SESNZ
SPORT & EXERCISE SCIENCE
NEW ZEALAND

The Journal of Sport and Exercise Science, Vol. 4, Issue 1, 1-10 (2020)

www.sesnz.org.nz

JSES
ISSN: 2703-240X

Lab and field $\dot{V}O2peak$ testing in highly trained cyclists

Phillip A. Smith[1, 2 *], Nicolas J. A. Berger[1]

[1]*School of Health & Life Sciences, Teesside University, United Kingdom*
[2]*Department of Sport & Exercise Sciences, Durham University, United Kingdom*

ARTICLE INFO

Received: 16.07.2019
Accepted: 18.11.2019
Online: 23.01.2020

Keywords:
Aerobic Endurance
Power Output
Performance

ABSTRACT

The issues with traditional maximal oxygen uptake ($\dot{V}O_{2max}$) testing include an inability to regulate intensity due to fixed resistance and a lack of conscious decision making during the test (Noakes, 2008). Depending on the test and conditions, some athletes do not reach $\dot{V}O_{2max}$ despite reaching volitional exhaustion, and in this case, the result is recorded as the highest, or peak oxygen uptake attained in this test, known as $\dot{V}O_{2peak}$. To investigate this, a study was conducted to determine if a field-based test would result in a higher $\dot{V}O_{2peak}$ value than a lab-based test. Twelve highly trained cyclists performed a 20w/minute ramp test on a cycle ergometer and a 3.2km hill climb on their own racing bike wearing a portable gas analyser (MetaMax 3b, Cortex GmbH, Leipzig, Germany). A paired t-test revealed that the hill climb resulted in a higher but not statistically significant absolute $\dot{V}O_{2peak}$: lab 5.49 ± 0.8 L·min^{-1} vs. field 5.59 ± 0.7 L·min^{-1}, $p = .189$ and relative $\dot{V}O_{2peak}$: lab 71.9 ± 10.0 ml·kg^{-1}·min^{-1} vs. field 74.0 ± 9.9 ml·kg^{-1}·min^{-1}, $p = .060$. Additionally, field testing resulted in a significantly higher RER_{max}: lab 1.07 ± 0.0 vs. field 1.16 ± 0.1, $p = .019$, end lactate: lab 9.24 ± 1.6 mmol·L^{-1} vs. field 11.99 ± 2.3 mmol·L^{-1}, $p = .039$, and 5-minute-post lactate: lab 7.56 ± 1.4 mmol·L^{-1} vs. field 11.87 ± 2.0 mmol·L^{-1}, $p < 0.001$. There was no difference in HR_{max} between tests: lab 187.9 ± 11.6 b·min^{-1} vs. field 187.6 ± 10.6 b·min^{-1}, $p = .952$. Slightly higher $\dot{V}O_{2peak}$ values recorded during the field test may be explained by the closed-loop format allowing riders to pace their effort better, the cooling effect of the wind outdoors, freedom to ride out-the-saddle (leading to greater muscle recruitment), or perhaps the sub-optimal length of the lab test 20.4 ± 3.0 mins vs 8.4 ± 1.2 mins field test. Findings suggest the increased ecological validity of field testing led to higher (but not statistically significant) VO_{2peak} values and can be considered a viable alternative to lab-based testing if a climb with suitable length and gradient is available.

1. Introduction

Maximal oxygen uptake ($\dot{V}O_{2max}$) is viewed as the gold standard measure for cardiorespiratory fitness (Williams *et al*., 2017), aerobic endurance (Bassett & Howley, 2000), and forms a key predictor of overall performance in endurance sports (McLaughlin *et al*., 2010).

Traditional $\dot{V}O_{2max}$ testing consists of an incremental increase in exercise intensity (Poole *et al*., 2008) until the participant reaches volitional exhaustion. This increase may be in the form of a constant ramp or longer steps of 2-5 minutes, which allow participants to reach a steady state of O_2 consumption. Tests are usually designed to last around 8-12 minutes as longer tests were found to result in lower $\dot{V}O_{2max}$ values in trained males (Yoon *et al*., 2007). This is likely due to premature local muscular fatigue before the maximum capacity of the cardiovascular system is reached (Buchfuhrer *et al*., 1983; Yoon *et al*., 2007).

Noakes (2008) describes further issues with the $\dot{V}O_{2max}$ testing process that affect the ecological validity and outcome; as the athlete is not aware of the endpoint, there is an open-loop scenario which leads to an inability to regulate intensity.

The fixed and progressive method of increasing pedalling resistance is unlike anything experienced while cycling outdoors, limiting the role of decision-making and conscious pacing control during the test. All an athlete is able to decide is when to terminate the test: maximum volitional exhaustion.

*Corresponding Author: Phillip A. Smith, School of Health & Life Sciences, Teesside University, UK, phillip.smith@tees.ac.uk

Recent studies have attempted to follow Noakes (2008) suggestions for a maximal test, which considers the role of the brain in exercise; for example, using rate of perceived exertion (RPE). RPE-clamped protocols that use a fixed length test of 10 minutes have been used. These are made up of 5x2-minute stages in which participants were instructed to target a specific incremental RPE value (11, 13, 15, 17, 20, Borg 6-20). This protocol was found to result in significantly higher $\dot{V}O_{2max}$ values than a traditional step test in untrained participants (RPE 40 ± 10 ml·kg^{-1}·min^{-1} vs Ramp 37 ± 8 ml·kg^{-1}·min^{-1}) (Mauger & Sculthorpe, 2012). In contrast, the same protocol in trained cyclists did not result in a significant difference (Ramp 3.86 ± 0.73 L·min^{-1} vs 3.87 ± 0.72 L·min^{-1} in the RPE-clamped) (Straub *et al*., 2014). It is notable, however, that trained cyclists did significantly better on the test format they favoured. Participants were divided between those who preferred not having to consciously regulate the intensity and those who preferred control over their pacing.

While RPE-clamped protocols improve ecological validity, as they allow conscious intensity regulation, this is still limited as ergometer cycling is biomechanically and physiologically different to riding outdoors due to differences in inertial load and muscle activation patterns (Fregly, Zajac & Dairaghi, 2000; Bertucci, Grappe & Groslambert, 2007). Outdoor cycling can feature greater total muscle activation when riding out the saddle (Ryschon & Stray-Gundersen, 1991; Hansen & Waldeland, 2008), and a cooling effect from the wind (Brito *et al*., 2017). Additionally, a known endpoint of exercise allows conscious control of pacing, which Noakes (2008) suggests may lead to greater motivation and ability to push harder; e.g. when athletes are capable of a final sprint to the line after a hard race.

Meyer *et al*. (2003) conducted a study in trained runners comparing a treadmill-based ramp protocol with an identical protocol performed on a running track (paced by a light system). While this protocol increased ecological validity by taking runners off the treadmill, and perhaps increasing the role of conscious pacing by asking them to match their running speed to light cues, the participants were not self-paced to the same extent as those in Mauger and Sculthorpe (2012) and Straub *et al*. (2014) were. The results found no significant difference in $\dot{V}O_{2max}$ between tests (lab 4.65 ± 0.51 L·min^{-1}, field 4.63 ± 0.55 L·min^{-1}, p = .71). HR_{max} was reported as significantly higher in the field (lab 188 ± 6 b·min^{-1}, field 189 ± 6 b·min^{-1}, p = .02). Finally, test duration was significantly longer in the field (lab 691 ± 39 seconds, field 727 ± 42 seconds, p < .001). This 5% increase in test duration, and therefore performance, was put down to greater running economy on the track leading to lower $\dot{V}O_2$ throughout.

Ricci and Leger (1983) performed a study examining the difference in $\dot{V}O_{2max}$ between cyclists riding on an ergometer, on a velodrome and on a treadmill. This study has several limitations, such as the type of participants (7 male, 1 female), the age of the participants (13-40 years), as well as the equipment and method for calculating $\dot{V}O_{2max}$ used (backwards extrapolation), especially during the velodrome test. In comparison, Meyer *et al*. (2003) used a MetaMax portable, breath-by-breath gas analyser. Ricci and Leger (1983) found a significantly higher $\dot{V}O_{2max}$ during ergometer testing compared to both treadmill and velodrome tests (ergometer 62.4 ± 8.2 ml·kg^{-1}·min^{-1}, treadmill 54.7 ± 6.3 ml·kg^{-1}·min^{-1}, velodrome 53.0 ± 7.8 ml·kg^{-1}·min^{-1}). Ricci and Leger (1983) struggled to explain the ~15% higher $\dot{V}O_{2max}$ during the ergometer test, but suggested cadence, fibre recruitment or mechanical efficiency may play a role.

While Bassett and Howley (2000) define $\dot{V}O_{2max}$ as the maximum amount of O_2 that can be taken in and utilised by the body during severe exercise, it is difficult to ascertain whether the value achieved during a test truly represents an athlete's $\dot{V}O_{2max}$. Hill and Lupton (1923) noted that past a certain running pace O_2 consumption ceased to rise with the increased workload. This plateau, defined by BASES (1997) as an increase of < 150 ml·min^{-1} or 2 ml·kg^{-1}·min^{-1} is often used to signify that an athlete has reached $\dot{V}O_{2max}$, although studies have found that this phenomenon can appear in 0-100% of tests (Midgley & Carroll, 2009) and at as low as 61% (Midgley *et al*., 2009) and 73% (Poole *et al*., 2008) of $\dot{V}O_{2max}$. Because of this, secondary criteria are used to help determine $\dot{V}O_{2max}$ attainment. BASES (1997) use 5-minute-post blood lactate (BLac) ≥ 8.0 mmol·L^{-1}, heart rate (HR) ≥ 10 beats of age predicted max (220-age), respiratory exchange ratio (RER) ≥ 1.15, along with subjective fatigue and volitional exhaustion. Other studies may be less strict with lower values of RPE ≥ 17-19 or RER ≥ 1.05-1.1 permitted, which may be influenced by the mode of exercise.

Some criteria have been found to be achieved at a submaximal workloads, for example RER ≥ 1.1 can be satisfied 27% below $\dot{V}O_{2max}$ and ≥ 1.15 at 16% below $\dot{V}O_{2max}$. (Poole, Wilkerson & Jones, 2008). While other criteria may be too rigorous for participants to achieve, as Poole *et al*. (2008) found that heart rate ≥ 10 b·min^{-1} of age predicted max led to the rejection of 3/8 participants' tests and BLac ≥ 8.0 mmol·L rejected 6/8 participants tests. Due to these uncertainties in determining $\dot{V}O_{2max}$ attainment, we prefer the term $\dot{V}O_{2peak}$ and report the highest, repeated values participants reached over a 30 second period.

Due to the issues described with traditional laboratory-based testing, we sought to determine if a real-life cycling event with (approximately) the optimal length and intensity of a $\dot{V}O_{2peak}$ test would be comparable to that of a traditional lab-based test. The course was chosen specifically because it hosts an annual hill climb race (our route was extended slightly, from 2.5km to 3.23km, to result in a duration of 8-12 minutes) (Buchfuhrer *et al*., 1983; Yoon *et al*., 2007), and featured a gradient that got progressively steeper towards the summit, with the intention of forcing an increase in participants' power output similar to a lab test.

It was hypothesised that due to the greater conscious control of pacing, closed-loop format with a known endpoint, and greater muscle recruitment (Ryschon & Stray-Gundersen, 1991, Hansen & Waldeland, 2008), $\dot{V}O_{2peak}$ would be significantly higher in field-based testing than lab based testing; both measured with participants wearing a portable breath-by-breath gas analyser.

2. Methods

2.1. Study design

To test the hypothesis that field testing would lead to a higher $\dot{V}O_{2peak}$ compared to lab testing a randomised, counterbalance study was conducted. Differences in $\dot{V}O_{2peak}$, maximal heart rate (HR_{max}), maximal respiratory exchange ratio (RER_{max}) and peak

BLac concentrations were compared between the lab and field tests.

2.2. Participants

Highly trained, competitive cyclists, with over two years racing experience, from the north east of England were recruited to complete a lab and field test. The study was approved by the Teesside University ethics committee and all participants gave written informed consent prior to testing. 12 participants undertook the lab and field tests. Mean ± SD age 28.4 ± 12 years, height 182.8 ± 7 cm, (lab) mass 76.99 ± 10.9 kg.

2.3. Procedures

Participants were randomly allocated to complete either the lab or field test first. They were instructed to avoid strenuous exercise for 24 hours prior to testing and a minimum of 24 hours was left between tests, which were conducted at the same time of day to minimize diurnal variations in performance. All participants were familiar with the course of the field-based test (hill climb), while half had previously completed a lab-based $\dot{V}O_{2peak}$ test.

2.4. Procedures for lab test

Participants rested for at least 5 minutes before BLac (YSI 2300 Yellow Springs, OH), height (Seca stadiometer, Birmingham, UK) and body mass (Seca 869, Birmingham, UK) were measured before commencing the 20w/min ramp test on a cycle ergometer (Lode Excalibur Sport, Groningen, The Netherlands) set up to match their road bike position, using their own pedals and shoes, and wearing bib shorts and either a vest or no top. All testing was completed in a well-ventilated laboratory at a temperature of 20°C. The use of a fan was not permitted to minimise any cooling effect associated with riding outdoors, and as it has been shown that the use of a fan can increase maximal oxygen uptake (Brito *et al*., 2017). Participants were instructed to ride at their normal cadence throughout and the test was terminated when they could not maintain a cadence ≥70 $rev{\cdot}min^{-1}$. A warm-up was not conducted prior to the lab test as it started from 0 w resistance therefore it was not until 10 minutes into the test that the participants reached 200 w (around the warm-up intensity for the field test).

At the point of failure peak power, HR (Polar H7, Polar Electro Oy, Kempele, Finland) and BLac were recorded, and a further BLac sample after 5 minutes of active recovery (cycling at 100 w) was taken. Breath-by-breath data was analysed using Microsoft Excel (Excel 2016, Microsoft, Redmond, WA, USA) and MetaSoft Studio (Cortex GmbH, Leipzig, Germany) to determine both absolute and relative $\dot{V}O_{2peak}$ and RER.

2.5. Procedures for field test

Body mass and resting BLac (Lactate Pro, Arkray KDK, Japan) were measured prior to participants commencing a standardised 10-minute warm-up at 55-60% of their current, self-reported functional threshold power (FTP). This was conducted using a turbo trainer and participants own bike and power meter, displayed in table 1. Following the warm-up, power meters were calibrated according to the manufacturer's instructions and participants were fitted with the portable gas analyser. The Lactate Pro used for field testing was found to be an accurate measure of BLac (Bonaventura *et al*., 2015). Pyne *et al*. (2000) reported a near-perfect correlation (r = .99) between the Lactate Pro and YSI 2300 (lab test analyser).

Table 1: Power meters used by participants

Power Meter	Number	Measurement Location	Manufacturer claimed accuracy
4iii Precision 2nd gen	1	Left crank	1%
Favero BePro	2	Pedals	2%
Quarq DZero	2	Crank spider	1.5%
Quarq Riken	1	Crank spider	1.5%
Stages Ultegra 2nd gen	1	Left crank	2%
Powertap P1	1	Pedals	1.5%
Rotor INpower	1	Left side, axle based	1%
SRM Dura Ace 9000	1	Crank spider	1%
TeamZwatt Zimanox	1	Left crank	No reported data

*One participant did not have a power meter and bottom bracket compatibility did not allow them to borrow an available left crank-based unit (4iiii Precision).

Participants were instructed to reach the summit of the 3.2 km, 173 m elevation gain, 5% average gradient route (Figure 3) as quickly as possible while recording HR (Polar H7, Polar Electro Oy, Kempele, Finland), power data and time on their personal cycle computer and wearing the portable gas analyser (Cortex MetaMax 3B, Cortex GmbH, Leipzig, Germany) to record expired gasses breath-by-breath. The system weighed 1.4 kg (Mcfarlane & Wong, 2011) and was worn in the same manner as during the lab test. Participants wore bib shorts and a cycling top, used the same shoes and pedals as during the lab test, but were also required to wear a helmet for the field test, which was not worn during lab testing. Environmental conditions stayed relatively stable over the testing period, with temperatures ranging from 15-23°C and with low wind speeds.

A support car followed each participant to monitor the ride and provide protection from upcoming traffic. Upon completion, BLac was sampled immediately afterward and 5 minutes post-test, following active recovery at a self-selected power. Participants power output, HR (Polar H7), speed and time data was downloaded from their cycle computer for analysis.

2.6. Statistical analyses

Field test cycling data were analysed using Training Peaks (Peaksware, Boulder, CO, USA) to determine average and HRmax and average power.

V̇O2peak was determined using Microsoft Excel (2016, Redmond, WA, USA) scatter graph function to determine the highest individual V̇O2 recorded over a 30 second period. This excluded any outlying breaths, and the highest value had to be agreed on by both authors, independent of each other. If there were any discrepancies a third person would be consulted, although this was not required.

SPSS for Windows (V25; IBM, Armonk, NY, USA) was used for a paired *t*-test to determine significant differences and confidence intervals for HRmax, RERmax, end BLac, 5-minutes-post BLac, absolute V̇O2peak, relative V̇O2peak and test duration between lab and field tests. The alpha level of significance was set at $p < 0.05$. 95% Confidence Intervals (C.I.) were calculated using a customised spreadsheet (Hopkins, 2006). Effect Size thresholds for Hedge's G are: 0.2=> small effect, 0.5=> medium effect, and 0.8=> large effect (Cohen, 2013).

3. Results

All twelve participants completed both the lab and field tests.

Field-based testing resulted in a higher value for all variables measured with the exception of HR_{max}, which was less than half a beat per minute higher in the lab, and body mass, which was half a kilo heavier in the lab. Of these results, statistically significant findings were reported for end BLac, 5-minute-post BLac and RER_{max} (Table 2). Despite a field test increase in absolute and relative $\dot{V}O_{2peak}$ of 100 ml·min^{-1} and 2.14 ml·kg^{-1}·min^{-1} respectively, they did not reach statistical significance.

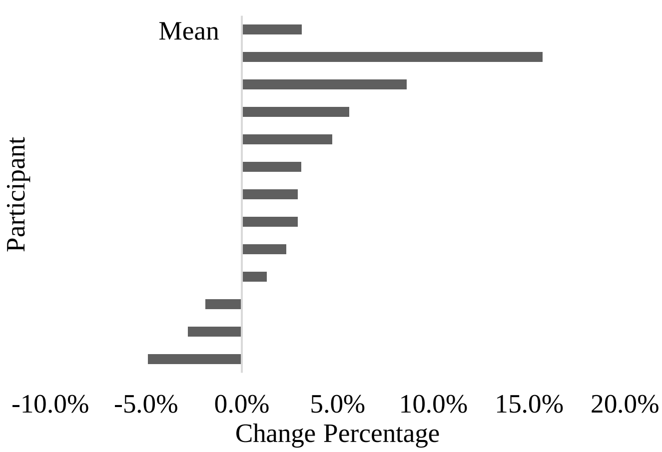

Figure 1: Individual changes in relative $\dot{V}O_{2peak}$ from the lab test to the field test

7/12 participants displayed a greater absolute $\dot{V}O_{2peak}$ in the field. Mean $\dot{V}O_{2peak}$ was 2.33% higher in the field than lab.

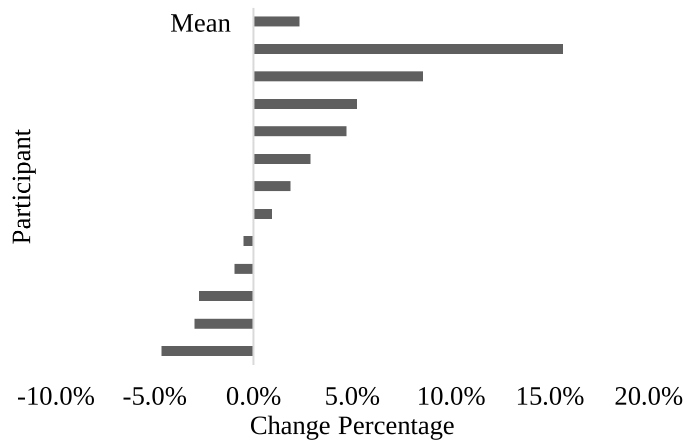

Figure 2: Individual changes in absolute $\dot{V}O_{2peak}$ from the lab test to the field test

Due to lower body mass recorded in the field 76.5 ± 11.7 kg compared to the lab 76.99 ± 10.9 kg, relative $\dot{V}O_{2peak}$ (figure 1) is higher for more participants than absolute $\dot{V}O_{2peak}$ (figure 2). Here 9/12 report a greater field $\dot{V}O_{2peak}$. Mean field $\dot{V}O_{2peak}$ was 3.13% higher than in the lab.

Figure 3 displays participants displayed a high peak power value at the start of the test as they accelerated up to speed. Peak power (1 second) was 774 ± 168 w. Power dropped over the next 500 m as they maintained speed on the flatter parts of the course, as gradient increased power output did too, with the exception of the penultimate 500 m where power output dropped, possibly due to fatigue and an unsustainable pacing strategy. As expected, power output increased over the last 230 m. Average power output sustained during the field test was 393 ± 50 w, which was 89.9% of lab test peak power (422 ± 60 w), defined as the power output achieved at the point of failure.

Figure 4 displays lab test power output, which was fixed at a linear increase of 20 w/minute. The highest peak power output was 503 w, the lowest 328 w and the mean was 422 ± 60 w.

Table 2: Participant results

n=12	Lab Test Mean ± SD	Lab Test 95% CI	Field Test Mean ± SD	Field Test 95% CI	Percentage Difference	Significance	Effect Size (Hedges' G)
Body Mass (kg)	76.99 ± 10.9	67.03 – 85.68	76.48 ± 11.7	65.69 – 85.69	-0.66	.135	0.04
HR_{max} ($b \cdot min^{-1}$)	187.91 ± 11.6	185 – 196.14	187.55 ± 10.6	186.14 – 195.89	-0.19	.952	0.03
End BLac ($mmol \cdot L^{-1}$)	9.24 ± 1.6	8.36 – 10.75	11.99 ± 2.3	11.3 – 13.29	29.76	.039*	1.43
5-Minute-Post BLac ($mmol \cdot L^{-1}$)	7.56 ± 1.4	6.47 – 8.72	11.87 ± 2.0	10.09 – 12.5	57.01	.000*	2.57
RER_{peak}	1.07 ± 0.0	1.04 – 1.09	1.16 ± 0.1	1.10 – 1.28	8.41	.019*	1.22
Absolute $\dot{V}O_{2peak}$ ($L \cdot min^{-1}$)	5.49 ± 0.8	4.87 – 6.04	5.59 ± 0.7	5.08 – 6.04	1.82	.189	0.13
Relative $\dot{V}O_{2peak}$ ($ml \cdot kg^{-1} \cdot min^{-1}$)	71.90 ± 10.0	65.45 – 80.93	74.04 ± 9.9	66.95 – 82.97	2.98	.060	0.21
Test Duration (s)	1266.75 ± 178.8	1156.71 – 1422.43	506.17 ± 69.1	438.6 – 531.71	-60.03	.000*	5.44

* Denotes significant difference ($p < 0.05$)

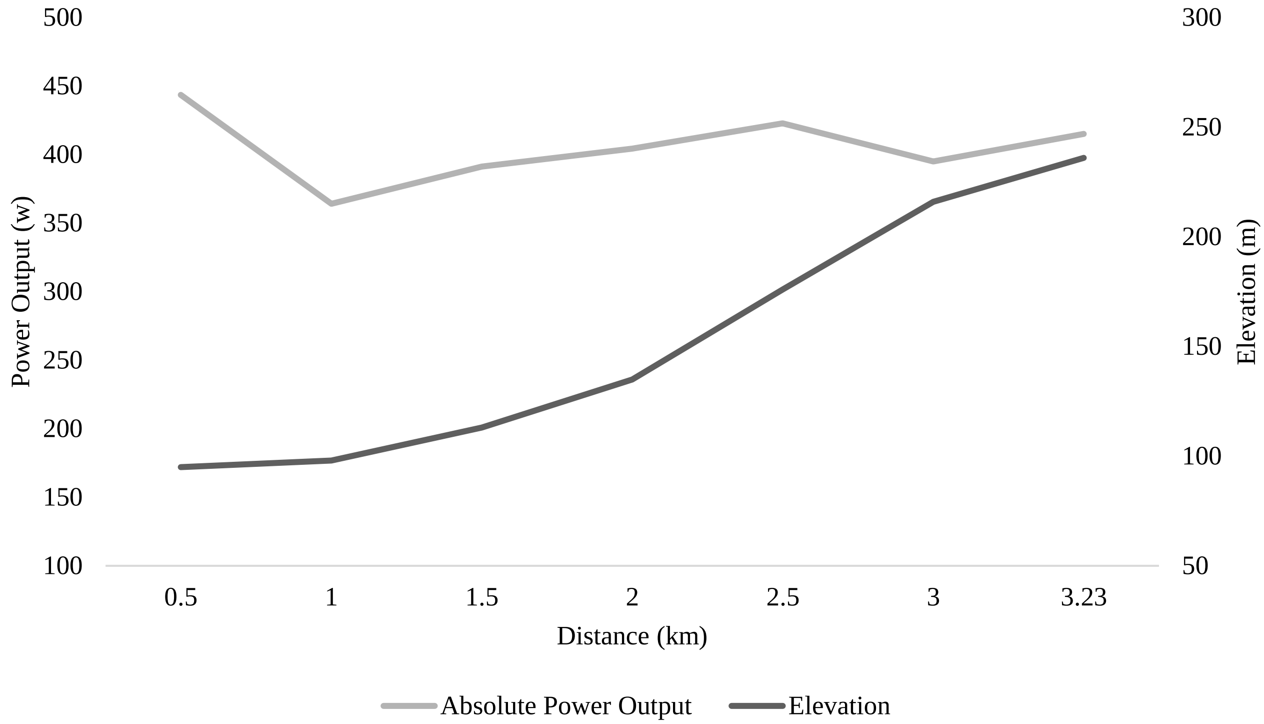

Figure 3: Mean power output in 500 m intervals during the field test

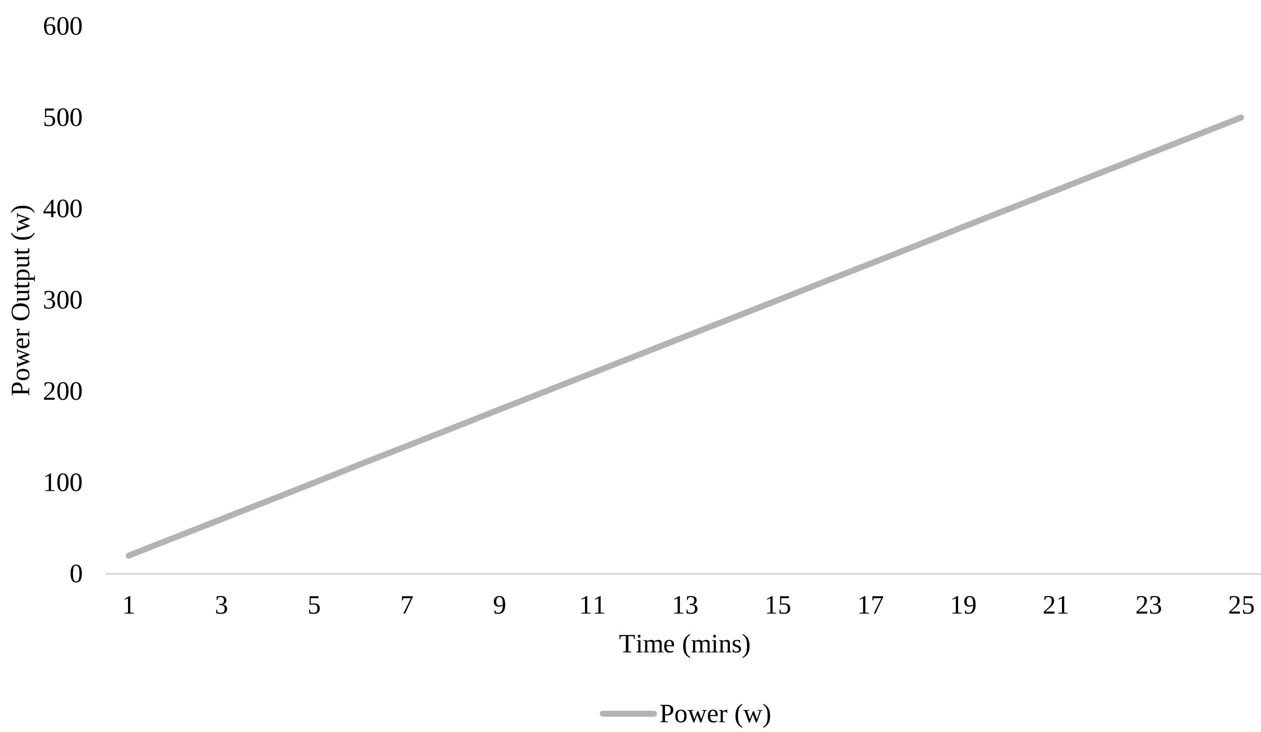

Figure 4: Lab test power output

4. Discussion

The aim of this study was to determine if there was a difference between $\dot{V}O_{2peak}$ measured through a conventional laboratory-based ramp test and a field test over a 3.2 km hill climb. It was hypothesised that the field test would lead to a greater $\dot{V}O_{2peak}$ as it accurately represents the real world, maximal effort cycling conditions the participants are used to. In addition, the field-based test has a set endpoint creating a closed-loop scenario which allows conscious control of pacing, thus increasing ecological validity; something lab-based tests lack (Noakes, 2008).

This study discovered that field-based testing resulted in a higher absolute and relative $\dot{V}O_{2peak}$ compared to the lab test, although this did not reach statistical significance, possibly due to the mixed responses to the field-based testing and the small sample size. This finding is similar to that of Straub *et al.* (2014) who did not find a difference in $\dot{V}O_{2max}$ between traditional $\dot{V}O_{2max}$ testing and a self-paced test in trained cyclists.

To our knowledge, this study was the first to use an actual competitive event of optimal duration to field-test cyclists $\dot{V}O_{2peak}$, free from any prescribed intensity regulation guidelines. A route with increasing gradient throughout was predicted to cause an increase in power output as the test progressed, although power data analysis revealed that all participants displayed steady pacing throughout. Despite the steady intensity, field testing still resulted in a higher (but not statistically significant) $\dot{V}O_{2peak}$.

The cooling effect of wind outdoors may have influenced performance. The ramp test was conducted in an air-conditioned laboratory without a fan for participants. Previous research showed that a 10km·h^{-1} airflow led to a lower $\dot{V}O_2$ at all stages of a maximal test, except for the last stage where maximal oxygen uptake was higher with a fan (Brito *et al*., 2017). This effect may be more significant outdoors, as during this test participants' average speed was 22.6 km·h^{-1}, resulting in increased airflow. Although it is important to point out that Brito *et al.* (2017) used a lower threshold to determine $\dot{V}O_{2max}$ attainment. Only one out of three criteria (O_2 plateau, RER ≥ 1.15 or HR ≥ 10 b·min^{-1} age predicted max) had to be satisfied, which has previously found to be met at an intensity as low as 61% $\dot{V}O_{2max}$ (Midgley *et al*., 2009), reducing the validity of their findings.

Despite this, previous research has found testing methods with higher ecological validity result in a lower $\dot{V}O_{2max}$. This was also the case with Ricci and Leger (1983), who found that a velodrome-based test resulted in a significantly lower $\dot{V}O_{2max}$ compared to cycling treadmill and ergometer tests. They did however not control for cadence, which resulted in significant differences between velodrome and ergometer (100 rev·min^{-1} vs 60 rev·min^{-1}). Moore *et al*. (2008) showed that O_2 consumption at 100 rev·min^{-1} was significantly higher than at 80 rev.min^{-1}. While the opposite effect was shown with Ricci and Leger (1983), it appears cadence causes variations in oxygen consumption. This study allowed participants to ride at a self-selected cadence for both lab and field tests which should have ensured cadence was matched during both tests, although it was not recorded during the lab test. Gearing did not limit participants cadence on this test (88.1 ± 10.5 rev·min^{-1}), although it may be an issue on longer and/or steeper climbs.

Biomechanical differences may explain changes in $\dot{V}O_{2peak}$ between lab and field tests. Cycle ergometers have a lower inertial load compared to road cycling (Fregley *et al*., 2000), which increases the torque production required at the top and bottom of the pedal stroke (Bertucci *et al*., 2007), resulting in a higher RPE and likely changes in muscle activation patterns. This is likened to riding in an extremely strong headwind in a very low gear (Fregley *et al*., 2000). In addition, the field test permitted out-the-saddle cycling, which has been found to result in a significantly higher O_2 consumption than seated pedaling. (Ryschon & Stray-Gunderson, 1991) Maximal oxygen uptake ($\dot{V}O_{2max)}$ is viewed as the gold standard measure for cardiorespiratory fitness (Williams *et al*., 2017), aerobic endurance (Bassett & Howley, 2000), and forms a key predictor of overall performance in endurance sports (McLaughlin *et al*., 2010).

Traditional $\dot{V}O_{2max}$ testing consists of an incremental increase in exercise intensity (Poole *et al*., 2008) until the participant reaches volitional exhaustion. This increase may be in the form of a constant ramp or longer steps of 2-5 minutes, which allow participants to reach a steady state of O_2 consumption. Tests are usually designed to last around 8-12 minutes as longer tests were found to result in lower $\dot{V}O_{2max}$ values in trained males (Yoon *et al*., 2007). This is likely due to premature local muscular fatigue before the maximum capacity of the cardiovascular system is reached (Buchfuhrer *et al*., 1983; Yoon *et al*., 2007).

Noakes (2008) describes further issues with the $\dot{V}O_{2max}$ testing process that affect the ecological validity and outcome; as the athlete is not aware of the endpoint, there is an open-loop scenario which leads to an inability to regulate intensity.

The fixed and progressive method of increasing pedaling resistance is unlike anything experienced while cycling outdoors, limiting the role of decision-making and conscious pacing control during the test. All an athlete is able to decide is when to terminate the test: maximum volitional exhaustion.

Recent studies have attempted to follow Noakes (2008) suggestions for a maximal test, which considers the role of the brain in exercise; for example, using rate of perceived exertion (RPE). RPE-clamped protocols that use a fixed length test of 10 minutes have been used. These are made up of 5x2-minute stages in which participants were instructed to target a specific incremental RPE value (11, 13, 15, 17, 20, Borg 6-20). This protocol was found to result in significantly higher $\dot{V}O_{2max}$ values than a traditional step test in untrained participants (RPE 40 ± 10 ml·kg^{-1}·min^{-1} vs Ramp 37 ± 8 ml·kg^{-1}·min^{-1}) (Mauger & Sculthorpe, 2012). In contrast, the same protocol in trained cyclists did not result in a significant difference (Ramp 3.86 ± 0.73 L·min^{-1} vs 3.87 ± 0.72 L·min^{-1} in the RPE-clamped) (Straub *et al*., 2014). It is notable, however, that trained cyclists did significantly better on the test format they favored. Participants were divided between those who preferred not having to consciously regulate the intensity and those who preferred control over their pacing.

While RPE-clamped protocols improve ecological validity, as they allow conscious intensity regulation, this is still limited as ergometer cycling is biomechanically and physiologically different to riding outdoors due to differences in inertial load and muscle activation patterns (Fregly *et al*., 2000; Bertucci *et al*., 2007). Outdoor cycling can feature greater total muscle activation when riding out the saddle (Ryschon & Stray-Gundersen, 1991; Hansen & Waldeland, 2008), and a cooling effect from the wind (Brito *et al*., 2017). Additionally, a known endpoint of exercise allows conscious control of pacing, which Noakes (2008) suggests may lead to greater motivation and ability to push harder;

e.g. when athletes are capable of a final sprint to the line after a hard race.

Meyer *et al*. (2003) conducted a study in trained runners comparing a treadmill-based ramp protocol with an identical protocol performed on a running track (paced by a light system). While this protocol increased ecological validity by taking runners off the treadmill, and perhaps increasing the role of conscious pacing by asking them to match their running speed to light cues, the participants were not self-paced to the same extent as those in Mauger and Sculthorpe (2012) and Straub *et al*. (2014) were. The results found no significant difference in $\dot{V}O_{2max}$ between tests (lab 4.65 ± 0.51 $L \cdot min^{-1}$, field 4.63 ± 0.55 $L \cdot min^{-1}$, $p = .71$). HR_{max} was reported as significantly higher in the field (lab 188 ± 6 $b \cdot min^{-1}$, field 189 ± 6 $b \cdot min^{-1}$, $p = .02$). Finally, test duration was significantly longer in the field (lab 691 ± 39 seconds, field 727 ± 42 seconds, $p < .001$). This 5% increase in test duration, and therefore performance, was put down to greater running economy on the track leading to lower $\dot{V}O_2$ throughout.

Ricci and Leger (1983) performed a study examining the difference in $\dot{V}O_{2max}$ between cyclists riding on an ergometer, on a velodrome and on a treadmill. This study has several limitations, such as the type of participants (7 male, 1 female), the age of the participants (13-40 years), as well as the equipment and method for calculating $\dot{V}O_{2max}$ used (backwards extrapolation), especially during the velodrome test. In comparison, Meyer *et al*. (2003) used a MetaMax portable, breath-by-breath gas analyser. Ricci and Leger (1983) found a significantly higher $\dot{V}O_{2max}$ during ergometer testing compared to both treadmill and velodrome tests (ergometer 62.4 ± 8.2 $ml \cdot kg^{-1} \cdot min^{-1}$, treadmill 54.7 ± 6.3 $ml \cdot kg^{-1} \cdot min^{-1}$, velodrome 53.0 ± 7.8 $ml \cdot kg^{-1} \cdot min^{-1}$). Ricci and Leger (1983) struggled to explain the ~15% higher $\dot{V}O_{2max}$ during the ergometer test, but suggested cadence, fibre recruitment or mechanical efficiency may play a role.

While Bassett and Howley (2000) define $\dot{V}O_{2max}$ as the maximum amount of O_2 that can be taken in and utilised by the body during severe exercise, it is difficult to ascertain whether the value achieved during a test truly represents an athlete's $\dot{V}O_{2max}$. Hill and Lupton (1923) noted that past a certain running pace O_2 consumption ceased to rise with the increased workload. This plateau, defined by BASES (1997) as an increase of < 150 $ml \cdot min^{-1}$ or 2 $ml \cdot kg^{-1} \cdot min^{-1}$ is often used to signify that an athlete has reached $\dot{V}O_{2max}$, although studies have found that this phenomenon can appear in 0-100% of tests (Midgley & Carroll, 2009) and at as low as 61% (Midgley *et al*., 2009) and 73% (Poole *et al*., 2008) of $\dot{V}O_{2max}$. Because of this, secondary criteria are used to help determine $\dot{V}O_{2max}$ attainment. BASES (1997) use 5-minute-post blood lactate (BLac) ≥ 8.0 $mmol \cdot L^{-1}$, heart rate (HR) ≥ 10 beats of age predicted max (220-age), respiratory exchange ratio (RER) ≥ 1.15, along with subjective fatigue and volitional exhaustion. Other studies may be less strict with lower values of RPE ≥ 17-19 or RER ≥ 1.05-1.1 permitted, which may be influenced by the mode of exercise.

Some criteria have been found to be achieved at a submaximal workload, for example RER ≥ 1.1 can be satisfied 27% below $\dot{V}O_{2max}$ and ≥ 1.15 at 16% below $\dot{V}O_{2max}$. (Poole *et al*., 2008). While other criteria may be too rigorous for participants to achieve, as Poole *et al*. (2008) found that heart rate ≥ 10 $b \cdot min^{-1}$ of age predicted max led to the rejection of 3/8 participants' tests and BLac ≥ 8.0 mmol·L rejected 6/8 participants tests. Due to these uncertainties in determining $\dot{V}O_{2max}$ attainment, we prefer the term $\dot{V}O_{2peak}$ and report the highest, repeated values participants reached over a 30 second period.

Due to the issues described with traditional laboratory-based testing, we sought to determine if a real-life cycling event with (approximately) the optimal length and intensity of a $\dot{V}O_{2peak}$ test would be comparable to that of a traditional lab-based test. The course was chosen specifically because it hosts an annual hill climb race (our route was extended slightly, from 2.5km to 3.23km, to result in a duration of 8-12 minutes) (Buchfuhrer *et al*., 1983; Yoon *et al*., 2007), and featured a gradient that got progressively steeper towards the summit, with the intention of forcing an increase in participants' power output similar to a lab test.

It was hypothesised that due to the greater conscious control of pacing, closed-loop format with a known endpoint, and greater muscle recruitment (Ryschon & Stray-Gundersen, 1991, Hansen & Waldeland, 2008), $\dot{V}O_{2peak}$ would be significantly higher in field-based testing than lab based testing; both measured with participants wearing a portable breath-by-breath gas analyser (Ryschon & Stray-Gundersen, 1991; Hansen & Waldeland, 2008).

This study showed some, but not all, secondary $\dot{V}O_{2max}$ determination criteria to be higher in the field; e.g. HR_{max} was only half a beat higher in the lab. End BLac, 5-minute-post BLac and RER_{max} were all significantly higher in the field. It is surprising that HR_{max} was not higher in the field given the hypothesis participants would increase their intensity in a final sprint to a known endpoint, especially since the higher end BLac suggests they finished the hill climb at a higher intensity, or spent a longer period above lactate threshold than during the lab test. This suggests if a higher HR was not responsible for the greater $\dot{V}O_{2peak}$ in the field, other physiological and biomechanical factors are responsible. Riding out the saddle leads to greater muscle activation (Ryschon & Stray-Gundersen, 1991; Hansen & Waldeland, 2008), causing a greater muscle pump action and venous return (Astorino *et al*., 2004), and thus a greater stroke volume and cardiac output (Faulkner *et al*., 2015). Other reasons may be a greater peripheral blood flow (Mauger *et al*., 2013) and/or oxygen extraction by the muscles (Faulkner *et al*., 2015).

A limitation of the study is the choice of lab test used. The 20w/minute ramp test was of a suboptimal length for $\dot{V}O_{2max}$ attainment in highly trained cyclists, as on average it took 20.4 minutes to complete compared to 8.4 minutes on the hill climb. It was found by Buchfuhrer *et al*., (1983) that tests lasting between 8-17 minutes led to a higher $\dot{V}O_{2max}$ than those of shorter or longer duration. Yoon *et al.* (2007) recommended tests of 8-10 minutes in length, as they found longer tests of 14-16 minutes resulted in a lower $\dot{V}O_{2max}$. Explanations include cardiac output reaching a peak during exercise of 5-9 minutes duration (Lepretre, Koralsztein & Billat, 2004), which is the same as the duration of the hill climb for most participants. Furthermore, the stronger a rider is, the shorter their time trial was likely to be (depending on mass), while their ramp test would last longer. The extended duration of the ramp test meant participants rode longer at high power outputs, possibly causing premature local muscular fatigue and failure to reach maximal workloads limiting lab test $\dot{V}O_{2peak}$ (Astorino *et al*., 2004).

We propose a modified version of this test be trialled in future research, based on the average duration of the field test (8.5 minutes), and literature identifying the 8-12-minute time-period as optimal for maximal testing (Buchfuhrer *et al*., 1983; Yoon *et* *al*., 2007). As well as the pre-existing 8-minute test used by cycling coaches (Carmichael & Rutberg, 2012), and training software, which consists of two 8-minute maximal efforts separated by 10 minutes of active recovery. The second 8-minute test would serve as a verification test for the first and allows training zones to be set based on the functional threshold power figure, calculated as 90% average power of the two 8-minute tests (Carmichael & Rutberg, 2012).

While this test is often carried out on a static turbo trainer, it would better be performed outdoors on a slight incline to provide resistance which leads to higher ecological validity due to the cooling effect of the wind (Brito *et al*., 2017), and also avoids the sub-optimal torque profile and muscle activation patterns of ergometer cycling (Fregley *et al*., 2000). Previous research conducted on the 8-minute field test (Klika *et al*., 2007; Sanders *et al*., 2017) has found that 8-minute power was strongly related to power at 4 $mmol{\cdot}L^{-1}$ BLac, commonly used as a physiological threshold. Sanders *et al.* (2017) warn of switching between lab and field testing as power measurement accuracy can vary, although the field test is likely more useful to an athlete as they are testing with the same equipment they train and compete with. This is supported by Klika *et al.* (2007) who report the 8-minute field test is a valid measure for changes in fitness and allows for the setting of training zones. However, changing environmental conditions may affect the accuracy of this. For example, hot conditions may result in lower than expected values if an athlete is not acclimatised to the heat.

5. Conclusion

Field-based testing likely resulted in a higher $\dot{V}O_{2peak}$ due to: greater familiarity with the course, known end-point allowing pacing (Noakes, 2008) and possibly higher motivation, optimal test duration (Buchfuhrer *et al*., 1983; Yoon *et al*., 2007) out-the-saddle riding allowing greater muscle recruitment (Ryschon & Stray-Gundersen, 1991; Hansen & Waldeland, 2008), and a cooling effect from the wind (Brito *et al*., 2017). Field-based testing can be considered a valid and likely more convenient alternative to laboratory testing for well-trained cyclists, assuming environmental conditions do not vary significantly between tests. Further testing with a larger sample size may result in a significantly higher $\dot{V}O_{2peak}$ in the field although it is difficult to predict as individual responses were mixed.

6. Practical Applications

Using a portable gas analyser to measure $\dot{V}O_{2max}/_{peak}$ is a valid alternative to lab-based testing as this study and previous studies (Mauger & Sculthorpe, 2012; Mauger *et al*., 2013; Straub *et al*., 2014; Hogg, Hopker & Mauger, 2015) have found it to result in similar if not higher $\dot{V}O_{2max}/_{peak}$ values. Field testing may be preferred by athletes due to their greater familiarity with the testing process and use of their own equipment. Care should be taken when comparing the results between lab tests, and future field tests with temperature, barometric pressure, humidity and wind potentially affecting results.

There is potential to combine this test with the 2x8 minute test (Klika *et al*., 2007; Carmichael & Rutberg, 2012; Sanders *et al*., 2017) to determine functional threshold power (FTP) in addition to $\dot{V}O_{2max}/_{peak}$ if a hill climb of 8-12 minutes length (Buchfuhrer *et al*., 1983; Yoon *et al*., 2007) and gradient is selected.

Conflict of Interest

The authors declare no conflict of interests.

Acknowledgments

We would like to thank Daniel Cooley, Georgia Campbell, Ben Sherwood and Stuart Brown for their technical expertise and help with data collection.

References

Astorino, T. A., Rietschel, J. C., Tam, P. A., Taylor, K., Johnson, S. M., Freedman, T. P., & Sakarya, C. E. (2004). Reinvestigation of optimal duration of Vo2max testing. *Journal of Exercise Physiology Online*, *7*(6).

Bassett, D. R., & Howley, E. T. (2000). Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Medicine and Science in Sports and Exercise*, *32*(1), 70-84.

Bertucci, W., Grappe, F., & Groslambert, A. (2007). Laboratory versus outdoor cycling conditions: differences in pedaling biomechanics. *Journal of Applied Biomechanics*, *23*(2), 87-92.

Bonaventura, J. M., Sharpe, K., Knight, E., Fuller, K. L., Tanner, R. K., & Gore, C. J. (2015). Reliability and accuracy of six hand-held blood lactate analysers. *Journal of Sports Science & Medicine*, *14*(1), 203-214.

Brito, J. P., Costa, A. M., Bento, P., Garrido, N., Reis, V. M., Conceição, A., & Louro, H. (2017). Air ventilation effects during the stationary roller bicycle test. *Journal of Physical Education and Sport*, *17*(1), 361-366.

Buchfuhrer, M. J., Hansen, J. E., Robinson, T. E., Sue, D. Y., Wasserman, K. A., & Whipp, B. J. (1983). Optimizing the exercise protocol for cardiopulmonary assessment. *Journal of Applied Physiology*, *55*(5), 1558-1564.

Cohen, J., 2013. *Statistical power analysis for the behavioural sciences*. Routledge.

Faulkner, J., Mauger, A. R., Woolley, B., & Lambrick, D. (2015). The efficacy of a self-paced VO2max test during motorized treadmill exercise. *International Journal of Sports Physiology and Performance*, *10*(1), 99-105.

Fregly, B. J., Zajac, F. E., & Dairaghi, C. A. (2000). Bicycle drive system dynamics: theory and experimental validation. *Journal of Biomechanical Engineering*, *122*(4), 446-452.

Hansen, E. A., & Waldeland, H. (2008). Seated versus standing position for maximization of performance during intense uphill cycling. *Journal of Sports Sciences*, *26*(9), 977-984.

Hill, A.V. and Lupton, H., 1923. Muscular exercise, lactic acid, and the supply and utilization of oxygen. *QJM: An International Journal of Medicine*, (62), 135-171.

Hogg, J. S., Hopker, J. G., & Mauger, A. R. (2015). The self-paced VO2max test to assess maximal oxygen uptake in highly trained runners. *International Journal of Sports Physiology and Performance*, *10*(2), 172-177.

Hopkins, W. G. (2006). Spreadsheets for analysis of controlled trials, with adjustment for a subject characteristic. *Sportscience*, *10*(1), 46-50.

Klika, R. J., Alderdice, M. S., Kvale, J. J., & Kearney, J. T. (2007). Efficacy of cycling training based on a power field test. *The Journal of Strength & Conditioning Research*, *21*(1), 265-269.

Lepretre, P. M., Koralsztein, J. P., & Billat, V. L. (2004). Effect of Exercise Intensity on Relationship between VO2max and Cardiac Output. *Medicine and Science in Sports and Exercise*, *36*, 1357-1363.

Macfarlane, D. J., & Wong, P. (2012). Validity, reliability and stability of the portable Cortex Metamax 3B gas analysis system. *European Journal of Applied Physiology*, *112*(7), 2539-2547.

Mauger, A. R., & Sculthorpe, N. (2012). A new VO2max protocol allowing self-pacing in maximal incremental exercise. *British Journal of Sports Medicine*, *46*(1), 59-63.

Mauger, A. R., Metcalfe, A. J., Taylor, L., & Castle, P. C. (2013). The efficacy of the self-paced VO2max test to measure maximal oxygen uptake in treadmill running. *Applied Physiology, Nutrition, and Metabolism*, *38*(12), 1211-1216.

Mclaughlin, J., Howley, E., Bassett, D., Thompson, D. And Fitzhugh, E. (2010). Test of the classic model for predicting endurance running performance. *Medicine & Science in Sports & Exercise*, *42*(5), 991-997.

Meyer, T., Welter, J. P., Scharhag, J., & Kindermann, W. (2003). Maximal oxygen uptake during field running does not exceed that measured during treadmill exercise. *European Journal of Applied Physiology*, *88*(4-5), 387-389.

Midgley, A. W., & Carroll, S. (2009). Emergence of the verification phase procedure for confirming 'true' VO2max. *Scandinavian Journal of Medicine & Science in Sports*, *19*(3), 313-322.

Midgley, A. W., Carroll, S., Marchant, D., McNaughton, L. R., & Siegler, J. (2009). Evaluation of true maximal oxygen uptake based on a novel set of standardized criteria. *Applied Physiology, Nutrition, and Metabolism*, *34*(2), 115-123.

Moore, J. L., Shaffrath, J. D., Casazza, G. A., & Stebbins, C. L. (2008). Cardiovascular effects of cadence and workload. *International Journal of Sports Medicine*, *29*(2), 116-119.

Noakes, T. D. (2008). Testing for maximum oxygen consumption has produced a brainless model of human exercise performance. *British Journal of Sports Medicine*, *42*(7), 551-555.

Poole, D. C., Wilkerson, D. P., & Jones, A. M. (2008). Validity of criteria for establishing maximal O2 uptake during ramp exercise tests. *European Journal of Applied Physiology*, *102*(4), 403-410.

Pyne, D.B., Boston, T., Martin, D.T., & Logan, A. (2000). Evaluation of the Lactate Pro blood lactate analyser. *European Journal of Applied Physiology*, *82*(1-2), 112-116.

Ricci, J., & Leger, L. A. (1983). Vo2 max of cyclists from treadmill, bicycle ergometer and velodrome tests. *European Journal of Applied Physiology and Occupational Physiology*, *50*(2), 283-289.

Ryschon, T.W., & Stray-Gundersen, J. (1991). The effect of body position on the energy cost of cycling. *Medicine and Science in Sports and Exercise*, *23*(8), 949-953.

Sanders, D., Taylor, R. J., Myers, T., & Akubat, I. (2017). A field-based cycling test to assess predictors of endurance performance and establishing training zones. *Journal of Strength and Conditioning Research*.

Straub, A. M., Midgley, A. W., Zavorsky, G. S., & Hillman, A. R. (2014). Ramp-incremented and RPE-clamped test protocols elicit similar VO 2max values in trained cyclists. *European Journal of Applied Physiology*, *114*(8), 1581-1590.

Williams, C. J., Williams, M. G., Eynon, N., Ashton, K. J., Little, J. P., Wisloff, U., & Coombes, J. S. (2017). Genes to predict VO 2max trainability: a systematic review. *BMC Genomics*, *18*(8), 831.

Yoon, B. K., Kravitz, L., & Robergs, R. (2007). VO2max, protocol duration, and the V O2 plateau. *Medicine & Science in Sports & Exercise*, *39*(7), 1186-1192.