TITLE: Evaluation of a graded exercise test to determine peak fat oxidation in individuals with low cardiorespiratory fitness.

AUTHORS: Oliver J. Chrzanowski-Smith¹, Robert M. Edinburgh¹, James A. Betts PhD¹, Keith A. Stokes PhD¹, Javier T. Gonzalez PhD¹.

CORRESPONDENCE: Dr Javier T. Gonzalez
Department for Health
University of Bath
Bath, BA2 7AY
United Kingdom
Tel: +44-1225-385-518
Email: J.T.Gonzalez@bath.ac.uk

AFFILIATIONS: ¹ Department for Health, University of Bath, Bath, BA2 7AY, UK
O.J.Chrzanowski-Smith@bath.ac.uk
R.M.Edinburgh@bath.ac.uk
J.Betts@bath.ac.uk
K.Stokes@bath.ac.uk
J.T.Gonzalez@bath.ac.uk

RUNNING HEAD: Estimating maximal capacity for fat oxidation
ABSTRACT

The maximal capacity to utilise fat (peak fat oxidation [PFO]) may have implications for health and ultra-endurance performance, and is commonly determined by incremental exercise tests employing 3-minute stages. However, 3-minute stages may be insufficient to attain steady-state gas kinetics, compromising test validity. We assessed whether 4-minute stages produce steady-state gas exchange and reliable PFO in adults with \( \dot{V}O_2 \text{peak} < 40 \text{ ml·kg}^{-1}·\text{min}^{-1} \). Fifteen participants (9F) completed a graded test to determine PFO and the intensity this occurred at (FAT\text{MAX}). Three short continuous exercise sessions (SCE) were then completed in a randomised order, involving completion of the graded test to the stage: 1) preceding; 2) equal to (SCE\text{equal}); or 3) after, PFO was attained, where participants then continued to cycle for 10-minutes at that respective intensity. Expired gases were sampled at minutes 3-4, 5-6, 7-8 and 9-10. Individual data showed steady-state gas exchange was achieved within 4-minutes during SCE\text{equal}. Mean fat oxidation rates were not different across time within SCE\text{equal}, nor compared to the graded test at FAT\text{MAX} (both \( p > 0.05 \)). However, the graded test displayed poor surrogate validity (SCE\text{equal} minute 3-4 vs 5-6, 7-8 and 9-10) and day-to-day reliability (minute 3-4 SCE\text{equal} vs graded test) to determine PFO, as evident by correlations (range: 0.47 – 0.83), and typical errors and 95% limits of agreement (range: 0.03 – 0.05 and \( \pm 0.09 – 0.15 \text{ g·min}^{-1} \), respectively). In conclusion, intra-individual variation in PFO is substantial despite four-minute stages establishing steady-state gas exchange in individuals with low fitness. Individual assessment of PFO may require multiple assessments.

KEY WORDS: Peak Fat Oxidation, FAT\text{MAX}, Exercise Metabolism, Gas Exchange Kinetics, Low Cardiorespiratory Fitness
INTRODUCTION

There is emerging interest into the importance of peak fat oxidation during exercise (PFO; a whole-body measure of the ‘maximal’ capacity to utilise fat) and identifying the exercise intensity that elicits PFO ($\text{FAT}_{\text{MAX}}$) (Dandanell et al. 2017a; Isacco et al. 2014; Mogensen et al. 2009; Robinson et al. 2015; Romain et al. 2012; Rosenkilde et al. 2010). For example, PFO may represent a relatively non-invasive and easily measureable determinant of endurance exercise performance (Frandsen et al. 2017), metabolic health (Rosenkilde et al. 2010) and long-term weight loss and/or maintenance (Dandanell et al. 2017a). However, the valid determination of PFO is paramount to allow an appropriate interpretation of the outcomes of such studies, and to infer any role in sports performance or metabolic disease.

Early studies characterised fat oxidation during exercise over few stages, sometimes only two (Friedlander et al. 1998; Sidossis et al. 1997; Steffan et al. 1999; Thompson et al. 1998), three (Bircher and Knechtle 2004; Romijn et al. 2000; Romijn et al. 1993; van Loon et al. 2001) or four (Bergman and Brooks 1999) progressive exercise intensities, limiting the resolution to detect PFO. Attempting to address this poor resolution, Achten and colleagues (2002) developed a practical, one-off graded protocol ($\text{GE}_{35\,3}$) in 2001. The test comprised three-minute stages commencing at 90 Watts (W) and progressively increasing by 35 W increments until volitional exhaustion. The protocol is widely advocated as a reliable and valid method to determine $\text{FAT}_{\text{MAX}}$ in trained cyclists (Achten et al. 2002).
Slight variations of the GE\textsubscript{35/3} protocol have since been widely employed to determine PFO and \(\text{FAT}_{\text{MAX}}\) in various populations. However, even within each standardised protocol, substantial inter-individual variability is apparent (0.18–1.01 g\(\cdot\)min\(^{-1}\) for PFO and 25-77 % \(\dot{\text{VO}}_{2}\text{max}\) for \(\text{FAT}_{\text{MAX}}\)) (Achten et al. 2002; Venables et al. 2005). Variations in the graded protocol typically involve alterations in work intensities, with early stage durations (i.e. three-minutes) held constant (Ara et al. 2011; Astorino et al. 2017; Blaize et al. 2014; Dandanell et al. 2017\(^{a}\); Dandanell et al. 2017\(^{b}\); Larsen et al. 2009; Nordby et al. 2015; Robinson et al. 2015; Rosenkilde et al. 2010; Venables et al. 2005). However, the delayed attainment of steady-state gaseous exchange (oxygen consumption and carbon dioxide production) in individuals with low vs high cardiorespiratory fitness (Poole and Jones 2012) raises the question as to whether three-minute stages are sufficient to produce valid and reliable estimates of PFO, or whether longer stages are required in populations with lower cardiorespiratory fitness (Bordenave et al. 2007; Brun et al. 2011).

To date, only one study has investigated the day-to-day reliability and surrogate validity of a graded test to exhaustion in determining PFO and \(\text{FAT}_{\text{MAX}}\) in individuals with low cardiorespiratory fitness (Dandanell et al. 2017\(^{b}\)). In sixteen obese adults (8 females), the graded protocol produced PFO values that were not in close agreement (e.g. 95 % limits of agreement ± 0.17 g\(\cdot\)min\(^{-1}\)) with estimates derived from three short continuous exercise bouts at three exercise intensities that were not workload matched \textit{per se} to the graded test (10-minutes at 35 %, 50 % and 65 % \(\dot{\text{VO}}_{2}\text{max}\)). The substantial intra-individual (i.e. day-to-day) variation in PFO between the protocols indicates further optimisation of
graded exercise tests is required to improve their validity (Dandanell et al. 2017b). Importantly, neither Dandanell and co-workers (2017b) nor others (Achten et al. 2002; Bordenave et al. 2007) have systematically evaluated whether steady-state gaseous exchange is attained with short-stage graded exercise protocols. Thus, the substantial intra-individual variability (Dandanell et al. 2017b) and reported overestimation of fat oxidation in graded protocols utilising three- vs six-minute stages (Bordenave et al. 2007), may be due to non-steady-state gas kinetics.

Therefore, the main aims of this study were to develop a graded exercise protocol employing four-minute initial stages to: (i) observe whether steady-state gaseous exchange is attained within four-minutes; and to assess the graded test’s (ii) surrogate validity (by comparing four-minute stages to six-, eight- and ten-minute stages); and (iii) day-to-day reliability, to determine PFO in individuals with low cardiorespiratory fitness. Four-minute stages were selected to ‘optimise’ the graded test (i.e. increase the likelihood that steady-state gas exchange would be achieved in the initial stages, in an attempt to improve the validity and reliability of determining PFO). The hypotheses were (i) that steady-state gaseous exchange would be achieved within four-minutes and (ii) the graded protocol would provide valid and reliable estimates of PFO.

MATERIALS AND METHODS

Participants

Following written informed consent, sixteen adults (10 females) with low cardiorespiratory fitness ($\dot{V}O_2$peak <40 mL·kg$^{-1}$·min$^{-1}$) were recruited and completed the
study protocol between March and July 2017 (Supplementary Material S1a). Exclusion criteria included; age < 18 or > 65 yrs; current or history of cardio-pulmonary, metabolic or musculoskeletal disease; a body mass index < 18.9 or > 35 kg·m⁻²; breastfeeding or was / potentially pregnant; or any conditions or concurrent behaviour (including medication) that may pose undue personal risk to the participant or introduce bias to the study. Menstrual cycle and use of oral contraceptives was recorded but not controlled. One participant was excluded from the analyses due to hyperventilation during expired gas collection in all exercise trials (RER > 1.00), leaving a total sample of \( n = 15 \) (9 females). Participant characteristics are presented in Table 1.

**Experimental design**

A cross-over design was adopted where participants completed four exercise trials (one graded exercise test and three short continuous exercise sessions) in a laboratory at the University of Bath, UK. Trials were completed at a similar time (± 1 h within participant) in the morning (0600 – 1100 h) after an overnight fast (11 ± 1 h) and all within 21 days. Over the 24-h preceding each trial, participants were asked to: a) abstain from alcohol and strenuous physical activity; and b) replicate their dietary intake and physical activity. Additionally, participants maintained their habitual lifestyle throughout their involvement in the study. These criteria were confirmed by verbal questioning. All trials (within-subject) were performed under similar laboratory conditions (ambient temperature, humidity and barometric pressure; all \( p > 0.05 \); data not shown) where *ad-libitum* water intake and use of fans was permitted. The study was performed in accordance with the

---

1 Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/apnm-2018-0098.
1964 Declaration of Helsinki and was approved by the Research Ethics Approval Committee for Health at the University of Bath (REF: EP 16/17 141).

**Graded exercise test**

Participants first completed an incremental graded cycling test on a mechanically-braked cycle ergometer (Monark Peak Bike Ergomedic 894E, Varberg, Sweden). Upon arrival to the laboratory, anthropometric measurements were taken before placement of a heart rate monitor and self-selected ergometer set-up (seat and handlebar height were recorded to standardise for subsequent sessions). Participants then completed the graded test which comprised: a) four-minute stages for the first seven stages and b) two-minute stages from the eighth stage onwards. Initial power output was set at ~40 W and increased by ~25 W over the next six stages and by ~50 W from stage seven onwards. Participants exercised up until volitional exhaustion. The two-phase test was developed upon unpublished observations of pilot testing and adopted to 1) increase the sensitivity of the graded test for determining PFO in the early stages and 2) to facilitate the achievement of \( \dot{V}O_2 \text{peak} \) and attainment of volitional exhaustion quickly to reduce the confounding effects of fatigue. Expired gases were collected in the final minute of the first seven stages and upon the participant’s signal of one-minute remaining before volitional exhaustion. The mouthpiece was removed in between stages and provided one-minute before each gas collection to ensure adequate ‘flushing’ of ambient air. Expired gas samples were analysed immediately after each collection. Heart rate (HR) via telemetry (Polar RS400 Heart Rate Monitor, Kempele, Finland) and ratings of perceived exertion (RPE) (Borg 1973) were recorded during each expired gas sample. The graded test was used to determine:
a) Peak fat oxidation [g·min\(^{-1}\); the stage with the highest recorded fat oxidation value (measured values approach)];

b) \(\text{FAT}_{\text{MAX}}\) (expressed as absolute \(\dot{V}O_2\); W and HR at stage eliciting PFO and as a % of \(\dot{V}O_2\)\(_{\text{peak}},\ PPO\) and \(HR_{\text{MAX}}\));

c) Peak power output (W; power output of the last completed stage, plus the fraction of time in the final non-completed stage, multiplied by the Watt increment of that stage);

d) An estimate of peak oxygen consumption [\(\dot{V}O_2\)\(_{\text{peak}}\); mL·kg\(^{-1}\)·min\(^{-1}\); the upper functional limit (i.e. highest \(\dot{V}O_2\) value) recorded in the single graded exercise test (Green and Askew, 2018)].

**Short continuous exercise sessions (SCE)**

Participants then completed three short continuous exercise sessions (SCE) in a randomised order on separate days (randomisation list created by O.C-S on [www.randomizer.org/](http://www.randomizer.org/) Date: 24/02/2017). Upon arrival, anthropometric measurements were taken before participants completed the graded test protocol up until either the stage: 1) preceding (SCE\(_{\text{pre}}\)); 2) equal to (SCE\(_{\text{equal}}\)); or 3) after (SCE\(_{\text{post}}\)), PFO was previously determined, where participants then continued to cycle for a further ten-minutes at that respective intensity. If PFO was measured in the first stage of the graded test, SCE\(_{\text{pre}}\) power output was lowered to \(\sim 30\) W (\(n = 13\)), where participants exercised for ten-minutes at this absolute intensity. One-minute expired gas samples were collected during the 10-
minute continuous period at minutes 3-4, 5-6, 7-8 and 9-10. Participants were provided the mouthpiece one-minute before the first collection and this was then kept in for remainder of the session. Heart rate and RPE were collected during each expired gas sample. Expired gas samples were analysed immediately at the end of the session in order of collection.

**Anthropometric measurements**

All anthropometric measurements were performed according to International Standards for Anthropometric Assessment 2007 (Marfell-Jones et al. 2007). Body stature was measured to the nearest 0.1 cm using a wall-mounted stadiometer (Holtain Ltd, Pembrokeshire, UK). Body mass (to nearest 0.1 kg) and body fat % (bioelectrical impedance analysis to nearest 0.1 %) were measured using electronic weighing scales (BC-543 Monitor, Tanita, Tokyo, Japan). Waist and hip circumference (cm) were additionally measured in the SCE sessions to the nearest 0.1 cm using a non-elastic measuring tape (SECA 201, Hamburg, Germany).

**Metabolic measurements**

Expired gas samples were collected into 100-150 L Douglas bags (Cranlea and Hans Rudolph, Birmingham, UK) via a mouthpiece connected to a two-way, T shaped non-rebreathing valve (Model 2700, Hans Rudolph Inc, Kansas City, USA) and falconia tubing (Hans Rudolph Inc, Kansas City, USA). Concentrations of O₂ and CO₂ were measured in a known volume of each sample via paramagnetic and infrared transducers, respectively (Mini MP 5200, Servomex Group Ltd., Crowborough, East Sussex, UK) and until values were stable. The sensor was calibrated to a two-point low and high calibration
of known gas concentrations (Low: 99.998 % Nitrogen, 0 % O₂ and CO₂; High: Balance nitrogen mix, 20.06 % O₂, 8.11 % CO₂) (BOC Industrial Gases, Linde AG, Munich, Germany). Concurrent measurements of inspired air composition were made during collection of each expired gas sample to adjust for changes in ambient O₂ and CO₂ concentrations (Betts and Thompson 2012). Indirect calorimetry was used to determine: \( \dot{V}O_2 \) (L·min⁻¹); \( \dot{V}CO_2 \) (L·min⁻¹); \( \dot{V}E \) (L·min⁻¹); respiratory exchange ratio (RER); and rate of fat oxidation [g·min⁻¹; estimated by stoichiometric equations (Frayn 1983) assuming urinary nitrogen excretion was negligible].

**Statistical analysis**

Normality and equal variance were explored via the Shapiro-Wilk test and visual inspection of histograms, Q-Q plots, P-P plots and scatterplots on raw data and residuals of comparisons. Assumption of normality and equal variance was assumed unless elsewise stated and sensitivity analyses were performed where applicable. The attainment of steady-state gas exchange was visually inspected and confirmed as described in Supplementary Material S1b². A range of *a priori* statistical analysis tests were performed to evaluate the graded test in determining primarily PFO as advocated (Atkinson and Nevill 1998). To test for systematic differences in metabolic variables within each SCE session and *versus* the graded test at FAT\(_{\text{MAX}}\), one-way repeated measures ANOVAs [time of expired gas sample as within-subject factor; irrespective of normality due to robustness of ANOVA (Maxwell 1990)] and paired sample *t*-tests or Wilcoxon Signed-Rank tests were performed, respectively. *Post-hoc* Bonferroni corrected *t*-tests

² Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/apnm-2018-0098.
were performed where significant main effects of time were detected. The surrogate validity of four-minute stages (i.e. minute 3-4) to estimate fat oxidation was further explored within SCE\textsubscript{pre} and SCE\textsubscript{equal} using Pearson correlation coefficient ($r$), Intra-Class Correlation [ICC; 3,1 absolute agreement, single measures (Shrout and Fleiss 1979)], Within-subject Coefficient of Variation (CV; root mean square method), Typical Error [TE; SD of difference between scores / $\sqrt{2}$ (Hopkins 2015)] and Bland-Altman plot with mean difference (bias) and 95 % Limits of Agreement (LoA) (Bland and Altman 1986). These tests were not performed at the stage after FAT\textsubscript{MAX} due to systematic differences in fat oxidation at this exercise intensity (data reported below). The day-to-day reliability of peak fat oxidation was also further explored by comparing fat oxidation estimates in the graded test at FAT\textsubscript{MAX} (minute 3-4) vs SCE\textsubscript{equal} minute 3-4 through the aforementioned intra-individual variation statistical tests. Mean difference in fat oxidation was calculated within SCE by subtraction of the minute 3-4 value from the value obtained at each ‘respective time point’, while comparisons between SCE\textsubscript{equal} and the graded test were calculated through subtraction of the graded test value from the value obtained at SCE\textsubscript{equal} minute 3-4. Individual data are plotted on all graphs. A maximum sample size of $n = 15$ was available for statistical analyses (see Supplementary Material S1c for description on sample sizes of analyses). Descriptive and statistical analyses were run on Microsoft Excel (2013) and IBM SPSS statistics version 22 for windows (IBM, New York, USA) and graphs were created on Graph Pad Prism 7 software (La Jolla, Calif., USA). Data are presented as means ± 95 % confidence intervals unless otherwise stated and statistical significance was accepted at $p \leq 0.05$. 

---

3 Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/apnm-2018-0098.
RESULTS

Gas exchange kinetics and surrogate validity of fat oxidation

There were no systematic differences detected across the four time points during SCE_{pre}, SCE_{equal} or SCE_{post}, for \( \dot{V}O_2 \) (Fig. 1a, 1c, 1g; \( F (3, 42) = 0.489, p = 0.692; F (3, 36) = 0.405, p = 0.750; F (3, 42) = 1.905, p = 0.143 \) respectively), \( \dot{V}CO_2 \) (Fig. 1b, 1d, 1f; \( F (3, 42) = 0.393, p = 0.759; F (3, 36) = 0.910, p = 0.446; F (3, 42) = 0.456, p = 0.714 \) respectively) or \( \dot{V}E \) (Table 2). Visual inspection shows no clear within-subject fluctuation in \( \dot{V}O_2 \) and \( \dot{V}CO_2 \) across time points during SCE_{pre} and SCE_{equal} in all but one participant, indicating steady-state gas kinetics were typically achieved by minute 3-4 (Fig. 1a – 1f).

There were no systematic differences detected across the four time points during SCE_{pre} or SCE_{equal} for fat oxidation (Fig. 2a – 2b; \( F (3, 42) = 0.191, p = 0.902 \) and \( F (3, 36) = 2.782, p = 0.055 \), respectively), or RER (Table 2). However, systematic differences were found for fat oxidation (Fig. 2c; \( F (3, 42) = 8.069, p < 0.001 \)) and RER (Table 2) during SCE_{post}. As depicted in Fig. 2c, fat oxidation was lower during minute 3-4 and 5-6 vs minute 9-10 (\( p = 0.010 \) and 0.002, respectively), whilst RER was higher during minute 3-4 and 5-6 vs minute 9-10 (Table 2; \( p = 0.016 \) and 0.004, respectively).

Pearson correlations and ICCs between PFO determined at minute 3-4 vs ‘longer stage durations’ in SCE_{equal} were high to excellent (ranged between 0.61 to 0.83), but coincided with large 95 % CIs (Table 3). There were large within-subject CVs and reasonably small typical errors [range between 19 to 25 % and 0.03 to 0.04 g.min\(^{-1}\),
respectively (Table 3) and the Bland-Altman analysis of agreements revealed a small bias (range of -0.01 to -0.04 g·min⁻¹) and moderately large 95% LoAs [range of ± 0.09 to 0.12 g·min⁻¹ (Fig. 3a – 3c)], in the determination of fat oxidation at minute 3-4 vs longer stages at SCE_{equal}. Reliability and intra-individual variation of fat oxidation at SCE_{pre} is presented in Table 3 and Supplementary Figure S1⁴.

**Day-to-day reliability**

There were no systematic differences in \( \dot{V}O_2 \) (Fig. 1g; \( T (11) = -0.167, p = 0.870 \)), \( \dot{V}CO_2 \) (Fig. 1h; \( Z = -0.78, p = 0.937 \)) or \( \dot{V}E \) (Supplementary Table S1⁴) between the graded protocol and SCE minute 3-4 at the stage equal to FAT\(_{MAX}\). Similarly, visual inspection shows little within-subject fluctuation in \( \dot{V}O_2 \) and \( \dot{V}CO_2 \) in all but one participant (Fig. 1g and 1h).

No systematic differences were found in fat oxidation (Fig. 2d; \( T (11) = 0.954, p = 0.361 \)), or RER (Supplementary Table S1⁴) between the graded test and SCE minute 3-4 at the stage equal to FAT\(_{MAX}\).

The Pearson correlation and ICC of the graded test vs SCE minute 3-4 at the stage equal to FAT\(_{MAX}\) to determine PFO were fair with large 95% CIs apparent [0.48 (95% CI: -0.14 – 0.82) and 0.47 (95% CI: -0.10 – 0.81], respectively]. There was a large within-subject CV [23% (95% CI: 6 – 32%)] and a moderately small typical error [0.05 g·min⁻¹ (95% CI: 0.04 – 0.09 g·min⁻¹)], whilst Bland-Altman analysis of agreement revealed a

---

⁴ Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/apnm-2018-0098.
small mean bias (-0.02 g·min\(^{-1}\)) and a considerably large 95 % LoA [± 0.15 g·min\(^{-1}\) (Fig. 3d)], in the determination of PFO between the graded test and SCE minute 3-4 at FAT\(_{\text{MAX}}\).

Sensitivity analysis performed on an outlier revealed no meaningful effect nor altered interpretation of SCE\(_{\text{equal}}\) and graded test minute 3-4 vs SCE minute 3-4 at FAT\(_{\text{MAX}}\) analyses, except for Pearson correlation and ICC on the latter analysis (\(r = 0.48\) vs 0.15 and ICC = 0.47 vs 0.14, respectively). Furthermore, outcomes of the SCE analyses were independent of the statistical analysis employed (Friedman vs repeated measures ANOVA; data not shown).

DISCUSSION

The capacity to oxidise fat may represent a determinant of metabolic health and ultra-endurance performance. Acknowledging the recent call to ‘optimise’ graded exercise tests for accurate determination of PFO and FAT\(_{\text{MAX}}\) (Dandanell et al. 2017b), this is the first study to examine the surrogate validity and day-to-day reliability of a graded test employing four-minute stages to determine PFO in individuals with low cardiorespiratory fitness. The novel data provide evidence that steady-state gas exchange is attained within four-minutes at intensities equal to or below FAT\(_{\text{MAX}}\). However, large intra-individual variability was apparent when determining PFO with four-minute stages both when compared to longer stage durations and across two separate days. This indicates that graded tests employing four-minute stages have poor surrogate validity and day-to-day reliability, respectively. Consequently, multiple testing is likely required to confidently assess PFO in individuals.
Incremental (graded) exercise tests are frequently used in a wide range of research studies to estimate PFO. However, there is currently no clear consensus (or justification) as to the most appropriate stage duration of these tests. We showed that steady-state gas exchange was achieved within four-minutes as there were no systematic differences alongside only minor fluctuations in the individual time course of gas kinetic variables during SCE\textsubscript{equal} (Fig. 1). The attainment of steady-state gas exchange is a key requirement for the use of indirect calorimetry to accurately assess substrate metabolism (Macfarlane 2017; Shephard and Aoyagi 2012). This suggests that despite previous uncertainty over utilising stage durations under six-minutes (Bordenave et al. 2007), graded exercise tests with as short as four-minute stages can be legitimately recommended for use in healthy populations with low cardiorespiratory fitness.

The current study is the first to comprehensively explore the surrogate validity of graded exercise tests employing various stage lengths (4-, 6-, 8- and 10-minute stages) in the determination of PFO. We observed no systematic differences in fat oxidation rates estimated across the various stage lengths at FAT\textsubscript{MAX} (i.e. during SCE\textsubscript{equal}; Fig. 2b). This is somewhat in contrast to Bordenave et al. (2007) who reported three-minute stages ‘over-estimated’ fat oxidation rates by \textasciitilde0.02 g·min\textsuperscript{−1} when compared to six-minute stages. However, in this previous study differences in fat oxidation rates were pooled across 4 x 6-minute submaximal exercise intensities and potential carry-over effects (i.e. previous stages influencing data from subsequent stages) were not controlled for, potentially limiting the interpretation of these results. The potential carry over effect of previous stages is also pertinent to previous reliability and validity investigations of graded tests.
(Achten et al. 2002; Dandanell et al. 2017b). Whilst numerically we report a slightly higher ‘over-estimation’ of PFO with four-minute vs six-minute stages (-0.03 g·min⁻¹), better agreement in PFO estimates were evident when stage durations of eight- or ten-minutes were compared to four-minute stages [-0.02 and -0.01 g·min⁻¹, respectively (Fig. 2b)]. Collectively, this suggests that stage lengths between four- to ten-minutes provide equally useful group level estimates of PFO.

Despite no evident systematic differences, we found considerable intra-individual variance in the determination of PFO when four-minute stages were compared to longer stage durations (i.e. graded tests with four-minute stages have poor surrogate validity). For example, the wide 95 % CIs surrounding the Pearson and Intra-Class correlation coefficients indicates large intra-individual variability exists when assessing PFO. This is further supported by the large within-subject CVs (Table 3) and 95 % LoAs. The latter suggest that estimates of PFO made with four-minute stages may vary by ± 0.09 to 0.12 g·min⁻¹ when compared to longer stages (Fig. 3a – 3c). This corresponds to the finding of Bordenave and colleagues (2007) who report a 95 % LoA of ± 0.06 to 0.10 g·min⁻¹ when three-minute stages were compared to six-minute stages. Notwithstanding that group level estimates of PFO are similar, researchers should acknowledge the large intra-individual variation in the determination of PFO with different stage durations. Thus, as graded tests currently stand, repeated testing with different stage lengths may be required to more confidently assess PFO.

The present study also assessed the day-to-day reliability of the graded test to determine PFO. We found no systematic difference at the group level when estimating
PFO with four-minute stages across two days at the same absolute power output (Fig. 2d). Alternatively, Dandanell and co-workers (2017b) reported a statistically significant systematic difference in the determination of PFO between a graded test employing three-minute stages and three ten-minute bouts at non-equivalent exercise intensities in sixteen adults (8 females) with obesity and low cardiorespiratory fitness (0.04 g·min⁻¹, p = 0.03). This slight contrast in finding may derive from methodological discrepancies such as the use of three- vs four-minute stages, as it is still unknown whether three-minute stages are sufficient to establish steady-state gas exchange in the population we studied. Unfortunately, the present study was unable to explore the comparison between three- vs four-minute stages due to the Douglas bag method preventing immediate collection of back-to-back samples. Notably, graded tests employing three-minute stages have reported larger group mean systematic differences in \( \dot{V}O_2 \), \( \dot{V}CO_2 \) and \( \dot{V}E \) when exercising at \( \text{FAT}_{\text{MAX}} \) on separate days (Achten et al. 2002; Dandanell et al. 2017b). Collectively, this suggests graded tests with three-minute stages may systematically produce greater variability in gas exchange kinetics and PFO estimates, advocating the use of four-minute stages in future studies.

Despite this, we observed large intra-individual variability in the day-to-day reliability of PFO (Fig. 3d and Table 3). Similarly large limits of agreement have previously been reported (\( \pm 0.13 \) g·min⁻¹) alongside a lower within-subject CV [11 % (95 % CI: 8 – 15 %)] in individuals with low cardiorespiratory fitness (Dandanell et al. 2017b). Moreover, corresponding high CVs (range of 23 to 26 %) and even greater 95 % LoAs (range of \( \pm 0.24 \) to 0.26 g·min⁻¹) have also been reported in PFO despite small systematic differences (range of 0.00 - 0.01 g·min⁻¹) between two identical sub-maximal cycling protocols utilising
five-minute stages in moderately trained males (Croci et al. 2014). Our results support and build upon this previous work by using the still considered gold standard method for expired gas sampling and analysis [i.e. the Douglas bag technique (Shephard 2017)] over the more frequently used breath-by-breath metabolic cart systems. We also adjusted for inspired air composition in substrate utilisation calculations (Betts and Thompson 2012) which is important due to (the largely overlooked) changes in gas composition of both atmospheric air and ambient laboratory air. Taken collectively, these data all suggest that there is poor day-to-day reliability at the individual level in the determination of PFO across a range of graded test protocols.

Numerous factors may contribute to the large intra-individual variation observed in the determination of PFO with graded exercise tests. While exercise intensity and duration are primary modulators of fat oxidation during exercise (Romijn et al. 1993; van Loon et al. 2001), nutrient status (Gonzalez et al. 2013), dietary macronutrient composition (Fletcher et al. 2017; Støa et al. 2016) and both acute and chronic physical activity levels (Venables et al. 2005) are also contributing factors. Standardising diet and physical activity for at least 48-h prior to testing is likely important for more reliable estimates of fat oxidation during exercise (Støa et al. 2016). However, prior research (including the present study) has rarely objectively ensured strict pre-trial standardisation of either diet or physical activity for ≥ 48-h (Astorino and Schubert 2018). While this may partially account for the day-to-day variability between protocols, similar intra-individual variability was apparent across different stage durations [i.e. within $\text{SCE}_{\text{pre}}$ and $\text{SCE}_{\text{equal}}$ (Fig. 3)] with metabolic estimates exposed to the same ‘standardisation’. Nevertheless, although
challenging to implement, greater standardisation may be integral to obtaining reliable estimates of PFO and should be further explored.

Like many previous studies, the current graded exercise protocols were based on absolute watt increments. Consequently, each stage would represent a different percentage of exercise capacity between participants (e.g. % \( \dot{V}O_2 \text{peak} \)) and thus elicit potentially divergent physiological responses. This may not only contribute to the variability in fat oxidation estimates, particularly as high intra-individual day-to-day variability in the metabolic milieu has previously been reported at FAT\text{MAX} (Dandanell et al. 2017b), but may account for the systematic unstable fat oxidation rates prevalent during SCE\text{post} (Fig. 2c). At intensities equal to or above lactate threshold, attainment of steady-state gas exchange may be delayed (for \(~\)15 minutes or more) due to contribution of the \( \dot{V}O_2 \) slow component and \( \dot{V}CO_2 \) from non-aerobic sources (e.g. buffering of \([H^+]\) ions by bicarbonate) (Rossiter 2011). It is plausible that the exercise intensity elicited by the stage after FAT\text{MAX} may have coincided or been above the lactate threshold for some individuals (Achten and Jeukendrup 2004; Michallet et al. 2008). An alternative approach would be to characterise the relative exercise intensities of each stage using the lactate threshold. However, this would require an additional testing day (and standardisation procedure), losing the one-off, time-effective practicality element of such graded tests. Our data also suggest caution should be applied when using more complex analytical approaches (i.e. curve modelling from multiple exercise intensities), which utilise estimates above FAT\text{MAX}, to characterise fat oxidation kinetics and PFO (Cheneviere et al. 2009; Croci et al. 2014).
In this work it should be acknowledged that cardiorespiratory fitness was estimated via $\dot{V}O_2$peak which while is an acceptable marker (Green and Askew, 2018) may underestimate participants ‘true’ maximal cardiorespiratory fitness (Poole and Jones, 2017). Additionally, PFO was recorded in the first stage of the graded test for thirteen participants. Consequently, PFO could have occurred at a lower intensity than the initial 40 W stage in the graded test with $SCE_{equa}$ also potentially not representing ‘FATMAX’. However, at the group level fat oxidation rates were lower in $SCE_{pre}$ than $SCE_{equal}$ (Fig. 2). Additionally, the considerable intra-individual day-to-day variation in PFO and FATMAX reported here and previously (Croci et al. 2014; Dandanell et al. 2017b), suggests large overlap in fat oxidation rates may exist between coinciding exercise intensities (Achten et al. 2004). The overlap in fat oxidation rates evident between $SCE_{pre}$ and $SCE_{equal}$ further suggest PFO could have occurred below 30 W. Acknowledging this shortcoming, practitioners may wish to commence graded tests at lower workloads than 30 / 40 W to increase the resolution of capturing PFO and FATMAX in individuals with low cardiorespiratory fitness.

Collectively, the above demonstrates that future work should seek to further optimise graded tests to more accurately capture and determine PFO. The practical relevance of this is reiterated by exercise training induced increases in PFO [e.g. + 0.12 to 0.22 g·min$^{-1}$ (Mogensen et al. 2009; Nordby et al. 2015)] that are within the realm of reported intra-individual variability. Consequently, to confidently ascribe any effects to an intervention, observed increases in PFO may need to exceed the reported intra-individual variability in PFO from the most similar protocol employed (Croci et al. 2014; Dandanell et al. 2017b; De Souza Silveira et al. 2016; Marzouki et al. 2014). That noted, only once
PFO has been quantified reliably can the question of what may constitute a meaningful change or difference in PFO for health and performance be addressed.

**Conclusion**

In summary, the present study shows for the first time that graded tests utilising four-minute stages are sufficient to establish steady-state gaseous exchange and group level peak fat oxidation rates compared to longer stage durations (≤ 10-minutes) in individuals with low levels of cardiorespiratory fitness. However, our data show that poor surrogate validity and substantial day-to-day variation are apparent when determining PFO. Consequently, graded tests with four-minute stages are advocated, but until further ‘optimisation’ is performed (particularly with a focus on strict pre-test physical activity and dietary standardisation), repeated assessments are prudent to more confidently and precisely determine ‘true’ peak fat oxidation rates.
ACKNOWLEDGMENTS

All participants are thanked for their involvement in the study. Dr Mark Thomas, Pippa Heath, Esther Punter, Max Simms and Aaron Hengist (Department for Health, University of Bath) are thanked for their help in the study.

AUTHORSHIPS

The study was designed by O.J.C-S., R.M.E., and J.T.G. Data collection was performed by O.J.C-S, and data analysis and interpretation assisted by J.A.B., and J.T.G. The manuscript was written by O.J.C-S. All authors revised the manuscript and approved the final article.

DECLARATIONS OF FUNDING SOURCES

Funding was provided by the Department for Health and University Research Studentship from the Graduate School, Faculty of Humanities and Social Sciences at the University of Bath. The funds were not involved in any part of the study except for financial and equipment support.

CONFLICTS OF INTEREST

The authors declare no conflict of interests.
REFERENCES


Green, S., and Askew, C. D. 2018. $\dot{V}O_2$peak is an acceptable estimate of cardiorespiratory fitness but not $\dot{V}O_2$max. J. Appl. Physiol. (1985). doi: 10.1152/japplphysiol.00850.2017. PMID: 29420148


### Table 1. Participant Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Total Sample</th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (Females/Males)</td>
<td>9/6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age, yrs</td>
<td>33.9 (27.4 – 40.4)</td>
<td>32.7 (22.9 – 42.4)</td>
<td>35.7 (27.7 – 43.6)</td>
</tr>
<tr>
<td>Stature, cm</td>
<td>170.6 (167.2 – 174.0)</td>
<td>166.6 (164.1 – 169.2)</td>
<td>176.5 (171.9 – 181.1)</td>
</tr>
<tr>
<td>Body Mass, kg</td>
<td>74.4 (66.7 – 82.1)</td>
<td>66.2 (62.3 – 70.2)</td>
<td>86.6 (73.1 – 100.1)</td>
</tr>
<tr>
<td>BMI, kg·m⁻²</td>
<td>25.4 (23.5 – 27.3)</td>
<td>23.9 (22.4 – 25.3)</td>
<td>27.7 (24.1 – 31.2)</td>
</tr>
<tr>
<td>Body Fat %</td>
<td>28.8 (25.3 – 32.3)</td>
<td>31.6 (28.5 – 34.8)</td>
<td>24.5 (18.4 – 30.6)</td>
</tr>
<tr>
<td>Waist Circumference, cm</td>
<td>83.9 (76.53 – 91.2)</td>
<td>75.5 (70.6 – 80.5)</td>
<td>96.4 (85.5 – 107.3)</td>
</tr>
<tr>
<td>Hip Circumference, cm</td>
<td>102.6 (99.2 – 105.9)</td>
<td>101.2 (98.1 – 104.2)</td>
<td>104.7 (97.5 – 111.8)</td>
</tr>
<tr>
<td>W:H ratio</td>
<td>0.81 (0.76 – 0.87)</td>
<td>0.75 (0.71 – 0.78)</td>
<td>0.92 (0.86 – 0.97)</td>
</tr>
<tr>
<td>V̇O₂peak, L·min⁻¹</td>
<td>2.4 (2.2 – 2.6)</td>
<td>2.2 (2.0 – 2.3)</td>
<td>2.8 (2.6 – 3.0)</td>
</tr>
<tr>
<td>V̇O₂peak, mL·kg⁻¹·min⁻¹</td>
<td>32.9 (30.8 – 35.1)</td>
<td>32.9 (30.6 – 35.2)</td>
<td>33.0 (28.6 – 37.3)</td>
</tr>
<tr>
<td>Peak Power Output, W</td>
<td>153 (142 – 163)</td>
<td>139 (132 – 146)</td>
<td>174 (164 – 183)</td>
</tr>
<tr>
<td>HRMAX, beats·min⁻¹</td>
<td>185 (176 – 193)</td>
<td>188 (176 – 200)</td>
<td>180 (171 – 189)</td>
</tr>
<tr>
<td>PFO a, g·min⁻¹</td>
<td>0.23 (0.19 – 0.27)</td>
<td>0.24 (0.19 – 0.29)</td>
<td>0.22 (0.15 – 0.28)</td>
</tr>
<tr>
<td>FATMAX a, %V̇O₂peak</td>
<td>36 (33 – 40)</td>
<td>39 (35 – 43)</td>
<td>33 (29 – 37)</td>
</tr>
<tr>
<td>FATMAX a, %PPO</td>
<td>29 (26 – 32)</td>
<td>31 (27 – 35)</td>
<td>25 (22 – 29)</td>
</tr>
<tr>
<td>FATMAX a, %HRMAX</td>
<td>55 (51 – 58)</td>
<td>58 (53 – 62)</td>
<td>51 (46 – 55)</td>
</tr>
</tbody>
</table>

\[a = n \text{ of 14, one participant excluded due to hyperventilation.}\]

BMI = Body Mass Index; W:H ratio = Waist-to-Hip ratio; V̇O₂peak = peak oxygen consumption; HRMAX = Maximum Heart Rate; PFO = Peak Fat Oxidation; FATMAX = the exercise intensity that elicits PFO; Data presented as mean (± 95 % CI).
Table 2. Metabolic variables during the short continuous exercise sessions (SCE)

<table>
<thead>
<tr>
<th></th>
<th>Minute 3-4</th>
<th>Minute 5-6</th>
<th>Minute 7-8</th>
<th>Minute 9-10</th>
<th>RM ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VE, L·min⁻¹</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCE_{pre}</td>
<td>17.94 (16.26 – 19.62)</td>
<td>17.41 (16.15 – 18.66)</td>
<td>17.94 (16.39 – 19.48)</td>
<td>18.13 (16.41 – 19.85)</td>
<td>F (1.959, 27.428) = 0.821, p = 0.448</td>
</tr>
<tr>
<td>SCE_{post}</td>
<td>28.17 (25.91 – 30.42)</td>
<td>28.00 (25.71 – 30.29)</td>
<td>28.37 (26.10 – 30.65)</td>
<td>28.43 (26.06 – 30.80)</td>
<td>F (3, 42) = 0.387, p = 0.763</td>
</tr>
<tr>
<td><strong>RER</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCE_{pre}</td>
<td>0.86 (0.82 – 0.89)</td>
<td>0.85 (0.83 – 0.88)</td>
<td>0.86 (0.86 – 0.88)</td>
<td>0.86 (0.83 – 0.89)</td>
<td>F (3, 42) = 0.145, p = 0.932</td>
</tr>
<tr>
<td>SCE_{equal}</td>
<td>0.85 (0.82 – 0.88)</td>
<td>0.87 (0.83 – 0.90)</td>
<td>0.87 (0.85 – 0.90)</td>
<td>0.86 (0.84 – 0.87)</td>
<td>F (3, 36) = 2.779, p = 0.055</td>
</tr>
<tr>
<td>SCE_{post}</td>
<td>0.93* (0.91 – 0.96)</td>
<td>0.93** (0.91 – 0.95)</td>
<td>0.91 (0.90 – 0.93)</td>
<td>0.91 (0.89 – 0.92)</td>
<td>F (3, 42) = 6.904, p = 0.001</td>
</tr>
</tbody>
</table>

SCE_{pre} n = 15; SCE_{equal} n = 13; SCE_{post} n = 15; **p ≤ 0.01 vs minute 9-10; *p < 0.05 vs minute 9-10. Data presented as mean (± 95 % CI).
Table 3. Surrogate validity of fat oxidation (g·min⁻¹) within the short continuous exercise sessions (SCE) at the stage preceding and equal to FAT$_{\text{MAX}}$.

<table>
<thead>
<tr>
<th>Fat Oxidation (g·min⁻¹)</th>
<th>Minute 3-4 vs 5-6</th>
<th>Minute 3-4 vs 7-8</th>
<th>Minute 3-4 vs 9-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCE$_{\text{pre}}$</td>
<td>$r$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>($n = 15$)</td>
<td>0.84 (0.58 – 0.95)</td>
<td>0.71 (0.30 – 0.89)</td>
<td>0.64 (0.18 – 0.87)</td>
</tr>
<tr>
<td>ICC</td>
<td>0.85 (0.60 – 0.95)</td>
<td>0.68 (0.27 – 0.88)</td>
<td>0.63 (0.19 – 0.86)</td>
</tr>
<tr>
<td>CV (%)</td>
<td>17 (10 – 22)</td>
<td>23 (17 – 27)</td>
<td>25 (13 – 32)</td>
</tr>
<tr>
<td>TE (g·min⁻¹)</td>
<td>0.03 (0.02 – 0.05)</td>
<td>0.04 (0.03 – 0.06)</td>
<td>0.04 (0.03 – 0.07)</td>
</tr>
<tr>
<td>SCE$_{\text{equal}}$</td>
<td>$r$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>($n = 13$)</td>
<td>0.82 (0.48 – 0.94)</td>
<td>0.83 (0.52 – 0.95)</td>
<td>0.65 (0.15 – 0.88)</td>
</tr>
<tr>
<td>ICC</td>
<td>0.77 (0.35 – 0.93)</td>
<td>0.72 (0.16 – 0.91)</td>
<td>0.61 (0.11 – 0.86)</td>
</tr>
<tr>
<td>CV (%)</td>
<td>25 (10 – 34)</td>
<td>19 (12 – 23)</td>
<td>19 (13 – 23)</td>
</tr>
<tr>
<td>TE (g·min⁻¹)</td>
<td>0.03 (0.02 – 0.06)</td>
<td>0.03 (0.02 – 0.05)</td>
<td>0.04 (0.03 – 0.07)</td>
</tr>
</tbody>
</table>

$r$ = Pearson correlation coefficient; ICC = Intra-Class Correlation; CV = Within-Subject Coefficient of Variation;

TE = Typical Error. Data presented as mean (± 95 % CI).
Figure 1a – 1h. Gas exchange kinetics ($\dot{V}O_2$ and $\dot{V}CO_2$) during the short continuous exercise bouts (SCE) at the stage before $FAT_{MAX}$ [$SCE_{pre} (a \ and \ b; \ n = 15)$], equal to $FAT_{MAX}$ [$SCE_{equal} (c \ and \ d; \ n = 13)$], after $FAT_{MAX}$ [$SCE_{post} (e \ and \ f; \ n = 15)$] and graded test (GT) vs SCE minute 3-4 at $FAT_{MAX}$ (g and h; $n = 12$). Fig. 1h displays Wilcoxon signed-rank test (due to the differences in $\dot{V}CO_2$ between the graded test (GT) vs SCE minute 3-4 at $FAT_{MAX}$ were non-normally distributed). A solid line represents the group mean $\pm$ 95% CI and dashed lines denote individual data. Overall group means are signified by triangles, females are denoted by stars and males are indicated by circles.
Figure 2a – 2d. Determination of fat oxidation (g-min⁻¹) during the short continuous exercise bouts (SCE) at the stage before FAT\textsubscript{MAX} [SCE\textsubscript{pre} (a; n = 15)], equal to FAT\textsubscript{MAX} [SCE\textsubscript{equal} (b; n = 13)], after FAT\textsubscript{MAX} [SCE\textsubscript{post} (c; n = 15)] and graded test (GT) vs SCE minute 3-4 at FAT\textsubscript{MAX} (d; n = 12). A solid line represents the group mean ± 95 % CI and dashed lines denote individual data. Overall group means are signified by triangles, females are denoted by stars and males are indicated by circles. ** \( p \leq 0.01 \).
Figure 3a – 3d. Bland-Altman plots of the agreement between fat oxidation (g·min⁻¹) in the short continuous exercise bout (SCE) at the stage equal to FAT\textsubscript{MAX} (SCE\textsubscript{equal}) minute 3-4 vs 5-6 (a; \( n = 13 \)), 7-8 (b; \( n = 13 \)) and 9-10 (c; \( n = 13 \)) and graded test (GT) vs SCE minute 3-4 (d; \( n = 12 \)) at FAT\textsubscript{MAX}. A solid line represents bias and the dashed lines represent lower and upper 95% limits of agreement. Females are denoted by stars and males are indicated by circles.
SUPPLEMENTARY MATERIAL S1

S1a. Detailed recruitment process

67 prospective participants expressed initial interest

Initial loss of prospective participants ($n = 41$)
- 17 self-excluded after reading participant information sheet
- 22 lost contact after expressing an initial interest
- 2 expressed interest, but after recruitment had closed

26 prospective participants had an initial meeting with O.C-S

Participant loss after initial meeting and informed consent obtained ($n = 2$)
- 2 excluded due to medical reasons

24 participants proceeded to complete graded test

Participant loss after graded test ($n = 6$)
- 1 participant withdrew for personal reasons
- 5 participants recorded a $\dot{V}O_2$ peak value above the cut-off threshold ($< 40 \text{ mL·kg}^{-1}·\text{min}^{-1}$)

16 participants met entire study inclusion criteria and completed the full protocol
**S1b. Criteria for attainment of steady-state gas exchange**

To assess whether steady-state gas exchange was achieved for each participant the following criteria was applied:

a) If the difference in $\dot{V}O_2$ and $\dot{V}CO_2$ between SCE minute 3-4 and the respective time point (i.e. SCE minute 5-6, 7-8, 9-10 or Graded Test minute 3-4) was within $1.96 \times$ Standard Deviation of the two concerned time-points.

And

b) If the Coefficient of Variation (CV) for $\dot{V}O_2$ and $\dot{V}CO_2$ was $\leq 10\%$ when comparing SCE minute 3-4 and each respective time point (i.e. SCE minute 5-6, 7-8, 9-10 or Graded Test minute 3-4). The CV was calculated via a slightly adapted version of Compher et al. (2006) guidelines on assessing whether steady-state conditions have been achieved to measure resting metabolic rate:

$$(\text{Standard Deviation of SCE minute 3-4 vs respective time point} \times \text{mean of SCE minute 3-4 and respective time point}) \times 100.$$
S1c. Sample sizes of statistical analysis tests

One participant hyperventilated in the short continuous exercise bout (SCE) at the stage equal to $\text{FAT}_{\text{MAX}}$ ($\text{SCE}_{\text{equal}}$) at all time points (ID 13) and was excluded from all analyses involving data at the stage equal to $\text{FAT}_{\text{MAX}}$ (i.e. $\text{SCE}_{\text{equal}}$ and vs graded test at $\text{FAT}_{\text{MAX}}$).

One participant hyperventilated at all time points in the graded test (ID 19) and was excluded from all analyses involving the graded test (i.e. graded test vs $\text{SCE}_{\text{equal}}$).

In one participant (ID 15) the expired gas sample at minute 3-4 was delayed by 40 seconds and thus was excluded from all analyses involving minute 3-4 data at the stage equal to $\text{FAT}_{\text{MAX}}$ (i.e. $\text{SCE}_{\text{equal}}$ and vs graded test at $\text{FAT}_{\text{MAX}}$ minute 3-4).

Repeated Measures ANOVA sample sizes:

- SCE at the stage before $\text{FAT}_{\text{MAX}}$ ($\text{SCE}_{\text{pre}}$), $n = 15$
- $\text{SCE}_{\text{equal}}$, $n = 13$ (Exclude IDs 13 and 15)
- SCE at the stage after $\text{FAT}_{\text{MAX}}$ ($\text{SCE}_{\text{post}}$), $n = 15$

Dependent Sample T-Tests or Wilcoxon Signed-Rank test sample sizes:

- Graded test at $\text{FAT}_{\text{MAX}}$ minute 3-4 vs $\text{SCE}_{\text{equal}}$ minute 3-4, $n = 12$ (Exclude IDs 13, 15 and 19)

Reliability and Intra-Individual Variability tests sample sizes:

- $\text{SCE}_{\text{pre}}$, $n = 15$
- $\text{SCE}_{\text{equal}}$, $n = 13$ (IDs 13 and 15 excluded)
- Graded test at $\text{FAT}_{\text{MAX}}$ minute 3-4 vs $\text{SCE}_{\text{equal}}$ minute 3-4, $n = 12$ (IDs 13, 15 and 19 excluded)
**Supplementary Table S1.** Metabolic variables in the graded test vs short continuous exercise session at stage equal to FAT\textsubscript{MAX}

<table>
<thead>
<tr>
<th></th>
<th>Graded Test</th>
<th>SCE minute 3-4</th>
<th>t-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>VE, L min(^{-1})</td>
<td>19.65 (17.23 – 22.07)</td>
<td>21.10 (18.75 – 23.45)</td>
<td>T (11) = -1.913, p = 0.082</td>
</tr>
<tr>
<td>RER</td>
<td>0.83 (0.80 – 0.86)</td>
<td>0.85 (0.81 – 0.88)</td>
<td>T (11) = -1.045, p = 0.319</td>
</tr>
</tbody>
</table>

Graded Test = minute 3-4; n = 12. Data presented as mean (± 95% CI).
**Supplementary Figure S1.** Bland-Altman plots of the agreement between fat oxidation (g·min⁻¹) in the short continuous exercise bout at the stage before FATMAX (SCEₚₑᵣ) minute 3-4 vs minute 5-6 (a; n = 15), 7-8 (b; n = 15) and 9-10 (c; n = 15). A solid line represents bias and the dashed lines represent lower and upper 95% limits of agreement. Females are denoted by stars and males are indicated by circles.