

Article

Supra-Versus Submaximal Cycle Ergometer Verification of VO_{2max} in Males and Females

Brandon J. Sawyer ^{1,*}, Nicholas McMahon ², Kirsten L. Thornhill ², Brett R. Baughman ², Jenny M. Mahoney ¹, Kai L. Pattison ², Kaitlin A. Freeberg ² and Ryan T. Botts ^{3,†}

¹ Departments of Kinesiology and Biology, Point Loma Nazarene University, San Diego, CA 92106, USA; jmahoney@pointloma.edu

² Department of Kinesiology, Point Loma Nazarene University, San Diego, CA 92106, USA; nmcMahon1993@pointloma.edu (N.M.); kthornhi@pointloma.edu (K.L.T.); bbaughman1215@pointloma.edu (B.R.B.); kpattison290@pointloma.edu (K.L.P.); kaitlin.freeberg@colorado.edu (K.A.F.)

³ Department of Mathematical, Information, and Computer Sciences, Point Loma Nazarene University, San Diego, CA 92106, USA; rbotts@pointloma.edu

* Correspondence: bsawyer@pointloma.edu

† This paper is dedicated to the memory of Dr. Ryan T. Botts who passed away suddenly on 22 September 2020.

Received: 19 November 2020; Accepted: 10 December 2020; Published: 12 December 2020



Abstract: This study was designed to determine the optimal intensity for verification phase testing (VP) in healthy, young adults. Thirty one young, active participants (16 females) completed a cycle ergometer graded exercise test (GXT) VO_{2max} test and 4 VP tests at 80, 90, 100, and 105% of the maximum wattage achieved during the GXT. GXT and VP VO_{2max} values showed a significant test x sex interaction ($p = 0.02$). The males elicited significantly higher VO_{2max} values during the GXT, 80%, and 90% when compared to the 105%, (105 vs. GXT: $p = 0.05$; 105% vs. 80%: $p < 0.01$; 105% vs. 90%: $p = 0.02$). There were no significant differences in VO_{2max} across the tests in the females ($p > 0.05$); 80% of the males achieved their highest VP VO_{2max} during a submaximal VP test compared to only 37.5% of the females. A secondary study conducted showed excellent reliability (ICCs > 0.90) and low variation (CVs $< 3\%$) for the 90% VP. Our findings show that a submaximal verification phase intensity is ideal for young healthy males to elicit the highest VO_{2max} during cycle ergometer testing. For females, a range of intensities (80–105%) produce similar VO_{2max} values. However, the 80% VP yields an unnecessarily high time to exhaustion.

Keywords: maximal oxygen uptake; true VO_{2max} ; criteria for VO_{2max} ; incremental exercise testing; sex difference

1. Introduction

Maximal oxygen uptake (VO_{2max}) represents the maximal rate at which oxygen delivery and utilization can occur. VO_{2max} is known as the gold standard measurement of cardiorespiratory fitness and is one of the most common tests in exercise physiology research [1]. VO_{2max} testing has many practical uses including: predicting mortality and disease risk [2], aiding in exercise prescription, tracking changes in cardiorespiratory fitness, testing the efficacy of exercise interventions, as well as many clinical uses in cardiology and pulmonology [3]. The broad use and importance of VO_{2max} testing underscores the necessity for accuracy in determining whether a true VO_{2max} was actually attained. The original primary criterion to ensure VO_{2max} had occurred was attainment of a plateau in oxygen uptake during the final stages of a maximal exercise test [4]. Subsequent studies have shown it to be unnecessary for VO_2 to plateau to ensure VO_{2max} [5,6]. Most researchers use secondary criteria,

including maximal values of blood lactate, heart rate (HR), and/or respiratory exchange ratio (RER) even though these have been shown to have major limitations [7–9], high individual variability [10], and may occur well before maximal effort is reached [9].

Due to these issues with the primary and secondary criteria of VO_{2max} attainment, a constant work rate, an exhaustive bout of exercise following the standard graded exercise test (GXT), has been strongly recommended in order to verify that VO_{2max} has been attained [11–13]. This verification phase test (VP) shows promise as a solution to the shortcomings of primary and secondary criteria. As has recently been reviewed by Schaun [13], the lack of standardization of VP protocols and criteria for achievement of verified VO_{2max} remain major limitations. The intensity of the VP test relative to the maximum work rate achieved on the GXT is one of the issues that still needs to be resolved.

The most common method of verification phase testing is using a work rate that exceeds that achieved on the GXT (IE: supramaximal work rate) [11–13] in order to extend the work rate– VO_2 relationship and construct a VO_2 plateau [11]. In contrast, some authors have used a maximal [14] or even submaximal VP work rate [5,15–18] for verification of VO_{2max} attainment. It is possible that submaximal work rates will lead to longer times to exhaustion and possibly higher VO_{2max} values in some individuals [19], especially those unaccustomed to high-intensity exercise [14]. Some individuals may reach exhaustion in less than 2 min during supramaximal VPs, which may not be enough time to elicit VO_{2max} [19]. A recent study using data from 109 subjects found supramaximal VPs to elicit a significantly lower mean VO_{2max} compared to the GXT [20]. Theoretically, any work rate that falls in the severe domain should lead to exhaustion with VO_{2max} attainment [21,22]. Furthermore, it is possible that some supramaximal work rates used in VP testing may actually exceed the upper limit of the severe domain and enter into the extreme domain where exhaustion occurs before VO_{2max} attainment [19,23]. A few studies have shown submaximal verification tests to elicit VO_{2max} values similar to the GXT [5,15–18]. While some of these studies have tested both sub- and supramaximal VPs together [16–18], only two directly compared VP VO_{2max} values to each other [16,17]. These two studies [16,17] did not find differences in VO_{2max} between the GXT and VP tests, but they had small sample sizes and were not designed to test for a sex difference. Identification of the work rate (whether submaximal, maximal, or supramaximal) that elicits the highest VO_{2max} in both males and females during a cycle ergometer VP test would be helpful in deciding which VP work rate should be used. Furthermore, no study has yet assessed test-retest reliability of a submaximal verification phase test.

Therefore, the primary purpose of this study was to test which VP work rate produces the highest VO_{2max} value. We compared submaximal, maximal, and supramaximal work rates in both males and females in an attempt to determine the optimal work rate for VP testing. Our secondary purpose was to determine whether or not the optimal VP work rate differs between males and females. Finally, as a follow up to our primary study, we also tested the reliability of the 90% VP test.

2. Materials and Methods

2.1. Subjects

Both the primary study and the reliability study were approved by the Point Loma Nazarene University Institutional Review Board (PLNU IRB ID #17609) and conformed to the ethical standards of the Declaration of Helsinki. All subjects provided written informed consent before participation. All subjects met the following inclusion criteria: free from known chronic disease, between 18 and 30 years of age, completion of the Physical Activity Readiness Questionnaire (PAR-Q) [24] without “yes” answers. For the primary study, on the basis of previously published data [14], using R [25] powerlmm package [26] we calculated that completing 15 subjects would yield 90% power to detect a 4% difference in VO_{2max} between tests (at a two-tailed alpha level of 0.05). For the reliability study we used G power [27] along with previously published data [14] and found that a sample size of 8 was needed to produce 90% power to detect a significant correlation in VO_{2max} between repeated VP tests. For the primary study we enrolled 34 physically active (between 100 and 300 min of moderate or

vigorous activity per week) subjects (15 males and 19 females) into this 5-visit randomized cross-over study. For the reliability study we enrolled 23 young healthy physically active (between 100 and 300 min of moderate or vigorous activity per week) males and females to complete a 3-visit study. In each study we enrolled enough subjects to have adequate power within each sex independently.

2.2. Experimental Design

Primary Study: All subjects completed the following 5 visits: 1) GXT: standard ramp-style $\text{VO}_{2\text{max}}$ test; 2–5) randomly assigned verification phase tests at either 80, 90, 100, or 105% of the maximal work rate achieved on the GXT. Each test included a 5 min warm up at 50 W (males) or 30 W (females), followed by the test. All participants were asked to arrive to the lab ~ 4 h having fasted and abstained from caffeine, alcohol, and dietary supplements for at least 24 h and to abstain from vigorous exercise or weight lifting for 48 h before each visit. All visits were conducted at the same time of day for each subject, separated by at least 48 h, and all testing was completed within 3 weeks for each subject.

Reliability Study: Participants completed 3 separate visits to the Exercise Physiology Laboratory. Each visit included a ramp-style GXT test followed by a 10-min active recovery period, then directly into a 90% power output VP test. The first visit was used as a familiarization visit. Therefore, we conducted our reliability analyses on the tests conducted on visits 2 and 3 (VP2 and VP3). All other testing procedures and subject controls matched the primary study.

2.3. Equipment/Instruments

All exercise tests were conducted on a Monark 839E electronically braked cycle ergometer (Varberg, Sweden), which can continuously adjust brake force in response to changes in pedaling cadence to keep power output constant. Seat height and handlebar position were recorded for each subject and was replicated for subsequent visits. Pulmonary ventilation and gas exchange were measured during all visits with a Parvomedics metabolic cart (Truemax2400, Sandy, UT, USA). The mean of the highest 2 consecutive 15 s averages was used to determine $\text{VO}_{2\text{max}}$ in all tests. Heart rate max was determined using the highest 15 s average at the end of the test. Standard 3-point calibration was performed before each test or every 4 h per manufacturer recommendation. Heart rate was continuously measured using a Polar heart rate monitor (Lake Success, NY, USA) and all blood lactate measurements were done with a Lactate Plus blood lactate analyzer (Sports Resource Group, Minneapolis, MN, USA).

2.4. Exercise Testing

The following procedures were used during all cycle ergometer tests. Ratings of perceived exertion were taken at the end of each minute. Capillary blood lactate was measured via finger-prick at rest prior to testing and then again immediately following exhaustion. Subjects were given verbal encouragement throughout the test and when pedal rate slowed below 50 revolutions per min (rpm) for more than 5 s, the test was stopped and time to exhaustion (TTE) was recorded.

The GXT protocol was chosen using an estimated $\text{VO}_{2\text{max}}$, estimated max wattage, and a grade designed to reach max wattage in approximately 10 min following the warm-up [17]. During the warm up, subjects were instructed to choose a 10 rpm range of their comfort between 60 and 90 rpm. Once the ramp-style GXT test began, the wattage increased from the warm up wattage continuously throughout the test. Max wattage was selected as the final set power output when pedal rate slowed below 50 rpm.

The max wattage on the GXT was used to determine the VP wattage based on the assigned percentage. In the primary study, the order of the 4 VP visits was randomly selected to avoid a sequence effect and the set wattage was hidden from the subjects. Subjects were notified of the oncoming increase in resistance of the cycle during the last 5 s of the warm-up and wattage was immediately increased to the set chosen VP wattage. Subjects were instructed to pedal as long as possible at an rpm range (60–70, 70–80, or 80–90 rpm) of their choosing. This rpm range was recorded after the first VP test and repeated for all subsequent tests.

2.5. Statistical Analyses

p values were two-tailed, and values of ≤ 0.05 were considered statistically significant. All data are presented as mean \pm SD unless otherwise noted. Descriptive statistics for the subjects who completed the primary study were calculated by sex (Table 1). Linear mixed effects models (LME) were used in lieu of repeated measures due to their superior ability to deal with unequal groups and their lack of reliance on sphericity, equal variance and covariance assumptions [28,29]. An LME model was fitted to each of the outcome measures to assess the effects of test, sex, and the interaction test \times sex, with a random effect to control for subject using the maximum likelihood method. Post-hoc tests were run using Bonferroni correction, comparing the means between each pair of verification tests and the GXT. Pearson's correlations were used to determine the relationships between time on each test and differences in VO_{2max} and HR_{max} , as well as the difference between the highest VO_{2max} achieved and each VP VO_{2max} with the TTE on each VP test. All tests were implemented in R [25]. Mixed effects models were fitted using the lme4 package [30] with lmerTest [31] to compute the *p*-values, while post-hoc tests were conducted using the emmean package [32].

Table 1. Descriptive and anthropometric data for all subjects who completed the primary study.

	Males (n = 15)	Females (n = 16)
Age (years)	22 \pm 2	21 \pm 2
Height (cm)	178.1 \pm 8.4	172.0 \pm 6.4
Weight (kg)	77.3 \pm 7.5	63.1 \pm 10.7
BMI (kg/m ²)	24.5 \pm 2.2	23.2 \pm 3.3
VO_{2max} (mL/kg/min)	48.82 \pm 6.32	39.12 \pm 5.96

Values represent mean \pm SD. VO_{2max} represents highest of all 5 tests for each subject.

For the reliability study, we used intraclass correlation coefficients (ICC, one way, consistency) and coefficient of variation (CV) to assess and examine reliability. A Bland-Altman Plot was used to determine the relationship between mean VO_{2max} during VP 2 and VP 3 and the highest VO_{2max} achieved. Paired samples t-tests were used to assess differences between VP 2 and VP 3 VO_{2max} , TTE, HR_{max} , RER_{max} , wattage, and the difference between GXT VO_{2max} and VP VO_{2max} .

3. Results

3.1. Primary Study

Of the 34 enrolled subjects (15 males and 19 females), 4 females dropped out due to time constraints or inability to comply with the 48 h of rest requirement between each test, resulting in 15 males and 16 females (See Table 1) completing all requirements of the study. All outcome variables during the GXT and VP tests are shown in Table 2 (males) and Table 3 (females).

Table 2. Outcome variables for the males for each test.

	VO_{2max} (L/min)	HR_{max} (bpm)	RER_{max}	Blood Lactate (mmol/L)	Peak Power (W)	Total Time (min)	RPE
GXT	3.64 * \pm 0.67	184 \pm 12	1.27 \pm 0.07	12.9 \pm 2.3	297 \pm 60	9.90 \pm 1.62	19 \pm 1
80%VP	3.67 * \pm 0.71	183 \pm 13	1.25 \pm 0.08	13.1 \pm 2.4	238 \pm 48	6.28 \pm 1.28	20 \pm 1
90%VP	3.66 * \pm 0.67	181 \pm 12	1.33 \pm 0.10	12.8 \pm 1.8	268 \pm 54	4.21 \pm 0.69	20 \pm 1
100%VP	3.56 \pm 0.51	180 \pm 12	1.41 \pm 0.08	13.8 \pm 1.9	297 \pm 60	2.70 \pm 0.53	20 \pm 1
105%VP	3.50 \pm 0.47	177 \pm 12	1.48 \pm 0.12	12.9 \pm 1.1	312 \pm 63	2.44 \pm 0.40	20 \pm 0

GXT = Initial graded exercise test; VP = verification phase test. Values represent mean \pm SD; * Significantly higher than 105% VP.

Table 3. Outcome variables for the females for each test.

	VO _{2max} (L/min)	HR _{max} (bpm)	RER _{max}	Blood Lactate (mmol/L)	Peak Power (W)	Total Time (min)	RPE
GXT	2.35 ± 0.33	186 ± 6	1.22 ± 0.06	11.2 ± 2.0	214 ± 26	10.44 ± 1.24	19 ± 1
80%VP	2.29 ± 0.34	187 ± 7	1.18 ± 0.12	11.2 ± 2.2	171 ± 21	9.24 ± 5.00	19 ± 1
90%VP	2.34 ± 0.33	185 ± 5	1.23 ± 0.09	12.4 ± 2.1	192 ± 23	5.33 ± 2.85	19 ± 1
100%VP	2.36 ± 0.32	185 ± 4	1.32 ± 0.08	11.8 ± 1.7	214 ± 26	2.86 ± 1.07	18 ± 2
105%VP	2.32 ± 0.32	184 ± 7	1.33 ± 0.09	12.0 ± 1.6	224 ± 27	2.19 ± 0.59	19 ± 1

GXT = Initial graded exercise test; VP = verification phase test. Values represent mean ± SD.

See Figure 1 and Tables 2 and 3 for the VO_{2max} data for both males and females. Results of the mixed effect model comparing all 4 VPs and the GXT showed a significant difference in VO_{2max} between males and females (Males: 3.61 ± 0.60, Females: 2.33 ± 0.32 L/min; $F(1,31) = 60.3, p < 0.001$) and a significant test × sex interaction ($F(4,124) = 3.10, p = 0.02$, See Tables 2 and 3, and Figure 1). There was no difference between the 5 mean VO_{2max} values with males and females combined ($F(4,124) = 1.97, p = 0.10$). When males and females were separated there was a significant main effect for test ($F(4,60) = 2.81, p = 0.03$) with the males, but not with the females ($F(4,64) = 1.49, p = 0.22$); 12 of the 15 males achieved their highest VO_{2max} during a submaximal VP, but the females highest achieved VO_{2max} values were more evenly distributed across the 4 VP work rates. Post-hoc tests for the males revealed the GXT elicited a significantly higher VO_{2max} compared to the VP at 105% (GXT –105%: mean difference = 0.14 ± 0.05, $t(133) = 2.83, p = 0.05$) but was not different from the other VPs. Additionally, there were significant differences between the 80% and 90% when compared to 105% (105–80%: mean difference = −0.17 ± 0.05, $t(133) = 3.51, p < 0.01$; 105–90%: mean difference = −0.16 ± 0.05, $t(133) = 3.21, p = 0.02$).

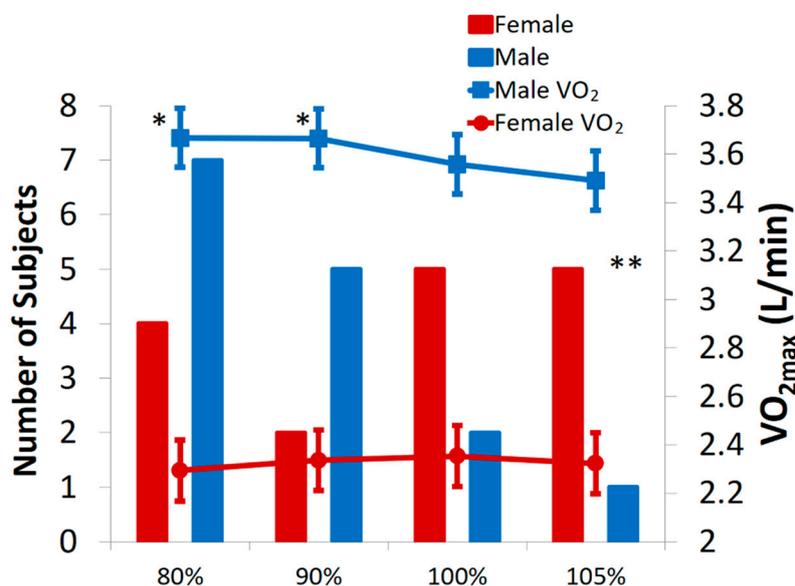


Figure 1. Primary axis = Number of subjects that elicited their highest VO_{2max} during each of the four verification phase tests. Secondary axis = Mean ± SD VO_{2max}. * Significantly higher than VP 105%. ** Significant interaction VP intensity × Sex ($p = 0.02$).

Pearson correlations revealed no relationship between TTE on the VP tests and VO_{2max} ($r = 0.01, p = 0.94$). There were significant relationships with the difference between the highest VO_{2max} and VP VO_{2max} correlated with TTE on each VP test (females: $r = 0.39, p = 0.002$; males: $r = -0.32, p = 0.01$).

3.2. Reliability Study

Of the 23 males and females enrolled in the reliability study, 20 (10 females, 10 males; BMI = 22.7 ± 2.9 kg/m²; Age = 25.5 ± 4.0 year) completed all the requirements for their data to be analyzed. The Bland-Altman plot comparing the highest VO_{2max} achieved and mean difference between VP 3 and VP 2 showed no bias and small limits of agreement (mean difference of 0.03 L/min, 95% limits of agreement: 0.17 and -0.11 L/min, see Figure 2). ICCs and CVs for males and females combined showed excellent consistency for VP VO_{2max} (ICC = 0.991; CV = $2.68 \pm 2.52\%$). ICC and CV for female participants VP VO_{2max} demonstrated excellent consistency (ICC = 0.987; CV = 2.5%). Male participants VP VO_{2max} displayed excellent consistency (ICC = 0.941; CV = 2.2%). ICCs and CVs for males and females combined showed excellent consistency for GXT VO_{2max} (ICC = 0.989; CV = 2.7%). Wattage and TTE did significantly differ between VP 2 and VP 3, but VO_{2max}, HR_{max}, RER_{max}, and the difference between GXT and VP VO_{2max} did not (See Table 4).

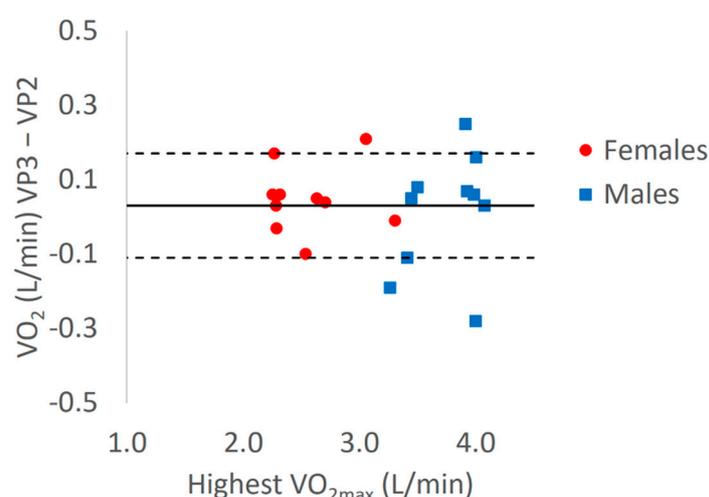


Figure 2. VP VO_{2max} Bland-Altman plot. Y axis = VO₂ (L/min) Verification Phase 3 – Verification Phase 2; X axis = Highest VO_{2max} (L/min); Dashed lines = Mean \pm 1.96 * SD; Solid line = mean of VP3 – VP2 (0.03 L/min).

Table 4. Outcome variables for all participants for each verification phase (VP) test for the submaximal VP reliability study.

	VP 2			VP 3			VP2 vs. VP3 All Subjects
	All	Males	Females	All	Males	Females	Significance
VO ₂ mL/kg/min	40.91 \pm 5.93	43.29 \pm 5.48	38.52 \pm 5.62	41.37 \pm 6.24	43.55 \pm 6.79	39.2 \pm 5.21	0.22
VO ₂ L/min	3.01 \pm 0.69	3.59 \pm 0.31	2.43 \pm 0.39	3.04 \pm 0.69	3.54 \pm 0.37	2.48 \pm 0.39	0.31
HR _{max}	168 \pm 41	176 \pm 11	161 \pm 57	179 \pm 10	178 \pm 10	180 \pm 10	0.27
RER _{max}	1.13 \pm 0.09	1.13 \pm 0.09	1.13 \pm 0.09	1.14 \pm 0.09	1.15 \pm 0.10	1.13 \pm 0.08	0.69
TTE	2.68 \pm 0.90	2.24 \pm 0.68	3.13 \pm 0.90	2.29 \pm 0.50	2.05 \pm 0.57	2.52 \pm 0.28	0.02 *
Wattage	235 \pm 54	275 \pm 40	195 \pm 32	240 \pm 55	280 \pm 43	200 \pm 33	0.04 *
VO ₂ (L/min): VP-GXT	-0.02 ± 0.13	-0.05 ± 0.16	0.00 ± 0.09	-0.01 ± 0.14	-0.03 ± 0.17	0.02 ± 0.09	0.55

* $p < 0.05$.

4. Discussion

The primary finding of this study is that the supramaximal verification phase test (105%) elicited a significantly lower VO_{2max} value compared to the submaximal VP tests (80 and 90%) and the GXT in males only. Conversely, there were no significant differences across the VP intensities and GXT for our female subjects and 80% of the males achieved their highest VP VO_{2max} during a submaximal VP test

compared to only 37.5% of the females. Our study is the first to compare four different VP intensities and to evaluate differences between males and females. Furthermore, our reliability study showed excellent consistency, no bias, and acceptable limits of agreement across a range of $\text{VO}_{2\text{max}}$ values for the 90% VP test in both males and females.

The results of the present study partially agree with those of Astorino [20]. Study number two in the Astorino [20] paper found a significant test \times gender interaction where the males had a higher $\text{VO}_{2\text{max}}$ on the incremental test compared to the supramaximal VP and the females did not. Conversely, study number one showed no test \times gender effect but overall showed a significantly higher $\text{VO}_{2\text{max}}$ in the incremental compared to the VP. Similar to study two above, we found a significant test \times sex interaction showing no difference across VP intensities for females but a significantly lower $\text{VO}_{2\text{max}}$ in the 105% VP compared to GXT, 80%, and 90% VP. Our results also agree with a previous study that used 4 square wave bouts of exercise to exhaustion in males and females in order to determine critical power [33]. These results show that the supramaximal square wave bout (118 and 122% of GXT power in males and females, respectively) only elicited 91% of $\text{VO}_{2\text{max}}$ achieved on the GXT in males, but 105% of GXT $\text{VO}_{2\text{max}}$ in females. Similarly, the highest mean $\text{VO}_{2\text{peak}}$ values for the males (97% of GXT $\text{VO}_{2\text{max}}$) were achieved during two submaximal square-wave bouts (P2 and P3, 80 and 71% of GXT peak power output, respectively). For the females, the highest $\text{VO}_{2\text{peak}}$ values during the square-wave bouts were achieved at power outputs equal to 90 and 122% of GXT peak power eliciting 108% and 105% of GXT $\text{VO}_{2\text{max}}$, respectively. This comparison should be viewed with caution due to the lower subject number (9 males and 4 females), a lack of statistically significant differences between $\text{VO}_{2\text{peak}}$ values, and differences in power outputs relative to GXT peak power.

The reason for a difference in males and females in optimal VP intensity is not apparent. One possible reason may be bias induced via calculation of VP work rate based on a percentage of GXT maximum work rate. Our mean peak wattage was 83 watts higher in males compared to females. The 5% above GXT added on average 15 watts for the males, but only 10 watts for the females and this extra added wattage could be enough to push male participants above the severe domain of exercise into the extreme domain where TTE is too short to elicit $\text{VO}_{2\text{max}}$ [19,34]. Along these same lines another explanation would be faster or similar VO_2 kinetics in females compared to males coupled with similar times to exhaustion. As expected in our study, the mean male $\text{VO}_{2\text{max}}$ was ~ 10 mL/kg/min higher than the female even though the TTE was similar. Accordingly, the males in our study may have not had sufficient time to reach their higher $\text{VO}_{2\text{max}}$ during the 105% VP, whereas the females did. An extreme example of this in our study was found in our subject with the highest $\text{VO}_{2\text{max}}$ (5.9 L/min) who had similar TTE on the 100 and 105% VP tests when compared to the other males, but those VP tests produced much lower $\text{VO}_{2\text{max}}$ values compared to his lower intensity VP tests. That subject's data can be seen as the 2 outliers in panel A in Figure 3 below. This is potentially supported by previous work showing no gender effect on gross efficiency or delta efficiency during submaximal cycling [35]. Importantly, the efficiency measured in that study [35] was submaximal and may not hold true at maximal and supramaximal work rates. Hill et al. [19] have suggested that times to exhaustion of at least 2 min may be necessary to elicit $\text{VO}_{2\text{max}}$. The average TTE during the 105% VP in our males was 2.44 min but did not elicit $\text{VO}_{2\text{max}}$. The study discussed above by Sawyer et al. [33] showed that $\text{VO}_{2\text{max}}$ reached in as little as 1.27 min in females, but 1.25 min only elicited 91% of GXT $\text{VO}_{2\text{max}}$ in males. Similarly, the current study shows a mean TTE during the 105% VP in females of 2.19 min, which produced a $\text{VO}_{2\text{max}}$ not different to GXT $\text{VO}_{2\text{max}}$. Interestingly (as shown in Figure 3), the correlations between the difference in the highest $\text{VO}_{2\text{max}}$ and each VP $\text{VO}_{2\text{max}}$ with TTE show a positive relationship for females ($r = 0.39$) and a negative one for males ($r = -0.32$). To clarify, for males the shorter the VP duration the further their elicited VO_2 was from $\text{VO}_{2\text{max}}$, but for females the shorter the VP duration the closer their elicited VO_2 was to $\text{VO}_{2\text{max}}$. These correlations are reflective of the maximal and supramaximal VPs producing the highest $\text{VO}_{2\text{max}}$ in the females and the submaximal VPs producing the highest $\text{VO}_{2\text{max}}$ in males. This sex difference could also be explained by the fact that our female subjects may have been more fit than the males and could withstand severe intensity

exercise for longer. This hypothesis is supported by the fact that for some of our female subjects the 80% VP was at or below critical power due to exhaustion never occurring in the 20 min test, but all males reached exhaustion in under 10 min. Furthermore, the average TTE during the 80% VP was 9.24 min in the females and only 6.28 min in the males. Comparison of our $\text{VO}_{2\text{max}}$ data to normative values does not support this hypothesis since the mean $\text{VO}_{2\text{max}}$ for females and males were both in the high end of the “fair” category [36]. Critical power would be a better indicator of fitness in terms of maximally sustainable VO_2 [34], but we did not measure critical power in this study.

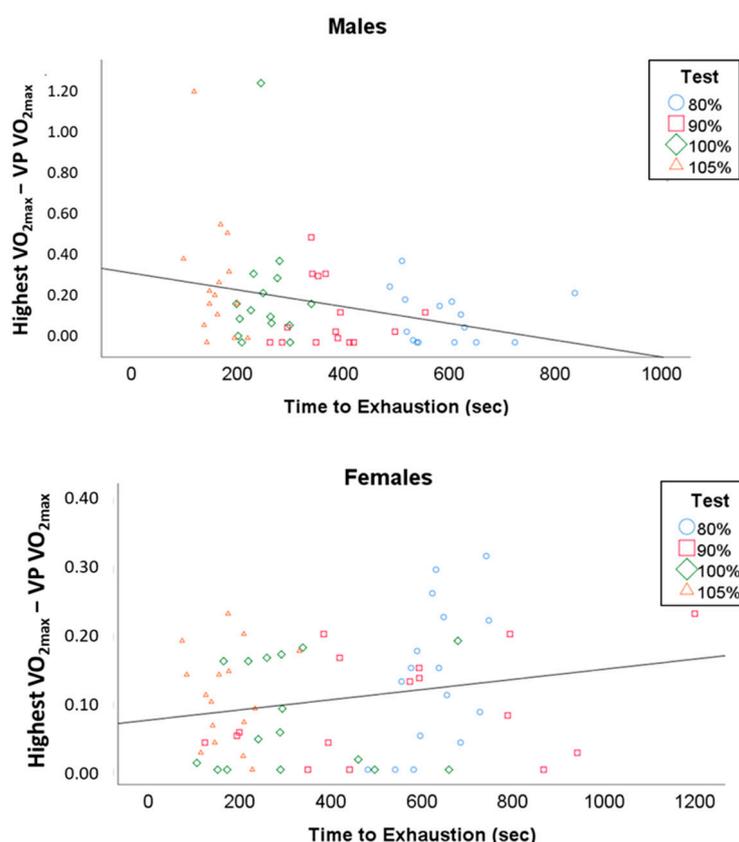


Figure 3. Scatterplots showing the relationship between time to exhaustion and the difference between the highest $\text{VO}_{2\text{max}}$ and each verification phase (VP) $\text{VO}_{2\text{max}}$. Each VP test is shown by a different color and marker type. Top panel = males ($r = -0.32$, $p = 0.01$); Bottom panel = females ($r = 0.39$, $p = 0.002$).

In addition, our finding that a submaximal VP test produces higher $\text{VO}_{2\text{max}}$ values in males may be limited to cycle ergometer testing. It has been shown that cycle ergometer $\text{VO}_{2\text{max}}$ is typically 5–20% lower than treadmill $\text{VO}_{2\text{max}}$ due to regional muscle fatigue [37]. The regional muscle fatigue is most likely why the supramaximal VP leads to shorter times to exhaustion. To our knowledge no study has shown supramaximal VP testing on a treadmill to elicit lower $\text{VO}_{2\text{max}}$ than submaximal or GXT. The larger muscle mass involved in locomotor exercise may allow for supramaximal VP testing on the treadmill to produce adequate times to exhaustion and therefore higher $\text{VO}_{2\text{max}}$ values compared to cycle ergometer.

The results of the reliability study further establishes support for using a 90% VP test by showing excellent reliability (ICC = 0.991), extremely similar $\text{VO}_{2\text{max}}$ values with repeat testing (VP2 $\text{VO}_{2\text{max}}$: 3.01 ± 0.69 L/min, VP 3 $\text{VO}_{2\text{max}}$: 3.04 ± 0.69 L/min; $p = 0.55$), a small coefficient of variation (overall CV = 2.68%), no bias across a wide range of $\text{VO}_{2\text{max}}$ values (mean difference between VP2 and VP3: 0.03 L/min), and acceptable limits of agreement. Previous work [5] has shown day to day variation in the measurement of $\text{VO}_{2\text{max}}$ to be ~5% (0.23 L/min with mean $\text{VO}_{2\text{max}}$ of 4.51 L/min), therefore our limits of agreement of ~3–6% of mean $\text{VO}_{2\text{max}}$ (See Figure 2; 0.17 and -0.11 L/min; ~4% of $\text{VO}_{2\text{max}}$)

show the random error between repeat tests to be in an acceptable range. Our ICC and CV values are similar to previous studies using supramaximal VP tests [15,17,38–41]. Furthermore, these results were replicated when we examined the males and females separately (See Table 4). To our knowledge, the current study is the first to examine the reliability of a 90% power output submaximal verification phase to confirm $\text{VO}_{2\text{max}}$.

Strengths and Limitations

One strength of this study is the inclusion of both males and females with adequate statistical power to draw conclusions for each sex independently. Furthermore, we are not aware of any other studies that have specifically compared as many different VP intensities as compared in this study. Moreover, using LME models to compare repeated measures within subjects afforded us the ability to compare multiple repeated tests without weakening the analyses due to inequality of variance, covariance, or sphericity. One possible limitation of this study is the fact that the VP tests were completed on separate days from the GXT. Most common use of VP testing would include GXT followed by a VP in one session, but completing the VP on a separate day has shown similar results [9,38]. The RER_{max} values for males in the primary study were higher than expected for maximal exercise, but not out of the range of values seen in previous work, especially with short times to exhaustion [9,42,43]. Importantly, use of a submaximal intensity for VP testing does not extend the VO_2 -power relationship to construct a VO_2 plateau as is recommended by Poole and Jones [11] with supramaximal VP testing. We did not attempt to extend the VO_2 -power relationship but to find a VP intensity that produces the highest $\text{VO}_{2\text{max}}$. In regards to the reliability study our results can only be applied to the VP test at 90% GXT wattage and therefore do not have implications about the reliability or bias in the other VP intensities. Lastly, this study has implications only for cycle ergometer testing in young healthy adults, whereas different conclusions may be reached with other modes of exercise in different populations.

5. Conclusions

In conclusion we found that males and females had different VO_2 responses to cycle ergometry verification phase testing. In males the submaximal tests elicited higher $\text{VO}_{2\text{max}}$ values compared to supramaximal testing. For females, a range of intensities (80–105%) produce similar $\text{VO}_{2\text{max}}$ values but the unnecessarily high times to exhaustion in the 80% VP do not warrant its use. We also established excellent reliability, low repeated test variability, and lack of bias across a range of $\text{VO}_{2\text{max}}$ values using the 90% VP in both males and females. Our results show that submaximal VP testing should be considered an acceptable method for verification of $\text{VO}_{2\text{max}}$ and potentially the preferable method in young healthy males during cycle ergometry.

Author Contributions: Conceptualization, B.J.S., K.L.T., and N.M.; methodology, B.J.S., K.L.T., N.M., and R.T.B.; validation, B.J.S., K.L.T., and R.T.B.; formal analysis, B.J.S., K.L.T., N.M., and R.T.B.; investigation, B.J.S., K.L.T., N.M., K.A.F., B.R.B., J.M.M., K.L.P., and R.T.B.; resources, B.J.S., K.L.T., N.M., K.A.F., B.R.B., J.M.M., K.L.P., and R.T.B.; data curation, B.J.S., N.M., K.L.T., K.A.F., J.M.M., K.L.P., B.R.B., and R.T.B.; writing—original draft preparation, B.J.S., N.M., and K.L.T.; writing—review and editing, B.J.S., K.L.T., N.M., K.A.F., B.R.B., J.M.M., K.L.P., and R.T.B.; visualization, B.J.S., R.T.B., K.L.T., and N.M.; supervision, B.J.S. and R.T.B.; project administration, B.J.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the participants of the study.

Conflicts of Interest: The authors have no conflict of interest to declare.

References

1. Levine, B.D. $\text{VO}_{2\text{max}}$: What do we know, and what do we still need to know? *J. Physiol.* **2008**, *586*, 25–34. [[CrossRef](#)] [[PubMed](#)]

2. Katzmarzyk, P.T.; Church, T.S.; Blair, S.N. Cardiorespiratory fitness attenuates the effects of the metabolic syndrome on all-cause and cardiovascular disease mortality in men. *Arch. Intern. Med.* **2004**, *164*, 1092–1097. [[CrossRef](#)] [[PubMed](#)]
3. Arena, R.; Myers, J.; Williams, M.A.; Gulati, M.; Kligfield, P.; Balady, G.J.; Collins, E.; Fletcher, G. Assessment of functional capacity in clinical and research settings: A scientific statement from the American Heart Association committee on exercise, rehabilitation, and prevention of the council on clinical cardiology and the council on cardiovascular n. *Circulation* **2007**, *116*, 329–343. [[CrossRef](#)] [[PubMed](#)]
4. Hill, A.V.; Lupton, H. Muscular exercise, lactic acid, and the supply and utilization of oxygen. *QJM* **1923**, *62*, 135–171. [[CrossRef](#)]
5. Day, J.R.; Rossiter, H.B.; Coats, E.M.; Skasick, A.; Whipp, B.J. The maximally attainable VO₂ during exercise in humans: The peak vs. maximum issue. *J. Appl. Physiol.* **2003**, *95*, 1901–1907. [[CrossRef](#)]
6. Howley, E.T. VO_{2max} and the Plateau-Needed or Not? *Med. Sci. Sport Exerc.* **2007**, *39*, 101–102. [[CrossRef](#)]
7. Howley, E.T.; Bassett, D.R.; Welch, H.G. Criteria for maximal oxygen uptake: Review and commentary. *Med. Sci. Sport Exerc.* **1995**, *27*, 1292–1301. [[CrossRef](#)]
8. Midgley, A.W.; McNaughton, L.R.; Polman, R.; Marchant, D. Criteria for determination of maximal oxygen uptake: A brief critique and recommendations for future research. *Sport Med.* **2007**, *37*, 1019–1028. [[CrossRef](#)]
9. Poole, D.C.; Wilkerson, D.P.; Jones, A.M. Validity of criteria for establishing maximal O₂ uptake during ramp exercise tests. *Eur. J. Appl. Physiol.* **2008**, *102*, 403–410. [[CrossRef](#)]
10. Astorino, T.A.; Robergs, R.A.; Ghiasvand, F.; Marks, D.; Burns, S. Incidence of the oxygen plateau at VO_{2max} during exercise testing to volitional fatigue. *J. Exerc. Physiol.* **2000**, *3*, 1–12.
11. Poole, D.C.; Jones, A.M. Measurement of the maximum oxygen uptake VO_{2max}: VO_{2peak} is no longer acceptable. *J. Appl. Physiol.* **2017**, *122*, 997–1002. [[CrossRef](#)] [[PubMed](#)]
12. Midgley, A.W.; Carroll, S. Emergence of the verification phase procedure for confirming ‘true’ VO_{2max}. *Scand. J. Med. Sci. Sports* **2009**, *19*, 313–322. [[CrossRef](#)] [[PubMed](#)]
13. Schaun, G.Z. The Maximal Oxygen Uptake Verification Phase: A Light at the End of the Tunnel? *Sports Med. Open* **2017**, *3*, 44. [[CrossRef](#)] [[PubMed](#)]
14. Sawyer, B.J.; Tucker, W.J.; Bhammar, D.M.; Gaesser, G.A. Using a Verification Test for Determination of VO_{2max} in Sedentary Adults with Obesity. *J. Strength Cond. Res.* **2015**, *29*, 3432–3438. [[CrossRef](#)]
15. Kirkeberg, J.M.; Dalleck, L.C.; Kamphoff, C.S.; Pettitt, R.W. Validity of 3 protocols for verifying VO_{2max}. *Int. J. Sports Med.* **2011**, *32*, 266–270. [[CrossRef](#)] [[PubMed](#)]
16. Rossiter, H.B.; Kowalchuk, J.M.; Whipp, B.J. A test to establish maximum O₂ uptake despite no plateau in the O₂ uptake response to ramp incremental exercise. *J. Appl. Physiol.* **2006**, *100*, 764–770. [[CrossRef](#)]
17. Sedgeman, D.; Dalleck, L.; Clark, I.E.; Jamnick, N.; Pettitt, R.W. Analysis of Square-wave Bouts to Verify VO_{2max}. *Int. J. Sport Med.* **2013**, *34*, 1058–1062. [[CrossRef](#)]
18. Murias, J.M.; Pogliaghi, S.; Paterson, D.H. Measurement of a True VO_{2max} during a Ramp Incremental Test Is Not Confirmed by a Verification Phase. *Front. Physiol.* **2018**, *9*, 143. [[CrossRef](#)]
19. Hill, D.W.; Poole, D.C.; Smith, J.C. The relationship between power and the time to achieve VO_{2max}. *Med. Sci. Sport Exerc.* **2002**, *34*, 709–714.
20. Astorino, T.A.; DeRevere, J. Efficacy of constant load verification testing to confirm VO_{2max} attainment. *Clin. Physiol. Funct. Imaging* **2018**, *38*, 703–709. [[CrossRef](#)]
21. Poole, D.C.; Ward, S.A.; Gardner, G.W.; Whipp, B.J. Metabolic and respiratory profile of the upper limit for prolonged exercise in man. *Ergonomics* **1988**, *31*, 1265–1279. [[CrossRef](#)] [[PubMed](#)]
22. Jones, A.M.; Wilkerson, D.P.; DiMenna, F.; Fulford, J.; Poole, D.C. Muscle metabolic responses to exercise above and below the “critical power” assessed using 31P-MRS. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2008**, *294*, R585. [[CrossRef](#)] [[PubMed](#)]
23. Burnley, M.; Jones, A.M. Oxygen uptake kinetics as a determinant of sports performance. *Eur. J. Sport Sci.* **2007**, *7*, 63–79. [[CrossRef](#)]
24. Adams, R. Revised Physical Activity Readiness Questionnaire. *Can. Fam. Physician* **1999**, *45*, 992–1005. [[PubMed](#)]
25. R Core Team. R: A Language and Environment for Statistical Computing. 2018. Available online: <https://www.r-project.org/> (accessed on 1 May 2019).
26. Magnusson, K. *Powerlmm: Power Analysis for Longitudinal Multilevel Models*; R Package Version 0.4.0; GitHub: San Francisco, CA, USA, 2018.

27. Faul, F.; Erdfelder, E.; Buchner, A.; Lang, A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods*. **2009**, *41*, 1149–1160. [[CrossRef](#)]
28. Baayen, R.H.; Davidson, D.J.; Bates, D.M. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **2008**, *59*, 390–412. [[CrossRef](#)]
29. Wang, Z.; Goonewardene, L.A. The use of MIXED models in the analysis of animal experiments with repeated measures data. *Can. J. Anim. Sci.* **2004**, *84*, 1–11. [[CrossRef](#)]
30. Bates, D.; Mächler, M.; Bolker, B.M.; Walker, S.C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [[CrossRef](#)]
31. Kuznetsova, A.; Brockhoff, P.B.C.R. lmerTest Package: Tests in Linear Mixed Effects. *J. Stat. Softw.* **2017**, *82*, 1–26. [[CrossRef](#)]
32. Searle, S.R.; Speed, F.M.; Milliken, G.A. Population marginal means in the linear model: An alternative to least squares means. *Am. Stat.* **1980**, *34*, 216–221. [[CrossRef](#)]
33. Sawyer, B.J.; Morton, R.H.; Womack, C.J.; Gaesser, G.A. VO_{2max} may not be reached during exercise to exhaustion above critical power. *Med. Sci. Sport Exerc.* **2012**, *44*, 1533–1538. [[CrossRef](#)] [[PubMed](#)]
34. Jones, A.M.; Vanhatalo, A.; Burnley, M.; Morton, R.; Poole, D.C. Critical Power: Implications for Determination of VO_{2max} and Exercise Tolerance. *Med. Sci. Sport Exerc.* **2010**, *42*, 1876–1890. [[CrossRef](#)] [[PubMed](#)]
35. Yasuda, N.; Gaskill, S.E.; Ruby, B.C. No gender-specific differences in mechanical efficiency during arm or leg exercise relative to ventilatory threshold. *Scand. J. Med. Sci. Sports* **2008**, *18*, 205–212. [[CrossRef](#)] [[PubMed](#)]
36. Kaminsky, L.A.; Arena, R.; Myers, J. Reference Standards for Cardiorespiratory Fitness Measured with Cardiopulmonary Exercise Testing. *Mayo Clin. Proc.* **2015**, *90*, 1515–1523. [[CrossRef](#)]
37. ACSM. *ACSM's Guidelines for Exercise Testing and Prescription*, 10th ed.; Wolters Kluwer: Philadelphia, PA, USA, 2018; Volume 10. [[CrossRef](#)]
38. Astorino, T.A.; White, A.C.; Dalleck, L.C. Supramaximal testing to confirm attainment of VO_{2max} in sedentary men and women. *Int. J. Sports Med.* **2009**, *30*, 279–284. [[CrossRef](#)] [[PubMed](#)]
39. Scharhag-Rosenberger, F.; Carlsohn, A.; Cassel, M.; Mayer, F.; Scharhag, J. How to test maximal oxygen uptake: A study on timing and testing procedure of a supramaximal verification test. *Appl. Physiol. Nutr. Metab.* **2011**, *36*, 153–160. [[CrossRef](#)]
40. Hawkins, M.N.; Raven, P.B.; Snell, P.G.; Stray-Gundersen, J.; Levine, B.D. Maximal oxygen uptake as a parametric measure of cardiorespiratory capacity. *Med. Sci. Sports Exerc.* **2007**, 103–107. [[CrossRef](#)]
41. Midgley, A.W.; Carroll, S.; Marchant, D.; McNaughton, L.R.; Siegler, J. Evaluation of true maximal oxygen uptake based on a novel set of standardized criteria. *Appl. Physiol. Nutr. Metab.* **2009**, *34*, 115–123. [[CrossRef](#)]
42. Buchfuhrer, M.J.; Hansen, J.E.; Robinson, T.E.; Sue, D.Y.; Wasserman, K.; Whipp, B.J. Optimizing the exercise protocol for cardio-pulmonary assessment. *J. Appl. Physiol.* **1983**, *55*, 1558–1564. [[CrossRef](#)]
43. Wasserman, K.; Hansen, J.E.; Sue, D.Y.; Whipp, B.J.; Casaburi, R. *Principles of Exercise Testing and Interpretation*, 2nd ed.; Lea and Febiger: London, UK, 1994.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).