

Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems

Scott E. Crouter · Amanda Antczak ·
Jonathan R. Hudak · Diane M. DellaValle ·
Jere D. Haas

Accepted: 25 May 2006 / Published online: 3 August 2006
© Springer-Verlag 2006

Abstract This study examined the accuracy and reliability of the MedGraphics VO2000 (VO2000) portable metabolic system and the ParvoMedics TrueOne 2400 (TrueOne 2400) metabolic cart against the criterion Douglas bag (DB) method. Ten healthy males (age 20 ± 1.7 years) had their gas exchange variables measured at rest and during cycling at 50, 100, 150, 200, and 250 W. Each stage was 10–12 min. For half of the stage gas exchange was measured with the DB and TrueOne 2400 simultaneously and for the other half of the stage gas exchange was measured with the VO2000. The testing was performed on two separate days and the order in which the equipment was used in each stage was randomized. Reliability between days for V_E (CV 7.3–8.8%) was similar among devices, however, for VO_2 , and VCO_2 the VO2000 (CV 14.2–15.8%) was less reliable compared to the DB (CV 5.3–6.0%) and TrueOne 2400 (CV 4.7–5.7%). The TrueOne 2400 was not significantly different from the DB at rest or any work rate for V_E , VO_2 , or VCO_2 ($P \geq 0.05$). The VO2000 was significantly different from the DB for V_E at 50–100 W, VO_2 at rest and 100–250 W, and VCO_2 at rest and 200–250 W (all, $P < 0.05$). The TrueOne 2400 provides accurate and reliable results for the measurement of gas exchange variables. The VO2000 portable metabolic system was less reliable for measur-

ing VO_2 and VCO_2 and generally overestimates VO_2 at most cycling work rates. Further research is needed to confirm the results found with the VO2000.

Keywords Oxygen consumption · Indirect calorimetry · Carbon dioxide production · Douglas bag

Introduction

Over the last two decades there has been an increase in the use of computerized metabolic systems. In addition, there has been an increase in the amount and types computerized systems available to researchers. The traditional method, and considered to be the “gold standard” of gas exchange measurements, is the Douglas bag (DB) method. While the DB method is considered the gold standard, it also has its own limitations, which provide a source of error that is difficult to account for. Limitations associated with the DB measurements include such things as: (1) the DB bag material being slightly permeable to room air allowing for the air collected in the DB to be mixed with external air, (2) difficulty in removing all the air from the DB and air leaking out during the removal process, (3) contents of the DB is representative of the entire sampling period, and (4) the DB method is time consuming and requires careful analysis by researchers to reduce errors while measuring the content of the bag. Over time with the advent of computerized systems and the development of new technology, the use of these systems has made gas exchange measurements easier and less time consuming, without compromising the accuracy provided by the DB method. In addition, as the technology becomes more sophisticated there is a movement

S. E. Crouter (✉) · D. M. DellaValle · J. D. Haas
Division of Nutritional Sciences, Cornell University,
279 MVR, Ithaca, NY 14853, USA
e-mail: sec62@cornell.edu

A. Antczak · J. R. Hudak
Department of Exercise and Sport Sciences, Ithaca College,
Ithaca, NY, USA

towards using portable gas exchange systems for the purpose of obtaining real life measurements rather than laboratory measurements. For a comprehensive review of portable metabolic systems and their application in sport science see Meyer et al. (2005).

The ParvoMedics TrueOne 2400 (TrueOne 2400) [previously called the ParvoMedics TrueMax 2400 (TrueMax 2400)] is a computerized metabolic cart for laboratory use. The TrueOne 2400, which uses a mixing chamber and is a non breath-by-breath system, has been shown to be an accurate device for the measurement of inspiratory and expiratory gas exchange variables (Bassett et al. 2001), however, its reliability has not been evaluated. Specifically, Bassett et al. (2001) found that compared to the DB the TrueMax 2400 provided highly accurate measures of minute ventilation (V_E) [mean error = -0.21 l min^{-1} , 95% prediction interval (95% PI), $-1.5, 0.09$], oxygen consumption (VO_2) (mean error = $-0.021 \text{ l min}^{-1}$, 95% PI, $-0.12, 0.08$), and carbon dioxide production (VCO_2) (mean error = $-0.011 \text{ l min}^{-1}$, 95% PI, $-0.07, 0.06$).

The MedGraphics VO2000 (VO2000) is a portable metabolic system, which to our knowledge, does not have a published validation on the accuracy and reliability of the device in comparison to DB measurements. There have been two studies that have compared the VO2000 and the MedGraphics CPX/D (CPX/D) (Byard and Dengel 2002; Olson et al. 2003). Olson et al. (2003) reported that there was no difference between the VO2000 and CPX/D for the measurement of V_E , VO_2 , and VCO_2 at rest. This study used a low-flow pneumotach designed by MedGraphics specifically for use during resting measures and is not recommended for exercise testing. A separate study by Byard and Dengel (2002) showed that there was no significant difference between the VO2000 and the CPX/D for the measurement of V_E , VO_2 , and VCO_2 during cycle ergometry. While these studies have shown the VO2000 to give similar values as the CPX/D they should be interpreted with caution because the CPX/D has been shown to underestimate VO_2 by approximately 11% in athletes compared to the DB method during cycling activity (Gore et al. 2003). This raises concern about the application of the VO2000 and the true validity of this portable device.

With new technology and devices emerging in the market it is important that researchers are aware of the accuracy and reliability of these devices. Therefore, the purpose of this study was to compare both the TrueOne 2400 and VO2000 against the DB method for the measurement of metabolic variables. In doing so the accuracy and test re-test reliability of the devices will be determined.

Methods

Subjects

Ten healthy male subjects from the Ithaca community volunteered to participate in the study. The procedures were reviewed and approved by the University Committee on Human Subjects at Cornell University, before the start of the study. Each participant signed a written informed consent and completed a physical activity readiness questionnaire (PAR-Q) before participating in the study. Participants were excluded from the study if they had any contraindications to exercise. Physical characteristics [mean (SD)] of the participants were; age, 20 (1.7) years; height, 179.3 (11.6) cm; weight, 78.2 (12.7) kg; and BMI, 24.2 (2.0).

Equipment

The VO2000 is a portable metabolic system that is worn on the chest with a harness and weighs 1.57 kg (includes metabolic unit, battery pack, harness, heart rate monitor, face mask, and breathing valve). The device employs a patented flow meter, which uses a proportional sampling valve and a 3-breath average for the measurement of VO_2 , VCO_2 , and V_E . Specifically, the VO2000 uses a galvanic fuel cell for the oxygen analyzer (range 0–96%), a non-dispersive infrared carbon dioxide analyzer (range 0–10%), and proprietary software. The flow device is a bi-directional differential pressure preVent™ pneumotach. According to the company it has a low flow pneumotach (range 2–30 1 min^{-1}) with an accuracy of $\pm 3\%$ of the absolute volume and a dead space of 8 ml for resting measures and an exercise pneumotach (range 20–200 1 min^{-1}) an accuracy of $\pm 3\%$ of the absolute volume, and a dead space of 39 ml. The flow device is connected to a neoprene facemask (with a silicone adapter) that covers the subject's mouth and nose. The dead space of the pneumotach and face mask is 39 ml, which is the value used when calculating the gas fractions. The flow device is connected to the portable metabolic unit with a flexible rubber hose and gas is sampled using a patented gas drying sample circuit contained within the portable unit. In addition, the system can also provide heart rate (HR), however, the current software program does not allow for the gas fractions ($F_{E}O_2$ and $F_{E}CO_2$) to be viewed. As with other portable systems the data can be transmitted to a computer or stored in the system and downloaded to a computer at a later time. Prior to testing, the VO2000 was calibrated according to the manufacturer's instructions. This consists of performing an auto-calibration routine that is

run through the software, which uses a proprietary room air calibration of the oxygen and carbon dioxide analyzers as well as an auto-calibration procedure for the pneumotach. The current software program utilized with the VO2000 does not allow for a two-point calibration using known gas volumes or manual calibration of the pneumotach with a calibration syringe.

The TrueOne 2400 computerized metabolic measurement system uses a Hans Rudolf 3813 (Kansas city, MO) pneumotachometer to measure ventilation. A limitation to this pneumotachometer is that it is non-linear at flow rates less than 80 l min^{-1} . Thus, Yeh and colleagues developed an algorithm to correct the linearity at low flow rates (Yeh et al. 1982, 1987). The TrueOne 2400 is a mixing chamber system, that uses a paramagnetic oxygen analyzer (range 0–25%) and an infrared, single beam, single wave-length carbon dioxide analyzer (range 0–10%). Prior to each testing session the gas analyzers were calibrated according to the manufacturer's recommendations. This consists of a room air auto-calibration routine and a two-point gas calibration with a single gas tank (15.09% O_2 , 6.01% CO_2). In addition, the flow meter was calibrated using a 3.000 l Hans Rudolf 5530 series syringe. This involves a five stroke calibration using different flow rates for each stroke.

Douglas bag collections were made simultaneously with the TrueOne 2400 measurements as described elsewhere (Bassett et al. 2001). Briefly, a three-way Warren E. Collins t-stop-cock and DB (200 l) were placed in series with the mixing chamber used by the TrueOne 2400 metabolic unit. Specifically, the t-stop-cock was attached to the air port on the back of the mixing chamber and the DB was attached to the t-stop-cock. During the last two minutes of each stage the t-stop-cock was turned so the air was directed into the DB for collection of the expired air. At the end of each collection period, the gas fractions from the DB were measured (over a two minute sampling period) using a ParvoMedics TrueMax 2400 with a Paramagnetic oxygen analyzer (range 0–25%) and an infrared, single beam, single wave-length carbon dioxide analyzer (range 0–10%). This was a different ParvoMedics TrueMax 2400 from the one used during the measurement of the metabolic variables at rest and exercise. The expired volume was determined by pushing the remaining collected air from the DB into a 120 l Tissot gasometer (Warren E. Collins, Braintree, MA). Corrections in volume were made for the volume of air removed for gas analysis. The TrueOne 2400 system samples expired air at the back of the mixing chamber with a flow rate of 0.36 l min^{-1} (flow rate determined by flow meter built into the TrueOne 2400 metabolic

cart), therefore there was a total of 0.72 l of air removed during the two minute sampling period. In addition, the flow rate of the TrueMax 2400 used to analyze the DB air sample (1 min sample period) had a flow rate of 0.28 l min^{-1} . Thus, the total correction for the volume of air removed was 1.0 l. Prior to each test the gas analyzers were calibrated using room air and a single gas tank (15.09% O_2 , 6.01% CO_2).

Experimental design

Initially, pilot tests were performed on four subjects to examine the effects of simultaneous gas exchange measurements with the DB in series with the VO2000. This was done by having the subject ride at 50 and 100 W with only the VO2000 connected for 5 min, then a two-way Hans-Rudolph valve (2700 series), hose and DB were connected in series with the VO2000 breathing valve and the subject cycled another 5 min. The addition of the Hans-Rudolph valve caused an increase in the VO_2 values reported by the VO2000. Specifically, at 50 W the VO_2 changed from 0.93 to 1.24 l min^{-1} when the Hans-Rudolph valve was added and at 100 W the VO_2 changed from 1.36 to 1.59 l min^{-1} when the Hans-Rudolph valve was added. Therefore, the study was designed to obtain non-simultaneous measurements of the gas exchange variables from the VO2000 and DB, while simultaneous measurement between the TrueOne 2400 and DB were obtained.

For the comparison of the metabolic systems all subjects performed two trials within 48 h at the same time of day. Prior to testing, subjects were fitted with a mouthpiece, nose clip, and headgear which were used with the TrueOne 2400 and DB measurements. They were also fitted with a neoprene face mask which would be used with the VO2000. A Polar HR strap was also fitted around the chest of the subject so a criterion for HR could be obtained with a Polar A1 HR monitor (Polar Electro Oy, Kempele, Finland). Each trial consisted of an 11 min rest period, followed by cycling on a Lode Corival (Groningen, The Netherlands) electronically braked cycle ergometer for 10 min at 50, 100, and 150 W, and 11 min at 200 and 250 W.

During one trial, the TrueOne 2400 and DB setup was connected to the subject for the first 6 min of rest, first 5 min at 50, 100, and 150 W and first 6 min at 200 and 250 W, after which the equipment was switched out and the VO2000 was connected to the subject for the remaining time of each stage. During the second trial the order in which the equipment was connected to the subjects was reversed. The order in which the equipment was connected to the subjects was randomized so that on day one half of the subjects were con-

nected to the TrueOne 2400 and DB first and half of the subjects were connected to the VO2000 first. For each device, the average of last 2 min of each stage, were used to calculate the gas exchange variables. After the subjects completed the 150 W stage they were given a 5 min rest period so the gas analyzers could be recalibrated and checked for analyzer drift. The last 2 min of each measurement period for the TrueOne 2400 and the VO2000 were averaged and compared with the DB 2 min collection period.

Statistical analyses

Statistical analyses were carried out using SPSS version 13.0 for Windows (SPSS Inc., Chicago, IL). For all analyses, an alpha level of 0.05 was used to indicate statistical significance. All values are reported as mean \pm standard deviation (SD). To examine the reliability of the devices several measurements were used: (1) A two-way repeated measures ANOVA (day of testing \times power output) for each device for V_E , VO_2 , and VCO_2 , was used to examine if there was a difference between days. (2) Pearson correlation coefficients (r) was used to compare the values obtained for V_E , VO_2 , and VCO_2 on the two separate days for rest and all work rates combined. (3) Bland–Altman plots were used to graphically show the variability in the individual error scores between days for rest and all work rates combined (Bland and Altman 1986). This allowed for the mean error score and the 95% prediction interval to be shown. Devices that are reliable will display a tight prediction interval around zero. Data points below zero signify an overestimation, while data points above zero signify an underestimation. (4) The data showed heteroscedasticity (i.e., as the work rate increased the difference between days increased), therefore the data was logarithmically (natural) transformed, as suggested by Atkinson and Nevill (1998). The logarithmic transformed data was then used to determine the mean bias and random error component (ratio limits of agreement), which more accurately describes the variability across the range of values reported. (5) Lastly, the coefficient of variation (CV) between the day-to-day measures was determined for V_E , VO_2 , and VCO_2 for rest and all work rates combined.

We choose several statistical tests to determine the reliability of the devices as suggested by Atkinson and Nevill (1998). While we used logarithmically transformed data to determine the mean bias and random error, we used the untransformed data for the graphical presentation in the Bland–Altman plots so readers could more easily understand the values. Since the use of ratio limits of agreement is fairly new to sport sci-

ence studies it is important to have an understanding about how to interpret the results. The ratio limits of agreement can be interpreted using the following example; the ratio limits of agreement for V_E , in the current study, was $0.98 \times / \div 1.38$, meaning that there was a 2% bias and for any individual from the population, two tests on different days will differ by no more than 38%. It is also important to point out that addition and subtraction (i.e., \pm) with the natural scale is equivalent to multiplying and dividing (i.e., \times / \div) using the ratio scale. Thus, when using the ratio scale the mean bias is multiplied and divided by the 95% prediction interval. For the example above this would give a mean bias of 2% and the 95% prediction interval would be 0.71, 1.35%. For a further discussion on the use of ratio limits of agreement in sport science see Atkins and Nevill (1998).

To examine if there was a device order effect we examined the ANOVA performed for the reliability testing. The results from the ANOVA showed no difference in V_E , VO_2 , or VCO_2 for any of the devices, thus we concluded that there was not an effect from the order of device use. For the remaining analysis the mean metabolic variables were averaged between the two days. Subsequently, two-way ANOVAs (device \times power output) were used to compare differences among the devices for each dependent variable. Additionally, because only nine subjects were able to complete 200 W and six subjects completed 250 W, these work rates were not included in this analysis, and a separate ANOVA was used to compare the three devices at 200 and 250 W, separately. Pairwise comparisons with Bonferroni adjustments were performed to locate significant differences between devices, when necessary. In addition, modified Bland–Altman plots were used to graphically show the variability in the individual error scores (DB variable minus device variable) (Bland and Altman 1986).

Results

None of the devices showed significant differences between days for V_E (DB, $F_{5,49} = 1.833$, $P = 0.124$; TrueOne 2400, $F_{5,49} = 1.840$, $P = 0.122$; VO2000, $F_{5,49} = 0.428$, $P = 0.827$), VO_2 (DB, $F_{5,49} = 1.286$, $P = 0.285$; TrueOne 2400, $F_{5,49} = 1.837$, $P = 0.123$; VO2000, $F_{5,49} = 1.600$, $P = 0.178$), or VCO_2 (DB, $F_{5,49} = 1.459$, $P = 0.220$; TrueOne 2400, $F_{5,49} = 1.313$, $P = 0.274$; VO2000, $F_{5,49} = 0.716$, $P = 0.614$). Figure 1a–i shows the Bland–Altman plots depicting the individual difference scores (difference between days) for V_E , VO_2 , and VCO_2 for each device. Additionally, the reliability

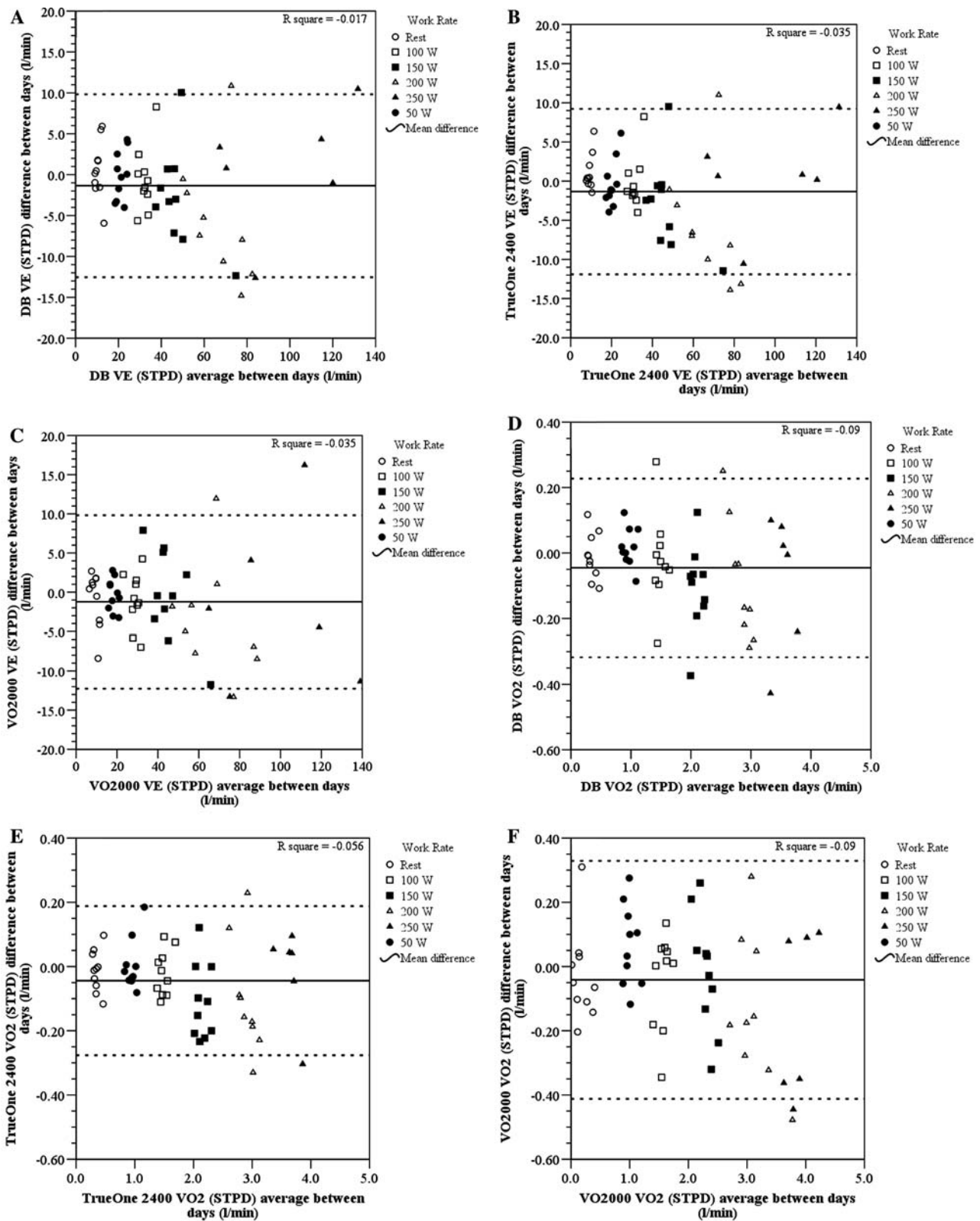


Fig. 1 Bland–Altman plots of the error scores (between testing days) for minute ventilation (V_E), oxygen consumption (VO_2) and carbon dioxide production (VCO_2). **a** Douglas bag (DB) V_E , **b** ParvoMedics TrueOne 2400 (TrueOne 2400) V_E , **c** MedGraphics VO2000 (VO2000) V_E , **d** DB VO_2 , **e** TrueOne 2400 VO_2 , **f**

VO2000 VO_2 , **g** DB VCO_2 , **h** TrueOne 2400 VCO_2 , **i** VO2000 VCO_2 . *Solid lines* represent the mean difference; *dashed lines* represent the 95% prediction interval. V_E , VO_2 and VCO_2 are reported in $l\ min^{-1}$ and STPD

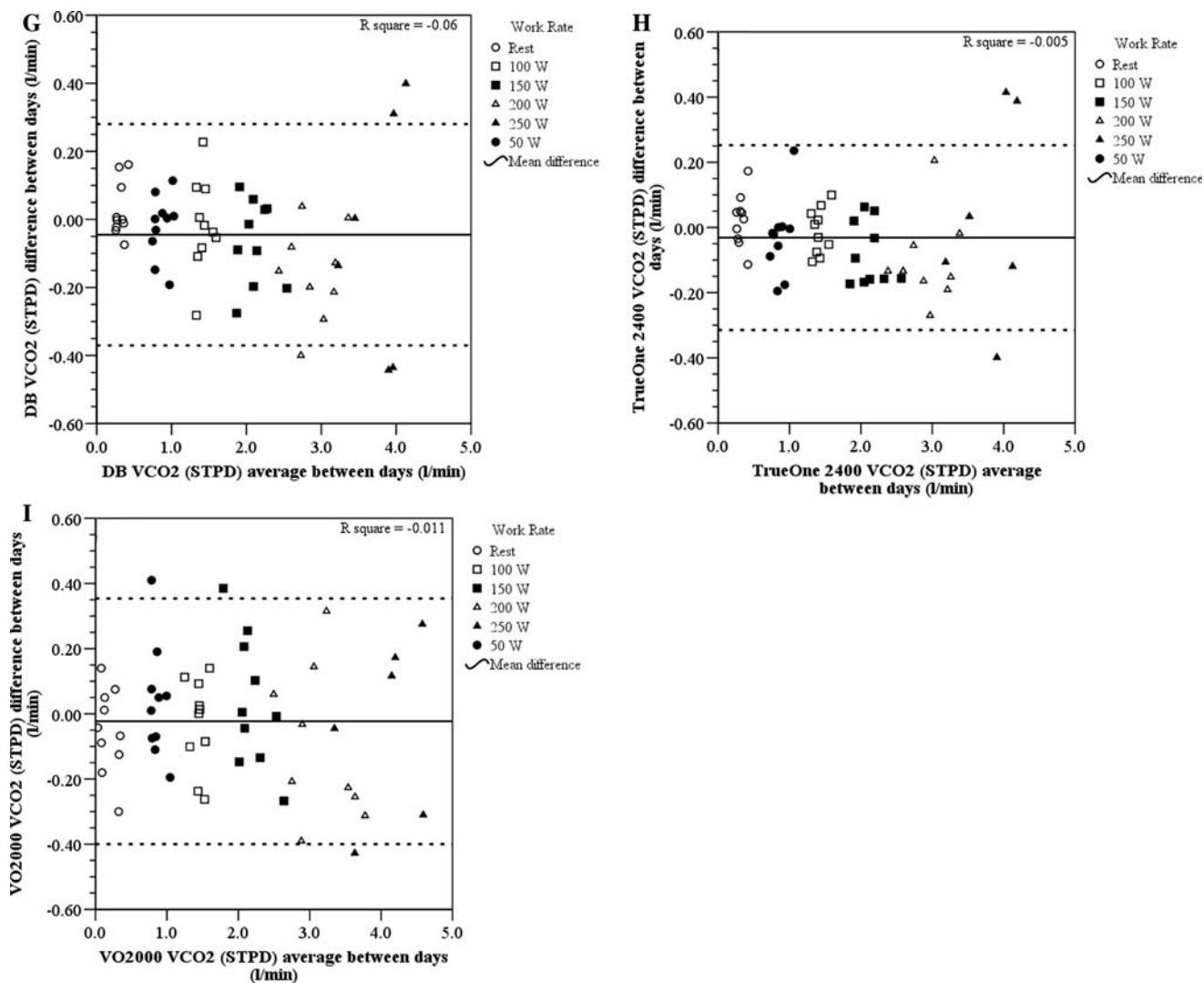


Fig. 1 continued

Table 1 Reliability of the Douglas bag (DB), ParvoMedics TrueOne 2400 (TrueOne 2400), and Medgraphics VO2000 (VO2000) for the measurement of minute ventilation (V_E), oxygen consumption (VO_2), carbon dioxide production (VCO_2). Pearson correlation coefficient (r), ratio limits of agreement, mean error and 95% prediction interval [mean error (95% PI)] and coefficient of variation (CV)

	r	Ratio limits of agreement	Mean error (95% PI)	CV (%)
V_E (STPD, $l \text{ min}^{-1}$)				
Douglas bag	0.970*	0.98 \times/\div 1.38	-1.36 (-12.54, 9.83)	8.5
TrueOne 2400	0.975*	0.99 \times/\div 1.33	-1.34 (-11.91, 9.22)	7.3
VO2000	0.983*	0.97 \times/\div 1.42	-1.22 (-12.27, 9.82)	8.8
VO_2 (STPD, $l \text{ min}^{-1}$)				
Douglas bag	0.990*	0.98 \times/\div 1.24	-0.05 (-0.32, 0.23)	5.3
TrueOne 2400	0.994*	0.98 \times/\div 1.19	-0.04 (-0.28, 0.19)	4.7
VO2000	0.989*	0.94 \times/\div 3.30	-0.04 (-0.41, 0.32)	14.2
VCO_2 (STPD, $l \text{ min}^{-1}$)				
Douglas bag	0.988*	0.99 \times/\div 1.29	-0.05 (-0.37, 0.28)	6.0
TrueOne 2400	0.991*	0.99 \times/\div 1.26	-0.03 (-0.32, 0.25)	5.7
VO2000	0.989*	0.90 \times/\div 5.57	-0.02 (-0.40, 0.35)	15.8

r pearson correlation coefficient, PI prediction interval, CV coefficient of variation, * $P < 0.01$

between days for V_E , VO_2 , and VCO_2 for each device determined by the Pearson correlation coefficient, ratio limits of agreement (determined from the logarithmically transformed data), and CV are shown in Table 1. All three devices appear to have good test re-test reliability between days for V_E , VO_2 , and VCO_2 according to the Pearson correlation coefficient. However, as the Bland–Altman plots show and is confirmed by the ratio limits of agreement and CV, the reliability of measuring V_E between days for the three devices is similar, but the VO2000 was less reliable compared to the DB and TrueOne 2400 for the measurement of VO_2 and VCO_2 .

Table 2 shows the physiological responses measured by each device and Fig. 2 shows the percent over- and/or underestimation for the TrueOne 2400 and VO2000 versus the DB for these physiological responses. The TrueOne 2400 metabolic cart was not significantly different from DB at any work rate for V_E , VO_2 , or VCO_2 ($P \geq 0.05$). Mean $F_{E}O_2$ however, was significantly lower at rest and 50 W and mean $F_{E}CO_2$ was significantly higher at rest, 50 and 100 W ($P < 0.05$). The VO_2 measured by the VO2000 was significantly different from DB and TrueOne 2400 at all work rates except 50 W, the V_E was significantly different from DB and TrueOne 2400 at 50 and 100 W, the VCO_2 was significantly different from DB and TrueOne 2400 at rest, 200, and 250 W. There were no differences in the respiratory rate (RR) between the TrueOne 2400 and VO2000. Lastly, the HR for the VO2000 was significantly different from the Polar HR monitor at rest and 200 W, while there were no significant differences between the Polar HR monitor and the TrueOne 2400 HR. In addition, when there were no statistically significant differences between the Polar HR monitor and the VO2000 the HR was still off on average by 10 to 18 bpm, indicating that there may be a malfunction with the device for measuring HR and further investigation is warranted.

Figures 3a–d and 4a–d are the Bland–Altman plots, showing the individual difference scores (DB minus computerized system) for the TrueOne 2400 and VO2000. Overall, there was close agreement between the DB method and the TrueOne 2400 metabolic system for all variables. The plots for the VO2000 show that it tends to underestimate VO_2 and VCO_2 at the lighter work rates and overestimate VO_2 and VCO_2 at the higher work rates.

Discussion

The main purpose of this study was to test the accuracy and reliability of the TrueOne 2400 metabolic cart and the VO2000 portable metabolic system. Under ideal cir-

Table 2 Physiological responses measured during rest and five work rates on a cycle ergometer using DB, a ParvoMedics TrueOne2400 metabolic cart and VO2000 portable metabolic system

	Douglas bag	ParvoMedics TrueOne 2400	MedGraphics VO2000
V_E (l min ⁻¹)			
Rest	10.6 ± 2.2	9.4 ± 1.7	9.3 ± 2.4
50 W	21.3 ± 2.7	20.3 ± 2.7	18.4 ± 2.1*#
100 W	32.3 ± 3.2	31.3 ± 2.9	29.0 ± 3.1*#
150 W	47.6 ± 11.1	47.2 ± 11.1	45.2 ± 9.4
200 W	66.5 ± 13.0	66.4 ± 13.3	67.2 ± 14.8
250 W	98.0 ± 26.9	98.1 ± 26.3	99.2 ± 27.8
VO_2 (l min ⁻¹)			
Rest	0.35 ± 0.08	0.35 ± 0.07	0.18 ± 0.15*#
50 W	0.96 ± 0.10	0.96 ± 0.10	1.00 ± 0.12
100 W	1.48 ± 0.10	1.49 ± 0.09	1.57 ± 0.12*#
150 W	2.10 ± 0.14	2.15 ± 0.13	2.30 ± 0.16*#
200 W	2.83 ± 0.21	2.90 ± 0.18	3.12 ± 0.32*#
250 W	3.52 ± 0.21	3.65 ± 0.18	3.88 ± 0.26*#
VCO_2 (l min ⁻¹)			
Rest	0.32 ± 0.07	0.32 ± 0.07	0.18 ± 0.14*#
50 W	0.87 ± 0.12	0.87 ± 0.12	0.86 ± 0.12
100 W	1.42 ± 0.11	1.42 ± 0.10	1.45 ± 0.12
150 W	2.11 ± 0.21	2.12 ± 0.22	2.19 ± 0.26
200 W	2.90 ± 0.32	2.94 ± 0.33	3.14 ± 0.44*#
250 W	3.77 ± 0.39	3.83 ± 0.42	4.08 ± 0.50*#
$F_{E}O_2$			
Rest	0.1761 ± 0.0081	0.1719 ± 0.0065*	
50 W	0.1647 ± 0.0035	0.1628 ± 0.0034*	
100 W	0.1634 ± 0.0040	0.1620 ± 0.0037	
150 W	0.1638 ± 0.0069	0.1626 ± 0.0065	
200 W	0.1655 ± 0.0069	0.1643 ± 0.0063	
250 W	0.1704 ± 0.0101	0.1694 ± 0.0096	
$F_{E}CO_2$			
Rest	0.0312 ± 0.0070	0.0355 ± 0.0051*	
50 W	0.0414 ± 0.0028	0.0431 ± 0.0027*	
100 W	0.0448 ± 0.0036	0.0458 ± 0.0031*	
150 W	0.0456 ± 0.0050	0.0468 ± 0.0057	
200 W	0.0448 ± 0.0054	0.0463 ± 0.0044	
250 W	0.0408 ± 0.0083	0.0410 ± 0.0078	
RR, (breaths min ⁻¹)			
Rest		13 ± 4	14 ± 7
50 W		21 ± 8	20 ± 4
100 W		25 ± 9	25 ± 6
150 W		30 ± 11	30 ± 8
200 W		33 ± 8	34 ± 8
250 W		45 ± 13	46 ± 15
HR, (beats min ⁻¹)			
Rest	78 ± 17	78 ± 17	62 ± 35*
50 W	97 ± 17	98 ± 17	85 ± 36
100 W	121 ± 21	121 ± 21	104 ± 45
150 W	146 ± 28	146 ± 28	127 ± 49
200 W	163 ± 28	160 ± 27	134 ± 54*
250 W	176 ± 19	176 ± 19	166 ± 34

V_E , VO_2 and VCO_2 are reported as STPD and $F_{E}O_2$ and $F_{E}CO_2$ are the mean O_2 and CO_2 fractions during the expiratory phase. Values are mean ± SD. V_E minute ventilation, $F_{E}O_2$ fraction of oxygen in expired air, $F_{E}CO_2$ fraction of carbon dioxide in expired air, VO_2 oxygen uptake, CO_2 carbon dioxide production, RR respiratory rate, HR heart rate. *Significantly different from the DB, #significantly different from the ParvoMedics TrueOne 2400

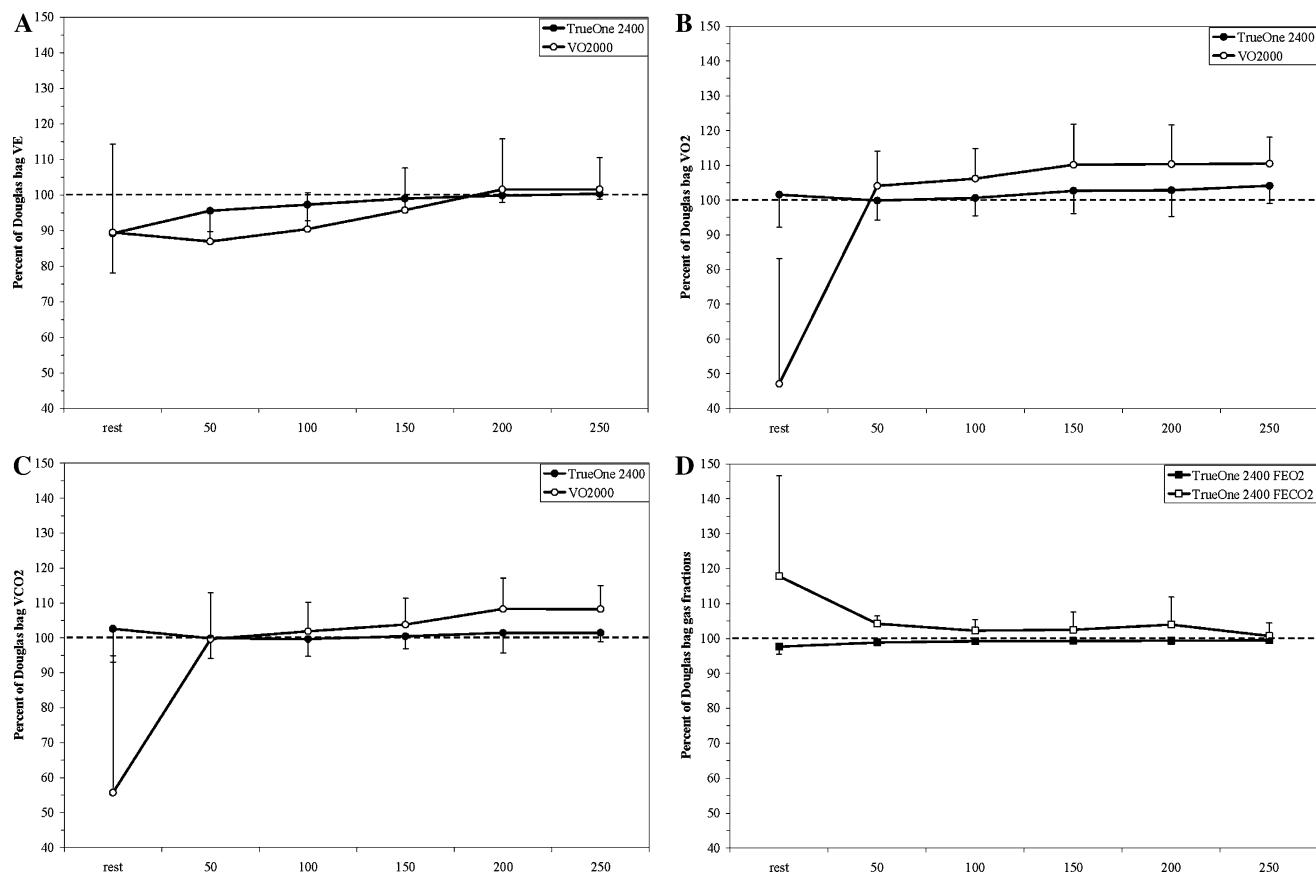


Fig. 2 Percent of the DB **a** ventilation (V_E), **b** oxygen consumption (VO_2), **c** carbon dioxide production (VCO_2), and **d** gas fractions for the TrueOne 2400 and VO2000 during rest and five work rates on the cycle ergometer. Values are means with standard

deviation bars. V_E , VO_2 and VCO_2 are reported as STPD and $F_{E}O_2$ and $F_{E}CO_2$ are the mean O_2 and CO_2 fractions during the expiratory phase

cumstances the metabolic measurements would have been made simultaneously (Bigard and Guezennec 1995; Lucia et al. 1993; Peel and Utsey 1993), however, the addition of the Hans-Rudolf valve to the breathing valve of the VO2000 altered the VO_2 values preventing us from making simultaneous measurements. Therefore, we tested the participants on two separate days (within 48 h) randomizing the order in which the devices were tested each day. The testing on separate days has been shown to result in a small (3–4%) variation in the day-to-day measurements of VO_2 (Stuart et al. 1981). Thus, the non-simultaneous measurements could have contributed to some of the differences seen between days and between the DB and VO2000, and should be considered a limitation of this study.

Reliability testing

The reliability results from this study should be interpreted with caution due to a few limiting factors; (1) the inclusion of resting measures into the analysis

increased the variability, and (2) we had a small subject pool. Given the limitations we still found differences between devices. The results from this study indicate that all three devices have acceptable reliability for the measurement of V_E based on the various statistical tests (e.g., $CV < 10$ for all devices). The DB and TrueOne 2400 had acceptable reliability for the measurement of VO_2 (ratio limits of agreement, $DB = 0.98 \times / \div 1.24$, TrueOne 2400 = $0.98 \times / \div 1.19$; CV , $DB = 5.3\%$, TrueOne 2400 = 4.7%) and VCO_2 (ratio limits of agreement, $DB = 0.99 \times / \div 1.29$, TrueOne 2400 = $0.99 \times / \div 1.26$; CV , $DB = 6.0\%$, TrueOne 2400 = 5.7%), however, the VO2000 showed poor reliability for the measurement of VO_2 (ratio limits of agreement, $0.94 \times / \div 3.30$; CV , 14.2%) and VCO_2 (ratio limits of agreement, $0.90 \times / \div 5.57$; CV , 15.8%).

There are only a few studies available that have examined the reliability of metabolic systems. Table 3 provides several reliability studies performed on portable and stationary metabolic systems. There are only a few studies which have examined the reliability of other metabolic systems. In addition, most studies only

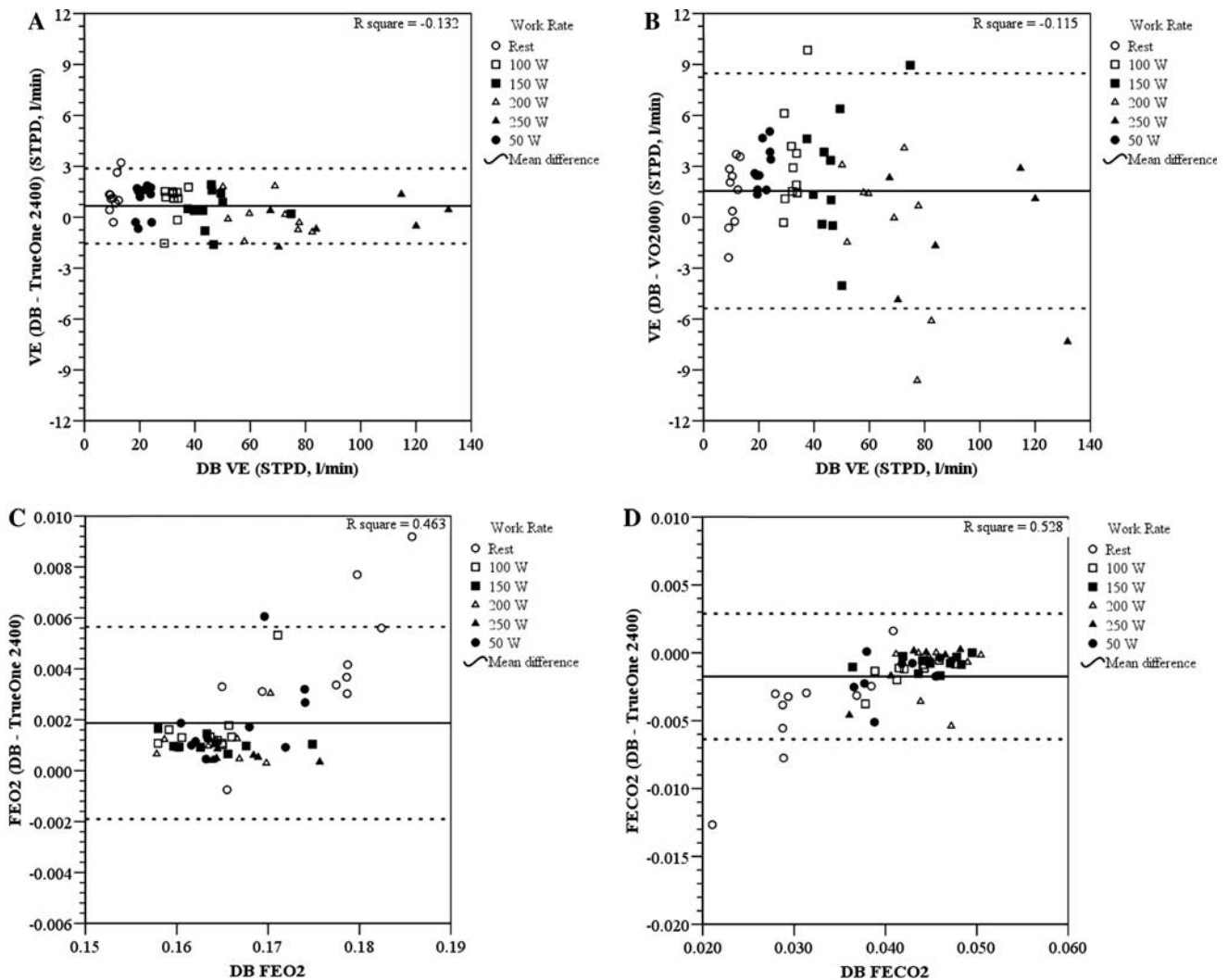


Fig. 3 Bland–Altman plots of the error scores (DB minus computerized system) for minute ventilation (V_E), fraction of expired oxygen ($F_{E}O_2$), and fraction of expired carbon dioxide ($F_{E}CO_2$). **a** TrueOne 2400 V_E , **b** VO2000 V_E , **c** TrueOne 2400 $F_{E}O_2$, and **d**

TrueOne 2400 $F_{E}CO_2$. Solid line represents the mean difference; dashed line represents the 95% prediction interval. V_E is reported in $l\ min^{-1}$ and STPD and $F_{E}O_2$ and $F_{E}CO_2$ are the mean O_2 and CO_2 fractions during the expiratory phase

examined one measure of reliability (i.e., intraclass correlation, coefficient of variation, or mean bias and 95% prediction interval), which makes it difficult to make comparisons between all the studies. Carter and Jeukendrup (Carter and Jeukendrup 2002) examined the DB method, Oxycon Alpha metabolic system, Pulmolab, metabolic system, and Oxycon Pro metabolic system for reliability during cycling at 100 and 150 W. Of importance in this study is that they also examine the reliability of the DB method, which are similar to the results found in the current study. However, they only examined two work rates, while we had five work rates and rest. The reliability of the TrueOne 2400, for measuring V_E , VO_2 , and VCO_2 , is similar to other systems currently available, which have been shown to have good reliability (Carter and Jeukendrup 2002;

Meyer et al. 2001). While the reliability of the VO2000 for measuring V_E is similar to other metabolic systems (Carter and Jeukendrup 2002; Larsson et al. 2004; Lucia et al. 1993; Meyer et al. 2001), the reliability for measuring VO_2 and VCO_2 is less than previously examined systems.

Accuracy

Table 4 provides results from previous validation studies examining portable and stationary systems which used the DB method as the criterion method, which will make for an easier comparison between studies. The mean bias and 95% prediction intervals for the TrueOne 2400 in the current study are similar to those

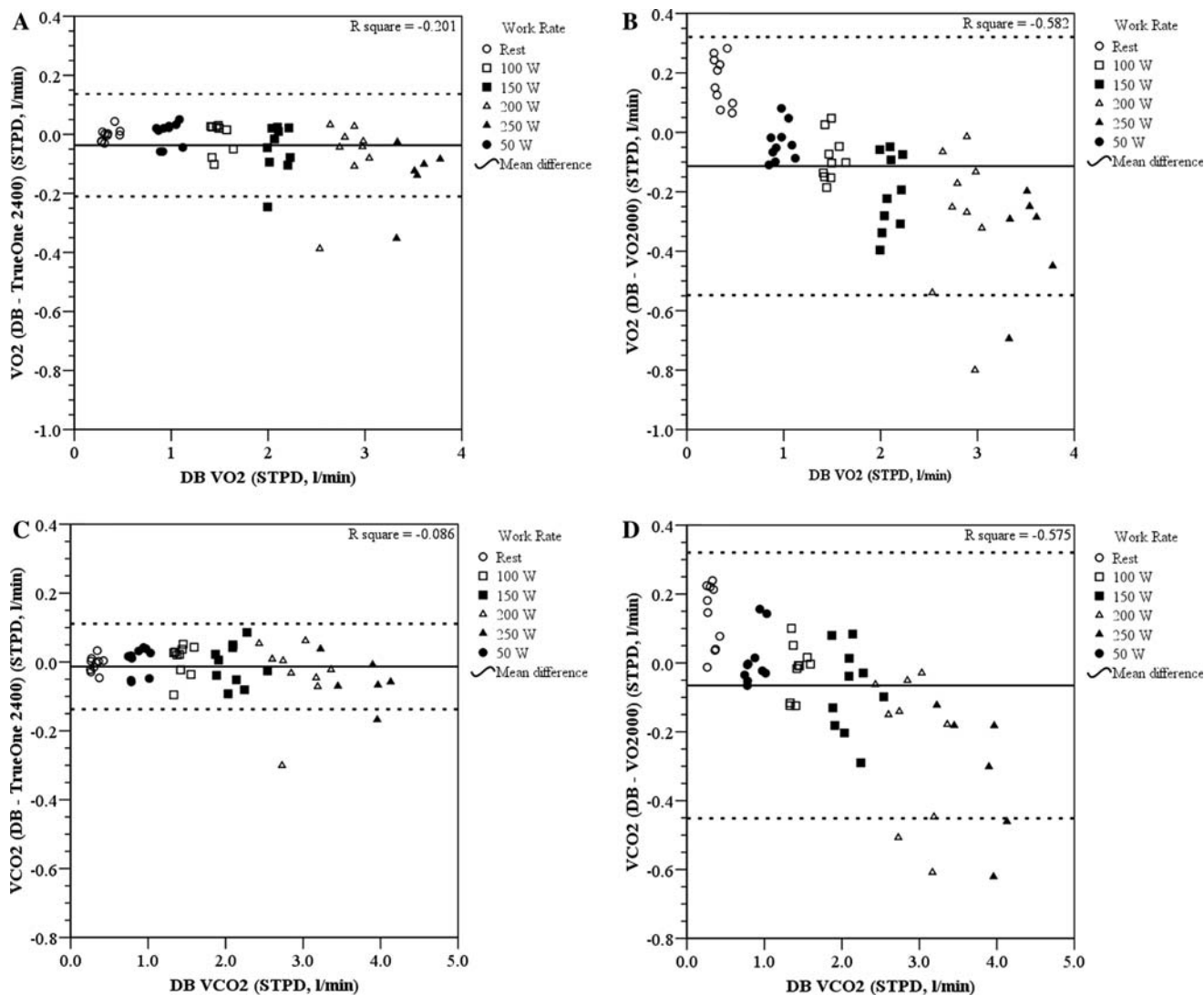


Fig. 4 Bland–Altman plots of the error scores (DB minus computerized system) for oxygen consumption (VO_2) and carbon dioxide production (VCO_2). **a** TrueOne 2400 VO_2 , **b** VO2000 VO_2 , **c** TrueOne 2400 VCO_2 , and **d** VO2000 VCO_2 . *Solid line*

represents the mean difference; *dashed line* represents the 95% prediction interval. VO_2 and VCO_2 are reported in $l\ min^{-1}$ and STPD

reported previously by Bassett et al. (2001). These results should be interpreted with caution, since the DB gas fractions were measured with a TrueMax 2400, which uses the same analyzers as the TrueOne 2400 which was being validated. However, our results are in agreement with those of Bassett et al. (2001), which used a Beckman LB2 CO_2 analyzer and an applied electrochemistry S3A O_2 analyzer to measure the gas fractions obtained from the DB measurements.

The VO2000 underestimated the DB values for V_E , VO_2 , and VCO_2 by 12, 53, and 46%, respectively, during rest. At work rates between 50 and 250 W the VO2000, on average, underestimated the DB values for V_E by 5% and overestimated VO_2 and VCO_2 by 8 and 4%, respectively. The mean bias and 95% limits of agree-

ment between the DB and VO2000 are 1.54 ± 6.93 , -0.11 ± 0.43 , and $-0.06 \pm 0.38\ l\ min^{-1}$ for V_E , VO_2 , and VCO_2 , respectively. Numerous studies have examined the accuracy of various portable metabolic systems. For example, McLaughlin et al. (2001) compared the DB and Cosmed K4b² portable system using non-simultaneous measurements during cycle ergometry (50–100 W) and found the mean bias and 95% limits of agreement to be $2.0 \pm 7.0\ l\ min^{-1}$ (V_E), $0.1 \pm 0.23\ l\ min^{-1}$ (VO_2) and $-0.1 \pm 0.35\ l\ min^{-1}$ (VCO_2). Given the available information the VO2000 has similar accuracy to other portable metabolic systems for the measurement of V_E , however, it is slightly less accurate for measuring VO_2 and VCO_2 when compared to the DB. Given that the direction and mag-

Table 3 Previous reliability studies of portable and stationary metabolic systems

Author	N	Conditions	System	ICC	CV (%)	Mean bias (95% PI) (l min ⁻¹)
Portable systems						
Lothian et al. (1993)	1	Incremental TM	Cosmed K2		VO ₂ , 3.0–14.0	
Lucia et al. (1993)	20	Sub-maximal and maximal TM tests	Cosmed K2			VO ₂ , 0.01 (NR) V _E , 1.05 (NR)
Meyer et al. (2001)	23	Incremental cycle test	MetaMax I	VO ₂ , 0.98 VCO ₂ , 0.99 V _E , 0.97		VO ₂ , 0.05 (–0.25, 0.35) VCO ₂ , 0.05 (–0.25, 0.35) V _E , 1.0 (–8.0, 10.0)
Larsson et al. (2004)	19	Cycle (100 (<i>n</i> = 9) and 200 W (<i>n</i> = 10))	MetaMax II	VO ₂ , 0.56–0.90 VCO ₂ , 0.66–0.88 V _E , 0.43–0.84		
Stationary online systems						
Lothian et al. (1993)	1	Incremental TM	Quinton		VO ₂ , 1.1–3.9	
Meyer et al. (2001)	23	Incremental cycle test	MetaLyzzer 3B	VO ₂ , 0.97 VCO ₂ , 0.96 V _E , 0.95		VO ₂ , –0.01 (–0.41, 0.39) VCO ₂ , –0.01 (–0.41, 0.39) V _E , 0.01 (–13.0, 13.0)
Carter and Jeukendrup (2002)	10	Cycle (100 and 150 W)	DB		VO ₂ , 3.3–5.1 VCO ₂ , 3.9–5.0 V _E , 5.1–5.7	
Carter and Jeukendrup (2002)	10	Cycle (100 and 150 W)	Oxycon Alpha		VO ₂ , 4.5–6.3 VCO ₂ , 4.8–5.3 V _E , 6.1–7.3	
Carter and Jeukendrup (2002)	10	Cycle (100 and 150 W)	Pulmolab EX670		VO ₂ , 26.8–33.6 VCO ₂ , 33.1–45.8 V _E , 7.6–12.5	
Carter and Jeukendrup (2002)	10	Cycle (100 and 150 W)	Oxycon Pro		VO ₂ , 4.7–6.5 VCO ₂ , 5.3–7.1 V _E , 6.6–7.4	

N number of subjects, *ICC* intraclass correlation, *CV* coefficient of variation, *95% PI* 95% prediction interval, *W* watt, *TM* treadmill

nitude of the V_E error generally translates into a similar error in VO_2 (i.e., each 1% underestimation in V_E translates into a 1% underestimation in VO_2) (Withers et al. 2000), the results from this indicate that V_E can only partially explain the errors seen in the VO_2 measurements. However, the current software available with the VO2000 does not have the option of reporting gas fractions, thus determining where the remainder error in VO_2 is coming from is not possible at this time.

The resting measurement taken in the current study using the VO2000 should be interpreted with caution. The differences seen at rest partly influence the high variability seen in the day-to-day metabolic measurements. For resting measurements MedGraphics provides a separate low flow pneumotach that is smaller in size (i.e., has a smaller dead space), however, we used the exercise pneumotach for all testing. We assume that during resting measurements the low flow pneumotach is recommended to increase the pressure inside the breathing valve so the correct measurements can be made. At rest the V_E measurements were similar to the DB values, however, in several instances there were VO_2 values of < 0.1 l min⁻¹ indicating a problem exists with the detection of the expired gas using the larger pneumotach at rest. While we did not test the

low flow pneumotach at rest, this raises concerns over the current setup of the device. Under many circumstances it is necessary to obtain resting measurements before starting an exercise test. In addition, the resting measurements provide a means for the researcher to check the system before an exercise test to ensure it is working properly. It is not always practical to obtain resting measurements with one pneumotach and then change pneumotachs and recalibrate before beginning the exercise test.

A limitation to the use of the VO2000 is that the current software does not provide the means to perform a manual calibration. The current software design allows only for an auto calibration, which uses current room conditions and proprietary software to perform the gas and flow calibrations. Therefore, it is possible that the device could be out of calibration and the user would not know this because there is no way to verify these values. In the current study, we relied on the auto calibration procedure because this is how the device will be used by researchers, and we did not feel it should be manipulated in a way that it could not be used in an actual research setting.

The findings from this study confirm previous results that the TrueOne 2400 is an accurate device for the

Table 4 Previous validation studies comparing portable and stationary metabolic systems to the DB method

Author	N	Conditions	System	Mean bias (95% prediction interval) (l min ⁻¹) and percent difference of DB value	
				VO ₂	VCO ₂
Portable systems					
Louhevaara et al. (1985)	6	TM walking and lifting	Oxylog	-4.1% (TM)* -6.4% (lifting)	
Lucia et al. (1993)	20	Incremental TM	Cosmed K2	-0.08 (-0.411, 0.251)	
King et al. (1999)	9	Cycle (50–250 W)	Aerosport KB1-C	0.01 (-0.49, 0.51)	1.0 (-9.0, 11.0)
McLaughlin et al. (2001)	10	Rest and cycle (50–250 W)	Cosmed K4b ²	-0.10 (-0.33, 0.15) +9.6–3.0% 50–200 W*	2.5 (-5.0, 10.0)
Prieur et al. (2003)	11	TM running	Sensormedics VmaxST	-0.017 (-0.52, 0.48)	-9.4 (-23.2, 4.4)
Brehm et al. (2004)	10	Cycle (80 W)	Sensormedics VmaxST	-0.04* (-0.14, 0.6)	0.4 (-2.6, 3.4)
Larsson et al. (2004)	29	Cycle (100 W (n = 9) 200 W (n = 10), 250 and 325 W (n = 10))	MetaMax II	0.05 (-0.18, 0.26)	1.80 (-4.0, 7.6)
Stationary online systems					
Rieijens et al. (2001) [#]	12	Incremental cycle test	Oxycon Pro	1.00 (0.88, 1.12)	0.98 (0.9, 1.0)
Bassett et al. (2001)	8	Rest and cycle (50–250 W)	ParvoMedics TrueMax 2400	0.01 (-0.08, 0.11)	0.2 (-0.8, 1.2)
Carter and Jeukendrup (2002)	10	Cycle (100 and 150 W)	Oxycon Alpha	± 4%	± 15.0%*
Carter and Jeukendrup (2002)	10	Cycle (100 and 150 W)	Pulmolab EX670	-35%*	+12.0%*
Carter and Jeukendrup (2002)	10	Cycle (100 and 150 W)	Oxycon Pro	± 3%	+13.0%*
Gore et al. (2003)	12	Cycle (100–300 W)	MedGraphics CPX/D	-10.7–11.0%*	-4.0–6.2%
Foss and Hallen (2005)	17	Incremental cycle test	Oxycon Pro	0.05 (-0.1, 0.15) 0.8% lower*	2.0 (-1.0, 4.9) 1.8% lower*

N number of subjects, DB Douglas bag, W watt, TM treadmill, *P < 0.05, [#]reported ratio limits of agreement

measurement of V_E , VO_2 , and VCO_2 , as well as extending those findings to show that it has good reliability. These findings also suggest that the VO2000 portable metabolic system may be less reliable for the measurement of VO_2 and VCO_2 than other systems. In addition, it tends to underestimate VO_2 at rest and overestimate VO_2 at most cycling work rates. Future work should be done to confirm these results and examine possible sources of error.

References

- Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 26:217–238
- Bassett DR Jr, Howley ET, Thompson DL, King GA, Strath SJ, McLaughlin JE, Parr BB (2001) Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. *J Appl Physiol* 91:218–224
- Bigard AX, Guezennec CY (1995) Evaluation of the Cosmed K2 telemetry system during exercise at moderate altitude. *Med Sci Sports Exerc* 27:1333–1338
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
- Brehm MA, Harlaar J, Groepenhof H (2004) Validation of the portable VmaxST system for oxygen-uptake measurement. *Gait Posture* 20:67–73
- Byard AD, Dengel DR (2002) Validity of a portable metabolic measurement system. *Med Sci Sports Exerc* 34:S149
- Carter J, Jeukendrup AE (2002) Validity and reliability of three commercially available breath-by-breath respiratory systems. *Eur J Appl Physiol* 86:435–441
- Foss O, Hallen J (2005) Validity and stability of a computerized metabolic system with mixing chamber. *Int J Sports Med* 26:569–575
- Gore CJ, Clark RJ, Shipp NJ, Van Der Ploeg GE, Withers RT (2003) CPX/D underestimates VO_2 in athletes compared with an automated Douglas bag system. *Med Sci Sports Exerc* 35:1341–1347
- King GA, McLaughlin JE, Howley ET, Bassett DR Jr, Ainsworth BE (1999) Validation of aerosport KB1-C portable metabolic system. *Int J Sports Med* 20:304–308
- Larsson PU, Wadell KME, Jakobsson EJI, Burlin LU, Henriksen-Larsen KB (2004) Validation of the MetaMax II portable metabolic system. *Int J Sports Med* 25:115–123
- Lothian F, Farrally MR, Mahoney C (1993) Validity and reliability of the Cosmed K2 to measure oxygen uptake. *Can J Appl Physiol* 18:197–206
- Louhevaara V, Ilmarinen J, Oja P (1985) Comparison of three field methods for measuring oxygen consumption. *Ergonomics* 28:463–470
- Lucia A, Fleck SJ, Gotshall RW, Kearney JT (1993) Validity and reliability of the Cosmed K2 instrument. *Int J Sports Med* 14:380–386
- McLaughlin JE, King GA, Howley ET, Bassett DR Jr, Ainsworth BE (2001) Validation of the COSMED K4b2 portable metabolic system. *Int J Sports Med* 22:280–284
- Meyer T, Georg T, Becker C, Kindermann W (2001) Reliability of gas exchange measurement from two different spiroergometry systems. *Int J Sports Med* 22:93–97
- Meyer T, Davison RC, Kindermann W (2005) Ambulatory gas exchange measurements—current status and future options. *Int J Sports Med* 26(Suppl 1):S19–27
- Olson TP, Tracy JE, Dengel DR (2003) Validity of a low-flow pneumotach and portable metabolic system for measurement of basal metabolic rate. *Med Sci Sports Exerc* 35:S143
- Peel C, Utsey C (1993) Oxygen consumption using the K2 telemetry system and a metabolic cart. *Med Sci Sports Exerc* 25:396–400
- Prieur F, Castells J, Denis C (2003) A methodology to assess the accuracy of a portable metabolic system (VmaxST). *Med Sci Sports Exerc* 35:879–885
- Rietjens GJ, Kuipers H, Kester AD, Keizer HA (2001) Validation of a computerized metabolic measurement system (Oxycon-Pro) during low and high intensity exercise. *Int J Sports Med* 22:291–294
- Stuart MK, Howley ET, Gladden LB, Cox RH (1981) Efficiency of trained subjects differing in maximal oxygen uptake and type of training. *J Appl Physiol* 50:444–449
- Withers RT, Gore CJ, Gass G, Hahn A (2000) Determination of maximal oxygen consumption (VO_{2max}) or maximal aerobic power. In: Gore CJ (eds) *Physiological tests for elite athletes*. Human kinetics, Champaign, pp 114–127
- Yeh MP, Gardner RM, Adams TD, Yanowitz FG (1982) Computerized determination of pneumotachometer characteristics using a calibrated syringe. *J Appl Physiol* 53:280–285
- Yeh MP, Adams TD, Gardner RM, Yanowitz FG (1987) Turbine flowmeter vs. Fleisch pneumotachometer: a comparative study for exercise testing. *J Appl Physiol* 63:1289–1295