Original Research

Validity and Reliability of the VO2 Master Pro for Oxygen Consumption and Ventilation Assessment

ALEXANDER H.K. MONTOYE‡1, JOSEPH D. VONDRASEK*1, and JAMES B. HANCOCK II‡2

1Department of Integrative Physiology and Health Science, Alma College, Alma, MI, USA; 2Department of Education, Alma College, Alma, MI, USA

*Denotes undergraduate student author, ‡Denotes professional author

ABSTRACT

International Journal of Exercise Science 13(4): 1382-1401, 2020. This study assessed validity and reliability of the VO_2 Master Pro portable metabolic analyzer for assessment of oxygen consumption (VO_2) and minute ventilation (VE). In Protocol 1, eight male participants (height: 182.6 ± 5.8 cm, weight: 79.6 ± 8.3 kg, age: 41.0 ± 12.3 years) with previous competitive cycling experience completed an hour-long stationary cycling protocol twice, progressing from 100-300 Watts every 10 minutes while wearing the VO_2 Master and a criterion measure (Parvomedics) for five minutes each, at each stage. In Protocol 2, 16 recreationally active male participants (height: 168.2 ± 8.4 cm, weight: 76.5 ± 13.3 kg, age: 23.0 ± 9.4 years) completed three incremental, maximal stationary cycling tests wearing one of three analyzers for each test (VO_2 Master version 1.1.1, VO_2 Master version 1.2.1, Parvomedics). For Protocol 1 and convergent validity, the VO_2 Master had mean absolute differences from the Parvomedics of <0.3 L/min for absolute VO_2 and <5 L/min for VE overall and at each exercise stage. Mean absolute percent differences (MAPD) for VO_2 and VE were <9% overall and <12% at each stage. Test-retest reliability of the VO_2 Master (MAPD: 8.9-10.9%) was somewhat poorer than the Parvomedics (MAPD: 5.3-7.6%). For Protocol 2, validity was similar for both VO_2 Master models (MAPD ~12% overall) compared to the Parvomedics for VO_2 and VE. The VO_2 Master had an acceptable validity and test-retest reliability for most intensities tested and may be an appealing option for field-based VO_2 and VE analysis.

KEY WORDS: Metabolism, indirect calorimetry, VO2max, metabolic analysis, portable metabolic analyzer

INTRODUCTION

Measurement of oxygen consumption (VO_2) during rest and exercise in laboratory and field settings is of fundamental importance for areas including general health, diagnosis of chronic disease, exercise training, and athletic performance. Early automated metabolic systems were not portable, limiting testing to either resting measures or measures that could be performed on a stationary ergometer (e.g., treadmill, cycle) in a laboratory setting. However, since the late 1980's several user-worn, portable systems capable of capturing VO_2 have emerged, allowing

for important metabolic and caloric expenditure analyses to be performed in field-based settings (15).

Brands of portable analyzers including COSMED, Oxycon, and MetaMax have received extensive validation and field use in a variety of settings, providing evidence of the importance of capturing such field-based data (19, 24). However, their cost, which at $30,000-$50,000+ is more than main brands of laboratory-based metabolic analyzers, may preclude their use in fitness settings, by individual sports teams, and at smaller colleges/universities. These types of organizations need a less costly option capable of performing high-quality metabolic analyses.

Full analysis of caloric expenditure and macronutrient fuel utilization requires measurement of both VO_2 and carbon dioxide consumption (VCO_2) for determination of respiratory exchange ratio (12). However, several handheld analyzers capable of measuring resting VO_2 (but not VCO_2) are available at a fraction of the cost of fully equipped analyzers. Even without VCO_2 measurement, such devices have been shown to provide more accurate measures of resting caloric expenditure than popular equations such as the Harris-Benedict (9) or Mifflin St. Jeor (21) equations (10, 27, 28). Accordingly, such devices may have utility for measuring resting caloric expenditure and metabolic rate in weight-loss or hospital settings (10). However, these devices are not designed to be able to assess energy expenditure during exercise and are, therefore, unsuitable for performance testing in fitness or athletic realms.

A recently developed, portable metabolic analyzer, the VO2 Master Pro (VO2 Master Health Sensors Inc., Vernon, British Columbia, CA), is reportedly capable of measuring VO_2 (but not VCO_2) across a variety of intensities from rest to maximal exercise. At a cost of <$5,000 (https://vo2master.com/shop/) and with a size smaller than the major portable metabolic analyzer brands, the VO2 Master Pro is an appealing option for VO_2 and minute ventilation assessment. According to the manufacturer, the VO2 Master Pro is accurate to ±3% O_2 and ±3% flow (ventilation) compared to a breathing simulator (17). However, we are not aware of independent validations of the VO2 Master analyzer. With development of such technologies, it is important for end-users to understand their validity and reliability. Therefore, the purpose of this study was to assess validity and reliability (both test-retest and inter-device) of the VO2 Master metabolic analyzer in a laboratory setting.

METHODS

Protocol 1: Hour test protocol

Participants
Eight males with previous cycling experience (range 2-40 years, average 18.5 years) completed what will hereafter be called the "hour test protocol." This protocol design allowed for assessment of VO2 Master validity and test-retest reliability. For this protocol, potential participants were invited to complete the study if they were confident they could complete 10 minutes of cycling at 250 Watts (W), with the ideal participant able to complete 10 minutes at 300 W. Participants had to be between the ages of 18-80 years and able to safely participate in

vigorous-intensity exercise. To ensure participant safety, participants completed a physical activity readiness questionnaire (PAR-Q). Any potential participants with risks identified by the PAR-Q were required to provide written physician approval for high-intensity exercise prior to participation. This research was carried out fully in accordance to the ethical standards of the International Journal of Exercise Science (22). Additionally, the college's Institutional Review Board approved all study protocols before testing began, and participants gave their written informed consent prior to participation. Demographic data for the hour test protocol are included in Table 1.

Table 1. Participant demographics for hour test protocol.

Age (years)	41.0 (12.3)
Height (cm)	182.6 (5.8)
Weight (kg)	79.6 (8.3)
Body mass index (kg/m²)	24.1 (2.0)

Data are shown as mean (standard deviation).

Protocol

Parvomedics analyzer: The Parvomedics TrueOne 2400 metabolic cart (Parvomedics, Inc., Salt Lake City, UT, USA) was used as a comparative measure of VO_2 and minute ventilation. Previous research has shown the Parvomedics to accurately measure both VO_2 and ventilation across a range of intensities compared to the gold-standard Douglas bag method, rendering the Parvomedics a suitable criterion measure for comparison in this study (1).

The Parvomedics was prepared using standard procedures, including a 30-minute warm-up period prior to calibration. Gas calibration was completed first and according to manufacturer settings. Ambient temperature, humidity, and barometric pressure were measured using a FanJu weather station (FanJu Electronics Technology Company, Ltd., Fujian, China) and input into the software, and gas calibration was completed with the machine sampling room air and then a known gas mixture of 16% O_2 and 4% CO_2. The gas calibration was completed until results were within ±1% for both CO_2 and O_2 with a total difference of ≤1%. Next, the flowmeter calibration was completed using a 3 L flow syringe, using one detection stroke, four flush strokes, and five calibration strokes at flow rates of 50-80 L/min, 100-199 L/min, 200-299 L/min, 300-399 L/min, and 400-599 L/min, as per manufacturer recommendations. The flowmeter calibration was completed until the results were within ±1%. A Hans Rudolph (Hans Rudolph, Inc., Shawnee, KS, USA) 2700B two-way non-rebreathing valve system was used, where all inhaled air came from the ambient environment, and all exhaled air flowed through a tube into the mixing chamber of the Parvomedics system. Participants were connected to the Hans Rudolph valve via a mouthpiece, and noseclips were used to ensure all exhaled air was captured by the system. A headgear setup (Model 2726; Hans Rudolph, Inc., Shawnee, KS, USA) held the mouthpiece and valve in place. Total dead space in the mouthpiece/valve setup was ~93 ml. The Parvomedics records breath-by-breath ventilation, VO_2, and VCO_2; however, VCO_2 was not used in the present study.

VO2 Master Pro analyzer: The VO2 Master Pro, version 1.1.1, was used for the hour test protocol. The unit and mask weigh 0.32 kg, with power provided by a single AAA battery. The analyzer

was connected to a Hans Rudolph 7450 V2, over-nose mask for wear using a mask adapter "user piece" (a plastic tube that connects the analyzer and mask, with an exhaust hole for air entry and exit), and a soft headgear was used to secure the unit to the participants' faces. There are two manufacturer-supplied user piece sizes, one that allows a flow rate of 30-160 L/min, and one that allows for a flow rate of 40-220 L/min. For all trials in both protocols, the 30-160 L/min user piece was used. A single-use filter was placed between the user piece and the analyzer, as per manufacturer instructions, and replaced following each test. Total dead space in the mask/user piece setup was ~125 ml.

The VO2 Master unit contains a passive, pump-less system for gas sampling, a galvanic fuel cell O_2 sensor, and a differential pressure flow sensor (personal communications with manufacturer). The VO2 Master automatically calibrated to the ambient air for gas concentrations, ambient temperature, humidity, and barometric pressure when it was turned on (the device does not calibrate to other O_2 concentrations). Following gas calibration, the VO2 Master was fitted to participants, and participants were instructed to take 10-15 deep breaths for flowmeter calibration. Once the protocol was started, the device recalibrated automatically at 5 minutes and 25 minutes (took ~30-45 seconds each time). The device measures breath-by-breath ventilation and VO_2 but does not have a CO_2 sensor and, therefore, cannot capture VCO_2. The VO2 Master transmitted data via Bluetooth to an iPod (Apple, Inc. Cupertino, CA, USA) equipped with the VO2 Master mobile application for storage and later download.

Cycling equipment: A Monark Ergomedic Stationary 894E cycle (Monark Exercise, Vansbro, Sweden) was used during the hour test protocol. This cycle does not need calibration as weights are added to a 1kg basket to adjust pedaling resistance. Participants had the option to use the standard pedals that came with the cycle or use their own (such as clipless pedals and shoes). The 894E has resistance increments of 0.1 kg. Power output (in W) was calculated as W = resistance (kg) * revolutions/min.

Other equipment: Participants wore a Polar H10 heart rate monitor (Polar Electro Oy, Kempele, Finland) on the chest, just below the xiphoid process. The monitor transmitted real-time heart rate data to a mobile phone (Samsung Galaxy S7; Samsung, Seoul, South Korea) equipped with the Polar Beat (version 3.3.6) mobile application. Participant weight was taken at the beginning of each visit, while wearing shoes and clothing to be worn during the cycling, using a Taylor Precision Scale (Taylor USA, Entrada Del Con, New Mexico).

Cycling protocol: Participants (N=8) completed two separate hour test protocols, spaced at least 48 hours apart and performed as close to the same time of day as possible. For each test, participants were asked to refrain from exercise for least 12 hours and caffeine/food for at least 3 hours prior to testing. Upon arrival, participants were fitted on the Monark 894E, attached their own pedals if desired, and completed a short (~10 minute), self-directed warm-up to ensure fit and comfort on the cycle. Following the warmup, the incremental test protocol started at 100 W and increased by 50 W (100, 150, 200, 250, 300 W) every 10 minutes. At every stage, participants wore the VO2 Master and the Parvomedics analyzer for five minutes each, with the order of analyzer wear randomized for each workload and counterbalanced between visits to

eliminate any potential ordering effects due to VO_2 drift or the slow component of VO_2. An overview of the hour test protocol can be found in Table 2. Research staff managed the switching of analyzers, which took ~45 seconds. While the analyzers were being switched, participants had the option of maintaining pedaling cadence or resting. If participants continued pedaling, the stage timer would continue; conversely, if they elected to stop pedaling during the switching process, the stage timer would be stopped until the switch was completed and the participants resumed the necessary pedaling cadence.

In the last ~30 seconds of wearing each analyzer at each stage (i.e., every 5 minutes), steady-state heart rate and rating of perceived exertion (Borg 6-20 scale) were recorded (3). Participants were allowed to choose their ideal cycling cadence (revolutions/min), and the resistance was adjusted accordingly so that the participants could stay as close as possible to their ideal cadence at each stage. For both visits, participants were required to use the same pedals and cadence/resistance settings at each stage. If participants deviated from the required cadence by more than ±2 revolutions/minute, they received verbal reminders by the research staff. Toward the end of the protocol, if participants could not maintain the necessary cadence after two consecutive reminders by the research staff, the protocol was ended.

Due to the extended length of the hour test protocol, participants were allowed to listen to music if they desired. Additionally, participants were given the option to have a box fan placed ~4 meters directly in front of them and set to the lowest setting. Pilot testing prior to the study starting showed that the fan had no apparent effect on VO_2 or ventilation for any fan speed when used at a distance of 3 or more meters. If participants chose to use music and/or fan for their first test, the same music/fan settings were used in the second test.

Table 2. Overview of the hour test protocol.

Time	Watts	Day 1 (N=8) analyzer#	Day 2 (N=8) analyzer
10-minute warm-up	Self-selected	None	None
0-5 minutes	100	Parvomedics	VO2 Master
5-10 minutes	100	VO2 Master	Parvomedics
10-15 minutes	150	VO2 Master	Parvomedics
15-20 minutes	150	Parvomedics	VO2 Master
20-25 minutes	200	Parvomedics	VO2 Master
25-30 minutes	200	VO2 Master	Parvomedics
30-35 minutes	250	VO2 Master	Parvomedics
35-40 minutes	250	Parvomedics	VO2 Master
40-45 minutes	300	Parvomedics	VO2 Master
45-50 minutes	300	VO2 Master	Parvomedics

#The order of analyzer wear was randomized and counterbalanced across participants. This table provides an example protocol for one participant.

Statistical analysis
Since gas sampling did not occur simultaneously for the two analyzers (which would allow for criterion validity assessment), we opted for assessing convergent validity by comparing outputs from analyzers during the same activity intensities but sampled at different times. We were also able to determine test-retest reliability for the VO2 Master analyzers using data from the hour test protocol. For the reliability analysis, the Parvomedics data served as a comparison to better

understand the extent to which the observed differences between analyzers were due to day-to-day variability in physiologic function vs. due to reliability of the devices.

Absolute VO_2 (L/min) and minute ventilation (L/min) data from the Parvomedics and VO2 Master analyzers were reintegrated to 30-second intervals for analysis, and both heart rate and RPE were measured at each stage to confirm equal physiologic and perceived workload when wearing the different monitors. Fraction of expired oxygen consumption and relative VO_2 data were also collected but, for conciseness, are mentioned but not fully presented. Using G Power (v 3.0.10), the minimum sample size needed to obtain a significant correlation of ≥0.80 with 80% power is n=7, so both test protocols were adequately powered for determination of high between-analyzer correlations.

To determine validity of the VO2 Master version 1.1.1, data from minutes 3-5 and 8-10 (the final two minutes of each stage, for each analyzer), were averaged across both visits and compared using paired-samples t-tests. Additionally, Pearson correlation coefficients were calculated overall, and mean absolute differences and mean absolute percent differences were calculated both overall and for each stage between the Parvomedics and VO2 Master version 1.1.1.

To determine test-retest (intra-device) reliability of the Parvomedics and VO2 Master version 1.1.1 analyzers, the final two minutes of data from each stage were first averaged for each analyzer. Then, Pearson correlation coefficients were calculated overall, and mean absolute differences and mean absolute percent differences were calculated overall and at each stage between the first and second protocols for each analyzer. Additionally, coefficients of variation were calculated for each analyzer and each outcome variable to determine stability in analyzer output.

Mean absolute percent differences were arbitrarily considered acceptable when they were <10%, in accordance with past work validating other field-based measurement tools (23). All statistical analyses were conducted in SPSS version 24.0 (IBM Corp., Armonk, NY, USA). A p value of p<0.05 paired t-tests analysis was used to denote statistical significance, and 0.05≤p≤0.10 was used to denote non-significant trends.

Protocol 2: Maximal test protocol

Participants
Sixteen recreationally active males (with or without cycling experience) completed what will hereafter be called the "maximal test protocol." This protocol design also allowed for assessment of VO2 Master validity as well as inter-monitor reliability. Demographic data for these participants are in Table 3. Participants had to be between the ages of 18-80 years and able to safely participate in vigorous-intensity exercise. To ensure participant safety, participants completed a PAR-Q, and participants selecting "Yes" to any of the questions were required to obtain written physician approval for high-intensity exercise prior to participation. This protocol was also approved by the college's Institutional Review Board approved and was carried out fully in accordance to the ethical standards of the International Journal of Exercise Science (22).

Table 3. Participant demographics for maximal test protocol.

Age (years)	23.1 (9.4)
Height (cm)	168.0 (8.4)
Weight (kg)	76.5 (13.3)
Body mass index (kg/m^2)	25.4 (3.3)

Data are shown as mean (standard deviation).

Protocol
The Parvomedics TrueOne 2400 metabolic cart was used as a comparative measure of VO$_2$ and minute ventilation. Calibration of the Parvomedics and mouthpiece use were identical to the hour test protocol. The VO2 Master Pro, versions 1.1.1 and 1.2.1, were used the maximal test protocol. Device specifications and calibration are identical for versions 1.1.1 and 1.2.1, and the calibration procedure was identical to that used in the hour test protocol.

A Monark Ergomedic Stationary 828E cycle (Monark Exercise, Vansbro, Sweden) was used for the maximal test protocol. The cycle was calibrated for resistance before testing began using 0, 2, and 4 kg weights. The participants used the stock pedals that came with the cycle and adjusted the standard pedal clips to match their foot size. The 828E has resistance increments of 0.5 kg. Power output (in W) was calculated as W = resistance (kg) * revolutions/min.

Participants wore a chest-based Polar H10 heart rate monitor, which transmitted real-time heart rate data to a mobile phone equipped with the Polar Beat application. Finally, participant weight was assessed in the same way as for the hour test protocol.

Participants (N=16) completed two maximal exercise tests, spaced at least 48 hours apart and performed as close to the same time of day as possible. For each test, participants were asked to refrain from exercise for at least 12 hours and caffeine/food for at least 3 hours prior to testing. When participants reported to the laboratory, they were fitted on the Monark 828E and completed a short (~5 minute), self-directed warm-up. The maximal exercise tests started at 50 W and increased by 50 W increments every three minutes until exhaustion. An overview of the maximal test protocol is in Table 4. With 30 seconds left in each stage, steady-state heart rate and rating of perceived exertion (Borg 6-20 scale) were recorded. Participants were allowed to choose their ideal cycling cadence (revolutions/min), and the resistance was adjusted accordingly so that the participants could stay close to their ideal cadence at each stage. If participants did not have an ideal cadence, the research staff helped them find a comfortable cadence. The same cadence was used for each test. If participants deviated from the required cadence by more than ±2 revolutions/minute, they received verbal reminders by the research staff. Toward the end of the protocol, if participants could not maintain the necessary cadence after two consecutive reminders by the research staff, the protocol was ended. The maximal time and workload were noted at the end of each test. For one maximal test, participants wore the Parvomedics, and for the other test participants wore the VO2 Master (version 1.1.1); the order of wearing these analyzers was randomized.

Near the end of data collection for the maximal test protocol, the VO2 Master version 1.2.1 was released, and we were able to conduct a third maximal exercise test on 13 of the 16 participants

with the newer version of the analyzer. Note that the test using the VO2 Master version 1.2.1 was the last test for all participants and, therefore, was not randomized for order as the first two tests were. The inclusion of version 1.2.1 allowed inter-device reliability of the VO2 Master to be determined.

Table 4. Overview of the maximal test protocol.

Time*	Watts	Day 1 (N=16) analyzer@	Day 2 (N=16) analyzer	Day 3 (n=13) analyzer
5-minute warm-up	Self-selected	None	None	None
0-3 minutes	50	Parvomedics	VO2 Master v1.1.1	VO2 Master v1.2.1
3-6 minutes	100	Parvomedics	VO2 Master v1.1.1	VO2 Master v1.2.1
6-9 minutes	150	Parvomedics	VO2 Master v1.1.1	VO2 Master v1.2.1
9-12 minutes	200	Parvomedics	VO2 Master v1.1.1	VO2 Master v1.2.1
12-15 minutes	250	Parvomedics	VO2 Master v1.1.1	VO2 Master v1.2.1
15-18 minutes	300	Parvomedics	VO2 Master v1.1.1	VO2 Master v1.2.1
18-21 minutes	350	Parvomedics	VO2 Master v1.1.1	VO2 Master v1.2.1

*Participants completed as many stages as possible, until exhaustion.
@The order of analyzer wear was randomized for the Parvomedics and VO2 Master v1.1.1, but all participants who completed the VO2 Master v1.2.1 protocol did so after the other protocols.

Statistical analysis
As with the hour test protocol, we assessed convergent validity of the VO2 Master analyzers with the Parvomedics. We were also able to determine inter-device reliability for the VO2 Master analyzers using data from the maximal test protocol. Absolute VO_2 (L/min) and minute ventilation (L/min) data from the Parvomedics and VO2 Master analyzers were reintegrated to 30-second intervals for analysis, and both heart rate and RPE were measured at each stage to confirm equal physiologic and perceived workload when wearing the different monitors. Data from minutes 2-3 (the last minute) of each stage were taken as steady-state values. The minute ventilation rates for >90% of the participants were <30 L/min during the 50 W stage, so data were not captured by the VO2 Master. Thus, the lowest intensity at which we analyzed data was 100 W.

To determine convergent validity of the VO2 Master analyzers, data from the three analyzers (Parvomedics, VO2 Master version 1.1.1, VO2 Master version 1.2.1) were compared using repeated-measures analysis of variance (RMANOVA) with a false discovery rate correction for multiple comparisons (7); a p value of p<0.05 was used to denote statistical significance, and 0.05≤p≤0.10 was used to denote non-significant trends. Additionally, Pearson correlation coefficients were calculated overall, and mean absolute differences and mean absolute percent differences were calculated between the Parvomedics and each VO2 Master analyzer both overall and at each stage. Inter-device reliability of the VO2 Master analyzers was determined by calculating Pearson correlation coefficients overall as well as mean absolute differences and mean absolute percent differences overall and at each stage between the version 1.1.1 and 1.2.1 of VO2 Master analyzer.

RESULTS

Hour test protocol: All eight participants were able to complete the 250 W stage in both hour-long tests, and five participants were able to complete the 300 W stage in both tests. Heart rate and RPE data (not shown, available upon request) were <3% different between analyzers and had between-analyzer correlations ≥0.99, suggesting no significant differences in physiologic load when wearing each analyzer at a given stage. Compared to the Parvomedics, the VO2 Master non-significantly trended ($0.05 \leq p \leq 0.10$) toward underestimating VO_2 at 100 and 150 W and either significantly overestimated or non-significantly trended toward overestimating VO_2 at the 250 and 300 W stages (Figure 1). Average absolute VO_2 was <0.2 L/min different between analyzers at all stages. Minute ventilation was significantly higher for the VO2 Master at 250 W and non-significantly trended toward being higher at 150 W, compared to the Parvomedics (Figure 2).
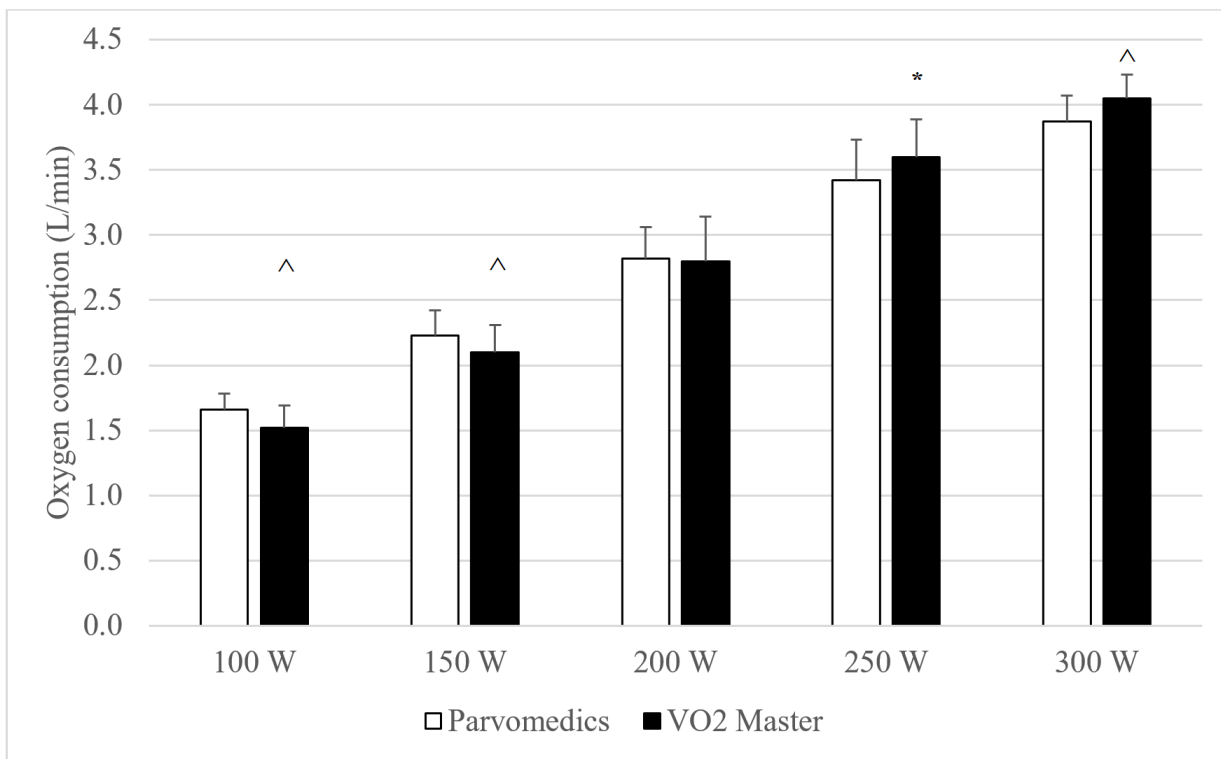

Figure 1. Parvomedics and VO2 Master (version 1.1.1) oxygen consumption measures during hour test protocol. Error bars represent standard deviations.
*Indicates significant difference from Parvomedics analyzer ($p<0.05$).
^Indicates non-significant trend toward difference from Parvomedics analyzer ($p=0.05-0.10$).
N=8 for 100-250 W and n=5 for 300 W.
Parvomedics: Parvomedics TrueOne 2400 metabolic analyzer; VO2 Master: VO2 Master Pro version 1.1.1 metabolic analyzer; W: Watts.
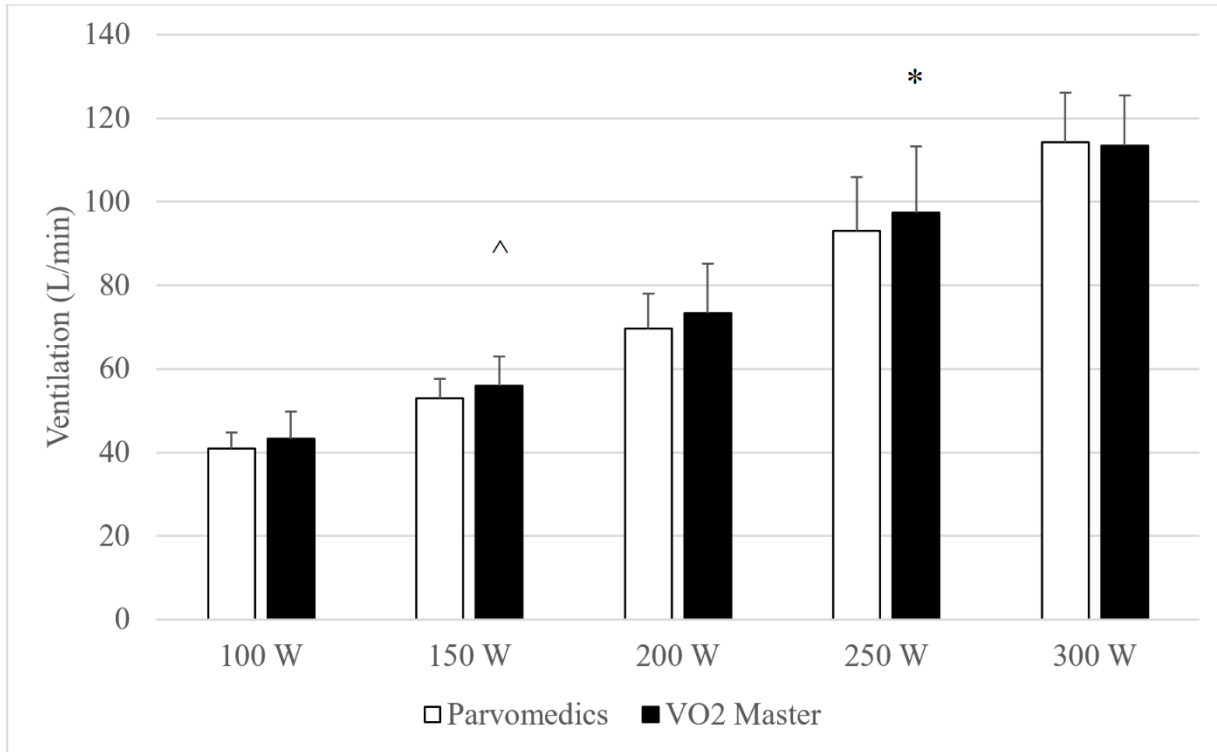
Figure 2. Parvomedics and VO2 Master (version 1.1.1) ventilation measures during hour test protocol.
Error bars represent standard deviations.
*Indicates significant difference from Parvomedics analyzer (p<0.05).
^Indicates non-significant trend toward difference from Parvomedics analyzer (p=0.05-0.10).
N=8 for 100-250 W and n=5 for 300 W.
Parvomedics: Parvomedics TrueOne 2400 metabolic analyzer; VO2 Master: VO2 Master Pro version 1.1.1 metabolic
analyzer; W: Watts.

Mean absolute and percent differences between analyzers are shown in Table 5, which allows
for assessment of convergent validity of the analyzers. At all stages, VO_2 was within 0.3 L/min,
which translated to being <1 MET (3.5 ml/kg/min) different between analyzers, and in all cases
other than 100 W the mean absolute percent difference for VO_2 and minute ventilation was
<10%, indicating acceptable accuracy.

Table 5. Mean absolute differences, mean absolute percent differences, and between-analyzer correlations for metabolic variables measured by the VO2 Master (version 1.1.1), compared to the Parvomedics, during the hour test protocol.

	VO_2 (L/min)	VE (L/min)
Overall		
MAD	0.21 (0.14)	4.9 (4.5)
MAPD (%)	8.2 (6.0)	7.3 (7.1)
Correlation	0.97	0.97
100 W		
MAD	0.17 (0.12)	4.6 (4.9)
MAPD (%)	11.1 (7.9)	10.6 (10.6)
150 W		
MAD	0.19 (0.11)	3.5 (3.3)
MAPD (%)	8.8 (5.4)	6.2 (5.7)
200 W		
MAD	0.27 (0.19)	7.1 (6.2)
MAPD (%)	9.5 (6.1)	9.6 (8.0)
250 W		
MAD	0.21 (0.19)	4.8 (4.3)
MAPD (%)	6.0 (5.4)	4.8 (3.9)
300 W		
MAD	0.17 (0.06)	4.2 (2.3)
MAPD (%)	4.4 (1.6)	3.7 (1.8)

Data are shown as mean (standard deviation).
N=8 for 100-250 W and n=5 for 300 W.
MAD: mean absolute difference between analyzers; MAPD: mean absolute percent difference between analyzers; W: Watts; VO_2: oxygen consumption; VE: minute ventilation.

Comparisons of Day 1 testing to Day 2 testing are shown in Table 6 and allow for assessment of test-retest reliability. Overall and in all stages, VO_2 varied less between days for the Parvomedics than the VO2 Master, indicating better test-retest reliability of the Parvomedics. Day-to-day variability in ventilation data were not consistently different between analyzers. Mean absolute percent differences were acceptable (<10%) for minute ventilation overall and at 150-300 W, but were <10% at only 250 and 300 W for VO_2. Overall and for 100-250 W, the Parvomedics had lower coefficients of variation than the VO2 Master VO_2, minute ventilation, indicating better test-retest reliability of the Parvomedics. These were reversed at 300 W, with higher coefficients of variation for the Parvomedics than the VO2 Master for VO_2 and minute ventilation.

Table 6. Mean absolute differences, mean absolute percent differences, and between-day correlations for metabolic variables measured by the VO2 Master (version 1.1.1) and Parvomedics on Day 1 vs. Day 2, during the hour test protocol.

	VO2 Master		Parvomedics	
	VO$_2$ (L/min)	VE (L/min)	VO$_2$ (L/min)	VE (L/min)
Overall				
MAD	0.26 (0.21)	5.8 (4.1)	0.14 (0.17)	5.5 (5.8)
MAPD (%)	10.9 (9.4)	8.9 (7.3)	5.3 (5.7)	7.6 (6.4)
Correlation	0.96	0.98	0.96	0.96
CV (%)	11.4	15.1	8.5	12.1
100 W				
MAD	0.23 (0.20)	5.7 (4.2)	0.09 (0.08)	4.2 (3.2)
MAPD (%)	15.1 (12.0)	13.1 (10.0)	5.2 (5.1)	9.8 (6.3)
CV (%)	14.6	16.0	7.9	10.7
150 W				
MAD	0.27 (0.21)	5.0 (3.9)	0.17 (0.23)	2.7 (2.2)
MAPD (%)	13.5 (11.4)	9.5 (8.2)	7.4 (8.8)	5.1 (4.2)
CV (%)	12.3	13.0	9.8	9.5
200 W				
MAD	0.32 (0.23)	6.5 (4.3)	0.10 (0.07)	5.9 (7.6)
MAPD (%)	11.3 (8.3)	9.2 (6.2)	3.5 (2.5)	7.8 (9.0)
CV (%)	13.3	16.8	8.7	13.2
250 W				
MAD	0.22 (0.20)	5.0 (4.2)	0.23 (0.26)	7.1 (6.5)
MAPD (%)	6.3 (6.0)	5.0 (3.9)	6.3 (6.4)	7.0 (5.8)
CV (%)	10.1	16.5	9.6	14.5
300 W				
MAD	0.28 (0.23)	7.5 (4.7)	0.12 (0.09)	9.2 (7.7)
MAPD (%)	6.8 (5.6)	6.9 (5.0)	3.2 (2.6)	8.1 (6.9)
CV (%)	5.6	12.4	5.7	12.9

Data are shown as mean (standard deviation).
N=8 for 100-250 W and n=5 for 300 W.
MAD: mean absolute difference between analyzers; MAPD: mean absolute percent difference between analyzers; CV: coefficient of variation; W: Watts; VO$_2$: oxygen consumption; VE: minute ventilation.

Maximal test protocol: Heart rate and RPE were <5% different and had correlations ≥0.93 overall across the three test conditions, suggesting similar physiologic load across visits for each stage. Both VO2 Master analyzers had significantly lower absolute VO$_2$ than the Parvomedics at 100 W, and the VO2 Master version 1.1.1 analyzer also had significantly lower VO$_2$ than the Parvomedics at 150 W (Figure 3). There were no differences in VO$_2$ between VO2 Master analyzers at any stage. Average absolute VO$_2$ was <0.3 L/min different among the three devices at the 100-250 W stages. The VO2 Master version 1.2.1 non-significantly trended toward higher minute ventilation than the other two analyzers at 100 W only (Figure 4).
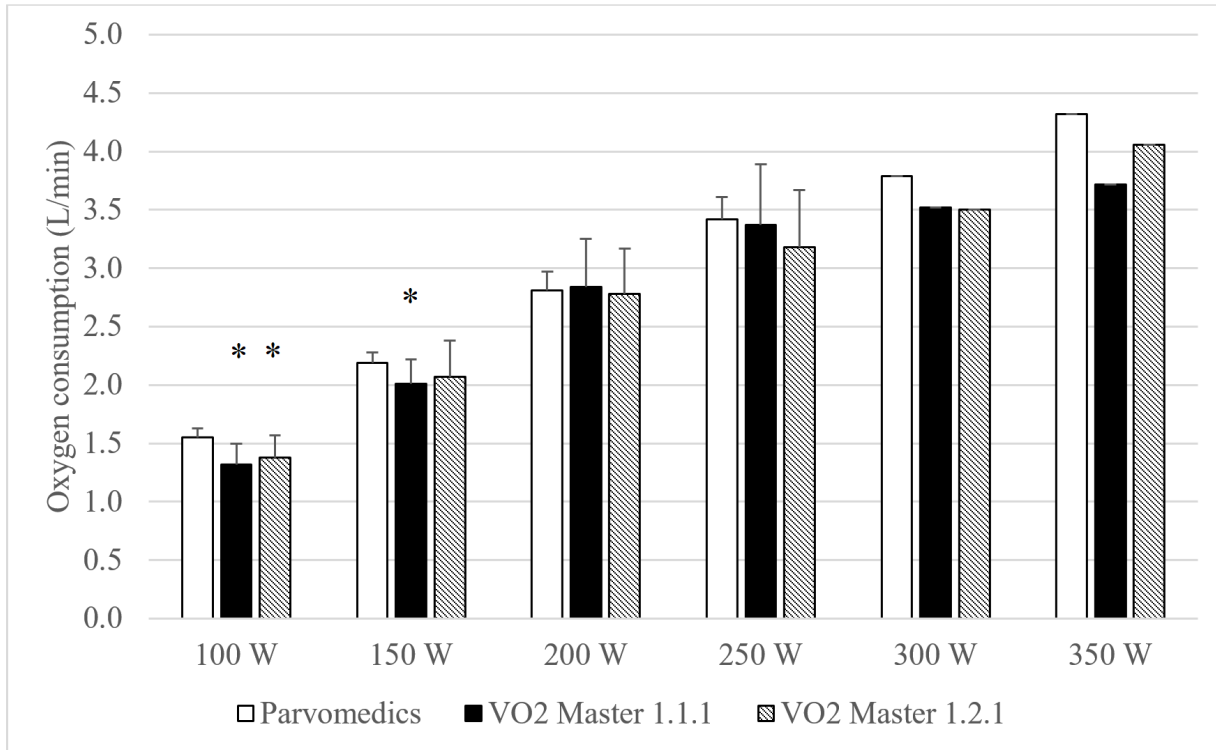
Figure 3. Parvomedics and VO2 Master (versions 1.1.1 and 1.2.1) oxygen consumption measures during maximal test protocol.

Error bars represent standard deviations. Note that 300 and 350 Watts do not have standard deviations, and statistical tests were not performed, due to sample size of n=1.

n=13 for 100-200 Watts, n=7 for 250 Watts, and n=1 for 300-350 Watts.

*Indicates significant difference from Parvomedics analyzer only (p<0.05).

Parvomedics: Parvomedics TrueOne 2400; VO2 Master 1.1.1: VO2 Master analyzer version 1.1.1. VO2 Master 1.2.1: VO2 Master analyzer version 1.2.1; W: Watts.
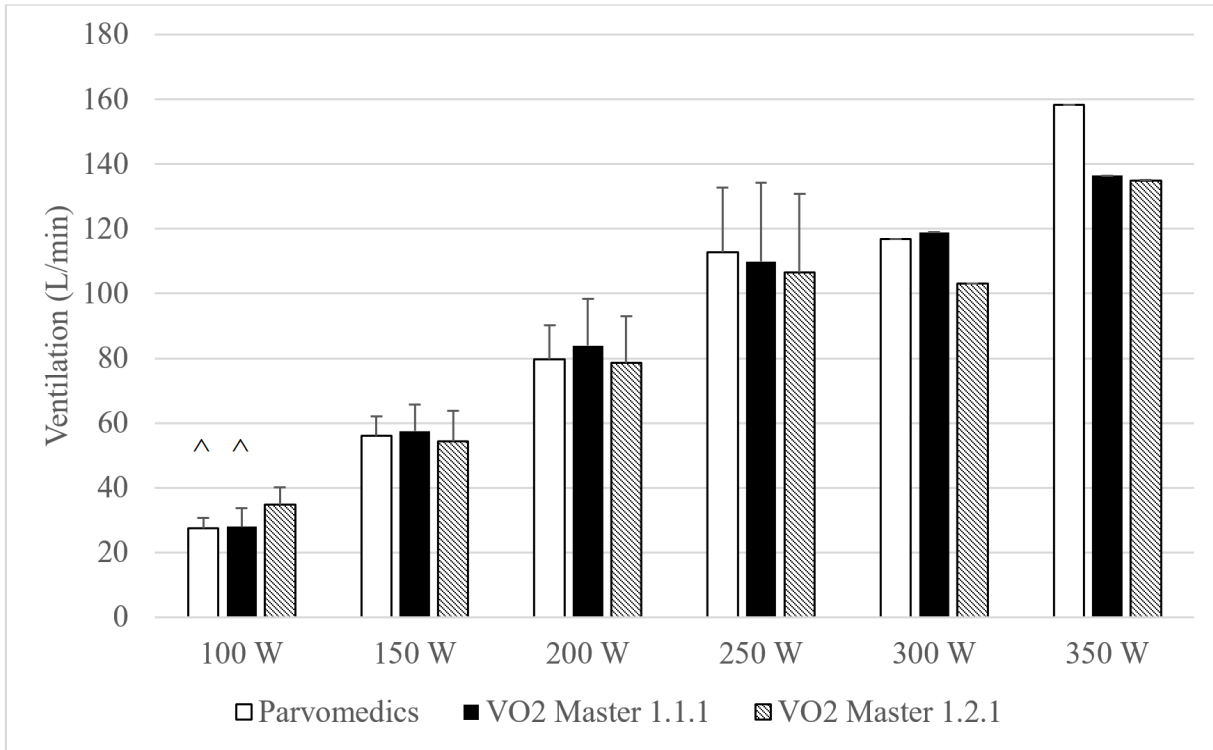
Figure 4. Parvomedics and VO2 Master (versions 1.1.1 and 1.2.1) ventilation measures during maximal test protocol.
Error bars represent standard deviations. Note that 300 and 350 Watts do not have standard deviations, and statistical tests were not performed, due to sample size of n=1.
n=13 for 100-200 Watts, n=7 for 250 Watts, and n=1 for 300-350 Watts.
^Indicates non-significant trend toward difference from VO2 Master version 1.2.1 analyzer (p=0.05-0.10).
Parvomedics: Parvomedics TrueOne 2400; VO2 Master 1.1.1: VO2 Master analyzer version 1.1.1. VO2 Master 1.2.1: VO2 Master analyzer version 1.2.1; W: Watts.

Overall and at each stage, absolute and percent differences in VO_2 and ventilation were similar when comparing each of the VO2 Master analyzers to the Parvomedics, indicating similar convergent validity of the different VO2 Master versions (Table 7). In assessing inter-device reliability, VO_2 and minute ventilation were 13-14% different between analyzers, with no apparent improvement in percent differences with increasing exercise intensity.

Table 7. Mean absolute differences, mean absolute percent differences, and between-analyzer correlations for metabolic variables measured by the Parvomedics and VO2 Master (versions 1.1.1 and 1.2.1) during the maximal test protocol.

	VO2 Master version 1.1.1 compared to Parvomedics		VO2 Master version 1.2.1 compared to Parvomedics		VO2 Master version 1.1.1 compared to version 1.2.1	
	VO_2 (L/min)	VE (L/min)	VO_2 (L/min)	VE (L/min)	VO_2 (L/min)	VE (L/min)
Overall						
MAD	0.27 (0.22)	6.9 (6.9)	0.26 (0.17)	6.8 (5.7)	0.32 (0.24)	8.7 (7.5)
MAPD (%)	12.0 (10.1)	10.3 (8.6)	11.9 (8.2)	10.3 (7.2)	14.0 (8.4)	13.2 (9.9)
Correlation	0.93	0.95	0.93	0.96	0.88	0.94
100 W						
MAD	0.24 (0.17)	3.8 (3.5)	0.19 (0.14)	3.8 (2.4)	0.20 (0.13)	4.8 (3.3)
MAPD (%)	17.9 (12.5)	10.0 (8.5)	13.7 (10.4)	11.2 (7.8)	14.4 (9.4)	13.4 (8.9)
150 W						
MAD	0.19 (0.17)	5.9 (5.0)	0.26 (0.14)	4.8 (4.0)	0.28 (0.15)	6.8 (5.3)
MAPD (%)	9.3 (8.8)	10.6 (9.2)	12.7 (7.4)	9.0 (8.3)	13.7 (7.0)	12.9 (11.2)
200 W						
MAD	0.27 (0.26)	8.2 (6.3)	0.26 (0.19)	8.2 (4.6)	0.36 (0.22)	9.4 (6.3)
MAPD (%)	9.4 (8.5)	10.2 (7.3)	9.8 (7.7)	10.7 (6.3)	12.8 (7.3)	12.0 (9.1)
250 W						
MAD	0.39 (0.30)	11.0 (11.4)	0.40 (0.22)	10.2 (8.4)	0.62 (0.37)	17.9 (11.3)
MAPD (%)	11.3 (8.6)	11.3 (12.2)	12.6 (7.5)	9.2 (7.3)	18.5 (10.3)	17.6 (12.3)
300 W#						
MAD	0.27	2.0	0.29	13.7	0.02	15.7
MAPD (%)	7.4	1.7	8.0	12.4	0.6	14.1
350 W#						
MAD	0.60	21.7	0.26	23.4	0.34	1.5
MAPD (%)	14.9	14.8	6.2	15.9	8.7	1.1

Data are shown as mean (standard deviation).
n=13 for 100-200 Watts, n=7 for 250 Watts, and n=1 for 300-350 Watts.
#300 and 350 Watts do not have standard deviations due to sample size of n=1.
MAD: mean absolute difference between analyzers; MAPD: mean absolute percent difference between analyzers; W: Watts; VO_2: oxygen consumption; VE: minute ventilation.

DISCUSSION

Technological advances in metabolic analysis equipment have allowed for new and innovative ways to assess human health and athletic performance through measurement of VO_2 and minute ventilation and associated caloric expenditure during rest and exercise. In the case of the VO2 Master Pro, its portability and low cost have the potential to allow metabolic and caloric analyses in populations and settings not previously accessible. However, new metabolic analysis technologies must first be validated against existing, industry standards before being suitable for widespread use. As such, our study assessed validity and reliability (both test-retest and inter-device) of the VO2 Master, compared to the Parvomedics metabolic analyzer.

Validity: Both protocols in our study generally indicated that the VO2 Master analyzers either trended toward or had significantly lower VO_2 than the Parvomedics at 100 and 150 W, whereas there was either no difference or slight overestimation by the VO2 Master (hour test protocol only) at 250 and 300 W. For version 1.1.1 of the VO2 Master, underestimation of VO_2 at the lower workloads was due to overestimation of the fraction of expired O_2 measured by the O_2 sensor (data not shown), and this appears to be true in version 1.2.1 for 100 W in the maximal test protocol as well. Despite differences reaching statistical significance at some stages, absolute differences between analyzers in the hour test protocol were <0.3 L/min between analyzers (translating to a difference in relative VO_2 of <3.0 ml/kg/min), which amounted to roughly 8% difference for both. Additionally, percent differences decreased as exercise intensity increased, suggesting that the VO2 Master may be better suited for VO_2 measures at vigorous or near-maximal intensities. Absolute differences were higher in the maximal test protocol than the hour test protocol, but we attribute this to the additional variability added by having tests on separate days for the maximal test protocol. Notably, differences from the Parvomedics were less for the VO2 Master version 1.2.1 vs. version 1.1.1 in the maximal protocol, suggesting that both versions likely measure VO_2 to within 1 MET (3.5 ml/kg/min) of Parvomedics measures across all tested exercise intensities. Additionally, correlations between the VO2 Master and Parvomedics were 0.97 in the hour test protocol and 0.92-0.93 in the maximal test protocol, demonstrating strong agreement between analyzer brands.

Minute ventilation data were (or trended toward being) significantly higher for the VO2 Master version 1.1.1 than the Parvomedics for 150 and 250 W but not at 100, 200, or 300 W in the hour test protocol. Similarly, minute ventilation was significantly higher for the VO2 Master version 1.2.1 compared to the other analyzers at 100 W during the maximal test protocol, but there were no other differences among analyzers at any stage. Absolute differences averaged <5 L/min (~7%) in the hour test protocol and <7 L/min (~10%) in the maximal test protocol, and correlations were ≥0.95 for both protocols, again suggesting strong agreement between both versions of VO2 Master with the Parvomedics.

To our knowledge, our study is the first to report validity data for the VO2 Master analyzer, so it is not possible to directly compare our findings to past work with the VO2 Master. However, comparisons to other metabolic analyzer brands is informative to provide indirect comparisons of device error across brands. Importantly, previous studies validating major brands including COSMED, Oxycon, and MetaMax indicate comparable degrees of error as the VO2 Master. Kawakami et al. (11) found that the COSMED K2 had absolute VO_2 measures 0.1-0.2 L/min different than a criterion during submaximal exercise, and McLaughlin et al. (18) found that the COSMED K4b^2 significantly overestimated absolute VO_2 by ~0.1 L/min at cycling workloads of 50-200W. In both studies, the authors advocated that these differences are of little practical significance. Other studies validating the COSMED have found little or no difference from criterion measures for VO_2 measurement (13, 14). Additionally, a study by Rosdahl et al. (26) validating the Oxycon Mobile have found slight overestimation of VO_2 during submaximal exercise but an underestimation of absolute VO_2 by ~0.2 L/min. Finally, the MetaMax 3B has been validated by several groups, generally finding that the MetaMax slightly overestimates absolute VO_2 compared to a criterion, by up to ~0.3 L/min, during submaximal and maximal

exercise (4, 16, 30). The similar magnitudes of differences found with other analyzer brands in these previous studies to those found with the VO2 Master, coupled with the popularity of these other analyzer brands, suggest that the differences in VO_2 found between the VO2 Master and Parvomedics in our study may be of little practical significance.

Reliability: Our study addressed test-retest reliability for the Parvomedics and VO2 Master (v1.1.1) across testing days in the hour test protocol. Mean absolute differences, mean absolute percent differences, and coefficients of variation for VO_2 and minute ventilation were consistently lower between days for the Parvomedics compared to the VO2 Master, indicating better test-retest reliability of the Parvomedics. Differences were also above 10% for VO_2 at 100-200 W and minute ventilation at 100-150 W but not at the highest intensities, signifying that reliability is better for higher or near-maximal intensities than lower intensities. Correlations for all variables were ≥0.96 for the VO2 Master and Parvomedics, indicating strong between-day agreement for both analyzers.

As with the validity data, we are unaware of past studies examining reliability of the VO2 Master and, therefore, focus our discussion on comparison to other portable analyzer brands. Past work with the COSMED K2 and $K4b^2$ found coefficients of variation of ~3-13% for VO_2 and minute ventilation variables during submaximal and maximal exercise (13, 14, 29), and these are slightly better than the 11.4-15.1% found in our study. However, the correlations in this study (r≥0.94) compare favorably to past work, with studies reporting correlations of 0.70-0.96 for metabolic variables assessed with various COSMED models (5, 6, 14, 29). Past studies assessing test-retest reliability for the Oxycon (coefficients of variation 2.8-5.8%) (26) and MetaMax (coefficients of variation 5.5-7.6%, correlations of r=0.77-0.99) (2, 20, 25) are similar to that of the COSMED. Together, these findings suggest that the VO2 Master has slightly higher variability in monitor output but equally strong between-day agreement as compared to other major portable analyzer brands.

Data on inter-monitor reliability of portable metabolic analyzers are sparse. One study by Guidetti et al. (8) with the COSMED K5 found mean absolute errors of <1% and correlations of r>0.99 when evaluating inter-unit reliability using a metabolic simulator. These numbers are similar to the metabolic simulator data comparison reported by the manufacturer of VO2 Master (17). Notably, data produced by a metabolic simulator under carefully controlled conditions represents the best possible scenario for analyzer performance, so it would be expected that reliability of the COSMED would be lower in field settings or when tested on humans, as was found in our study for the VO2 Master.

Aside from little past research examining inter-device reliability of portable metabolic analyzers, we are cautious to make firm conclusions regarding inter-device reliability from our analysis. The maximal test protocols took place on different days, introducing the confounding variable of inter-day variability of metabolism. While randomization of visit order would eliminate some of the potential bias, visits were not randomized between VO2 Master versions 1.1.1 and 1.2.1 because version 1.2.1 did not become available for purchase until testing with version 1.1.1 was nearly complete. We elected to include these data, even though imperfect, to gain preliminary

insights into inter-device reliability. However, further testing, perhaps with a protocol similar to our hour test protocol, would yield better data for determination of inter-device reliability.

Study strengths and limitations: This study had several notable strengths. First, our study assessed both validity and reliability using two different testing protocols and in different populations, which gives a better indication of how the VO2 Master device functions in a variety of populations and intensities. Heart rate and RPE data were included to confirm similar physiologic workloads across tests, and the counterbalanced design within each stage (hour test protocol) eliminated potential effects of VO_2 drift or the slow component of VO_2. Finally, a fan was placed in front of participants in the hour test protocol but was not used in the maximal test protocol; use of the fan did not lower analyzer validity or reliability, providing preliminary evidence that the device functionality is not impacted by moderate headwind.

Several study limitations must also be acknowledged. Our sample size, while comparable to past validation studies, was small and consisted entirely of male participants. Validation across a larger, more diverse population is warranted. The lack of randomization for the VO2 Master version 1.2.1 in the maximal test protocol is another limitation, which precludes us from decisively indicating inter-device reliability of this version of the VO2 Master. Of note, inter-device reliability is of interest only if a team/lab/clinic is using more than one analyzer and/or is looking to compare data to that of other teams/labs/clinics. Settings in which the same analyzer is used to collect data should be more concerned with validity and test-retest reliability of the analyzer. The use of different mouthpieces is a study limitation in that it could theoretically change breathing patterns or physiologic load of activity. However, the heart rate and RPE data suggest little difference in physiologic load, and breathing frequency and tidal volume (data not shown but available upon request) were not different between analyzers. Finally, while the VO2 Master version 1.2.1 has the ability to assess resting VO_2, we did not include resting measures in our study since data collection was nearly complete for the VO2 Master version 1.1.1 (which does not measure resting VO_2) and Parvomedics by the time the VO2 Master version 1.2.1 was released. Future research should validate resting VO_2 and minute ventilation data from the VO2 Master. Finally, as a major use of portable analyzers is measurement in field settings, the VO2 Master should be further validated in a variety of environmental (e.g., temperature, humidity, wind) conditions.

Conclusions: In conclusion, our study found acceptable convergent validity and test-retest reliability of the VO2 Master Pro analyzer for assessment of exercise VO_2 and minute ventilation, compared to the Parvomedics analyzer. Validity and reliability measures of the VO2 Master found in this study were comparable to those of other major portable metabolic analyzer brands (e.g., COSMED, Oxycon, MetaMax) in past research. While further work is needed to confirm these findings, our results suggest that the VO2 Master Pro is likewise a suitable option for those looking to measure VO_2 and minute ventilation during moderate- and vigorous-intensity exercise in field settings.

ACKNOWLEDGEMENTS

REFERENCES

1. Bassett DR Jr, Howley ET, Thompson DL, King GA, Strath SJ, McLaughlin JE, Parr BB. Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. J Appl Physiol 91(1): 218-224, 2001.

2. Blessinger J, Sawyer B, Davis C, Irving BA, Weltman A, Gaesser G. Reliability of the VmaxST portable metabolic measurement system. Int J Sports Med 30(1): 22-26, 2009.

3. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 14(5): 377-381, 1982.

4. Brehm MA, Harlaar J, Groepenhof H. Validation of the portable VmaxST system for oxygen-uptake measurement. Gait Posture 20(1): 67-73, 2004.

5. Darter BJ, Rodriguez KM, Wilken JM. Test-retest reliability and minimum detectable change using the K4b2: Oxygen consumption, gait efficiency, and heart rate for healthy adults during submaximal walking. Res Q Exerc Sport 84(2): 223-231, 2013.

6. Duffield R, Dawson B, Pinnington HC, Wong P. Accuracy and reliability of a Cosmed K4b2 portable gas analysis system. J Sci Med Sport 7(1): 11-22, 2004.

7. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J Clin Epidemiol 67(8): 850-857, 2014.

8. Guidetti L, Meucci M, Bolletta F, Emerenziani GP, Gallotta MC, Baldari C. Validity, reliability, and minimum detectable change of Cosmed K5 portable gas exchange system in breath-by-breath mode. PLoS One 13(12): e0209925, 2018.

9. Harris JA, Benedict FG. A biometric study of human basal metabolism. Proc Natl Acad Sci USA 4(12): 370-373, 1918.

10. Hipskind P, Glass C, Charlton D, Nowak D, Dasarathy S. Do handheld calorimeters have a role in assessment of nutrition needs in hospitalized patients? A systematic review of literature. Nutr Clin Pract 26(4): 426-433, 2011.

11. Kawakami Y, Nozaki D, Matsuo A, Fukunaga T. Reliability of measurement of oxygen uptake by a portable telemetric system. Eur J Appl Physiol Occup Physiol 65(5): 409-414, 1992.

12. Kenney W, Wilmore J, Costill D. Physiology of sport and exercise. Champaign, IL: Human Kinetics; 2012.

13. Lothian F, Farrally MR, Mahoney C. Validity and reliability of the Cosmed K2 to measure oxygen uptake. Can J Appl Physiol 18(2): 197-206, 1993.

14. Lucia A, Fleck SJ, Gotshall RW, Kearney JT. Validity and reliability of the Cosmed K2 instrument. Int J Sports Med 14(7): 380-386, 1993.

15. Macfarlane DJ. Open-circuit respirometry: A historical review of portable gas analysis systems. Eur J Appl Physiol 117(12): 2369-2386, 2017.

16. Macfarlane DJ, Wong P. Validity, reliability, and stability of the portable Cortex MetaMax 3b gas analysis system. Eur J Appl Physiol 112(7): 2539-2547, 2012.

17. Master VO$_2$. Validity and reliability of the VM Pro. In: https://vo2master.com/validation-studies/; 2019.

18. McLaughlin JE, King GA, Howley ET, Bassett DR Jr, Ainsworth BE. Validation of the Cosmed K4b2 portable metabolic system. Int J Sports Med 22(4): 280-284, 2001.

19. Meyer T, Davison RC, Kindermann W. Ambulatory gas exchange measurements-Current status and future options. Int J Sports Med 26 (Suppl 1): S19-27, 2005.

20. Meyer T, Georg T, Becker C, Kindermann W. Reliability of gas exchange measurements from two different spiroergometry systems. Int J Sports Med 22(8): 593-597, 2001.

21. Mifflin MD, St Jeor ST, Hill LA, Scott BJ, Daugherty SA, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. Am J Clin Nutr 51(2): 241-247, 1990.

22. Navalta JW, Stone WJ, Lyons TS. Ethical issues relating to scientific discovery in exercise science. Int J Exerc Sci 12(1): 1-8, 2019.

23. Nelson MB, Kaminsky LA, Dickin DC, Montoye AHK. Validity of consumer-based physical activity monitors for specific activity types. Med Sci Sports Exerc 48(8): 1619-1628, 2016.

24. Overstreet BS, Bassett DR Jr, Crouter SE, Rider BC, Parr BB. Portable open-circuit spirometry systems. J Sports Med Phys Fitness 57(3): 227-237, 2017.

25. Perkins CD, Pivarnik JM, Green MR. Reliability and validity of the VmaxST portable metabolic analyzer. J Phys Act Health 1(4): 413-422, 2004.

26. Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P, Schantz P. Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. Eur J Appl Physiol 109(2): 159-171, 2010.

27. Rubenbauer JR, Johannsen DL, Baier SM, Litchfield R, Flakoll PJ. The use of a handheld calorimetry unit to estimate energy expenditure during different physiological conditions. J Parenter Enteral Nutr 30(3): 246-250, 2006.

28. St-Onge MP, Rubiano F, Jones A Jr, Heymsfield SB. A new hand-held indirect calorimeter to measure postprandial energy expenditure. Obes Res 12(4): 704-709, 2004.

29. Thomas SS, Buckon CE, Schwartz MH, Russman BS, Sussman MD, Aiona MD. Variability and minimum detectable change for walking energy efficiency variables in children with cerebral palsy. Dev Med Child Neurol 51(8): 615-621, 2009.

30. Vogler AJ, Rice AJ, Gore CJ. Validity and reliability of the Cortex MetaMax3b portable metabolic system. J Sports Sci 28(7): 733-742, 2010.