Verification Testing to Confirm VO₂max Attainment in Inactive Women with Obesity

TODD A. ASTORINO‡, ANNIE BRYERS DE LA ROSA†, AMY CLARK†, and JAMIE L. DE REVERE*

Department of Kinesiology, California State University-San Marcos, San Marcos, CA, USA

*Denotes undergraduate student author, †Denotes graduate student author, ‡Denotes professional author

ABSTRACT

International Journal of Exercise Science 13(4): 1448-1458, 2020. Incidence of obesity is increasing worldwide which is deleterious to health due to its association with increased risk of cardiovascular disease, diabetes, and some cancers. Completion of regular physical activity in individuals with obesity increases maximal oxygen uptake (VO₂max). However, whether individuals with obesity can exhibit ‘true’ VO₂max is unresolved. This study examined efficacy of verification testing (VER) to identify ‘true’ VO₂max in 17 inactive women with obesity (age, body fat, and VO₂max = 37 ± 10 yr, 48.7 ± 3.5 %, and 19.4 ± 3.0 mL/kg/min, respectively). They performed ramp exercise (RAMP) to volitional fatigue followed by VER at 105 percent peak power output (%PPO) at baseline and after 3 and 6 wk of high intensity interval training. Results showed no difference in ramp and verification-derived VO₂max (1.99 ± 0.37 L/min vs. 1.98 ± 0.32 L/min, 2.00 ± 0.40 L/min vs. 2.04 ± 0.38 L/min, and 2.08 ± 0.34 L/min vs. 2.08 ± 0.32 L/min at 0, 3 and 6 wk of training), although in 40 % of VER tests, VO₂max was greater than the RAMP value. Overall, verification testing may be adopted as an additional approach to confirm ‘true’ VO₂max attainment in obese women as ramp exercise frequently underestimates VO₂max in this population.

KEY WORDS: Maximal oxygen uptake, ramp testing, supramaximal exercise, sedentary

INTRODUCTION

Obesity is defined as excess body fat or a body mass index above 30 kg/m² (6). In the United States, there has been a marked increase in obesity, as 30 % of adults were obese in 2000 which increased to 40 % in 2017 (7). Obesity reduces health status as it increases risk of cardiovascular disease, diabetes, and some cancers (19). In addition, the economic costs of obesity are staggering since $147 billion were spent in the treatment of obesity in 2008 (12).

Adults with obesity are recommended to complete a minimum of 150 min/wk of moderate intensity continuous training to elicit various health benefits including improvements in maximal oxygen uptake (VO₂max), which benefits activities of daily life and reduces mortality risk (5). Despite the benefits of regular physical activity, obese individuals’ participation in habitual activity is low (31) which makes reversing the co-morbidities associated with obesity challenging for clinicians.
VO$_2$max is assessed during progressive exercise to exhaustion (RAMP), and various primary and secondary criteria are commonly used to verify incidence of ‘true’ VO$_2$max (3, 20). However, these criteria have been criticized as they do not distinguish between persons who do and do not attain a ‘true’ VO$_2$max (15). Poole et al. (27) reported that maximal respiratory exchange ratio (RER > 1.10 or 1.15) occurs at intensities prior to attainment of VO$_2$max. There are over 30 separate criteria to confirm VO$_2$max attainment using predicted maximal heart rate (HR) (27), which diminishes the validity of this criterion. Recently, the verification test (VER) performed subsequent to RAMP was developed, and data acquired from older adults (9), active individuals (26, 30), inactive adults (2), and athletes (16, 20) show that this test yields similar estimates of VO$_2$max versus RAMP, hence confirming ‘true’ VO$_2$max.

In 135 overweight or obese adults with VO$_2$max between 27 – 42 mL/kg/min, Wood et al. (32) investigated attainment of various VO$_2$max criteria in response to treadmill RAMP exercise and a subsequent verification test. Data showed similar maximal values of VO$_2$ and HR between tests. Sawyer et al. (29) showed that verification testing at 100 percent peak power output (%PPO) elicited ‘true’ VO$_2$max in adults with obesity, as mean values were similar between protocols. However, many participants’ VER-derived VO$_2$max was higher than the RAMP value. Moreover, authors used a work rate equivalent to PPO for the verification test rather than above PPO as recommended (21), which does not confirm that a higher intensity does not elicit a higher VO$_2$. 

Recently, Barry et al. (5) emphasized the importance of increasing fitness rather than reducing fatness in adults, as the former seems to elicit superior health-related benefits than the latter. Moreover, Gaesser et al. (14) stated that increasing fitness rather than promoting weight loss should be the primary goal of most exercise regimens due to the difficulty in promoting long-term weight loss in most adults. Enhancing cardiorespiratory fitness in populations including the obese requires an accurate measure of fitness to monitor training-induced changes, and it is plausible that an erroneous measure of fitness may cause inaccurate decisions regarding subsequent health care.

This study examined use of verification testing to confirm VO$_2$max attainment in inactive women with obesity. We compared VO$_2$max values between RAMP and VER which required supramaximal work rates as previously recommended (20). It was hypothesized that mean VO$_2$max would not differ between protocols, but many participants would reveal higher VO$_2$max with VER versus RAMP. Identifying a ‘true’ VO$_2$max is important in every individual who completes exercise testing, as this value can be used to prescribe exercise training, assess efficacy of exercise training, and classify health risks.

**METHODS**

**Participants**

Women who were insufficiently active (< 2 h/wk of moderate to vigorous activity in the previous 12 mo) and obese (body mass index > 35 kg/m$^2$) participated in this study. All met this physical activity criterion, although some women engaged in infrequent physical activity;
whereas, others performed none. At baseline, their age, body mass index, and percent body fat were equal to 37 ± 10 yr, 39 ± 4 kg/m², and 48.7 ± 3.6 %, respectively. They were recruited from flyers placed on campus as well as word-of-mouth. All women were healthy non-smokers, were not taking medications known to alter metabolism, and lacked joint pain. Participants completed the Physical Activity Readiness Questionnaire (1) and a health history questionnaire before initiating the study. Written informed consent was obtained from participants prior to the study, and the protocol was approved by the University Institutional Review Board. This research was carried out fully in accordance to the ethical standards of the International Journal of Exercise Science (25). Data concerning changes in VO₂max, body composition, and muscular force in response to this intervention have already been published (8).

Protocol
Assessment of VO₂max: Initially, height and body mass were determined using a balance scale and stadiometer, and then body composition was measured using air displacement plethysmography (BodPod, COSMED USA Inc., Chicago, IL). Participants performed RAMP exercise on an electrically-braked cycle ergometer (Velotron DynaFit Pro, RacerMate, Seattle, WA). Power output started at 40 W for 2 min and was increased by 20 W/min until exhaustion which was identified by cadence < 50 rev/min. Gas exchange data were acquired every 15 s with a metabolic cart (ParvoMedics TrueOne, Sandy, UT) which was calibrated pre-exercise according to manufacturer’s recommendations. Change in VO₂ at VO₂max was identified as the difference between the highest consecutive VO₂ values during the last 60 s of exercise (29). During exercise, HR was continually measured via telemetry (Polar Electro, Woodbury, NY). PPO was identified as the work rate consequent with volitional fatigue. After a 10 min active recovery at 20 W, women were told to increase their cadence, and VER ensued at 105 %PPO to confirm incidence of ‘true’ VO₂max. This test was stopped when pedal cadence dropped below 50 rev/min. Maximal oxygen uptake was determined from both tests as the mean of the two highest consecutive values within the last four measurements. Data presented here are from a total of 51 assessments of VO₂max, which were acquired at baseline, during, and after cessation of training.

High intensity interval training: Sessions were performed in the lab 2 d/wk on the same cycle ergometer as well as 1 d/wk on their own, with at-home training following the structure of lab-based training for that week. Each session was preceded by a 5 min warm up at 20 %PPO. Training structure was implemented to cater to a very unfit population with likely negative attitudes to exercise training. Consequently, it could not require a huge amount of time or be so exhaustive that would potentially reduce compliance. Moreover, training structure changed frequently to provide the women with a novel stimulus that may enhance their compliance to training. During lab-based training, HR was continuously assessed using telemetry (Polar, Lake Success, NY); whereas, at-home training intensity (> 85 %HRmax) was confirmed using a downloadable monitor (Polar, Lake Success, NY). Training intensities were adjusted after session 9 based on a second VO₂max assessment. The training regimens are shown in Table 1.
Identifying individual differences in VO$_2$max and HRmax between protocols: We used a typical error score acquired from repeated testing of VO$_2$max using the identical exercise protocol 3 ($n = 14$) and 6 wk apart ($n = 7$) in young sedentary women (age and VO$_2$max = 23 ± 2 yr and 29 ± 3 mL/kg/min) to ascertain individual differences in RAMP and VER-derived VO$_2$max. The typical error was equal to 0.03 and 0.06 L/min for ramp and verification testing, respectively. Consequently, we used a conservative difference in VO$_2$max between protocols < 0.06 L/min to identify ‘true’ VO$_2$max. Difference in HRmax between protocols of < 2 b/min (7) was used to determine that the ramp-derived value was maximal.

Statistical Analysis
Data are expressed as mean ± SD and were analyzed using SPSS 24.0 (Armonk, NY). Paired t-test was used to compare maximal HR and gas exchange variables between RAMP and VER. One-way analysis of variance with repeated measures was used to examine differences in gas exchange variables across time (0, 3, and 6 wk). If a significant F ratio was obtained, Tukey’s post hoc test was used to identify differences between means. The Pearson product moment coefficient was used to examine pairwise associations between variables, and intraclass correlation coefficients (ICC) were used to examine the reliability of VO$_2$ and HR between protocols. Significance was set at $p < 0.05$.

RESULTS
Differences in VO$_2$max and HRmax between RAMP and VER: Table 2 shows differences in VO$_2$max and HRmax between protocols. Data revealed no difference in VO$_2$max between RAMP and VER at 0 ($p = 0.90$), 3 ($p = 0.34$), and 6 wk ($p = 0.80$) of training. Also, the absolute difference in VO$_2$max between protocols at 0 (-0.01 ± 0.14 L/min), 3 (0.04 ± 0.17 L/min), and 6 wk (0.01 ± 0.11 L/min) was insignificant ($p = 0.61$). Change in VO$_2$ at VO$_2$max was unchanged during the

<p>| Table 1. Description of the high intensity interval training performed in the study. |
|---------------------------------|-----------------|-----------------|-----------------|---------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Traditional</th>
<th>Number of Bouts</th>
<th>Bout Duration (seconds)</th>
<th>Rest Duration (seconds)</th>
<th>Intensity (% PPO)</th>
<th>Warm Up (minutes)</th>
<th>Total Time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>10</td>
<td>60</td>
<td>60</td>
<td>70</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Week 2</td>
<td>10</td>
<td>60</td>
<td>60</td>
<td>75</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Week 3</td>
<td>10</td>
<td>60</td>
<td>60</td>
<td>80</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Week 4</td>
<td>10</td>
<td>60</td>
<td>60</td>
<td>75</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Week 5</td>
<td>10</td>
<td>60</td>
<td>60</td>
<td>80</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Week 6</td>
<td>10</td>
<td>60</td>
<td>60</td>
<td>85</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Periodized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 1</td>
<td>10</td>
<td>60</td>
<td>60</td>
<td>70</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Week 2</td>
<td>6</td>
<td>20</td>
<td>120</td>
<td>105</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>Week 3</td>
<td>7</td>
<td>120</td>
<td>60</td>
<td>60</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td>Week 4</td>
<td>10</td>
<td>60</td>
<td>60</td>
<td>75</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Week 5</td>
<td>6</td>
<td>20</td>
<td>120</td>
<td>110</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>Week 6</td>
<td>7</td>
<td>120</td>
<td>60</td>
<td>65</td>
<td>5</td>
<td>26</td>
</tr>
</tbody>
</table>
study \((p = 0.06)\) and was equal to \(0.08 \pm 0.13, -0.01 \pm 0.11,\) and \(-0.01 \pm 0.08 \text{ L/min}\) at 0, 3, and 6 wk. Verification duration increased during the study \((p = 0.04)\) and was higher at 3 and 6 wk of training versus baseline. Intraclass correlation coefficients were equal to 0.96, 0.95, and 0.97 between RAMP and VER-derived VO\(_2\)max. Figure 1 demonstrates a positive and significant \((r = 0.92, p < 0.001)\) association between VO\(_2\)max values from both protocols across all timepoints of the study \((n = 51)\). The mean difference was equal to 0.01 L/min across all tests, with limits of agreement equal to -0.26 to 0.28 L/min, respectively (Figure 2).

Maximal HR was not different at any time point between RAMP and VER \((p = 0.25, 0.46, \text{ and } 0.41)\), respectively, and difference in HRmax was similar \((p = 0.20)\) between RAMP and VER at 0 \((-1.9 \pm 5.7 \text{ b/min})\), 3 \((1.4 \pm 7.0 \text{ b/min})\), and 6 wk \((-1.3 \pm 5.1 \text{ b/min})\) of training. The ICC for HRmax between protocols was equal to 0.97, 0.92, and 0.96, respectively.

### Table 2. Maximal gas exchange data from ramp and verification testing in women with obesity (mean ± SD).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Baseline</th>
<th>3 wk</th>
<th>6 wk</th>
</tr>
</thead>
<tbody>
<tr>
<td>VO(<em>2)max(</em>\text{ramp}) (L/min)</td>
<td>1.99 ± 0.37</td>
<td>2.00 ± 0.40</td>
<td>2.08 ± 0.34*</td>
</tr>
<tr>
<td>VO(<em>2)max(</em>\text{VER}) (L/min)</td>
<td>1.98 ± 0.32</td>
<td>2.04 ± 0.38</td>
<td>2.08 ± 0.32*</td>
</tr>
<tr>
<td>HRmax(_\text{ramp}) (b/min)</td>
<td>177.3 ± 14.6</td>
<td>171.8 ± 15.8</td>
<td>175.2 ± 12.0</td>
</tr>
<tr>
<td>HRmax(_\text{VER}) (b/min)</td>
<td>175.8 ± 13.9</td>
<td>173.2 ± 12.1</td>
<td>174.1 ± 12.2</td>
</tr>
<tr>
<td>RERmax(_\text{ramp})</td>
<td>1.33 ± 0.08</td>
<td>1.24 ± 0.07</td>
<td>1.26 ± 0.06</td>
</tr>
<tr>
<td>RERmax(_\text{VER})</td>
<td>1.17 ± 0.12*</td>
<td>1.13 ± 0.10*</td>
<td>1.14 ± 0.06*</td>
</tr>
<tr>
<td>V(<em>E)max(</em>\text{ramp}) (L/min)</td>
<td>87.7 ± 18.7</td>
<td>81.3 ± 19.7</td>
<td>86.7 ± 16.1</td>
</tr>
<tr>
<td>V(<em>E)max(</em>\text{VER}) (L/min)</td>
<td>87.0 ± 17.0</td>
<td>82.2 ± 17.4</td>
<td>87.0 ± 16.6</td>
</tr>
<tr>
<td>Duration(_\text{VER}) (min)</td>
<td>1.38 ± 0.27</td>
<td>1.60 ± 0.34*</td>
<td>1.54 ± 0.30*</td>
</tr>
</tbody>
</table>

\(a = p < 0.05\) versus baseline; VER = verification test; HR = heart rate; RER = respiratory exchange ratio; \(*= p < 0.05\) versus ramp within time; \(V_E\) = ventilation

Differences in gas exchange data between RAMP and VER: No difference occurred in V\(_E\)max between protocols \((p > 0.76)\). Yet, RERmax was higher \((p < 0.001)\) in response to RAMP versus VER (Table 2).

Individual differences in VO\(_2\)max and HRmax between RAMP and VER: Despite no mean differences in VO\(_2\)max between protocols, many participants ‘true’ VO\(_2\)max was not attained via RAMP. At 0, 3, and 6 wk, 5 (30 %), 9 (53 %), and 7 (41 %) women revealed a verification VO\(_2\)max value that was \(> 0.06 \text{ L/min}\) higher than obtained from RAMP. Three women (mean VO\(_2\)max = 18.3 mL/kg/min) exhibited consistent underestimations of VO\(_2\)max in response to RAMP, as their verification-derived value was 0.09 ± 0.02 L/min (4.5 %) higher at 0, 3, and 6 wk of training. Across all tests, differences in VO\(_2\)max between RAMP and VER ranged from \(-0.45 \text{ to } -0.35 \text{ L/min}\). There was no association between VO\(_2\)max \((r = -0.20, p = 0.15)\) or verification duration \((r = 0.19, p = 0.18)\) and the difference in VO\(_2\)max between protocols.

Data show that 4 (23 %), 7 (41 %), and 5 (30 %) participants exhibited a difference in HRmax > 2 b/min in response to VER versus RAMP. This difference ranged from \(-9 \text{ to } -18 \text{ b/min}\) across all tests.
Figure 1. Association between VO\textsubscript{2}\text{max} achieved on the incremental trial compared to VO\textsubscript{2}\text{max} achieved on the verification trial.
Figure 2. Bland-Altman plot with mean VO\textsubscript{2}max on the x-axis and the difference between ramp and verification-derived VO\textsubscript{2}max on the y-axis. Solid line = mean difference between verification and ramp VO\textsubscript{2}max = 0.01 L/min; dashed line = mean ± 1.96 SD.

DISCUSSION

Low VO\textsubscript{2}max is associated with enhanced morbidity and mortality (24) which is why improving this outcome is a primary goal of many exercise regimens. Nevertheless, it is often difficult to confirm ‘true’ VO\textsubscript{2}max using established criteria as they are unable to distinguish between individuals who do and do not reveal a ‘true’ VO\textsubscript{2}max (2, 27). We tested the efficacy of repeated verification testing in obese women with a VO\textsubscript{2}max almost 40 % below age-matched norms (18) who performed chronic interval training (8). Data show that despite no difference in mean VO\textsubscript{2}max or HR\textsubscript{max} between protocols at any time point, a large amount of RAMP tests underestimate VO\textsubscript{2}max which may warrant verification testing in populations with obesity when a ‘true’ VO\textsubscript{2}max value is needed.
Our data support others (9, 13, 16, 29) showing that at a group level, there is no difference in VO2max between RAMP and VER, which suggests that the value acquired from ramp testing is indeed maximal. Yet, VO2max is measured on an individual level to measure cardiorespiratory fitness and in turn, classify mortality risk, prescribe exercise regimens, and monitor responses to exercise training, so only comparing mean values is impractical. Our cutoff value acquired from repeated maximal exercise testing in inactive women demonstrated that 40 % of RAMP tests underestimated VO2max. A 2 % cutoff value was previously used (29) which did not appear to be acquired from repeated testing, and results showed that 70 % of participants revealed a higher VO2max in VER versus RAMP. If we used this less conservative criterion in our sample, 52 % of VER tests exhibit a higher VO2max (> 0.04 L/min) versus RAMP. Rather than utilizing some arbitrary value, we recommend that authors develop and report typical error values for both ramp and verification-derived VO2max measured in their own lab, as previously utilized (15, 30), to better characterize individual differences in VO2max between protocols.

Muriyas et al. (23) reported that RAMP was able to confirm VO2max attainment in active young (age and VO2max 25 ± 4 yr and 49 ± 8 mL/kg/min) and older men (age and VO2max = 65 ± 5 yr and 31 ± 7 mL/kg/min), as data showed no difference in mean VO2max between RAMP and verification testing at 85 and 105 %PPO. However, their study has a few methodological differences versus ours. First, their participants had a VO2max that was 50 to > 100 % higher than our population, and well above average for men this age (18). Previous work shows that persons with lower fitness are more apt to show a significant underestimation of VO2max in response to ramp testing than more fit individuals (4), so it is possible that their conclusions stem from the high fitness level of their sample. It is plausible that persons with low cardiorespiratory fitness terminate ramp exercise prematurely due to onset of leg pain, breathlessness, or general discomfort which may elicit underestimation in VO2max, yet when allowed to perform a subsequent verification test, they are more familiar with the effort required and may actually reveal a ‘true’ VO2max. In addition, Jones et al. (17) stated that incremental testing may augment oxygen kinetics in the subsequent verification test, leading to a higher VO2 value. Second, Muriyas et al. (23) only presented mean data and no individual data were reported. Our results and others (2, 15, 29) show that many participants show higher VO2max from verification testing compared to RAMP, despite no aggregate differences in VO2max.

Results from Misquita et al. (22) in 108 postmenopausal women (age, VO2max, and BMI = 60 ± 6 yr, 19 ± 3 mL/kg/min, and 33 ± 4 kg/m2) showed low incidence of a VO2 plateau as well as attainment of secondary criteria including maximal HR (220 – age) and RER > 1.10. Their data also indicated that women with a higher VO2max tended to attain these criteria versus those with lower VO2max, which is supported by results acquired from active older adults (9). Similarly, other data (10) reported low (< 60 %) attainment of VO2max criteria in obese women, and recommended that alternative criteria be used to verify incidence of VO2max. All our participants except one exhibited RER > 1.10 in all tests, but in only 9 of 51 tests (18 %) did they attain the HRmax criterion equal to 10 b/min within 220 – age, which suggests that this is not able to confirm maximal effort in this population.
This study has a few limitations. First, VO₂max was tested only once at each timepoint, and recent data acquired in young individuals (11) show small but significant increases in VO₂max across repeated testing. Second, our data apply to healthy women with obesity and not those with comorbidities including diabetes or cardiovascular disease who may be intolerant of supramaximal exercise required in verification testing. Third, our cutoff value signifying a meaningful difference in VO₂max between protocols was developed from inactive women whose VO₂max is higher than that of our sample. However, its magnitude of approximately 3% is similar to that reported in older adults (9), but lower than a value used in younger adults (29). Further study is merited to identify test-retest reliability of verification testing in individuals with obesity, as to our knowledge, this value is unknown.

In conclusion, our findings obtained in insufficiently active obese women undergoing incremental exercise testing show that 40% of tests reveal a higher VER-derived VO₂max value, which supports its implementation in this population as another criterion to confirm attainment of VO₂max, especially when monitoring changes in cardiorespiratory fitness acquired in response to training. Verification testing completed at supramaximal workloads was well-tolerated and did not cause any adverse events, so this test appears to be safe in obese women with low cardiorespiratory fitness. However, the magnitude of difference in VO₂max is small (~0.7 mL/kg/min), so the clinical relevance of this finding is unknown and requires further study.

ACKNOWLEDGEMENTS

This study was funded by a University Professional Development grant. We thank the participants for their dedication to the study and its requirements.

REFERENCES


6. Centers for Disease Control and Prevention. Division of Nutrition, Physical Activity, and Obesity, National Center for Chronic Disease Prevention and Health Promotion; 2017.


23. Murias JM, Pogliaghi S, Paterson DH. Measurement of a true VO2max during a ramp incremental test is not confirmed by a verification phase. Front Physiol 9: 143, 2018.


