SCIENTIFIC JOURNAL OF SPORT AND PERFORMANCE

ORIGINAL ARTICLE

The validity of the VO_2 master pro for assessing VO_{2max} and running economy in trained runners

Trace A. Dominy . *Department of Health, Exercise Science, and Recreational Management. University of Mississippi. Oxford, MS, United States of America.*
Asha D. Phillips. *Department of Health, Exercise Science, and Recreational Management. University of Mississippi. Oxford, MS, United States of America.*
Katherine L. Tibbs. *Department of Health, Exercise Science, and Recreational Management. University of Mississippi. Oxford, MS, United States of America.*
Brooke M. Gilmore. *Department of Nutrition and Hospitality Management. University of Mississippi. Oxford, MS, United States of America.*
Julia P. Phillips. *Department of Health, Exercise Science, and Recreational Management. University of Mississippi. Oxford, MS, United States of America.*
Eric K. O'Neal. *Cole Honors College. University of North Alabama. Florence, AL, United States of America.*
Thomas L. Andre. *Department of Health, Exercise Science, and Recreational Management. University of Mississippi. Oxford, MS, United States of America*

ABSTRACT

Purpose: This study assessed agreement between the portable VO_2 Master Pro (VM) and a non-portable metabolic cart (NP) for measuring VO2max and running economy (RE). Methods: Fourteen endurance-trained runners (9 female, 5 male; age 24.7 ± 5.0 y) completed four randomized crossover trials: two VO2max tests and two RE sessions using VM or NP. RE consisted of four 5-min bouts at 100%, 90%, 80%, and 70% of personal best 5-km pace. Agreement was evaluated using paired t-tests, intraclass correlation coefficients (ICC), standard error of measurement (SEM), coefficient of variation (CV), and Bland–Altman analysis. Results: VM significantly overestimated relative VO2max compared with NP (65.4 ± 7.3 vs 53.5 ± 5.2 ml·kg⁻¹·min⁻¹, $p < .001$), with weak reliability (ICC = 0.46), high SEM (2.6 ml·kg⁻¹·min⁻¹), and unacceptable CV (14.1%). Mean bias was 11.9 ± 7.1 ml·kg⁻¹·min⁻¹. Maximal ventilation (VE) showed moderate agreement (ICC = 0.81; CV = 8.3%). Similar trends were observed across all RE intensities, with modest VO_2 agreement (ICC = 0.60–0.72) and stronger VE agreement (ICC = 0.87–0.93). Conclusion: VM substantially overestimates VO_2 and demonstrates poor agreement with NP, limiting its utility as a replacement device. Greater agreement for VE suggests oxygen sensing as a primary source of error.
Keywords: Sport medicine, Performance analysis, Minute ventilation, Intraclass correlation, Bland-Altman.

Cite this article as:
Dominy, T. A., Phillips, A. D., Tibbs, K. L., Gilmore, B. M., Phillips, J. P., O'Neal, E. K., & Andre, T. L. (2026). The validity of the VO_2 master pro for assessing VO_{2max} and running economy in trained runners. *Scientific Journal of Sport and Performance*, *5*(4), 598-611. https://doi.org/10.55860/MZVI6493

Corresponding author. *Department of Health, Exercise Science, and Recreational Management. University of Mississippi. Oxford, MS, United States of America.*
E-mail: tadominy@olemiss.edu
Submitted for publication April 29, 2026.
Accepted for publication June 23, 2026.
Published July 10, 2026.
Scientific Journal of Sport and Performance. ISSN 2794-0586.

doi: https://doi.org/10.55860/MZVI6493

INTRODUCTION

Metabolic analysers use respired flow volume and fractional concentrations of oxygen (O_2) and carbon dioxide (CO_2), from which variables such as minute ventilation (V_E), respiratory frequency (Rf), tidal volume (Tv), and rates of oxygen consumption (VO_2) and carbon dioxide production (VCO_2) are derived (Thiessen et al., 2025). These variables are widely used to assess cardiorespiratory health and function, evaluate fitness status, and guide exercise prescription. Early iterations of metabolic analysers were not portable, and testing could only be performed in laboratory settings. Since the 1980s, multiple portable metabolic systems have been developed enabling field-based assessments (Montoye et al., 2020). Portable systems can cost upwards to $50,000, but a newer model, the VO_2 Master Pro (VM) (VO_2 Master Pro Model 1.6.1, Vernon, BC, Canada) offers a user-friendly interface and price point (~ $7,000) allowing a wider range of adoption across clinical, research, and fitness industry settings.

The VM manufacturer claims to be accurate to ± 3% for both O_2 and V_E assessment (Master VO_2, 2019), but mixed results have been reported in independent laboratory validation studies. Van Hooren et al., (2024) tested the accuracy of the VM during simulated (Relitech Systems BV, Nijkerk, The Netherlands) exercise intensities of 1.0 – 4.0 L/min and VO_2 was underestimated by an average of 12% with percent error increasing with higher V_E rates. In a more recent study, the same VM device was seen to overestimate VO_2 by 7.9% during simulated exercise intensities of 1.0 – 4.0 L/min compared to the same metabolic simulator (Relitech Systems BV, Nijkerk, The Netherlands). The difference in VO_2 estimation between these two studies may be attributed to an update in the VM algorithm for gas exchange assessment (Van Hooren et al., 2026). (Van Hooren et al., 2026). Conversely, Thiessen et al., (2025) assessed the VM during 2-min of steady state simulated (model 17056, VacuMed, Ventura, CA, USA) exercise at intensities of 1.0 – 4.0 L/min and saw an average error 0.6% for VO_2 and 1.1% for V_E across all intensities and a between day average difference of 1.2% for VO_2. This may suggest that under controlled laboratory conditions the VM provides gas exchange ventilatory values within an acceptable range (<5% error) for metabolic testing equipment. However, Thiessen et al., (2025) noticed the relative humidity, temperature, and barometric pressure of the VM were not identical to those of the metabolic simulator and derived correction factors for the VM, which could explain the reduced error seen between the devices (Van Hooren et al., 2026).

Davis et al., (2024) compared VM to a non-portable metabolic cart (Cosmed Quark) at 50%, 70%, and 80% of critical power and found no differences in VO_2 (35.1 ± 10.4 mL/kg/min vs. 37.7 ± 8.7 mL/kg/min) or V_E (73.1 ± 21.2 L/min vs 73.9 ± 21.8 L/min) across two visits. During a maximal cycling protocol (100 – 350 Watts) the VM significantly underestimated VO_2 by 0.24 L/min at 100 Watts, whereas during a 1-hour cycling protocol (100 – 300 Watts in 10-min stages) the VM significantly overestimated VO_2 by 0.21 L/min at 250 Watts compared to Parvomedics TrueOne 2400 (Montoye et al., 2020). Additionally, VM has been shown to significantly underestimate VO_2 during walking (3.2 km/h) (11.42 mL/kg/min vs. 7.07 mL/kg/min) and jogging (7.2 km/h) (30.54 mL/kg/min vs. 24.47 mL/kg/min) with the lowest mean absolute percent error occurring during jogging (20.05%) compared to a Cosmed Quark (Kim et al., 2025). The authors did note the discrepancies between devices at lower exercise intensities could be attributed to limited ventilatory flow capture or sensor responsiveness of the VM during submaximal exercise (Kim et al., 2025).Another study reported the VM did not report walking (4.82 km/h) data for most participants and underestimated VO_2 during treadmill running (47.86 ± 3.94 vs. 29.56 ± 4.15 mL/kg/min) and V_E (71 vs 51 mL/min) compared to an independently validated portable metabolic system, the Cosmed K5 (Toulouse et al., 2022). A simulated airflow test while connected to a 3L calibration syringe at stroke paces of 15, 25, and 35 strokes per minute revealed the VM to underestimate tidal volume by 5% - 8% to the known value of 3L, which may explain the discrepancies between the Cosmed K5 and VM (Toulouse et al., 2022).

The VM offers a cost-effective instrument for running coaches to assess changes in running economy (RE) and maximal oxygen consumption (VO_{2max}). Running economy (RE) is defined as the oxygen or metabolic cost at a given submaximal velocity and is a key determinant of long-distance running performance (Barnes et al., 2015; Joubert et al., 2020). Metabolic cost calculation requires the measurement of both VO_2 and VCO_2 which is not possible on the VM due to lack of a carbon dioxide sensor. However, for trained runners, RE assessed by VO_2 only has been suggested to be a strong predictor of performance among athletes with a similar VO_{2max} (Conley et al., 1980; Daniels et al., 1992; Arampatzis et al., 2006; Saunders et al., 2004; Barnes et al., 2015).

Accurate assessment of RE and VO_{2max} requires precise measurements of VO_2 at moderate to high intensities, where even small differences in measurement error could impact the interpretation of an individual's training effectiveness. Given the intensity-dependent discrepancies observed with the VM for VO_2 measurements, its suitability for evaluating RE or VO_{2max} is warranted. Therefore, the purpose of this study is to assess the absolute agreement between the VM and a non-portable metabolic cart (NP) for VO_{2max} and RE assessments.

MATERIAL AND METHODS

Participants

In total, fourteen endurance trained runners (male: n = 5; female: n = 9; age = 24.7 ± 5.0 years), free from injury and reporting at least three training sessions per week (>30 min per session) over the previous three months, completed this investigation. Participants were recruited through convivence sampling and no a priori power analysis was conducted, however the final sample size is consistent with previous research examining similar variables. The study sample included more females than males because more females were available and willing to participate during the recruitment period and we acknowledge how this could limit our study's generalizability. Participants were assessed for body mass (64.75 ± 9.14 kg) (Seca, Hamburg, Germany) and height (170 ± 6 cm) (Stadiometer, Hamburg, Germany) and were screened for apparent health risks associated with a physical activity questionnaire (PAR-Q). Following anthropometric measurements, participants reported their current estimated 5-km performance under ideal conditions and then rounded to the nearest whole minute (21.86 ± 2.80 min). All participants provided written informed consent prior to any experimental procedures. This study was approved by the local institutional review board and all procedures conformed to the standards set forth by the Declaration of Helsinki.

Measures

VM calibration procedures

The VM (VO_2 Master Pro Model 1.6.1, Vernon, BC, Canada) is a small compact, head mounted portable metabolic analyser that weighs approximately 0.32 kg and uses a single AAA battery. The analyser utilizes an interchangeable adapter referred to as a "*user piece*." The "*user piece*" is a small plastic tube with an exhaust port that is used to assess VO_2 and V_E during exercise. The VM has three different "*user piece*" sizes to accommodate for different V_E ranges varying across exercise intensities: resting (V_E range: 5 – 50 L/min), medium (V_E range: 30 – 160 L/min), and large (V_E range: 40 - 220 L/min). For the current study the medium "*user piece*" was used for all trials and no participant exceeded the 160 L/min V_E range. Prior to any data collection the gas and flowmeter calibration were performed using room air. The VM was turned on and was kept motionless for at least 10 s until the dark blue light turned white indicating that the analyser was powered on and ready for flowmeter and gas calibration. A known value 3 L syringe was used for both gas and flowmeter calibration and all calibration procedures were performed within the minimal error range, which was confirmed by "*Calibration Successful/Good*" message within the VM interface. All VM data was recorded with the default standard temperature and pressure dry (STPD) correction enabled within the application. The device self-calibrates to the ambient air at the time of testing using its internal default mode. Barometric pressure and calibration procedures were performed before each trial. Following completion of the calibration procedures, participants were fitted with a Hans Rudolph 7450 V2 over the nose mask (Hans Rudolph Inc., Shawnee, KS, USA), which connected to the device and medium "*user piece*" and was held in place by the provided headgear. Data collected utilizing the VM during experimental trials were transmitted via Bluetooth to the VM mobile application installed on an iPad (Apple Inc., Cupertino, California, United States) and data was exported as a CSV file after data collection was completed.

NP calibration procedures

The NP (TrueOne 2400, Parvomedics, Sandy, Utah, United States) is a stationary laboratory based metabolic analyser, was used as the criterion reference measure in this study. Expired gas exchange was measured using a mixing chamber that calculates oxygen consumption and carbon dioxide production from averaged expired gas concentrations. In accordance with the manufacturer guideline, the NP was calibrated prior to each trial to ensure accurate measurement. Prior to calibration, the system and the heated pneumotach were powered on and were allowed to warm-up for a minimum of 30-minutes. Ambient temperature, barometric pressure, and humidity were entered into the system before beginning flowmeter calibration procedures. Flowmeter calibration was conducted using a known value 3 L syringe that was connected via a two-way non-rebreathing valve (Hans Rudolph T-Shape non-rebreathing valve, Kansas City, Missouri, United States) and a series of repeated strokes were performed across a range of flow rates (slow, moderate, and fast) to capture expected ventilatory responses at different exercise intensities. Calibration was deemed successful when measured volumes were within ± 1%. Gas analyser calibration was conducted using a gas mixture of known concentration of 16.0% O_2 and 4.0% CO_2. Researchers followed the prompts by the systems automated calibration routine, which adjusts analyser gain and offset to match known fractional concentrations of O_2 and CO_2. Calibration was considered successful when measured concentrations were within ± 1%. Upon completion of the calibration procedures participants were fitted for a mask and were connected to a two-way non-rebreathing valve held in place by provided headgear. The mask is strapped around the head to create a tight seal and to ensure all inhaled and exhaled air passes through the sensor. Data was collected through ParvoMedics OSUW software (version 4.3.4) and were stored and exported as a CSV file after completion of data collection.

VO_{2max} test sessions

Upon arrival to the laboratory, anthropometric measurements were collected and participants completed a 5-min self-selected warm-up. After the 5-min warm-up participants were fitted for a mask for both the VM and NP trials and a two-way non-rebreathing valve for the NP trial only. Device start order was randomized and counterbalanced between participants. The VO_{2max} test completed was the Bruce Protocol which is a multi-stage graded exercise test that starts at 2.7 km/h and a 10% graded. Speed and incline increased by 1.2 km/h and 2% every three minutes until stage 5 was reached. Once stage 5 was reached speed increased by 0.8 km/h and grade continued to increase by 2% every three minutes. Both VO_{2max} tests were completed on a Trackmaster Treadmill (Trackmaster Treadmill; Full Vision, Fulton, Kansas, United States). The VO_{2max} test was conducted with all participants reaching volitional exhaustion between 12 – 16 mins. Indirect calorimetry data was collected and VO_2 was then manually searched for the highest 60s VO_2 average during the final 2 minutes of the test.

Running economy bouts

Participants completed four 5-minute RE bouts at 100% 5-km pace, 70%, 80%, and 90% of their 5-km personal best performance pace. Each RE bout was 5-minutes followed by 5-minutes of passive recovery between each bout. RE bouts took place on a lower body positive pressure treadmill (Boost Inc, Waukesha, WI, USA) at 100% body weight (Vincent et al., 2022). Two different treadmills were used due to technical capabilities for grade % changes with the Bruce protocol. VO_2 and V_E were analysed from each RE bout using the average of the final 2-minutes of each 5-minute bout.

Procedures

A randomized counter-balanced crossover design was implemented over four experimental trials. Experimental sessions differed only in whether participants completed a VO_{2max} test or four, 5-min submaximal RE bouts while connected to either the VM or the NP. Visits 1 and 3 consisted of a VO_{2max} test and visits 2 and 4, participants completed four 5-min RE bouts at 70%, 80%, 90% and 100% of their estimated personal best 5-km performance within the last year. The only difference between odd and even trial sessions was the instrument (VM or NP) used to assess VO_2 and V_E. The primary dependent variables were VO_{2max}, V_E, and RE. The four experimental trials were separated by at least 48 h but no more than 1 week. Each trial was conducted at approximately the same time of day (± 1 h) for each participant. Participants were asked to wear similar clothing and the same shoes for each visit to help standardize the trials. Instructions were given to avoid strenuous exercise 24 h before each visit and to keep any caffeine consumption consistent between trials. A 5-min self-selected warm-up was completed at the start of each session.

Analysis

Data were initially assessed for normality using the Shapiro Wilk test. Using SPSS statistical package 31 (SPSS Inc, Chicago, IL, USA) paired t-tests were conducted to assess differences in time to completion for VO_{2max} testing between trials.

Validity and reliability assessment

A five- step procedure was used to assess the validity and reliability of VM and NP for VO_2 and V_E data. First, paired t tests were used to compare VM and NP outcomes. Next intraclass correlations (ICC) were conducted. ICC estimates and their 95% confidence intervals (CI) were based on mean rating and absolute agreement (Koo et al., 2016). ICC model choice was also determined by ANOVA F test outcomes ($p < .05$). If the F test was not significant, a two-way random model was chosen, and if the F test was significant, a two-way mixed model was incorporated to account for systematic error (Koo et al., 2016). ICC were determined to be classified as weak at <0.80, moderate at <0.90 and strong at ≥ 0.90 (Weir, 2005).

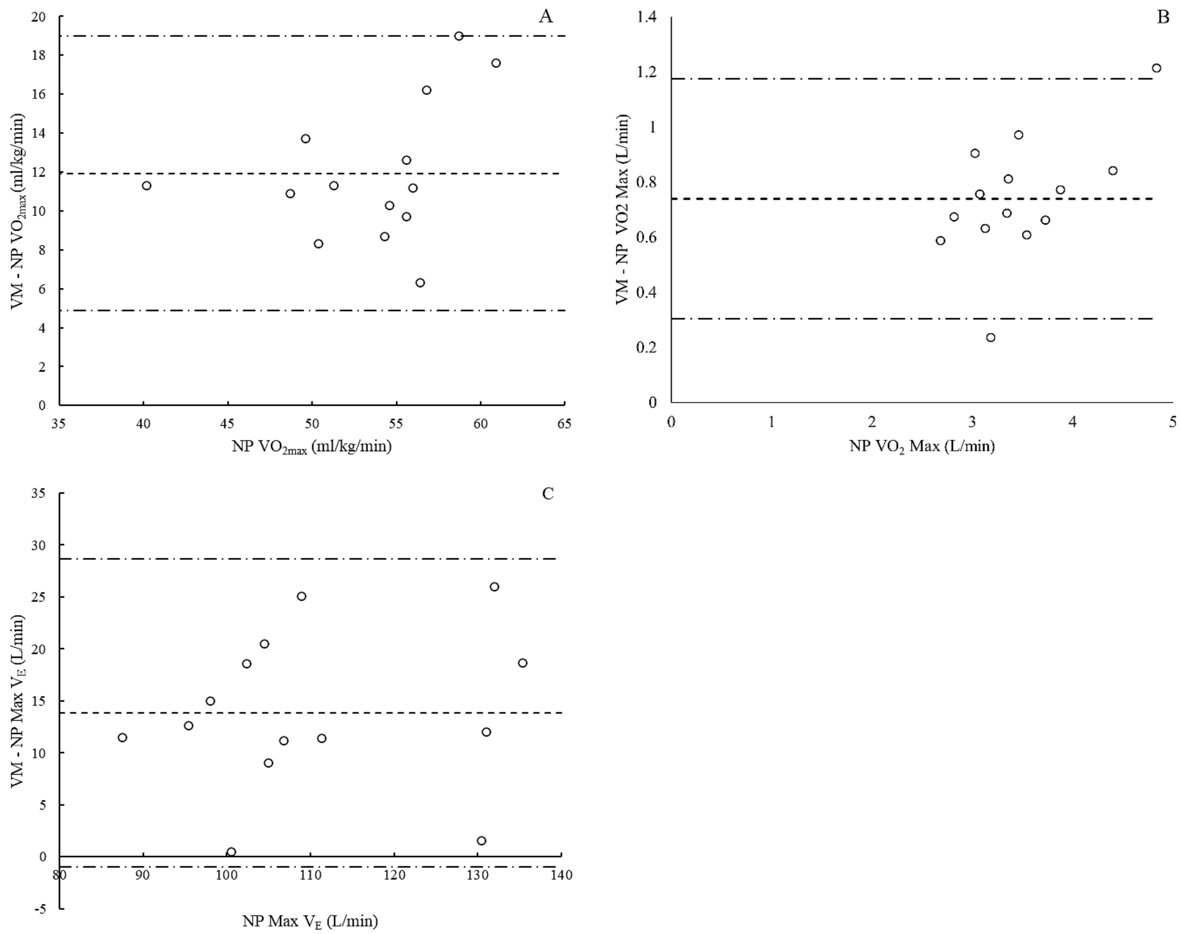
To provide an absolute unit-based estimate, the standard error of the measurement (SEM) was calculated by taking the root of the mean square of the residual as suggested by Vincent and Weir (2005). The 95% CI of the SEM was calculated by multiplying the SEM by 1.96. Within subject coefficient of variation (CV) (SD/mean) was calculated to present reliability in terms of percentage.

The CV for each individual was determined (SD/mean) x100) with all CV averaged for a grand CV%. Krouwer et al., (2008) suggested that if a new instrument was being tested against a criterion measurement, that the criterion measurement instead of the average of the criterion and new measurement be presented on the x-axis of a Bland-Altman plot (Bland & Altman et al., 1986). As such, modified Bland-Altman figures were developed to exhibit agreement with NP serving as the criterion measurement. Because of limitations for number of tables and figures, Bland-Altman plots were only visually depicted for relative VO_{2max} and relative VO_2 for the four running economy bouts. The mean agreement and 95% upper and lower limits of agreement were presented in tabular form for absolute VO_2.

RESULTS

VO_2max trials

There were no significant differences in time to completion (VM = 14:14 ± 1.03 mins; NP = 14:20 ± 1.05 mins; p = .54) between devices for the VO_{2max} tests. Relative and absolute VO_{2max} and maximal V_E all exhibited significant differences ($p < .001$) between treatments (Table 1). A two-way mixed effects model was used for all ICC as determined by a significant ANOVA F-test ($p < .05$). ICC for VO_{2max} outcomes were revealed to be weak with unacceptable levels of SEM and CV for VM to be considered a valid surrogate for NP (Table 1). The Bland-Altman plots in Figure 1 illustrate that all participants' VO_{2max} outcomes were biased in favouring higher outcomes with VM including 10 of 14 participants exceeding 10 ml/kg/min differences. V_E produced more promising results with a moderate ICC classification level, but SEM and CV% remained unideal (Table 1). Bland-Altman plots demonstrated more favourable agreement for V_E however, consistent overestimation bias was observed (Figure 1).



Note. A positive value for difference indicates a higher VO_{2max} for $VO_{2master}$ (VM) versus stationary metabolic NP (NP). Mean difference in agreement (short dashes) and ± 95% upper and lower levels of agreement (patterned dashes) for relative VO_{2max} (A) absolute VO_{2max} (B) and maximal V_E values were 11.9 ± 7.1 ml/kg/min, 0.74 ± 0.44 L/min, and 13.8 ± 14.8 L/min, respectively.

Figure 1. Modified Bland-Altman plots for relative VO_{2max} (A), absolute VO_{2max} (B), and maximal V_E (C).

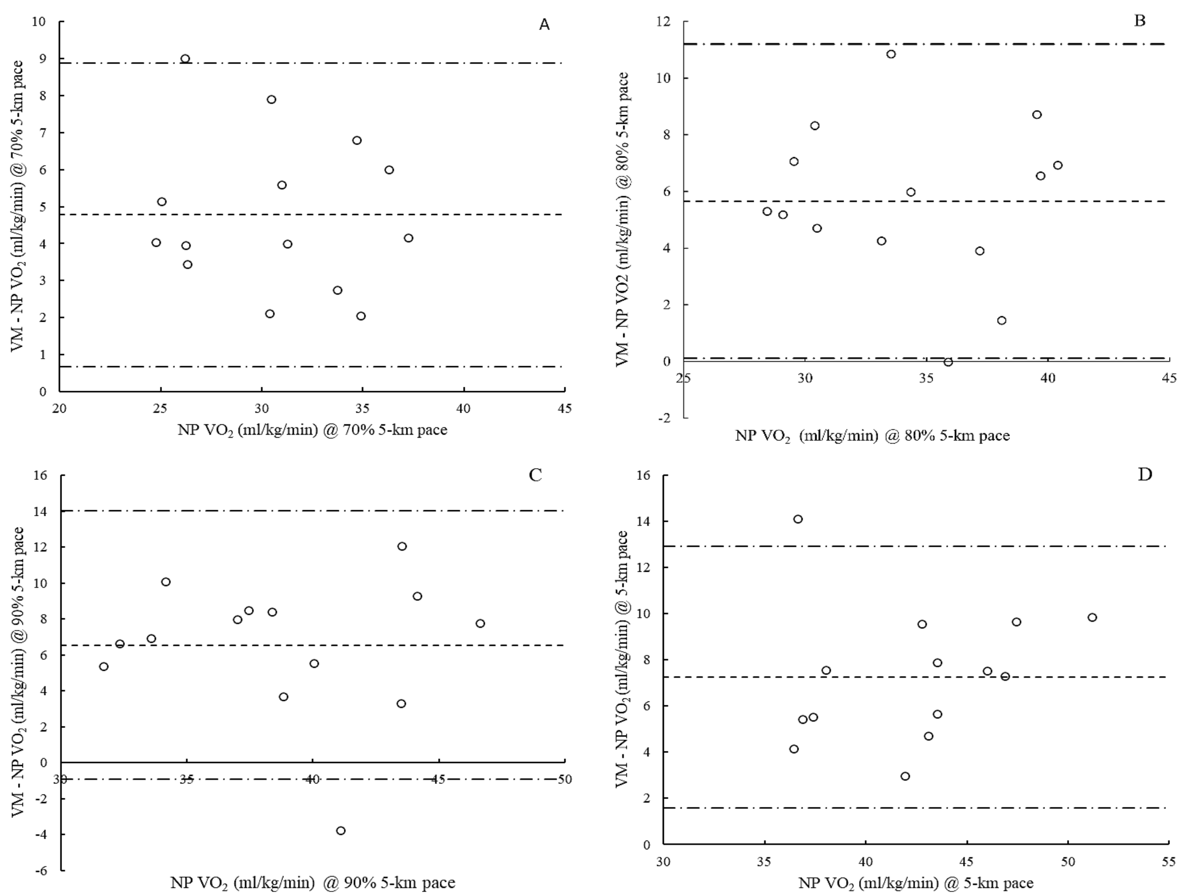
Table 1. Agreement between non-portable metabolic cart (NP) and VO_2 Master Pro (VM) for VO_{2max} and V_E (n = 14; mean ± SD).

	VO_{2max} (ml/kg/min)	VO_{2max} (L/min)	Max V_E (L/min)
NP	53.5 ± 5.2	3.46 ± 0.59	110.7 ± 15.3
VM	65.4 ± 7.3†	4.20 ± 0.74†	124.5 ± 18.1†
SEM	2.6	0.2	5.3
ICC [95% CI]	0.46 [-0.08-0.84]	0.74 [-0.08-0.95]	0.81 [-0.18-0.96]
CV	14.1%	13.6%	8.3%

Note. SEM = standard error of the measurement; ICC = intraclass correlations; CI = confidence interval; CV = coefficient of variation. † = statistical significance between NP and VM (p < .001).

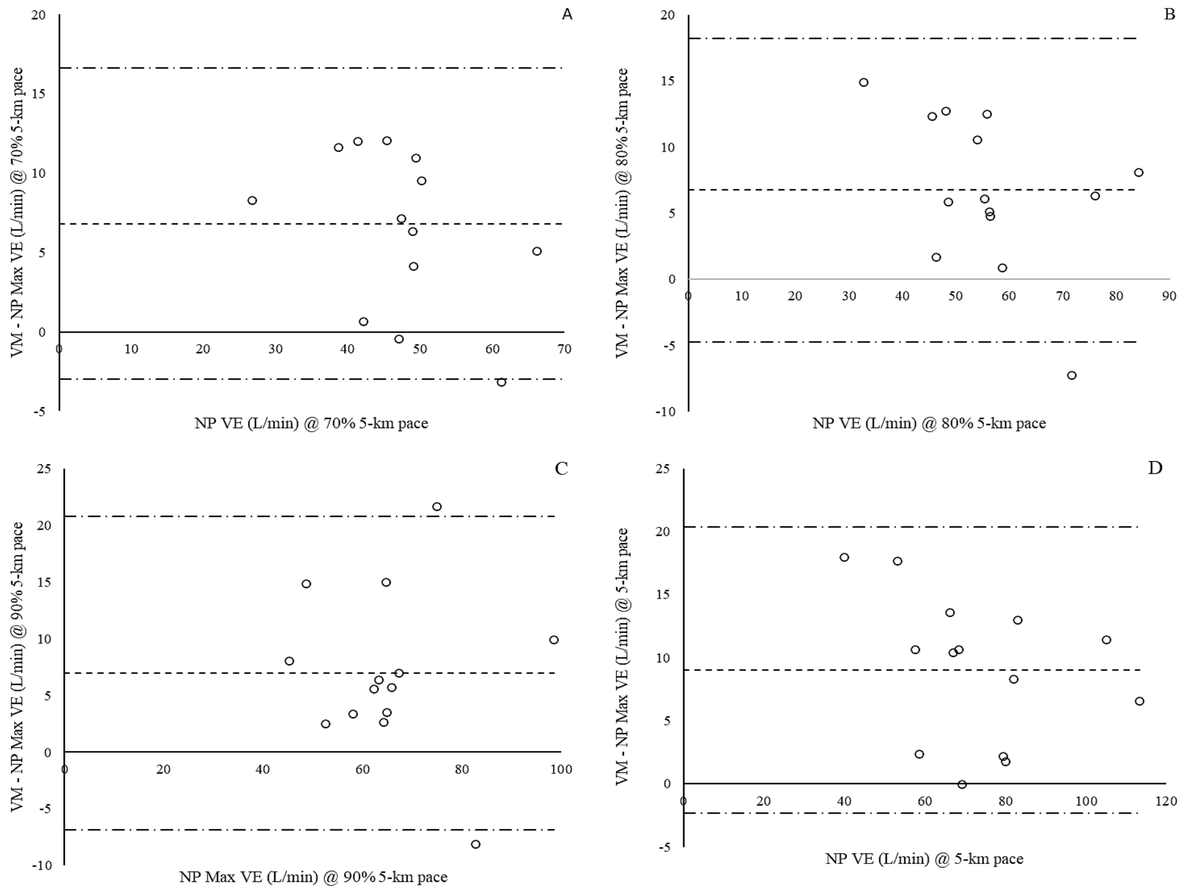
Running economy trials

RE outcome trends replicated those for VO_{2max}. All VO_2 and V_E outcomes at all intensities differed significantly (p < .001) between VM and NP (Table 2). No ICC for VO_2 reached the moderate classification, but V_E met or exceeded values of 0.87 at all 4 intensity levels (Table 2). Collectively, the magnitudes of SEM and CV (8.3-11.9%) in Table 2 and lack of strong agreement with nearly universal overestimation bias displayed by VM (Figures 2 and 3) do not support VM as an acceptable substitute for NP. The mean agreement and 95% upper and lower limits of agreement for absolute VO_2 are presented in Table 3.



Note. A positive value for difference indicates a higher VO_2 for $VO_{2master}$ (VM) versus stationary metabolic NP (NP). Mean difference in agreement represented by short dashes and 95% upper and lower levels of agreement represented by patterned dashes.

Figure 2. Modified Bland-Altman plots for relative VO_2 at (A) 70%, (B) 80%, (C) 90%, and (D) 100% of 5-km race pace.



Note. A positive value for difference indicates a higher V_E for $VO_{2master}$ (VM) versus stationary metabolic NP (NP). Mean difference in agreement represented by short dashes and 95% upper and lower levels of agreement represented by patterned dashes.

Figure 3. Modified Bland-Altman plots for V_E at (A) 70%, (B) 80%, (C) 90%, and (D) 100% of 5-km race pace.

Table 2. Agreement of cardiorespiratory outcomes measured via non-portable metabolic cart (NP) and portable VO_2 Master Pro (VM) at four running economy bout intensity levels (n = 14; mean ± SD).

		Relative VO_2 (ml/kg/min)		
	70% 5-km pace	80% 5-km pace	90% 5-km pace	100% 5-km pace
NP	30.6 ± 4.3	34.3 ± 4.3	38.8 ± 4.7	42.3 ± 4.7
VM	35.4 ± 4.6†	39.9 ± 4.9†	45.3 ± 6.0†	49.5 ± 5.9†
SEM	1.5	2.0	2.7	2.0
ICC [95% CI]	0.72 [-0.15-0.94]	0.63 [-0.19-0.90]	0.60 [-0.23-0.89]	0.62 [-0.14-0.90]
CV	10.3%	10.8%	11.9%	11.1%
		V_E (L/min)		
	70% 5-km pace	80% 5-km pace	90% 5-km pace	100% 5-km pace
NP	49.0 ± 11.4	56.4 ± 13.3	65.2 ± 13.7	73.1 ± 19.6
VM	55.8 ± 11.6†	63.2 ± 11.9†	72.2 ± 14.9†	82.1 ± 18.5†
SEM	3.5	4.1	5.0	4.1
ICC [95% CI]	0.87 [-0.05-0.97]	0.88 [0.12-0.97]	0.88 [0.28-0.97]	0.93 [0.01-0.98]
CV	10.3%	9.8%	8.3%	9.0%

Note. VE = minute ventilation; RE = running economy; SEM = standard error of measurement; ICC = intraclass correlation; CI = confidence interval; CV = coefficient of variation. † = statistical significance between NP and VM ($p < .001$).

Table 3. Bland-Altman agreement analyses outcomes for absolute VO_2 (not visually plotted as a figure) between non-portable metabolic cart (NP) and VO_2 Master Pro at four intensity levels (n = 14; mean ± SD).

	Absolute VO_2 (L/min)			
	70% 5-km pace	80% 5-km pace	90% 5-km pace	100% 5-km pace
Mean Difference	0.31 ± 0.14	0.36 ± 0.18	0.42 ± 0.24	0.47 ± 0.19
95% ULoA	0.59	0.70	0.89	0.84
95% LLoA	0.03	0.01	-0.04	0.09

Note. ULoA = upper limits of agreement; LLoA = lower limits of agreement.

DISCUSSION

This study evaluated the agreement of the VM versus a highly utilized NP for two common laboratory measurement scenarios VO_{2max} and RE testing in trained runners. There were three main findings. The first is VM demonstrated poor agreement with NP across all measured variables and exhibited weak ICC (ICC = 0.46 – 0.74) outcomes for VO_{2max} and VO_2 across the four RE intensities. Second, the error seems likely to be mostly attributed to O_2 sensor precision versus flow metrics as V_E demonstrated stronger consistency between devices (ICC range of 0.81 – 0.93 across all V_E scenarios). However, systematic bias remained evident, suggesting that while ventilatory measurements were more reliable, agreement with the NP was still unideal. Therefore, the improved ICC estimates for VE should be interpreted as evidence of greater measurement consistency rather than evidence of criterion validity. Finally, nearly all agreement bias for individual participants favoured VM reporting higher outcomes versus NP. These consistent and significant overestimation of variables of interest do not support the use of VM as a substitute for NP.

Our study is the first study to report VO_{2max} outcomes while wearing the VM versus an non-portable system (NPS) finding VM overestimated relative VO_{2max} by almost 12 mL/kg/min (Table 2) and 10 of the 14 participants exceeding 10 mL/kg/min difference compared to NP (Figure 1A). These results combined with wide limits of agreement suggest that relative VO_{2max} differ considerably from criterion values. In applied settings this variability could meaningfully alter interpretations of aerobic fitness, training adaptation, or exercise prescription. For example, overestimation of VO_{2max} could lead to coaches or practitioners to prescribe workloads that exceed an athlete's physiological capacity or lead to incorrect interpretations of longitudinal fitness changes. Montoye et al., (2020) had 16 male recreational cyclists complete an incremental maximal test using the same NPS as the current study, but reported outcomes by stage, not actual VO_{2max} (i.e. each stage produced different sample sizes due to attrition). It is difficult to compare these results data with such discrepancies in sample demographics and since less than half the sample was able to make it through the 4th stage. Seven subjects made it to the end of the 250 W stage. In this sub-group, mean absolute and mean absolute percent difference for the NPS were approximately 0.4 L/min and 12% higher than the VM depending on VM software edition incorporated (v. 1.1.1 and 1.2.1). Similar trends were found for V_E. These differences are impactful and trend in the opposite direction of the findings of the current study. It is possible the proprietary software (v. 1.6.1.) was altered based on the reporting of Montoye et al. (2020) resulting in an overcorrection. Previous research has demonstrated that VO_{2max} may be higher (5% - 15%) during treadmill exercise compared to cycle ergometry (Shepard, 1984; Loftin et al., 2004), due to a larger exercising muscle mass (Boileau et al., 1977), which may explain some discrepancies between our findings and that of Montoye et al., (2020).

Accurate evaluation of RE depends on precise measurements of VO_2 at moderate to high exercise intensities Previous research utilizing the VM for submaximal VO_2 and V_E measurements have also reported inconsistent findings (Kim et al., 2025; Davis et al., 2024; Thiessen et al., 2024). Thiessen et al., (2025) observed no differences between the VM and criterion devices (Cosmed Quark and Douglas Bags) for relative VO_2 while cycling at 100 Watts and 160 Watts. However, the sample size was relatively small with only 5 participants making interpretations difficult. During slow walking (3.2 km/h), brisk walking (5.6 km/h), and jogging (7.2 km/h) the VM consistently underestimated relative VO_2 by 6.0 ml/kg/min with a percent difference between 20.0% - 39.6% when compared to a Cosmed Quark. Lastly, Davis et al., (2024) reported no differences in between day relative VO_2 (35.1 ± 10.4 vs 35.4 ± 10.2 ml/kg/min) and V_E (72.4 ± 20.5 vs 73.8 ± 22.3 L/min) during submaximal cycling at 50%, 70%, and 85% of critical power for the VM. However, when compared to their criterion device (Cosmed Quark) the VM was not equivalent for VO_2 (37.7 ± 8.7 ml/kg/min vs. 35.2 ± 10.2 ml/kg/min) and V_E (74.0 ± 21.8 L/min vs. 73.1 ± 21.2 L/min). In the current study we observed consistent overestimation bias for both relative VO_2 and V_E (Table 2) during the four 5-min RE bouts and poor agreement was demonstrated across all intensities for relative VO_2 and V_E (Figure 2 and 3). Additionally, as exercise intensity increased measurement error increased suggesting the VM may become less accurate under higher ventilatory and metabolic demands. This aligns with previous research that reported larger discrepancies in VO_2 at higher exercise intensities (Montoye et al., 2020; Van Hooren et al., 2026). This is particularly important for endurance athletes and coaches since RE assessments are commonly performed at moderate to high submaximal intensities where small changes in VO_2 may meaningfully influence performance interpretations. These results highlight the need for caution when interpreting maximal and sub-maximal VO_2 and V_E outcomes using VM. Kim et al. (2025) made similar conclusions while reiterating that the mean difference seen between the VM and criterion metabolic systems should be included in individual outcome interpretations.

Table 4. Primary summaries of methodologies and related outcomes in studies assessing the validity and reliability of VO_2 and V_E using the VO_2 Master Pro (VM) during submaximal or maximal running or cycling exercise versus non-portable (NPS) and portable metabolic cart systems (PS).

	Sample	Protocol	Criterion instrument	Key findings
Montoye 2020	Recreational cyclists	Submax & max cycling	Parvomedics	Acceptable validity & reliability; overestimation at high intensity
Toulouse 2022	Recreational runners	Walking, jogging, running	Cosmed K5	Large underestimation of VO_2 & V_E
Van Hooren 2024	Simulated	1.0 L/min – 4.0 L/min	Relitech MS	~ 12% underestimation of VO_2
Davis 2024	Recreational cyclists	50 – 85% CP	Cosmed Quark	VM reliable vs. itself, not equivalent vs criterion
Kim 2025	Healthy adults	Walking, jogging, running	Cosmed Quark	Poor agreement & large underestimation
Thiessen 2025	Simulated + cycling	Cycling at 20, 100, 150 Watts	VacuMed MS, Cosmed Quark & Douglas Bags	High reliability (limited human data)
Van Hooren 2026	Simulated + cycling	Cycling 5% below VT	Relitech MS	~ 8% overestimation & mixed validity

Note. VO_2 = oxygen consumption; V_E = minute ventilation; VT = ventilatory threshold; km/h = kilometres per hour; ICC = intraclass correlation coefficient; CPET = cardiopulmonary exercise testing; MS = metabolic simulator; NPS = non-portable metabolic system; PS = portable metabolic system.

Previous validation studies using popular portable metabolic systems (Cosmed K2, Cosmed $K4b^2$, Oxycon Mobile, MetaMax 3B) have reported small overestimations of VO_2 (3% - 7%) during low to moderate exercise (McLaughlin et al., 2001; Brehm et al., 2004; Crouter et al., 2006; Ross et al., 2019) and larger discrepancies (5% - 9%) at high or maximal intensities (Rosdahl et al., 2010; Perez-Suarez et al., 2018) when compared to NPS. Toulouse et al., (2022) evaluated the validity of VM compared to an independently validated portable metabolic system (Cosmed K5) during indoor and outdoor walking (4.82 km/h), jogging (8.05 km/h), and running (11.29 km/h) and observed significant underestimations in both VO_2 and V_E across all intensity levels and conditions (Table 4).VM may also be influenced by both within and between day variability. Van Hooren et al., (2026) had 6 participants cycle at 5% below ventilatory threshold and found that the VM overestimated VO_2 by 8.4% despite demonstrating good between-day reliability for V_E (% error <3.0%). Additionally, the VM outperformed two other portable metabolic systems (KORR and Calibre) but not the Oxycon Pro or Vyntus CPX. Thiessen et al., (2025) reported high within and between device reliability for VO_2 (CV = 6.4% - 6.9%) during submaximal cycling at 100 Watts and 160 Watts compared to a Cosmed Qurak and Douglas Bags. Similarly, Davis et al., (2024) observed high within-device and between day reliability for VO_2 (35.1 ± 10.4 vs 35.4 ± 10.2 ml/kg/min) and V_E (72.4 ± 20.5 vs 73.8 ± 22.3 L/min) during submaximal cycling while using the VM. However, despite strong within device reliability, the VM was not equivalent to the Cosmed Quark for either VO_2 or V_E indicating strong within-device but weak between-device reliability (Davis et al., 2024).While RE typically exhibits low day to day variability (CV = 1 -2%) in highly trained long-distance runners (Morgan et al., 1994; Williams et al., 1991; Oehlert et al., 2025) the higher variability (CV = 10.3% - 11,1%) observed in the current study may impair the ability to detect small but meaningful training induced changes in long-distance runners, where changes in RE are often subtle. Future research should incorporate both within and between day reliability testing for both VO_{2max} and RE using the VM.

This study is not without limitations. First, the relatively small sample size and sex distribution may limit the generalizability of our findings. Secondly, because of the relatively small sample size the ICC estimates should be interpreted cautiously, since small samples may produce wider confidence intervals and less precise agreement estimates. Future studies should include larger and more sex-balanced samples to improve generalizability and confirm the current findings. Thirdly, accurate indirect calorimetry requires direct measurement of oxygen consumption and carbon dioxide production (Delsoglio et al., 2019). Notably, the VM lacks a carbon dioxide (CO2) sensor, instead estimating CO_2 production, which may explain why we observed more favourable agreement and reliability outcomes for V_E (Table 2; Figure 1C; Figure 3). V_E is directly derived from ventilatory flow measurements, whereas VO_2 requires additional estimation procedures. While VM proprietary algorithms for VO_2 estimation remains undisclosed, previous research has suggested VM relies on a simplified calculation assuming a fixed fraction of inspired oxygen (~ 21%) and it directly incorporates V_E into its VO_2 estimations (Kim et al., 2025) meaning that any error in V_E may propagate into VO_2 estimations, which may explain the consistent overestimation bias of VO_2 in the current study. The lack of direct CO_2 measurement may further increase estimation inaccuracies, particularly during high intensity exercise where ventilatory and metabolic demands are increased. Additional factors contributing to VO_2 error may include sensor calibration inaccuracies, STPD adjustment errors, and environmental assumptions, (Kim et al., 2025; Toulouse et al., 2022; Van Hooren et al., 2024; Thiessen et al., 2025). Lastly, the two systems rely on different gas analysis methods, which may also lead to slight variations in reported measurements.

CONCLUSION

In conclusion, the consistent and significant overestimation of VO_2 and V_E do not support VM as a viable replacement for NPS when accurate VO_2 assessment is required. However, the relatively stronger agreement observed for V_E and unidirectional bias may suggest that the VM may still have practical utility for monitoring general physiological trends when criterion laboratory systems are unavailable. Based on the current and previous findings researchers and practitioners should interpret VO_2 with caution when using the VM. Future research should explore correction factors for VO_2 estimations and determine which "*user piece*" estimates VO_2 and V_E most accurately during maximal and submaximal exercise.

AUTHOR CONTRIBUTIONS

All authors meet the criteria for authorship in accordance with established ethical guidelines. Contributions are specified according to the CRediT (Contributor Roles Taxonomy) as follows: Conceptualisation: Trace A. Dominy and Dr. Thomas L. Andre. Methodology: Trace A. Dominy and Dr. Thomas L. Andre. Formal analysis: Trace A. Dominy. Investigation: Trace A. Dominy, Asha D. Phillips, Katherine L. Tibbs, Brooke M. Gilmore, Julia P. Phillips, Dr. Thomas L. Andre. Data curation: Trace A. Dominy, Dr. Eric K. O'Neal. Writing – original draft: Trace A. Dominy. Writing – review & editing: Trace A. Dominy, Asha D. Phillips, Brooke m. Gilmore, Julia P. Phillips, Dr. Eric K. O'Neal, Dr. Thomas L. Andre. Supervision: Dr. Thomas L. Andre.

All authors have critically reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

AI USE DISCLOSURE

In accordance with current publishing ethics and transparency recommendations, artificial intelligence (AI) tools were used solely to assist with translation and language editing, with the aim of improving clarity and readability. No AI tools were used in the generation of scientific content, including the study design, data collection, analysis, interpretation of results, or the formulation of conclusions. The authors retain full responsibility for the content of the manuscript and confirm its originality, integrity, and accuracy.

REFERENCES

Arampatzis, A., De Monte, G., Karamanidis, K., Morey-Klapsing, G., Stafilidis, S., & Brüggemann, G. P. (2006). Influence of the muscle-tendon unit's mechanical and morphological properties on running economy. Journal of Experimental Biology, 209(17), 3345-3357. https://doi.org/10.1242/jeb.02340

Barnes, K. R., & Kilding, A. E. (2015). Running economy: Measurement, norms, and determining factors. Sports Medicine - Open, 1(1), 1-15. https://doi.org/10.1186/s40798-015-0007-y

Bland J. M., Altman D. G. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986 Feb 8;1(8476):307-10. PMID: 2868172. https://doi.org/10.1016/S0140-6736(86)90837-8

Blessinger, J., Sawyer, B., Davis, C., Irving, B. A., Weltman, A., & Gaesser, G. (2009). Reliability of the VmaxST portable metabolic measurement system. International Journal of Sports Medicine, 30(1), 22-26. https://doi.org/10.1055/s-2008-1038744

Boileau, R. A., Bonen, A., Heyward, V. H., & Massey, B. H. (1977). Maximal aerobic capacity on the treadmill and bicycle ergometer of boys 11-14 years of age.

Brehm, M. A., Harlaar, J., & Groepenhof, H. (2004). Validation of the portable VmaxST system for oxygen-uptake measurement. Gait & posture, 20(1), 67-73. https://doi.org/10.1016/S0966-6362(03)00097-3

Conley, D. L., & Krahenbuhl, G. S. (1980). Running economy and distance running performance of highly trained athletes. Medicine & Science in Sports & Exercise, 12(5), 357-360. https://doi.org/10.1249/00005768-198025000-00010

Crouter, S. E., Antczak, A., Hudak, J. R., DellaValle, D. M., & Haas, J. D. (2006). Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems. European journal of applied physiology, 98(2), 139-151. https://doi.org/10.1007/s00421-006-0255-0

Daniels, J., & Daniels, N. (1992). Running economy of elite male and elite female runners. Medicine & Science in Sports & Exercise, 24(4), 483-489. https://doi.org/10.1249/00005768-199204000-00015

Delsoglio, M., Achamrah, N., Berger, M. M., & Pichard, C. (2019). Indirect calorimetry in clinical practice. Journal of Clinical Medicine, 8(9), 1387. https://doi.org/10.3390/jcm8091387

Guidetti, L., Meucci, M., Bolletta, F., Emerenziani, G. P., Gallotta, M. C., & Baldari, C. (2018). Validity, reliability and minimum detectable change of COSMED K5 portable gas exchange system in breath-by-breath mode. PLOS ONE, 13(12), e0209925. https://doi.org/10.1371/journal.pone.0209925

Joubert, D. P., Guerra, N. A., Jones, E. J., Knowles, E. G., & Piper, A. D. (2020). Ground contact time imbalances strongly related to impaired running economy. International Journal of Exercise Science, 13(4), 427. https://doi.org/10.70252/DBRF4451

Kenney, W. L., Wilmore, J. H., & Costill, D. L. (2012). Physiology of sport and exercise (5th ed.). Human Kinetics.

Kim, J., Hong, K. R., Hwang, I. W., Wen, X., Shen, J. H., Kim, H. J., et al. (2025). The validity of the VO_2 Master Pro for measuring oxygen consumption during sedentary activity and treadmill walking and jogging. Applied Physiology, Nutrition, and Metabolism, 50, 1-11. https://doi.org/10.1139/apnm-2025-0095

Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun;15(2):155-63. doi:10.1016/j.jcm.2016.02.012. Epub 2016 Mar 31. Erratum in: J Chiropr Med. 2017 Dec;16(4):346. https://doi.org/10.1016/j.jcm.2017.10.001

Krouwer, J. S. (2008). Why Bland-Altman plots should use X, not (Y + X)/2 when X is a reference method. Statistics in Medicine, 27(5), 778-780. https://doi.org/10.1002/sim.3086

Loftin, M., Sothern, M., Warren, B., & Udall, J. (2004). Comparison of VO2 peak during treadmill and cycle ergometry in severely overweight youth. Journal of sports science & medicine, 3(4), 554.

Master VO2. (2019). Validity and reliability of the VM Pro. Retrieved from [Accessed 2026, 29 June]: https://vo2master.com/validation-studies/

McLaughlin, J. E., King, G. A., Howley, E. T., Bassett Jr, D. R., & Ainsworth, B. E. (2001). Validation of the COSMED K4 b2 portable metabolic system. International journal of sports medicine, 22(04), 280-284. https://doi.org/10.1055/s-2001-13816

Meyer, T., Davison, R. C. R., & Kindermann, W. (2005). Ambulatory gas exchange measurements: Current status and future options. International Journal of Sports Medicine, 26(S1), S19-S27. https://doi.org/10.1055/s-2004-830507

Meyer, T., Georg, T., Becker, C., & Kindermann, W. (2001). Reliability of gas exchange measurements from two different spiroergometry systems. International Journal of Sports Medicine, 22(8), 593-597. https://doi.org/10.1055/s-2001-18523

Montoye, A. H., Vondrasek, J., & Hancock, J. B. (2020). Validity and reliability of the VO_2 Master Pro for oxygen consumption and ventilation assessment. International Journal of Exercise Science, 13(4), 1382. https://doi.org/10.70252/THJT1177

Morgan, D. W., Craib, M. W., Krahenbuhl, G. S., Woodall, K., Jordan, S., Filarski, K., et al. (1994). Daily variability in running economy among well-trained male and female distance runners. Research Quarterly for Exercise and Sport, 65(1), 72-77. https://doi.org/10.1080/02701367.1994.10762210

Oehlert, G. M., Jones, E. J., Savala, P., & Joubert, D. P. (2025). Within-day reliability of submaximal running economy testing: Implications for footwear testing. Footwear Science, 17(3), 143-151. https://doi.org/10.1080/19424280.2025.2488755

Overstreet, B. S., Bassett, D. R., Jr., Crouter, S. E., Rider, B. C., & Parr, B. B. (2016). Portable open-circuit spirometry systems. Journal of Sports Medicine and Physical Fitness, 57(3), 227-237. https://doi.org/10.23736/S0022-4707.16.06049-7

Perez-Suarez, I., Martin-Rincon, M., Gonzalez-Henriquez, J. J., Fezzardi, C., Perez-Regalado, S., Galvan-Alvarez, V., et al. (2018). Accuracy and precision of the COSMED K5 portable analyser. Frontiers in Physiology, 9, 1764. https://doi.org/10.3389/fphys.2018.01764

Perkins, C. D., Pivarnik, J. M., & Green, M. R. (2004). Reliability and validity of the VmaxST portable metabolic analyzer. Journal of Physical Activity and Health, 1(4), 413-422. https://doi.org/10.1123/jpah.1.4.413

Rosdahl, H., Gullstrand, L., Salier-Eriksson, J., Johansson, P., & Schantz, P. (2010). Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. European Journal of Applied Physiology, 109(2), 159-171. https://doi.org/10.1007/s00421-009-1326-9

Ross, R., ALDuhishy, A., & González-Haro, C. (2020). Validation of the cosmed K4b2 portable metabolic system during running outdoors. The Journal of Strength & Conditioning Research, 34(1), 124-133. https://doi.org/10.1519/JSC.0000000000003050

Saunders, P. U., Pyne, D. B., Telford, R. D., & Hawley, J. A. (2004). Factors affecting running economy in trained distance runners. Sports Medicine, 34, 465-485. https://doi.org/10.2165/00007256-200434070-00005

Shephard, R. J. (1984). Tests of maximum oxygen intake a critical review. Sports medicine, 1(2), 99-124. https://doi.org/10.2165/00007256-198401020-00002

Thiessen, J. S., Guluzade, N. A., Faricier, R., & Keir, D. A. (2025). Reassessment of the COSMED Quark CPET and VO_2 Master Pro systems for measuring pulmonary gas exchange and ventilation. Scandinavian Journal of Medicine & Science in Sports, 35(2), e70019. https://doi.org/10.1111/sms.70019

Toulouse, A., Joubert, D., Oden, G., & Davis, P. R. (2022). Comparison of the VO_2 Master Pro and COSMED K5 during walking, jogging, and running. Journal of Science in Sport and Exercise, 4(2), 119-127. https://doi.org/10.1007/s42978-021-00146-w

Van Hooren, B., Souren, T., & Bongers, B. C. (2024). Accuracy of respiratory gas variables, substrate, and energy use from 15 CPET systems during simulated and human exercise. Scandinavian Journal of Medicine & Science in Sports, 34(1), e14490. https://doi.org/10.1111/sms.14490

Vincent, H. K., Madsen, A., & Vincent, K. R. (2022). Role of antigravity training in rehabilitation and return to sport after running injuries. Arthroscopy, Sports Medicine, and Rehabilitation, 4(1), e141-e149. https://doi.org/10.1016/j.asmr.2021.09.031

Ward, S. A. (2018). Open-circuit respirometry: Real-time, laboratory-based systems. European Journal of Applied Physiology, 118, 875-898. https://doi.org/10.1007/s00421-018-3860-9

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. The Journal of Strength & Conditioning Research, 19(1), 231-240. https://doi.org/10.1519/00124278-200502000-00038

Williams, T. J., Krahenbuhl, G. S., & Morgan, D. W. (1991). Daily variation in running economy of moderately trained male runners. Medicine & Science in Sports & Exercise, 23(8), 944-948. https://doi.org/10.1249/00005768-199108000-00010