptive stimulation (Mauger 2014), and afferent feedback (Amann et al. 2009). As a result, Renfree et al. (2014) suggest that this may engender athletes to adopt more heuristic decision-making processes. This is because the overbearing discomfort and negatively oriented sensations/perceptions—as seen in this study (Fig. 5)—that arise due to harder intensity exercise may cause athletes to make decisions based on more select pieces of information to save effort (Gigerenzer and Gaissmaier 2011). Therefore, responses become more 'primal' and 'instinctive', meaning that they may be more easily replicated as they are based on stable trait-like factors.

On the other hand, exercise at $\mathrm{RPE}_{\mathrm{GET}}$ is expected to occur entirely within the moderate intensity domain whereby metabolite production equals metabolite clearance (Gaesser and Poole 1996). Therefore, the athlete experiences fewer negative sensations and perceptions, such as discomfort and pain. Consequently, Renfree et al. (2014) suggest that this would endear the athlete to employ more rational-based decision-making. As a result, more situational factors are considered when regulating exercise intensity, which could translate into more variances in behaviour overall. However, as this study did not monitor the underlying decision-making processes during the fixed effort exercise, firmer conclusions cannot be drawn. Nonetheless, recent studies have employed the use of a novel "Think-Aloud" protocol which allows researchers to understand the underlying thought and decision-making processes that are articulated during an endurance event (Whitehead et al. 2018). In line with this, future research may wish to consider the use of Think-Aloud approaches to begin to discern how effort is consciously regulated and the concomitant changes to psychophysiological processes as a result.

Finally, it is interesting to note the differences in the trajectory of responses between conditions during this study. Although the study aims primarily focussed on the reliability measures associated with novel fixed perceived effort cycling trials, some discussion can also be generated around the potential mechanisms that underpin the changes in performance, physiological, and psychological indices that were measured in this study. For instance, all performance (W), physiological (HR, $\dot{V}O_2$max, $\dot{V}_E$, BF, $[\mathrm{La}^-]_b$), and psychological (affect and self-efficacy) measures were significantly different between conditions at all TZ/time points and overall. In particular, responses for affect were negative throughout the entire fixed effort exercise in the $\mathrm{RPE}_{+15\%\mathrm{GET}}$ condition compared to a gradual decrease from positive to neutral in the $\mathrm{RPE}_{\mathrm{GET}}$ condition (Fig. 5).

Numerous studies have highlighted that affective valence may be a useful indicator of future exercise uptake and adherence (Brand and Ekkekakis 2021). To illustrate, studies have exhibited that when individuals completed exercise in line with a positive affect (Parfitt et al. 2012a), individuals were more likely to continue engaging in exercise compared to a fixed power output/velocity exercise. Interestingly, this was despite there being no actual differences in the actual physical intensity of the exercise between conditions (Parfitt et al. 2012a, b). Results from these studies demonstrate that a fixed effort exercise at lower RPE values (e.g., $\mathrm{RPE}_{\mathrm{GET}}$) is reliable and elicits more positive/neutral affective responses may provide a useful method for future studies focussing on exercise prescription and adherence.

## Conclusion

Overall, this study has demonstrated that recreationally active cyclists can execute reliable fixed effort exercise cycling trials which are aligned to physiological thresholds/domains. It appears that the harder the RPE intensity, the more reliably exercises can be conducted at both within and between individual levels. However, the underpinning factors for this remain unknown and yet to be fully explored. Some possible avenues for exploration may be the underlying decision-making processes that influence exercise behaviours during fixed effort cycling. Finally, this study also noted a significant difference in all performance, physiological, and psychological variables between conditions. Notably, affect was continually negative throughout the more intense $\mathrm{RPE}_{+15\%\mathrm{GET}}$ compared to the less-intense $\mathrm{RPE}_{\mathrm{GET}}$ condition. This may be of benefit to studies within the exercise rehabilitation domain as comparative findings suggest exercising at lower fixed perceived intensities that maintain positive affect may be better for exercise uptake and adherence. However, a continued exploration of this topic is required.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00421-022-05094-z.

**Author contribution** CAO, CLF, and ARM were responsible for the design and planning of the study. CAO was responsible for the data collection, analysis, and write-up of the manuscript. CAO, CLF, and ARM were responsible for the proof-reading and editing of the manuscript. All persons designated as authors qualify for authorship, and all those who qualify for authorship are listed. All authors have read and approved of the final manuscript submitted for publication. Authors are

accountable for the work and will ensure that all questions pertaining to the accuracy and authenticity of the work are appropriately investigated and resolved.

**Funding** No funds, grant, or otherwise were received to support this research.

**Data availability** Raw data are available upon request from the corresponding author.

## Declarations

**Conflict of interest** None.

**Ethical approval** The School of Sport and Exercise Sciences Research Ethics Advisory Group (Prop 31_2019_20) approved all procedures and protocols in accordance with the Declaration of Helsinki.

**Consent to participate** All participants provided written informed consent prior to any participation in the study.

## References

Amann M, Proctor LT, Sebranek JJ et al (2009) Opioid-mediated muscle afferents inhibit central motor drive and limit peripheral muscle fatigue development in humans. J Physiol 587:271–283. https://doi.org/10.1113/jphysiol.2008.163303

Amann M, Blain GM, Proctor LT et al (2010) Group III and IV muscle afferents contribute to ventilatory and cardiovascular response to rhythmic exercise in humans. J Appl Physiol 109:966–976. https://doi.org/10.1152/japplphysiol.00462.2010.-We

Astokorki AHY, Mauger AR (2017) Tolerance of exercise-induced pain at a fixed rating of perceived exertion predicts time trial cycling performance. Scand J Med Sci Sports 27:309–317. https://doi.org/10.1111/sms.12659

Bandura A (1997) Self-efficacy: the exercise of control. W.H. Freeman, New York

Beaver WL, Wasserman K, Whipp BJ, Whipp BJA (1986) A new method for detecting anaerobic threshold by gas exchange. J Appl Physiol 60:2020–2027

Bergstrom HC, Housh TJ, Cochrane KC et al (2015) Factors underlying the perception of effort during constant heart rate running above and below the critical heart rate. Eur J Appl Physiol 115:2231–2241. https://doi.org/10.1007/s00421-015-3204-y

Borg G (1982) Ratings of perceived exertion and heart rates during short-term cycle exercise and their use in a new cycling strength test. Int J Sports Med 3:153–158

Borg E, Borg G (2002) A comparison of AME and CR100 for scaling perceived exertion. Acta Psychol 109:157–175

Borg DN, Osborne JO, Stewart IB et al (2018) The reproducibility of 10 and 20 km time trial cycling performance in recreational cyclists, runners and team sport athletes. J Sci Med Sport 21:858–863. https://doi.org/10.1016/j.jsams.2018.01.004

Brand R, Ekkekakis P (2021) Exercise behavior change revisited: Affective-reflective theory. in: Essentials of exercise and sport psychology: an open access textbook. society for transparency, openness, and replication in kinesiology 62–92

Cochrane KC, Housh TJ, Bergstrom HC et al (2015a) Physiological responses during cycle ergometry at a constant perception of effort. Int J Sports Med 36:466–473. https://doi.org/10.1055/s-0034-1396826

Cochrane KC, Housh TJ, Hill EC et al (2015b) Physiological responses underlying the perception of effort during moderate and heavy intensity cycle ergometry. Sports 3:369–382. https://doi.org/10.3390/sports3040369

Cochrane-Snyman KC, Housh TJ, Smith CM et al (2016) Inter-individual variability in the patterns of responses for electromyography and mechanomyography during cycle ergometry using an RPE-clamp model. Eur J Appl Physiol 116:1639–1649. https://doi.org/10.1007/s00421-016-3394-y

Cochrane-Snyman KC, Housh TJ, Smith CM et al (2019) Treadmill running using an RPE-clamp model: mediators of perception and implications for exercise prescription. Eur J Appl Physiol 119:2083–2094. https://doi.org/10.1007/s00421-019-04197-4

de Morree HM, Klein C, Marcora SM (2012) Perception of effort reflects central motor command during movement execution. Psychophysiol 49:1242–1253. https://doi.org/10.1111/j.1469-8986.2012.01399.x

de Pauw K, Roelands B, Cheung SS et al (2013) Guidelines to classify subject groups in sport-science research. Int J Sports Physiol Perform 8:111–122

Dempsey JA, Amann M, Romer LM, Miller JD (2008) Respiratory system determinants of peripheral fatigue and endurance performance. Med Sci Sports Exerc 40:457–461. https://doi.org/10.1249/MSS.0b013e31815f8957

Elferink-Gemser MT, Hettinga FJ (2017) Pacing and self-regulation: Important skills for talent development in endurance sports. Int J Sports Physiol Perform 12:831–835. https://doi.org/10.1123/ijspp.2017-0080

Eston RG, Williams JG (1988) Reliability of ratings of perceived effort regulation of exercise intensity. Br J Sports Med 22:153–155. https://doi.org/10.1136/bjsm.22.4.153

Faulkner J, Parfitt G, Eston R (2007) Prediction of maximal oxygen uptake from the ratings of perceived exertion and heart rate during a perceptually-regulated sub-maximal exercise test in active and sedentary participants. Eur J Appl Physiol 101:397–407. https://doi.org/10.1007/s00421-007-0508-6

Gaesser GA, Poole DC (1996) The slow component of oxygen uptake kinetics. Exerc Sport Sci Rev 24:35–70

Gescheider GA (1997) Psychophysics: the fundamentals, 3rd edn. Erlbaum, London

Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. Annu Rev Psychol 62:451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigliotti F (2010) Mechanisms of dyspnea in healthy subjects. Multidiscip Respir Med 5:195–201

Halperin I, Emanuel A (2020) Rating of perceived effort: methodological concerns and future directions. Sports Med 50:679–687

Hardy CJ, Rejeski WJ (1989) Not what, but how one feels: the measurement of affect during exercise. J Sport Exerc Psychol 11:304–317

Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med 30:1–15

Jones AM, Burnley M, Black MI et al (2019) The maximal metabolic steady state: Redefining the 'gold standard.' Physiol Rep. https://doi.org/10.14814/phy2.14098

Lamb KL, Eston RG, Corns D (1999) Reliability of ratings of perceived exertion during progressive treadmill exercise. Br J Sports Med 33:336–339. https://doi.org/10.1136/bjsm.33.5.336

Laviolette L, Laveneziana P (2014) Dyspnoea: a multidimensional and multidisciplinary approach. Eur Respir J 43:1750–1762

Lim W, Lambrick D, Mauger AR et al (2016) The effect of trial familiarisation on the validity and reproducibility of a field-based self-paced VO2max test. Biol Sport 33:269–275. https://doi.org/10.5604/20831862.1208478

Marcora SM (2008) Do we really need a central governor to explain brain regulation of exercise performance? Eur J Appl Physiol 104:929–931

Marcora SM (2009) Perception of effort during exercise is independent of afferent feedback from skeletal muscles, heart, and lungs. J Appl Physiol 106:2060–2062

Marcora SM (2010) Counterpoint: afferent feedback from fatigued locomotor muscles is not an important determinant of endurance exercise performance. J Appl Physiol 108:454–456

Marcora SM, Bosio A, de Morree HM (2008) Locomotor muscle fatigue increases cardiorespiratory responses and reduces performance during intense cycling exercise independently from metabolic stress. Am J Physiol Regul Integr Comp Physiol 294:874–883. https://doi.org/10.1152/ajpregu.00678.2007.-Locomotor

Mauger AR (2014) Factors affecting the regulation of pacing: current perspectives. Open Access J Sports Med 5:209–214. https://doi.org/10.2147/oajsm.s38599

Mauger AR, Metcalfe AJ, Taylor L, Castle PC (2013) The efficacy of the self-paced VO2max test to measure maximal oxygen uptake in treadmill running. Appl Physiol Nutri Metabol 38:1211–1216. https://doi.org/10.1139/apnm-2012-0384

Meeusen R, Roelands B (2018) Fatigue: is it all neurochemistry. Eur J Sport Sci 18:37–46. https://doi.org/10.1080/17461391.2017.1296890

Mauger AR, Huntley T, Fletcher IM (2014) RPE-derived work rates can be accurately produced without external feedback or reference to the RPE scale. Percept Mot Skills 118:507–521. https://doi.org/10.2466/27.06.PMS.118k19w7

O'Donnell DE, Ora J, Webb KA, Laveneziana P, Jensen D (2009) Mechanisms of activity-related dyspnea in pulmonary diseases. Respir Phsyiol Neurobiol 167:116–132. https://doi.org/10.1016/j.resp.2009.01.010

O'Donnell DE, Milne KM, James MD, de Torres JP, Neder JA (2020) Dyspnea in COPD: new mechanistic insights and management implications. Adv Ther 37:41–60. https://doi.org/10.1007/s12325-019-01128-9

O'Grady C, Passfield L, Hopker JG (2021) Variability in submaximal self-paced exercise bouts of different intensity and duration. Int J Sports Physiol Perform 16:1824–1833. https://doi.org/10.1123/ijspp.2020-0785

Okuno NM, Soares-Caldeira LF, Milanez VF, Perandini LAB (2015) Predicting time to exhaustion during high-intensity exercise using rating of perceived exertion. Sci Sports 30:e155–e161. https://doi.org/10.1016/j.scispo.2015.05.003

Pageaux B (2014) The psychobiological model of endurance performance: an effort-based decision-making theory to explain self-paced endurance performance. Sports Med 44:1319–1320

Pageaux B (2016) Perception of effort in exercise science: definition, measurement and perspectives. Eur J Sport Sci 16:885–894

Parfitt G, Alrumh A, Rowlands AV (2012a) Affect-regulated exercise intensity: Does training at an intensity that feels "good" improve physical health? J Sci Med Sport 15:548–553. https://doi.org/10.1016/j.jsams.2012.01.005

Parfitt G, Evans H, Eston R (2012b) Perceptually regulated training at RPE13 is pleasant and improves physical health. Med Sci Sports Exerc 44:1613–1618. https://doi.org/10.1249/MSS.0b013e31824d266e

Paton C, Hopkins W (2001) Tests of cycling performance. Sports Med 31:489–496

Renfree A, Martin L, Micklewright D, St Clair Gibson A (2014) Application of decision-making theory to the regulation of muscular work rate during self-paced competitive endurance activity. Sports Med 44:147–158

Seiler KS, Kjerland GØ (2006) Quantifying training intensity distribution in elite endurance athletes: Is there evidence for an "optimal" distribution? Scand J Med Sci Sports 16:49–56. https://doi.org/10.1111/j.1600-0838.2004.00418.x

Tate RF, Klett GW (1959) Optimal confidence intervals for the variance of a normal distribution. J Am Statist Assoc 54:674–682

Tucker R (2009) The anticipatory regulation of performance: the physiological basis for pacing strategies and the development of a perception-based model for exercise performance. Br J Sports Med 43:392–400

Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 19:231–240

Whitehead AE, Jones HS, Williams EL et al (2018) Investigating the relationship between cognitions, pacing strategies and performance in 16.1 km cycling time trials using a think aloud protocol. Psychol Sport Exerc 34:95–109. https://doi.org/10.1016/j.psychsport.2017.10.001

Yoon BK, Kravitz L, Robergs R (2007) VO2max, protocol duration, and the VO2 plateau. Med Sci Sports Exerc 39:1186–1192. https://doi.org/10.1249/mss.0b13e318054e304