



The Reliability and Validity of a Four-Minute Running Time-Trial in Assessing $\dot{V}O_2\text{max}$ and Performance

Kerry McGawley*

Swedish Winter Sports Research Centre, Department of Health Sciences, Mid Sweden University, Östersund, Sweden

OPEN ACCESS

Edited by:

Gary W. Mack,
Brigham Young University, USA

Reviewed by:

Ben Rattray,
University of Canberra, Australia
Kai Petteri Savonen,
Kuopio Research Institute of Exercise
Medicine, University of Eastern
Finland, Finland

*Correspondence:

Kerry McGawley
kerry.mcgawley@miun.se

Specialty section:

This article was submitted to
Exercise Physiology,
a section of the journal
Frontiers in Physiology

Received: 24 November 2016

Accepted: 13 April 2017

Published: 03 May 2017

Citation:

McGawley K (2017) The Reliability and Validity of a Four-Minute Running Time-Trial in Assessing $\dot{V}O_2\text{max}$ and Performance. *Front. Physiol.* 8:270. doi: 10.3389/fphys.2017.00270

Introduction: Traditional graded-exercise tests to volitional exhaustion (GXTs) are limited by the need to establish starting workloads, stage durations, and step increments. Short-duration time-trials (TTs) may be easier to implement and more ecologically valid in terms of real-world athletic events. The purpose of the current study was to assess the reliability and validity of maximal oxygen uptake ($\dot{V}O_2\text{max}$) and performance measured during a traditional GXT (STEP) and a four-minute running time-trial (RunTT).

Methods: Ten recreational runners (age: 32 ± 7 years; body mass: 69 ± 10 kg) completed five STEP tests with a verification phase (VER) and five self-paced RunTTs on a treadmill. The order of the STEP/VER and RunTT trials was alternated and counter-balanced. Performance was measured as time to exhaustion (TTE) for STEP and VER and distance covered for RunTT.

Results: The coefficient of variation (CV) for $\dot{V}O_2\text{max}$ was similar between STEP, VER, and RunTT (1.9 ± 1.0 , 2.2 ± 1.1 , and $1.8 \pm 0.8\%$, respectively), but varied for performance between the three types of test (4.5 ± 1.9 , 9.7 ± 3.5 , and $1.8 \pm 0.7\%$ for STEP, VER, and RunTT, respectively). Bland-Altman limits of agreement (bias \pm 95%) showed $\dot{V}O_2\text{max}$ to be 1.6 ± 3.6 mL \cdot kg $^{-1}\cdot$ min $^{-1}$ higher for STEP vs. RunTT. Peak HR was also significantly higher during STEP compared with RunTT ($P = 0.019$).

Conclusion: A four-minute running time-trial appears to provide more reliable performance data in comparison to an incremental test to exhaustion, but may underestimate $\dot{V}O_2\text{max}$.

Keywords: graded-exercise test, maximal oxygen uptake, reproducibility, testing, verification phase

INTRODUCTION

Maximal oxygen uptake ($\dot{V}O_2\text{max}$) testing is widely used to assess aerobic fitness, with procedures typically involving a graded-exercise test (GXT) to volitional exhaustion. While guidelines have been published for common exercise modes such as cycling, running, and rowing (Gore, 2000; Winter et al., 2008; Cooke, 2009), standardized GXT protocols do not exist for individual sports. Instead, starting workloads, stage durations, and step increments all need to be established prior to the test in an effort to produce a final test duration that lies within the recommended range of 5–10 min (Jones, 2008). Open-ended GXT protocols (i.e., without a fixed end-point) lead to variations

in test duration that significantly affect the $\dot{V}O_2\text{max}$ measurement (Yoon et al., 2007) and produce unreliable measures of performance (Currell and Jeukendrup, 2008). As such, alternative protocols may be preferable.

Beltrami et al. (2012) demonstrated higher $\dot{V}O_2\text{max}$ values with a decremental test (i.e., progressive reductions in treadmill speed, rather than increases) compared with a traditional GXT. However, a decremental test suffers from similar practical drawbacks as an incremental test in terms of needing to select starting workloads, stage durations, and step decrements, as well as being open-ended. Another alternative to the traditional GXT has shown significantly higher $\dot{V}O_2\text{max}$ values in both cycling and running when “clamping” five, 2-min incremental stages according to rating of perceived exertion (RPE) scores (Mauger and Sculthorpe, 2012; Mauger et al., 2013; Hogg et al., 2015). While this type of protocol does overcome the issues associated with open-ended tests, since there is a fixed end-point, the progressive increase in RPE clamps (starting at 11 then increasing to 13, 15, 17, and finally 20) is still incremental and therefore not ecologically valid in terms of real-world athletic events.

All-out cycling tests lasting between 90 and 180 s have been shown to elicit similar $\dot{V}O_2\text{max}$ values to those attained during typical GXT protocols (Williams et al., 2005; Burnley et al., 2006; Rossiter et al., 2006; Sperlich et al., 2011; Chidnok et al., 2013). However, all-out tests are characterized by an immediate acceleration phase to peak power output followed closely by a reduction in power output as fatigue ensues (Abbiss and Laursen, 2008). Since individual athletic events lasting >90 s are typically characterized by a pacing component, with individuals self-selecting their distribution of physical effort (Abbiss and Laursen, 2008), all-out tests also lack a degree of ecological validity.

Short time-trials (TTs) may be a preferable alternative to all of the aforementioned $\dot{V}O_2\text{max}$ protocols since they have a fixed end-point and are self-paced. As well as a higher degree of ecological validity, TTs elicit far lower coefficient of variation (CV) scores for performance compared to time-to-exhaustion (TTE) tests (Jeukendrup et al., 1996; Currell and Jeukendrup, 2008). Previous findings have shown that the $\dot{V}O_2\text{max}$ attained during a 1-mile (~5 min) running TT performed on a 200-m indoor track (Crouter et al., 2001), a 4-km (~5 min) cycle-ergometer TT (Ansley et al., 2004), and a 600-m (~3 min) skate roller-ski TT performed on a treadmill (Losnegard et al., 2012) was similar to that attained during a standard GXT. In some cases, higher $\dot{V}O_2\text{max}$ values have been produced during TTs compared with GXT protocols (Foster et al., 1993; McGawley and Holmberg, 2014). Despite these findings, and the similarity between TTs and real-world competition, TT protocols have not become standard procedure for assessing $\dot{V}O_2\text{max}$ in studies with athlete populations. In addition, the validity and reliability of using laboratory-based treadmill running TTs for assessing $\dot{V}O_2\text{max}$ and performance have not been investigated.

The aims of the current study were (i) to compare the $\dot{V}O_2\text{max}$ attained during a standard open-ended GXT to volitional exhaustion on a treadmill (STEP) with that attained during a 4-min, self-paced, laboratory-based running TT (RunTT), and (ii) to assess the reliability of the STEP and the RunTT protocols. It was hypothesized that the $\dot{V}O_2\text{max}$ attained during the RunTT

would be similar to that attained during the STEP and that the RunTT would generate more reliable performance data than the STEP.

MATERIALS AND METHODS

Participants

Five males and five females (mean \pm SD: age 32 ± 7 years, body mass 69 ± 10 kg) were recruited from local running, triathlon, and multi-sport clubs. During the test period participants were completing 4 ± 1 run training sessions per week as well as a mix of swimming, cycling, paddling, cross-country skiing, and gym training. Immediately prior to the start of the study participants were carrying out high-intensity interval run training weekly or bi-weekly and they competed in a mixture of running and/or multi-sport competitions locally and internationally. Best reported 5- and 10-km run times for the group were mean \pm SD: 19.5 ± 1.8 and 40.1 ± 3.3 min, respectively. All participants were fully informed about the study before providing written consent to participate and the study was pre-approved by the Regional Ethical Review Board, Umeå University, Umeå, Sweden.

Study Overview

Each participant visited the laboratory on 11 occasions. On the first visit individuals completed familiarization sessions for the STEP and RunTT. The following 10 visits involved participants completing either a STEP, which included a verification phase (VER) or a RunTT. The type of test completed was alternated on each visit and counterbalanced, such that five participants commenced with the STEP + VER and five commenced with the RunTT.

Equipment

All run tests were completed on a motor-driven treadmill (Rodby RL 3500, Rodby, Vänge, Sweden) with speed externally controlled by the test leader for the STEP and VER trials, and controlled by the participant during the RunTTs. The treadmill was fitted with a laser system allowing participants to automatically increase or decrease the speed during the RunTTs by moving to the front or the rear of the belt, respectively, maintaining a constant speed otherwise (Swarén et al., 2013). Heart rate (HR) was monitored continuously throughout each trial using a Polar system (RS800CX, Polar Electro Oy, Kempele, Finland) and maximal HR (HR_{max}) was calculated as a peak 5-s value. Blood lactate concentration was measured from fingertip blood samples (Biosen 5140, EKF diagnostic GmbH, Magdeburg, Germany). Respiratory variables were measured using a mixed expired air procedure with an ergospirometry system (AMIS 2001 model C, Innovision A/S, Odense, Denmark) equipped with a flow meter. The gas analyzers were calibrated with a high-precision mixture of 16.0% O_2 and 4.0% CO_2 (Air Liquide, Kungsängen, Sweden) and the flow meter was calibrated at three rates with a 3-L air syringe (Hans Rudolph, Kansas City, USA). The $\dot{V}O_2\text{max}$, $\dot{V}CO_2$, and $\dot{V}E$ were monitored continuously and $\dot{V}O_2\text{max}$ -values were calculated and reported as 30-s averages. The associated RER-values were calculated as averages for the same time points as those used to calculate $\dot{V}O_2\text{max}$.

Standardized Procedures

Participants reported to the laboratory prior to each trial in a fed state and had abstained from alcohol for at least 24 h prior to testing and from caffeine on the day of the trial. Participants arrived rested and had not completed any intense training on the day before testing. Prior to the first experimental trial (visit 2) participants recorded their diet on the day before and the day of testing and were instructed to consume the same diet prior to all subsequent trials. The 10 experimental trials were completed at the same time of day \pm 1 h for each individual, in order to control for circadian variance (Reilly and Brooks, 1982). Following the measurement of body mass in minimal clothing participants performed an individualized 10-min warm-up that was standardized prior to each trial and consisted of 5 min of low-intensity running, 3 \times 30-s high-intensity intervals separated by 30 s of low-intensity running and finally 2 min of low-intensity running (Watkins et al., 2017). Three minutes after completing the warm-up participants commenced the STEP + VER or RunTT and HR and expired air were measured continuously. Immediately after completing the test participants reported their RPE (Borg, 1990). Fingertip blood samples were taken 1, 2, 3, and 4 min after each test for the measurement of [LAC] and the highest value was recorded as peak [LAC]. Each of the 10 experimental trials were separated by a minimum of 3 days.

Incremental Step Test to Exhaustion (STEP)

Participants completed a familiarization of the STEP on their first laboratory visit and the five experimental STEP trials followed the same procedures. Following the 10-min warm-up described previously and a 3-min break participants commenced the STEP at a gradient of 1% and either 12, 13, 14, or 15 km·h⁻¹, depending on running ability. The gradient of the treadmill was increased by 1% every minute while speed remained constant and participants, receiving standardized encouragement, continued until volitional exhaustion. No feedback was provided during the test regarding performance (e.g., time or gradient) or physiological (e.g., HR or $\dot{V}O_2\max$) responses. Time to exhaustion (TTE) was recorded automatically by the treadmill software. Upon completion of the test participants reported their RPE score and fingertip blood samples were collected.

Verification Phase (VER)

Nine minutes after reaching exhaustion in the STEP participants commenced the VER, which consisted of an additional run test to exhaustion (Midgley et al., 2006; Rossiter et al., 2006). The VER was completed at the same gradient as that reached during the STEP (on that particular day) and at 105% of the STEP speed. Expired air and HR were collected throughout the VER trial and TTE was recorded as the performance variable. Upon completion of the test participants again reported their RPE score and fingertip blood samples were collected.

Four-Minute Running Time-Trial (RunTT)

Approximately 12 min after the end of the STEP familiarization participants completed a familiarization of the RunTT, which followed the same procedures as the subsequent experimental RunTT trials. Following the 10-min warm-up described

previously and a 3-min break (or the STEP familiarization and a 12-min break in the case of the familiarization trial) participants commenced the RunTT at a gradient of 1% and a default starting speed that matched each individual's STEP speed. Participants were able to modify the speed of the treadmill immediately by moving to the front of the belt to speed up (with an acceleration rate of 0.5 km·h⁻¹·s⁻¹) or the rear to slow down (with a deceleration rate of 0.4 km·h⁻¹·s⁻¹). The runners were able to see their position on the treadmill in relation to the front, middle and rear zones on a screen in front of them and were instructed to control the speed in order to cover as much distance as possible in 4 min, thereby producing a maximal effort. Participants could see elapsed time on the screen but received no performance (e.g., distance covered) or physiological (e.g., HR or $\dot{V}O_2\max$) feedback. Standardized encouragement was provided throughout each RunTT, as well as verbal information when only 5 s remained. Distance covered was recorded automatically by the treadmill software.

Criteria for the Attainment of $\dot{V}O_2\max$

The following criteria outlined by Cooke (2009) were used to assess the attainment of $\dot{V}O_2\max$ during the STEP, VER and RunTT tests: (i) a $\dot{V}O_2$ increase of <2.0 mL·kg⁻¹·min⁻¹ or 3% (i.e., a plateau), (ii) an RER value ≥ 1.15 , (iii) a HR value within 10 beats·min⁻¹ of the age-predicted maximum (220-age), (iv) a peak [LAC] of ≥ 8.0 mmol·L⁻¹, (v) an RPE of 19 or 20. The $\dot{V}O_2\max$ plateau was evaluated over the 1-min period preceding the attainment of the highest 30-s $\dot{V}O_2\max$ value.

Data Analyses

Data are expressed as mean \pm SD unless stated otherwise and the level of significance was set at $P \leq 0.05$. $\dot{V}O_2\max$ and peak HR, lactate, RER, and RPE responses were compared for each test type and trial using two-way ANOVAs with repeated measures. Sphericity was checked using Mauchly's test and the Greenhouse-Geisser correction was used when the assumption of sphericity was violated. *Post-hoc* tests with Bonferroni adjustments for multiple comparisons were used to identify pairwise differences. All ANOVA and associated *post-hoc* tests were carried out using the Statistical Package for the Social Sciences (SPSS Inc., Chicago, USA). The intra-individual (trial-to-trial) CV was calculated as SD/mean and intraclass correlation coefficients (ICCs) with 95% confidence intervals (CIs) were calculated across the five repeated trials for both $\dot{V}O_2\max$ and performance. The bias \pm 95% limits of agreement between STEP, VER, and RunTT for $\dot{V}O_2\max$ were evaluated using the Bland-Altman method with multiple observations per individual (Bland and Altman, 2007). The ICCs and Bland-Altman calculations were carried out using MedCalc statistical software (MedCalc Software, Ostend, Belgium).

RESULTS

The STEP and VER trials lasted 481 ± 56 s (range: 372–606 s) and 108 ± 16 s (range: 79–148 s), respectively, and distance covered during the RunTT was 1114 ± 90 m (range: 967–1306 m). $\dot{V}O_2\max$ was attained after 469 ± 59 , 107 ± 17 , and 227 ± 17 s during the STEP, VER, and RunTT trials, respectively.

The frequencies of participants fulfilling the criteria for attaining $\dot{V}O_2\text{max}$ during trials 1–5 for STEP, VER, and RunTT are displayed in **Table 1**.

Reliability of $\dot{V}O_2\text{max}$ and Performance

The $\dot{V}O_2\text{max}$ and performance data for the five STEP, VER, and RunTT trials are shown in **Tables 2, 3**, respectively, together with the trial-to-trial CV- and ICC-values. The CV and ICC for $\dot{V}O_2\text{max}$ showed similar levels of repeatability between STEP, VER, and RunTT, while reliability for performance was substantially improved for RunTT vs. STEP and VER, as well as for STEP vs. VER.

Validity of $\dot{V}O_2\text{max}$

One missing data point during trial 2 resulted in $n = 9$ for the statistical comparisons associated with the $\dot{V}O_2\text{max}$ data. There was a significant effect of test type on $\dot{V}O_2\text{max}$, with higher values recorded during STEP compared with both VER ($P = 0.013$) and RunTT ($P = 0.008$). However, no significant differences were identified between VER and RunTT ($P = 0.455$). Bland-Altman limits of agreement showed a bias $\pm 95\%$ of $1.1 \pm 3.6 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for STEP vs. VER (**Figure 1A**), $1.6 \pm 3.6 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for STEP vs. RunTT (**Figure 1B**), and $0.4 \pm 3.7 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for VER vs. RunTT (**Figure 1C**).

Physiological and RPE Responses

There was a significant effect of test type on peak HR ($n = 6$ due to a number of corrupt HR files), with significantly higher values recorded during STEP compared with both VER ($P = 0.004$)

and RunTT ($P = 0.019$; **Figure 2A**). Peak lactate was significantly lower for STEP vs. VER ($P = 0.001$) and there was a significant trial effect, with higher peak lactate values recorded after trial 1 vs. trial 5 ($P = 0.039$; **Figure 2B**). Peak RER was significantly higher for both STEP and RunTT compared with VER ($P < 0.001$; **Figure 2C**), while peak RPE did not differ between any of the test protocols or trials (**Figure 2D**).

DISCUSSION

The first aim of the current study was to compare the $\dot{V}O_2\text{max}$ attained during a standard GXT to volitional exhaustion with that attained during a self-paced, laboratory-based running TT. Findings showed that $\dot{V}O_2\text{max}$ values attained during STEP were significantly higher than those attained during RunTT. In addition, Bland-Altman limits of agreement showed a general bias of $1.6 \pm 3.6 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for STEP vs. RunTT. The second aim of the study was to assess the reliability of the STEP and RunTT protocols. Findings showed similar levels of variation for $\dot{V}O_2\text{max}$ in both the STEP and the RunTT; however, the RunTT generated more reliable performance data than the STEP.

Graded-Exercise Tests vs. Time-Trials

Traditional GXTs have been criticized in the literature for a number of reasons. Firstly, no standardized protocol exists and selecting different starting workloads, stage durations, and step increments can significantly affect measures of sub-maximal [LAC], peak HR, and maximal power output (Bishop et al., 1998;

TABLE 1 | The number of participants (out of 10) who fulfilled the criteria for attaining $\dot{V}O_2\text{max}$ during the five STEP, VER, and RunTT trials (T1–T5).

| | | T1 | T2 | T3 | T4 | T5 | Median |
|--|-------|----|----|----|----|----|--------|
| $\dot{V}O_2$ ($<2.0 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ or 3% increase) | STEP | 6 | 8 | 4 | 5 | 5 | 5 |
| | VER | 0 | 0 | 0 | 0 | 0 | 0 |
| | RunTT | 7 | 5 | 5 | 6 | 6 | 6 |
| RER (≥ 1.15) | STEP | 1 | 2 | 3 | 1 | 1 | 1 |
| | VER | 0 | 0 | 0 | 0 | 0 | 0 |
| | RunTT | 6 | 5 | 4 | 4 | 2 | 4 |
| HR ($\leq 10 \text{ beats}\cdot\text{min}^{-1}$ of age-predicted max) | STEP | 8 | 6 | 5 | 6 | 7 | 6 |
| | VER | 3 | 2 | 3 | 2 | 2 | 2 |
| | RunTT | 4 | 4 | 4 | 5 | 5 | 4 |
| Peak [LAC] ($\geq 8.0 \text{ mmol}\cdot\text{L}^{-1}$) | STEP | 8 | 8 | 8 | 8 | 6 | 8 |
| | VER | 8 | 8 | 8 | 9 | 7 | 8 |
| | RunTT | 9 | 8 | 9 | 8 | 7 | 8 |
| RPE (19 or 20) | STEP | 7 | 7 | 7 | 5 | 7 | 7 |
| | VER | 7 | 5 | 4 | 7 | 7 | 7 |
| | RunTT | 6 | 4 | 7 | 5 | 5 | 5 |
| ≥ 3 criteria attained | STEP | 5 | 6 | 5 | 5 | 6 | 5 |
| | VER | 3 | 2 | 1 | 2 | 2 | 2 |
| | RunTT | 8 | 6 | 6 | 6 | 5 | 6 |

TABLE 2 | Mean \pm SD of $\dot{V}O_2\max$ ($\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) for each trial for the incremental test to volitional exhaustion (STEP), the verification phase (VER), and the four-minute running time-trial (RunTT) with associated coefficient of variation (CV) and intraclass correlation coefficient (ICC) data.

| $n = 9$ | T1 | T2 | T3 | T4 | T5 | CV (%) | ICC (95% CI) |
|---------|----------------|----------------|----------------|----------------|----------------|---------------|------------------|
| STEP | 58.9 \pm 6.0 | 59.5 \pm 7.0 | 58.4 \pm 6.6 | 59.2 \pm 7.2 | 59.8 \pm 7.1 | 1.9 \pm 1.0 | 0.97 (0.92–0.99) |
| VER | 58.1 \pm 6.4 | 58.8 \pm 7.3 | 57.3 \pm 6.7 | 58.4 \pm 6.3 | 57.8 \pm 6.7 | 2.2 \pm 1.1 | 0.97 (0.92–0.99) |
| RunTT | 57.5 \pm 5.6 | 57.9 \pm 6.4 | 57.8 \pm 5.8 | 57.5 \pm 6.1 | 57.5 \pm 6.6 | 1.8 \pm 0.8 | 0.97 (0.93–0.99) |

TABLE 3 | Mean \pm SD of performance for each trial for the incremental test to volitional exhaustion (STEP), the verification phase (VER), and the four-minute running time-trial (RunTT) with associated coefficient of variation (CV) and intraclass correlation coefficient (ICC) data.

| | T1 | T2 | T3 | T4 | T5 | CV (%) | ICC (95% CI) |
|-----------|---------------|----------------|---------------|---------------|---------------|---------------|------------------|
| STEP (s) | 475 \pm 44 | 488 \pm 57 | 472 \pm 61 | 486 \pm 62 | 483 \pm 66 | 4.5 \pm 1.9 | 0.83 (0.65–0.95) |
| VER (s) | 116 \pm 19 | 110 \pm 18 | 102 \pm 16 | 109 \pm 15 | 104 \pm 13 | 9.7 \pm 3.5 | 0.59 (0.32–0.85) |
| RunTT (m) | 1123 \pm 98 | 1108 \pm 101 | 1117 \pm 94 | 1115 \pm 89 | 1106 \pm 85 | 1.8 \pm 0.7 | 0.95 (0.88–0.98) |

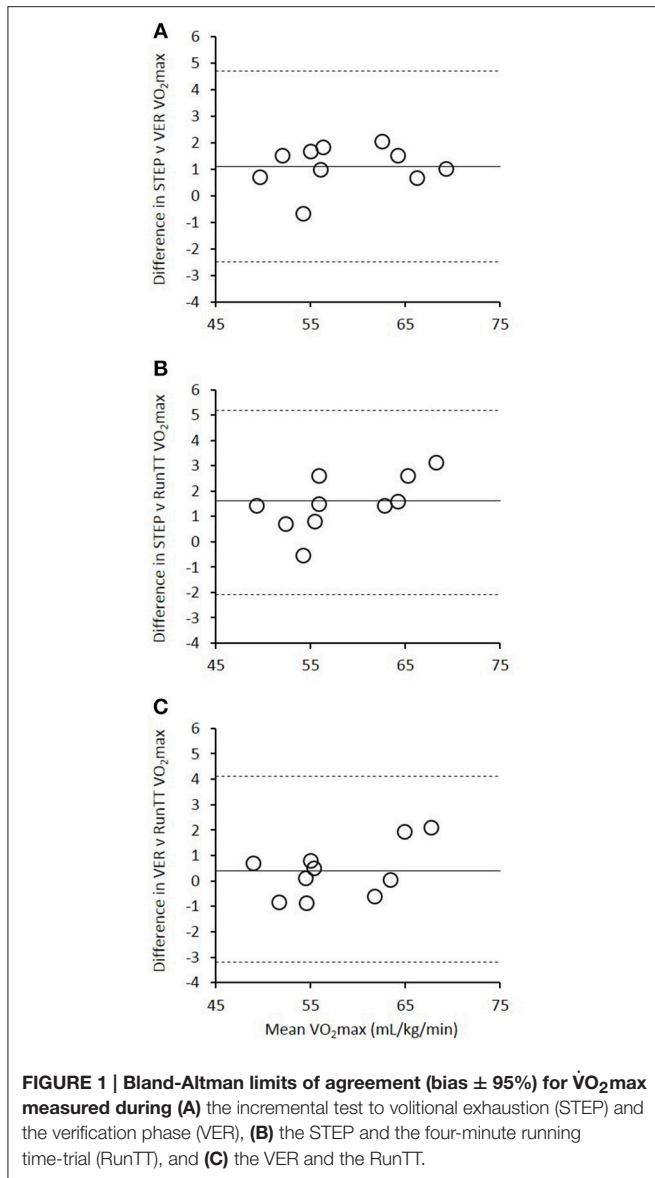
Kuipers et al., 2003). Secondly, the point of volitional exhaustion in an open-ended test is subjective and variable, which has been shown to lead to variations in $\dot{V}O_2\max$ (Yoon et al., 2007) and performance (Currell and Jeukendrup, 2008). Thirdly, the ecological validity of a GXT has been questioned, with Noakes (2008) stating that progressive increases in exercise intensity up to a maximum is not how humans “usually” exercise. For these reasons, the current study was designed to evaluate both the reliability and validity of assessing $\dot{V}O_2\max$ and performance in the laboratory using a more ecologically valid test.

Self-paced maximal efforts lasting ~ 4 min have been shown to be effective in eliciting $\dot{V}O_2\max$ across a variety of exercise modes, including indoor track running, cycle ergometry, and treadmill roller-skiing (Crouter et al., 2001; Ansley et al., 2004; Losnegard et al., 2012; McGawley and Holmberg, 2014). Standardizing time rather than distance, while subtly different from a typically distance-based athletic event (Abbiss et al., 2016), allows for direct comparisons between exercise modes. Treadmill running under laboratory conditions was chosen in the current study in order to provide a greater degree of environmental control compared with running in the field (i.e., on a track). Furthermore, the automated treadmill system used in the current study allowed participants to truly self-pace and produce a maximal effort. While a limited number of studies have attempted to validate alternative run-based protocols for the assessment of $\dot{V}O_2\max$ (Beltrami et al., 2012; Mauger et al., 2013; Hogg et al., 2015; Scheadler and Devor, 2015), only one study has prescribed a truly self-paced running TT (Crouter et al., 2001). In their study, Crouter et al. (2001) concluded that $\dot{V}O_2\max$ was not different between a traditional incremental test to exhaustion (where the gradient was fixed at 1% and speed was increased every 2 min) and a 1-mile running TT performed on a 200-m indoor track. This is contrary to the findings of the current study where lower $\dot{V}O_2\max$ scores were demonstrated during self-paced running.

Reasons for the conflicting findings in the current study may be explained by differences in study design and methods of data analysis. Crouter et al. (2001) performed only one incremental

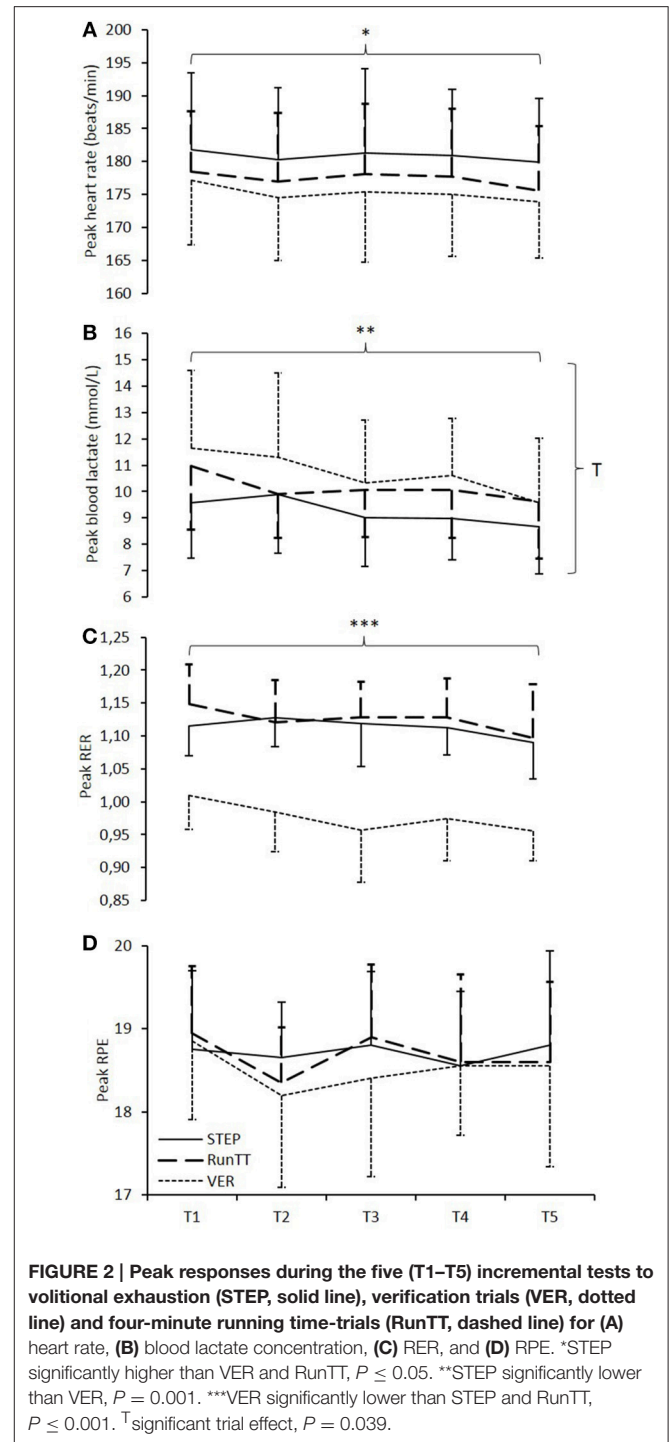
test and one, 1-mile TT. This design is similar to the self-paced RPE-clamped running studies, which have concluded from a single comparison that a similar or higher $\dot{V}O_2\max$ is achieved during a self-paced trial compared with a traditional GXT (Mauger et al., 2013; Hogg et al., 2015). However, the current study involved a series of repeated measures. Interestingly, comparing tests from the first trial only (using a one-way repeated-measures ANOVA with Bonferroni-adjusted *post-hoc* comparisons) would have revealed no significant differences between STEP, VER, and/or RunTT. Therefore, if a series of comparisons, as well as the Bland-Altman calculations, had not been conducted then the conclusion would also have been one of no difference. This suggests that one-off comparisons may not be sufficient to conclude that $\dot{V}O_2\max$ will be systematically similar (or different) between tests. This is the first study to have compared test protocols over a series of five repeated trials and the significant difference identified between test type, as well as the bias reflected in the Bland-Altman limits of agreement, suggests that $\dot{V}O_2\max$ may be underestimated in the RunTT compared with the STEP.

Another likely explanation for the differences between previous and current findings relates to the specific nature of the GXT protocol. In the current study a fixed speed was used during the STEP and the treadmill gradient was increased by 1% every minute until volitional exhaustion. This method was selected in an attempt to elicit a true $\dot{V}O_2\max$ for each individual, with higher values known to occur during uphill versus flat running (Sloniger et al., 1997; Pringle et al., 2002). By contrast, the RunTT was completed at a fixed gradient of 1% and this was chosen to simulate “normal” middle-distance-type racing conditions. This difference in gradient (i.e., an uphill STEP compared with a flat RunTT) may explain the lower $\dot{V}O_2\max$ in the RunTT. In support of this notion, the similar $\dot{V}O_2\max$ values observed by Crouter et al. (2001) were measured using a GXT fixed at 1% with increases in speed and a 1-mile TT completed on a flat, indoor running track (i.e., similar gradients between tests). Furthermore, matched gradients during a run-based GXT and an RPE-clamped self-paced test (both 3%) led to similar $\dot{V}O_2\max$



scores, while a significantly higher $\dot{V}O_2\text{max}$ was observed during an RPE-clamped self-paced test with a final incline of $11.0 \pm 3.2\%$ (Hogg et al., 2015). Additionally, a higher end-exercise gradient during a run-based GXT ($\sim 10\%$) compared with a self-paced test at an 8% gradient led to higher $\dot{V}O_2\text{max}$ scores during the GXT (Scheidler and Devor, 2015). These results highlight the importance of running gradient when comparing $\dot{V}O_2\text{max}$ values from different types of test. In practice, the choice of gradient may potentially depend upon the specific purpose of the test.

Most previous studies aiming to validate an alternative running test for assessing $\dot{V}O_2\text{max}$ have performed only *t*-tests or repeated-measures ANOVAs. However, in a recent study Hogg et al. (2015) presented Bland-Altman plots and a bias of $\sim 1 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for the $\dot{V}O_2\text{max}$ elicited during a GXT compared with a self-paced test, but no significant difference between the means (both tests were completed at a 3% gradient). Despite this, Figure 3 in their study illustrates individual differences that vary



by $\sim 16 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ (from $\sim -10 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for one individual to $\sim 6 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for another). These large inter-individual variations in test data indicate an additional need for test-retest reliability analyses, since an unreliable test is of little practical use. The Bland-Altman analyses in the present study used a modified approach (Bland and Altman, 2007), due to multiple observations made per individual (i.e., each participant performed the same test five times). As such, Figure 1 in the

current study does not display individual data points, rather shows grouped data for each of the 10 participants. However, when examining the individual data points there was a somewhat improved range of inter-individual differences between the STEP and RunTT $\dot{V}O_2\text{max}$ values compared to the results of Hogg et al. (2015), with differences between any single comparison of tests ranging from $-2.6 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for one individual to $+4.9 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ for another (i.e., a total range for mean differences of $7.5 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$). In addition, the CV for $\dot{V}O_2\text{max}$ in both the STEP and RunTT protocols demonstrated a high level of reproducibility, with mean values for both protocols of $<2\%$. This is lower than the 3.7% CV reported by Mauger et al. (2013) for two GXT trials in a sub-group of five well-trained runners and similar to the 1.4% value reported by Rollo et al. (2008) for three 60-min treadmill-based TTs in a similar group of 10 runners.

Reliability of Time-Trial Performance

The present study is the first to have investigated the reliability of self-paced treadmill performance using a laser system of this kind and the CV and ICC data reflect high levels of repeatability, particularly in comparison to both the STEP- and VER-tests. This is perhaps not surprising, given that time-trials are known to be more reproducible than TTE-tests (Jeukendrup et al., 1996; Currell and Jeukendrup, 2008). The low CV and high ICC scores for RunTT performance may be explained by the training status of the participants, who were recreationally—to well-trained according to the classification system of De Pauw et al. (2013), and the pre-experimental familiarization procedures, both of which lead to improved reliability (Currell and Jeukendrup, 2008). While a self-paced treadmill of this type is uncommon it is not unique, with other research groups describing similar systems within laboratory settings (Stöggl et al., 2007; Losnegard et al., 2012). With reliable performance data, a range of novel laboratory-based studies are possible using running or cross-country roller-skiing, as opposed to the more common mode of cycling, whereby pacing and intervention strategies may be examined in relation to maximal performance (Andersson et al., 2016; Stocks et al., 2016; Watkins et al., 2017).

Criteria for the Attainment of $\dot{V}O_2\text{max}$

While there appears to be no universal agreement regarding the criteria used to identify the attainment of $\dot{V}O_2\text{max}$ during a GXT (Howley et al., 1995; Midgley et al., 2007), those recommended by the British Association of Sports and Exercise Sciences cited by Cooke (2009) were used in the current study. These include a $\dot{V}O_2\text{max}$ plateau (measured as an increase of $<2.0 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ or 3% from the previous minute), a peak RER ≥ 1.15 , a peak HR within $\leq 10 \text{ beats}\cdot\text{min}^{-1}$ of age-predicted maximum, a peak [LAC] $\geq 8.0 \text{ mmol}\cdot\text{L}^{-1}$ and an RPE of 19 or 20. The number of participants attaining at least three of these criteria during each of the maximal trials was analyzed and the median value from the five trials was presented in **Table 1** for each type of test. The findings support previous suggestions that the traditional criteria for the attainment of $\dot{V}O_2\text{max}$ are not consistently met during GXT protocols (Poole et al., 2008; Midgley et al., 2009), with a median of only five out of ten participants achieving at least three of the criteria during the

STEP in the present study. The count for the RunTT was slightly higher, but still not compelling, with a median of six out of ten participants attaining at least three of the criteria. Consistent with previous findings (Snoza et al., 2016), the most frequently attained criterion was a peak [LAC] $\geq 8.0 \text{ mmol}\cdot\text{L}^{-1}$ (median: 8 out of 10 participants for both STEP and RunTT), while the least commonly attained criterion was an RER ≥ 1.15 .

Verification Phase

In an attempt to validate the attainment of a true $\dot{V}O_2\text{max}$, a verification stage has been introduced to GXT protocols involving a shorter, exhaustive test performed at an exercise intensity slightly higher than the final stage of the preceding GXT. Previous studies have shown similar $\dot{V}O_2\text{max}$ values elicited during a verification phase as in a preceding GXT (Rossiter et al., 2006; Foster et al., 2007). However, results from the current study suggest that the VER may underestimate the $\dot{V}O_2\text{max}$ measured during the STEP, with a significant difference identified between the two tests and Bland-Altman limits of agreement demonstrating a bias of $1.1 \pm 3.6 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$. Again, higher $\dot{V}O_2\text{max}$ values attained during the STEP compared with the VER would not have been identified, had five repeated trials not been performed. Results from the present study therefore support the concerns of Midgley and Carroll (2009), who claim that a lack of convincing evidence exists for using standard verification protocols to confirm the attainment of a “true” $\dot{V}O_2\text{max}$.

Conclusion

The current study has shown that $\dot{V}O_2\text{max}$ may be underestimated during a self-paced running TT performed on a treadmill at a gradient of 1% , when compared with the $\dot{V}O_2\text{max}$ measured during a GXT performed at increasing gradients until exhaustion. If obtaining a true estimate of $\dot{V}O_2\text{max}$ is of greatest concern then the RunTT used in the current study would not be recommended. However, if an ecologically valid measure of $\dot{V}O_2\text{max}$ for running on flat terrain is a priority (i.e., for track or road runners) then the RunTT may be more relevant. If performance is also of interest then a 4-min self-paced running TT provides a more reliable measure (i.e., distance covered) compared with a GXT (i.e., time to exhaustion). In addition, a TT (unlike a GXT) is also able to provide detailed information regarding pacing strategies, as well as the metabolic cost and anaerobic energy contribution during performance (provided a suitable sub-maximal pre-test is completed; McGawley and Holmberg, 2014; Andersson et al., 2016). Therefore, it is important to understand the limitations of each test and select the most appropriate protocol based on the target population and the outcome variables of greatest interest.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Regional Ethical Review Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Regional Ethical Review Board, Sweden.

AUTHOR CONTRIBUTIONS

KM conceived and designed the study and made substantial contributions to the acquisition, analysis, and interpretation of the data. KM drafted and revised the work, approves the final version to be published, and agrees to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

The author would like to thank Dr. Martyn Beaven, Dr. Glenn Björklund, Dr. Erik Andersson, and Dr. Andrew Govus for their valuable input relating to the planning, data collection, statistical analyses, and manuscript preparation for this study. In addition, the author would like to thank Mr. Simon Platt for his help in the lab during the data collection phase.

REFERENCES

- Abbiss, C. R., and Laursen, P. B. (2008). Describing and understanding pacing strategies during athletic competition. *Sports Med.* 38, 239–252. doi: 10.2165/00007256-200838030-00004
- Abbiss, C. R., Thompson, K. G., Lipski, M., Meyer, T., and Skorski, S. (2016). Pacing differs between time- and distance-based time trials in trained cyclists. *Int. J. Sports Physiol. Perform.* 11, 1018–1023. doi: 10.1123/ijsp.2015-0613
- Andersson, E., Holmberg, H. C., Ørtenblad, N., and Björklund, G. (2016). Metabolic responses and pacing strategies during successive sprint skiing time trials. *Med. Sci. Sports Exerc.* 48, 2544–2554. doi: 10.1249/MSS.0000000000001037
- Anslley, L., Schabert, E., St Clair Gibson, A., Lambert, M. I., and Noakes, T. D. (2004). Regulation of pacing strategies during successive 4-km time trials. *Med. Sci. Sports Exerc.* 36, 1819–1825. doi: 10.1249/01.MSS.0000142409.70181.9D
- Beltrami, F. G., Froyd, C., Mauger, A. R., Metcalfe, A. J., Marino, F., and Noakes, T. D. (2012). Conventional testing methods produce submaximal values of maximum oxygen consumption. *Br. J. Sports Med.* 46, 23–29. doi: 10.1136/bjsports-2011-090306
- Bishop, D., Jenkins, D. G., and Mackinnon, L. T. (1998). The effect of stage duration on the calculation of peak $\dot{V}O_2$ during cycle ergometry. *J. Sci. Med. Sport* 1, 171–178. doi: 10.1016/S1440-2440(98)80012-1
- Bland, J. M., and Altman, D. G. (2007). Agreement between methods of measurement with multiple observations per individual. *J. Biopharm. Stat.* 17, 571–582. doi: 10.1080/10543400701329422
- Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scand. J. Work Environ. Health* 16, 55–58. doi: 10.5271/sjweh.1815
- Burnley, M., Doust, J. H., and Vanhatalo, A. (2006). A 3-min all-out test to determine peak oxygen uptake and the maximal steady state. *Med. Sci. Sports Exerc.* 38, 1995–2003. doi: 10.1249/01.mss.0000232024.06114.a6
- Chidnok, W., Dimenna, F. J., Bailey, S. J., Burnley, M., Wilkerson, D. P., Vanhatalo, A., et al. (2013). $\dot{V}O_2\max$ is not altered by self-pacing during incremental exercise. *Eur. J. Appl. Physiol.* 113, 529–539. doi: 10.1007/s00421-012-2478-6
- Cooke, C. B. (2009). “Maximal oxygen uptake, economy and efficiency,” in *Kinanthropometry and Exercise Physiology Laboratory Manual: Tests, Procedures and Data, 3rd Edn*, eds R. Eston and T. Reilly (London: Routledge), 174–212.
- Crouter, S., Foster, C., Esten, P., Brice, G., and Porcari, J. P. (2001). Comparison of incremental treadmill exercise and free range running. *Med. Sci. Sports Exerc.* 33, 644–647. doi: 10.1097/00005768-200104000-00020
- Currell, K., and Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Med.* 38, 297–316. doi: 10.2165/00007256-200838040-00003
- De Pauw, K., Roelands, B., Cheung, S. S., de Geus, B., Rietjens, G., and Meeusen, R. (2013). Guidelines to classify subject groups in sport-science research. *Int. J. Sports Physiol. Perform.* 8, 111–122. doi: 10.1123/ijsp.8.2.111
- Foster, C., Green, M. A., Snyder, A. C., and Thompson, N. N. (1993). Physiological responses during simulated competition. *Med. Sci. Sports Exerc.* 25, 877–882. doi: 10.1249/00005768-199307000-00018
- Foster, C., Kuffel, E., Bradley, N., Battista, R. A., Wright, G., Porcari, J. P., et al. (2007). $\dot{V}O_2\max$ during successive maximal efforts. *Eur. J. Appl. Physiol.* 102, 67–72. doi: 10.1007/s00421-007-0565-x
- Gore, C. J. (ed.). (2000). *Physiological Tests for Elite Athletes*. Champaign, IL: Human Kinetics.
- Hogg, J. S., Hopker, J. G., and Mauger, A. R. (2015). The self-paced $\dot{V}O_2\max$ test to assess maximal oxygen uptake in highly trained runners. *Int. J. Sports Physiol. Perform.* 10, 172–177. doi: 10.1123/ijsp.2014-0041
- Howley, E. T., Bassett, D. R. J., and Welch, H. G. (1995). Criteria for maximal oxygen uptake: review and commentary. *Med. Sci. Sports Exerc.* 27, 1292–1301. doi: 10.1249/00005768-199509000-00009
- Jeukendrup, A., Saris, W. H. M., Brouns, F., and Kester, A. D. M. (1996). A new validated endurance performance test. *Med. Sci. Sports Exerc.* 28, 266–270. doi: 10.1097/00005768-199602000-00017
- Jones, A. M. (2008). “Middle- and long-distance running,” in *Sport and Exercise Physiology Testing Guidelines*, eds E. M. Winter, A. M. Jones, R. C. Richard Davison, P. D. Bromley, and T. H. Mercer. (London: Routledge), 147–154.
- Kuipers, H., Rietjens, G., Verstappen, F., Schoenmakers, H., and Hofman, G. (2003). Effects of stage duration in incremental running tests on physiological variables. *Int. J. Sports Med.* 24, 486–491. doi: 10.1055/s-2003-42020
- Losnegard, T., Myklebust, H., and Hallén, J. (2012). Anaerobic capacity as a determinant of performance in sprint skiing. *Med. Sci. Sports Exerc.* 44, 673–681. doi: 10.1249/MSS.0b013e3182388684
- Mauger, A. R., Metcalfe, A. J., Taylor, L., and Castle, P. C. (2013). The efficacy of the self-paced $\dot{V}O_2\max$ test to measure maximal oxygen uptake in treadmill running. *Appl. Physiol. Nutr. Metab.* 38, 1211–1216. doi: 10.1139/apnm-2012-0384
- Mauger, A. R., and Sculthorpe, N. (2012). A new $\dot{V}O_2\max$ protocol allowing self-pacing in maximal incremental exercise. *Br. J. Sports Med.* 46, 59–63. doi: 10.1136/bjsports-2011-090006
- McGawley, K., and Holmberg, H. C. (2014). Aerobic and anaerobic contributions to energy production among junior male and female cross-country skiers during diagonal skiing. *Int. J. Sports Physiol. Perform.* 9, 32–40. doi: 10.1123/ijsp.2013-0239
- Midgley, A., Carroll, S., Marchant, D., McNaughton, L. R., and Siegler, J. (2009). Evaluation of true maximal oxygen uptake based on a novel set of standardized criteria. *Appl. Physiol. Nutr. Metab.* 34, 115–123. doi: 10.1139/H08-146
- Midgley, A., McNaughton, L., Remco, P., and Marchant, D. (2007). Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. *Sports Med.* 37, 1019–1028. doi: 10.2165/00007256-200737120-00002
- Midgley, A. W., and Carroll, S. (2009). Emergence of the verification phase procedure for confirming “true” $\dot{V}O_2\max$. *Scand. J. Med. Sci. Sports* 19, 313–322. doi: 10.1111/j.1600-0838.2009.00898.x
- Midgley, A. W., McNaughton, L. R., and Carroll, S. (2006). Verification phase as a useful tool in the determination of the maximal oxygen uptake of distance runners. *Appl. Physiol. Nutr. Metab.* 31, 541–548. doi: 10.1139/h06-023
- Noakes, T. D. (2008). How did A. V. Hill understand the $\dot{V}O_2\max$ and the “plateau phenomenon”? Still no clarity? *Br. J. Sports Med.* 42, 574–580. doi: 10.1136/bjism.2008.046771
- Poole, D. C., Wilkerson, D. P., and Jones, A. M. (2008). Validity of criteria for establishing maximal $\dot{V}O_2$ uptake during ramp exercise tests. *Eur. J. Appl. Physiol.* 102, 403–410. doi: 10.1007/s00421-007-0596-3
- Pringle, J. S., Carter, H., Doust, J. H., and Jones, A. M. (2002). Oxygen uptake kinetics during horizontal and uphill treadmill running in humans. *Eur. J. Appl. Physiol.* 88, 163–169. doi: 10.1007/s00421-002-0687-0
- Reilly, T., and Brooks, G. A. (1982). Investigation of circadian rhythms in metabolic responses to exercise. *Ergonomics* 25, 1093–1107. doi: 10.1080/00140138208925067

- Rollo, I., Williams, C., and Nevill, A. (2008). Repeatability of scores on a novel test of endurance running performance. *J. Sports Sci.* 26, 1379–1386. doi: 10.1080/02640410802277452
- Rossiter, H. B., Kowalchuk, J. M., and Whipp, B. J. (2006). A test to establish maximum O_2 uptake despite no plateau in the O_2 uptake response to ramp incremental exercise. *J. Appl. Physiol.* 100, 764–770. doi: 10.1152/jappphysiol.00932.2005
- Scheidler, C. M., and Devor, S. T. (2015). $\dot{V}O_2$ max measured with a self-selected work rate protocol on an automated treadmill. *Med. Sci. Sports Exerc.* 47, 2158–2165. doi: 10.1249/MSS.0000000000000647
- Sloniger, M. A., Cureton, K. J., Prior, B. M., and Evans, E. M. (1997). Anaerobic capacity and muscle activation during horizontal and uphill running. *J. Appl. Physiol.* 83, 262–269.
- Snoza, C. T., Berg, K. E., and Slivka, D. R. (2016). Comparison of $\dot{V}O_2$ peak and achievement of $\dot{V}O_2$ peak criteria in three modes of exercise in female triathletes. *J. Strength Cond. Res.* 30, 2816–2822. doi: 10.1519/JSC.0000000000000710
- Sperlich, B., Haegle, M., Thissen, A., Mester, J., and Holmberg, H. C. (2011). Are peak oxygen uptake and power output at maximal lactate steady state obtained from a 3-min all-out cycle test? *Int. J. Sports Med.* 32, 433–437. doi: 10.1055/s-0031-1271770
- Stocks, B., Betts, J. A., and McGawley, K. (2016). Effects of carbohydrate dose and frequency on metabolism, gastrointestinal discomfort, and cross-country skiing performance. *Scand. J. Med. Sci. Sports* 26, 1100–1108. doi: 10.1111/sms.12544
- Stöggel, T., Lindinger, S., and Müller, E. (2007). Analysis of a simulated sprint competition in classical cross country skiing. *Scand. J. Med. Sci. Sports* 17, 362–372. doi: 10.1111/j.1600-0838.2006.00589.x
- Swarén, M., Supej, M., Eriksson, A., and Holmberg, H. C. (2013). “Treadmill simulation of olympic cross-country ski tracks,” in *Science and Nordic Skiing II*, eds A. Hakkarainen, V. Linnamo, and S. J. Lindinger, (Oxford: Meyer and Meyer Verlag), 237–242.
- Watkins, J., Platt, S., Andersson, E., and McGawley, K. (2017). Pacing strategies and metabolic responses during 4-minute running time-trials. *Int. J. Sports Physiol. Perform.* 1–24. doi: 10.1123/ijsp.2016-0341. [Epub ahead of print].
- Williams, C. A., Ratel, S., and Armstrong, N. (2005). Achievement of peak $\dot{V}O_2$ during a 90-s maximal intensity cycle sprint in adolescents. *Can. J. Appl. Physiol.* 30, 157–171. doi: 10.1139/h05-112
- Winter, E. M., Jones, A. M., Davison, R. C. R., Bromley, P. D., and Mercer, T. H. (eds.). (2008). *Sport and Exercise Physiology Testing*. Oxon: Routledge.
- Yoon, B. K., Kravitz, L., and Robergs, R. (2007). $\dot{V}O_2$ max, protocol duration, and the $\dot{V}O_2$ plateau. *Med. Sci. Sports Exerc.* 39, 1186–1192. doi: 10.1249/mss.0b13e318054e304

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 McGawley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.