Baltic Journal of Sport and Health Sciences
Online first
DOI: https://doi.org/10.33607/bjshs.v3i140.2145

eISSN 2538-8347

LITHUANIAN SPORTS UNIVERSITY

Baltic Journal of Sport and Health Sciences

Journal homepage: https://journals.lsu.lt/baltic-journal-of-sport-health

Validity and Test–Retest Reliability of the VO2 Master Metabolic Analyser: A Laboratory and Artistic Gymnastics-Specific Evaluation

Alexander Seemann-Sinn[1, 2*], Peter Rüdrich[3], Falk Naundorf[2], Ingo Sandau[4]

[1] *Department of Sports Medicine, Humboldt University of Berlin, Berlin, Germany*
[2] *Department Team, Combat & Acrobatic Sports, Institute for Applied Training Science, Leipzig, Germany*
[3] *Department of Sports Medicine, Institute for Applied Training Science, Leipzig, Germany*
[4] *Department Strength, Power and Technical Sports, Institute for Applied Training Science, Leipzig, Germany*
* *Correspondence: seemann-sinn@iat.uni-leipzig.de; +49 341 4945-322*

ABSTRACT

Background: This study assessed the validity and reliability of oxygen uptake ($\dot{V}O_2$) and minute ventilation ($\dot{V}E$) of the mobile metabolic analyser VO2 Master under laboratory and gymnastics-specific conditions.

Methods: For validity testing six metabolic rates were reproduced by a metabolic simulator (VacuMed) and $\dot{V}E$ and $\dot{V}O_2$ were measured by two VO2 Masters (3209 and 3212). To evaluate test–retest reliability nine gymnasts (age 16.4 ± 1.0 years; height 168.2 cm; weight 59.4 kg) performed a circle test (CT) and nine athletic adults (age 29.7 ± 4.9 years; height 177.8 ± 7.4 cm; weight 71.3 ± 6.6 kg) performed a floor-specific test (FST). Validity was assessed by concordance correlation coefficient (CCC) and mean percentage error (MPE). Test–retest reliability was analysed using standard error of measurement (%SEM) and intra-class correlation coefficient ICC (2,1) for $\dot{V}O_2$mean, $\dot{V}O_2$peak, O_2 deficit and O_2 integral.

Results: The VO2 Masters showed a MPE of 10.44 ± 1.48%, 7.56 ± 1.97%, 6.73 ± 1.93% and 7.56 ± 1.45% (3209) for $\dot{V}O_2$ and $\dot{V}E$. The CCCs of 0.97 and 0.99 for $\dot{V}O_2$ and 0.98 for $\dot{V}E$ indicate a substantial to almost perfect agreement between VacuMed and VO2 Masters. In the CT, $\dot{V}O_2$mean, $\dot{V}O_2$peak, O_2 deficit, and O_2 integral showed ICCs between 0.74 and 0.95 and SEM of ≤10% and ≤5%. In the FST, $\dot{V}O_2$mean, $\dot{V}O_2$peak, O_2 deficit, and O_2 integral showed ICCs between 0.76 and 0.94 and values of ≤10% and ≤5% for %SEM.

Conclusion: The VO2 Master is a practical, reliable and valid metabolic analyser for short term exercises like gymnastics routines.

Keywords: Metabolic analyser, oxygen uptake, quality criteria, error of measurement, short term exercises

1. INTRODUCTION

The use of mobile metabolic analyser to measure oxygen uptake ($\dot{V}O_2$) and minute ventilation ($\dot{V}E$) has become an essential tool for the analysis of physical performance and clinical diagnoses (Guidetti et al., 2018). For example, mobile metabolic analyser can be used to determine energy profiles and thus clarify key questions of a sport-specific needs analysis.

There are now a variety of mobile metabolic analysers available on the market. The most common models are the K5 (*COSMED S.r.l., Rome, Italy*), the MetaMax 3B (*Cortex Biophysik GmbH,* *Leipzig, Germany*), or the Oxycon Mobile (*Vyaire Medical, Mettawa, USA*). In recent years, a large number of accuracy studies have been carried out with these analysers, so that they can be considered as valid and reliable measurement systems for assessing $\dot{V}O_2$ and $\dot{V}E$ (Guidetti et al., 2018; Perret & Mueller, 2006; Vogler et al., 2010; Winkert et al., 2020). From a technical point of view, these analysers consist of a breathing mask with a turbine and a connected measuring unit on the back or chest. This technical design is suitable for the use in endurance sports such as running, cycling, or rowing. In acrobatic or combat sports with an increased risk of falling or fighting situations on the ground, the use of such a metabolic analyser with the measuring unit on the back or chest is limited. Furthermore, the measuring unit on the back can lead to balance irritations in acrobatic sports such as artistic gymnastics.

The VO2 Master (*VO2 Master Health Sensors Inc., Vernon, Canada*) is a novel, patented mobile $\dot{V}O_2$ analyser that is fully integrated into the breathing mask worn on the subject's face, eliminating the limitations of a back or chest measurement unit (Webber, 2019). This technical design of the VO2 Master makes it interesting for diagnostics in acrobatic and combat sports and ensures a sport-specific practicability. In addition to practicability, good validity and reliability of such new technologies is crucial for the application in research, especially under the specific conditions of the sport being examined. According to the manufacturer, typical percent error of the VO2 Master is less than 2% for all metrics and intra- and inter-reliability is excellent with intra-class correlation coefficients (ICC) greater than 0.99, and minimum detectable changes less than 3.5% for $\dot{V}O_2$ and 1.5% for $\dot{V}E$ (Webber, 2019). In addition to the manufacturer's specifications regarding the measurement accuracy of their measuring instruments, independent peer review studies are important for verifying the manufacturer's specifications. For this reason, a study by Montoye et al. (2020) examined the validity of the VO2 Master compared to the Parvo Medics' TrueOne 2400 (*Parvo Medics, Salt Lake City, USA*) metabolic analyser as well as the test–retest and inter-device reliability using long-term cycling exercises in a laboratory setting. Acceptable validity with a mean absolute percent difference of 8.2% and test–retest reliability with a mean absolute percent difference between 8.2 and 10.9% across five metabolic rates of the VO2 Master analyser were found for the assessment of $\dot{V}O_2$ in long term cycling exercises (Montoye et al., 2020). Davis (2024) also examined the validity and test–retest reliability of the VO2 Master on a bicycle ergometer, where the test subjects exercised for 10 minutes at 50%, 70%, and 85% of their critical power. The results of the study show almost identical $\dot{V}O_2$ values with 35.25 ± 10.23 $ml \cdot kg^{-1} \cdot min^{-1}$ in the first test and 37.68 ± 8.73 $ml \cdot kg^{-1} \cdot min^{-1}$ in the second test, demonstrating good reliability of the VO2 Master (Davis, 2024). Thiessen et al. (2025) also reported good validity (mean errors of 2.3–2.7% for $\dot{V}O_2$ and 1.5–2.4% for $\dot{V}E$) and reliability (between-day differences in $\dot{V}O_2$ of −32, 21, 21, and 61 $ml \cdot min^{-1}$) of two VO2 Master devices in their study. In a major study, Van Hooren et al. (2024) examined the accuracy of respiratory gas variables, substrate, and energy consumption of 15 Cardio Pulmonary Exercise Test (CPET) systems and found different results regarding the validity of the VO2 Master. The results showed that the VO2 Master had an average percentage error of −17.9 ± 13.3% in $\dot{V}O_2$ measurement at all tested intensities, significantly underestimating $\dot{V}O_2$ compared to a metabolic simulator (Van Hooren et al., 2024). However, the results should be interpreted with caution, as only one device was tested and it appears that individual devices have different degrees of validity and the transferability of the results is unclear (Thiessen et al., 2025).

Looking at the studies cited, it can be concluded that the results of the studies cannot be transferred directly to other types of exercises, such as short-term exercises in acrobatic sports like artistic gymnastics. Furthermore, there are indications that the results regarding the reliability and validity of a device cannot be transferred without restriction from one device to another (Winkert et al., 2020). Therefore, it is useful to check the validity of one's own mobile metabolic analyser, as was done in this study, in order to confirm the validity of the data.

The aim of the study was therefore first to assess the concurrent validity of two VO2 Master devices (serial number 3209 and 3212) against a metabolic simulator and second to test the test−retest reliability of $\dot{V}O_2$ parameters in gymnastics-specific short-term exercises using the two VO2 Master devices. Based on the earlier results showing mean errors between 2.3% and 17.9% in the validity of $\dot{V}O2$ measurements, we hypothesised (H1) that the VO2 Master devices reach acceptable validity (mean percentage error ≤10%; concordance correlation coefficient (CCC) < 0.95). The second hypothesis (H2) was that the VO2 Master devices show at least moderate reliability (standard error of measurement (%SEM) 5−10%; ICC 0.9−0.75) in gymnastic specific short-term exercises.

2. METHODS

2.1. Experimental approach

Validity testing was performed under laboratory conditions using a metabolic simulator (*VacuMed 17056, USA*). The VacuMed uses a motor-drive syringe able to vary tidal volume and respiratory frequency to reproduce different $\dot{V}E$, and a gas tank containing air with 79% N_2 and 21% CO_2 that is used to reproduce the $\dot{V}O_2$ and $\dot{V}CO_2$ in the exhaled air at different metabolic rates by diluting the gas tank with room air inside the piston pump (Guidetti et al., 2018). Therefore, $\dot{V}O_2$ and $\dot{V}CO_2$ are proportional to the gas flow of calibration gas from the tank. The VacuMed has been used in other studies to test the validity of other mobile metabolic analysers (Guidetti et al., 2018; Winkert et al., 2020).

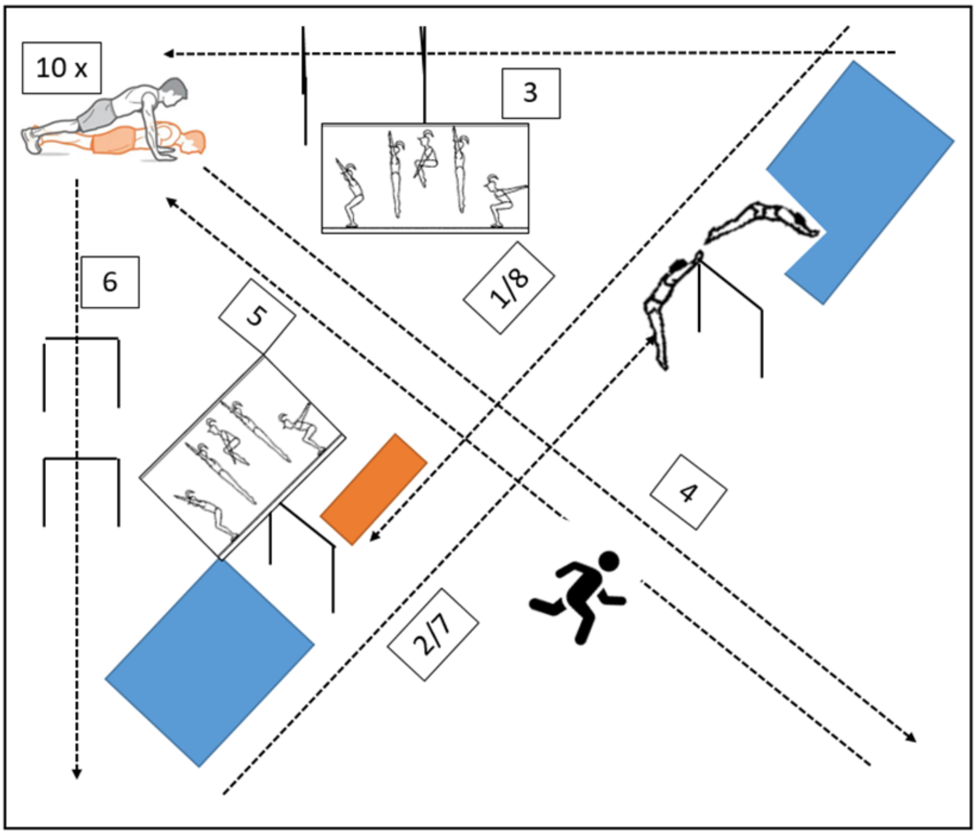
The validity of the two VO2 Master metabolic analysers for assessing $\dot{V}O_2$ and $\dot{V}E$ was tested using the test protocol shown in Table 1 (Rüdrich et al., 2022). This test protocol was chosen because previous studies showed that elite gymnasts achieve $\dot{V}O_2$ values of 4,000 ml/min and respiratory rates of 100 1/min during their exercises on the rings and pommel horse (Seemann-Sinn et al., 2023; Seemann-Sinn et al., 2024). The data for each level (metabolic rate) were measured ‘breath by breath’ over a period of 70 s. For later analysis, the data of the first 10 s of the measurement were removed to eliminate data related to the wash-out of the simulator’s gas-filled dead space, and averages were calculated over the remaining breaths within the 60 s.

Table 1. **Test protocol used for the VacuMed simulator to assess validity**

Level	Respiratory Rate [1/min]	Tidal Volume [l]	Ventilation [L]	Oxygen Uptake [ml]
1	25	1.5	37.5	1,000
2	30	2.0	60.0	2,000
3	40	2.5	100.0	3,000
4	50	2.5	125.0	4,000
5	70	1.5	105.0	4,000
6	80	1.5	120.0	4,000

The sport-specific reliability of the VO2 Master was tested using two gymnastics-specific tests in a controlled test−retest design. A circle test (CT) on the pommel horse and a floor-specific test (FST) were performed for this purpose. The test and retest on the floor and pommel horse were performed on two consecutive days (at the same time of day). In the CT, the gymnasts had to complete 50 circles in their own preferred speed mode. With an exercise duration of 50−55 s, this reflects the demands of high-performance pommel horse routines (Seemann-Sinn et al., 2021). The FST consisted of a total of eight exercise lines with two 12 metre sprints, multiple jumps over obstacles, and ten push-ups in a 12 × 12-metre area (Figure 1). This test simulated the load situation during a floor routine that includes final jumps, reactive jumps during acrobatic connections, and non-acrobatic elements (element group I of the “Code of Points”, (Fédération Internationale de Gymnastique, 2020)). The FST exercise time was 62.9 ± 2.3 s, slightly below the required routine duration of 65−75 s for high-performance floor routines (Fédération Internationale de Gymnastique, 2020).

Figure 1. **Graphical presentation of the floor-specific test**



2.2. Subjects

Nine German junior national squad gymnasts (age 16.4 ± 1.0 years; height 168.2 cm; weight 59.4 kg) performed the CTs. The FST was performed with nine (7 male, 2 female) athletic adults (age 29.7 ± 4.9 years; height 177.8 ± 7.4 cm; weight 71.3 ± 6.6 kg). All subjects were informed in the run-up to the study about the procedure, benefits, and possible risks of the study and gave their consent to participate in the study (written informed consent or cooperation agreement between the authors' research institute and the German Gymnastics Federation). The experimental protocol was approved by the Ethics Committee of the Institute for Applied Training Science (ER_2022.30.09_28) and was conducted in accordance with the Declaration of Helsinki.

2.3. Data collection and data evaluation

$\dot{V}O_2$ and heart rate (HR) were measured 'breath-by-breath' during CT and FST using VO2 Master metabolic analyser and a paired HR sensor (*H10, Polar, Finland*) during the 1-minute prestart phase, during the tests, and up to eight minutes after the tests. Before each test, the VO2 Master was calibrated using room air, exhaled air, and a defined 3-litre air volume according to the manufacturer's instructions. To determine blood lactate concentration, 20 µl of capillary blood were taken from the hyperaemic earlobe using end-to-end glass capillary before the prestart phase and 1, 3, and 5 minutes after the tests. Subsequently, the capillary blood was analysed using a blood analysis device (*SuperGl, Dr. Müller Gerätebau, Germany*) to determine the blood lactate concentration. Before and after the tests, the subjects had to indicate the degree of perceived exertion (RPE) using the 6–20 scale (Borg, 1982). The Δ lactate and Δ RPE were determined from the differences between prestart lactate and peak lactate and prestart RPE and peak RPE.

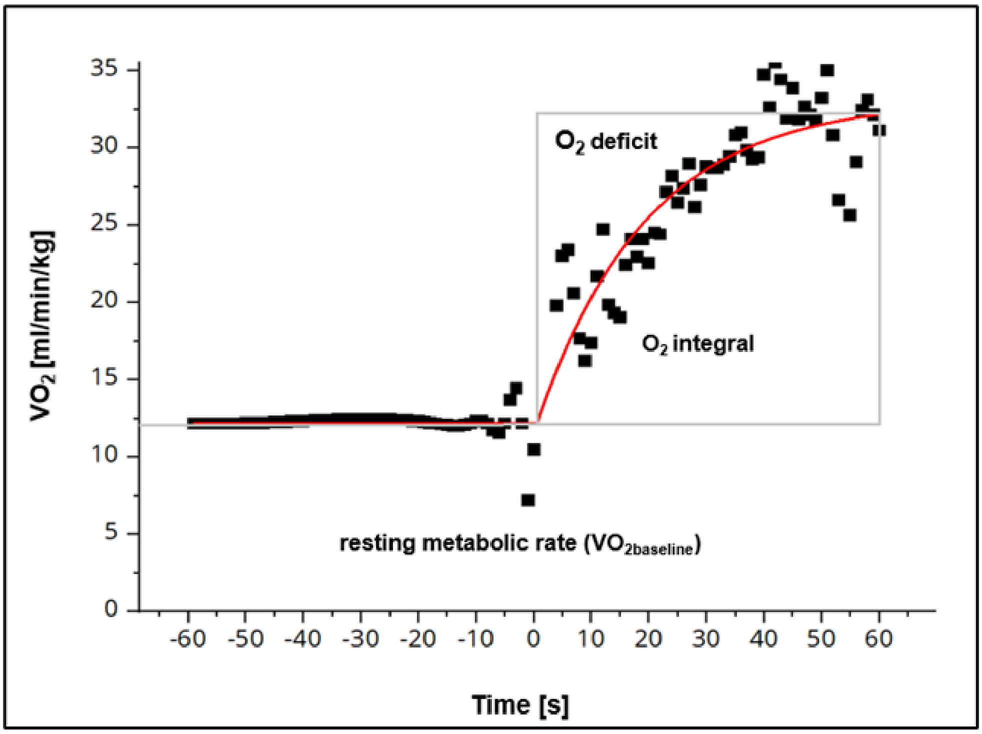
For the $\dot{V}O_2$ and HR data analysis, the raw data were interpolated to 1-second values (*Origin, 2016, OriginLab, USA*). After the first mono-exponential curve fit (equation 1) of the interpolated $\dot{V}O_2$ data, erroneous values that deviated from the predicted regression value by more than ±6 (CT) or ±10 (FST) residuals were considered as erroneous breaths and eliminated. The different levels of residuals, which were set as the limit for erroneous values, result from the fact that circles on pommel horse represent rhythmic movements and therefore the breath-by-breath $\dot{V}O_2$ values vary as less than in the FST. Further smoothing procedures were not carried out to avoid a reduction in the number of data points available for parameter estimation (Rodriguez et al., 2016). A second mono-exponential curve fit (also equation 1) was then carried out with the smoothed $\dot{V}O_2$ values, where $\dot{V}O_2$ (t) is the $\dot{V}O_2$ at time t, $\dot{V}O_2$baseline is the $\dot{V}O_2$ of the resting metabolic rate, A_1 is the amplitude of the exponential term, τ is the associated time constant, and TD is the time delay of the exponential term.

$$\dot{V}O_2(t) = \dot{V}O_{2baseline} + A_1 \times (1-e^{-(t-TD)/\tau})$$

The parameters O_2 deficit and O_2 integral were then determined using the fitted $\dot{V}O_2$ curve. O_2 integral was defined as the area below the fitted $\dot{V}O_2$ curve minus the resting metabolic rate ($\dot{V}O_2$baseline) and calculated using the trapezoidal method (Figure 2). O_2 deficit was defined as the difference between the O_2 that would have been consumed if a steady state had been reached immediately at the beginning of the exercise and the O_2 consumed during the exercise duration (ED), which corresponds to the O_2 integral (Bearden & Moffatt, 2001) (equation 2).

$$O_2\ Deficit = \left(ED \times \dot{V}O_{2peak-fit} - \int_{Start}^{End} EQ1\right) - (ED \times \dot{V}O_{2baseline})$$

Figure 2. **Graphical presentation of $\dot{V}O_2$ data of one gymnast during the circle test and the calculated parameters O_2 deficit and O_2 integral**



The parameters mean $\dot{V}O_2$ ($\dot{V}O_2$mean) and peak $\dot{V}O_2$ ($\dot{V}O_2$peak) were calculated from the smoothed $\dot{V}O_2$ data after the first mono-exponential curve fit. $\dot{V}O_2$peak was defined as the mean value of the last 5 s of the CT or FST.

2.4. Statistical analysis

The statistical analyses were conducted using statistical software (*Microsoft Excel 2016* and *Jamovi 2.3.24*). Descriptive statistics were performed and the data are presented as mean ± standard deviation (SD). To test validity, normal distribution of the differences (VacuMed vs VO2 Master 3212 and VacuMed vs VO2 Master 3209 for $\dot{V}O_2$ and $\dot{V}E$) was examined and confirmed using the Shapiro–Wilk test. Relative validity was assessed using Lin's concordance correlation coefficient (CCC) with 95% confidence intervals (CI). Strength-of-agreement criteria for CCC were categorised as poor (CCC < 0.9), moderate (CCC < 0.95), substantial (CCC < 0.99), and nearly perfect (CCC ≥ 0.99) (McBride, 2005). To assess absolute validity, a Deming regression was first performed to check for constant and proportional bias between the procedures. A significant constant bias was indicated if the range of the 95% CI of the intercept did not include the value 0, and a significant proportional bias was indicated if the range of the 95% CI of the slope did not include the value 1 (Payne, 1997). In case of significant proportional and/or constant bias, the regression equation was used to correct (i.e. recalibrate) the measurements (Ungerer & Pretorius, 2017). After correction, a Bland–Altman analysis was performed for the recalibrated measurements to calculate the 95% limits of agreement (LoA) and the standard deviation of measurement differences (SDD; random measurement error). Furthermore, the mean absolute and percentage error were calculated.

To assess the test–retest reliability, mean differences between test and retest were tested using paired t-tests, and the effect size *d* was assessed based on Cohen (1992). In all analyses, a significance level of 5% ($p < 0.05$) was assumed. In addition, the absolute and relative error, the standard error of measurement (SEM), and the ICC (ICC(2,1); tested for absolute agreement) were determined. The SEM was classified as good at a value of 0–5%, moderate at 5–10%, and poor >10% based on the recommendations of Gómez-Carmona et al. (2021). ICCs were rated as excellent at a value of >0.9, good from 0.9 to 0.75, moderate from 0.75 to 0.5, and poor <0.5 (Koo & Li, 2016). Furthermore, Bland–Altman analysis was carried out. Normality of the data was tested using the Shapiro–Wilk test.

3. RESULTS

3.1. Validity

Table 2 shows the recorded $\dot{V}O_2$ and $\dot{V}E$ values of the VacuMed and the two VO2 Masters (serial number 3212 and 3209). Both VO2 Masters overestimate $\dot{V}O_2$ and $\dot{V}E$. With increasing $\dot{V}O_2$, the percentage overestimation of $\dot{V}O_2$ by the two VO2 Masters reduces (Figure 3).

Table 2. **$\dot{V}O_2$ and $\dot{V}E$ values of VacuMed and VO_2 Masters 3212 and 3209**

Level	VacuMed		VO_2 Master 3212		VO_2 Master 3209	
	$\dot{V}O_2$ [ml/min]	$\dot{V}E$ [l/min]	$\dot{V}O_2$ [ml/min]	$\dot{V}E$ [l/min]	$\dot{V}O_2$ [ml/min]	$\dot{V}E$ [l/min]
1	1,006	37.5	1,136	41.7	1,098	40.8
2	2,015	60.0	2,226	64.7	2,181	65.4
3	3,017	100.0	3,326	105.7	3,235	107.0
4	4,030	125.0	4,464	133.7	4,303	135.8
5	4,030	105.0	4,366	127.2	4,223	127.0
6	4,030	120.0	4,429	113.0	4,199	111.4

Note: ($\dot{V}O_2$) oxygen uptake; ($\dot{V}E$) ventilation.

The mean absolute error (MAE) for $\dot{V}O_2$ is 303.17 ± 114.80 [ml/min] for VO2 Master 3212, which corresponds to a mean percentage error (MPE) of 10.44 ± 1.48%. For $\dot{V}E$, the MAE is 6.43 ± 1.84 l/min, which corresponds to a MPE of 7.56 ± 1.97%. The VO2 Master 3209 has a MAE of 185.17 ± 60.27 ml/min and a MPE of 6.73 ± 1.93% for $\dot{V}O_2$ and a MAE of 6.64 ± 2.47 l/min and a MPE of 7.56 ± 1.45% for $\dot{V}E$.

Figure 3. **Percentage of difference of the two VO_2 Masters compared to the VacuMed during $\dot{V}O_2$ and $\dot{V}E$ measurement**

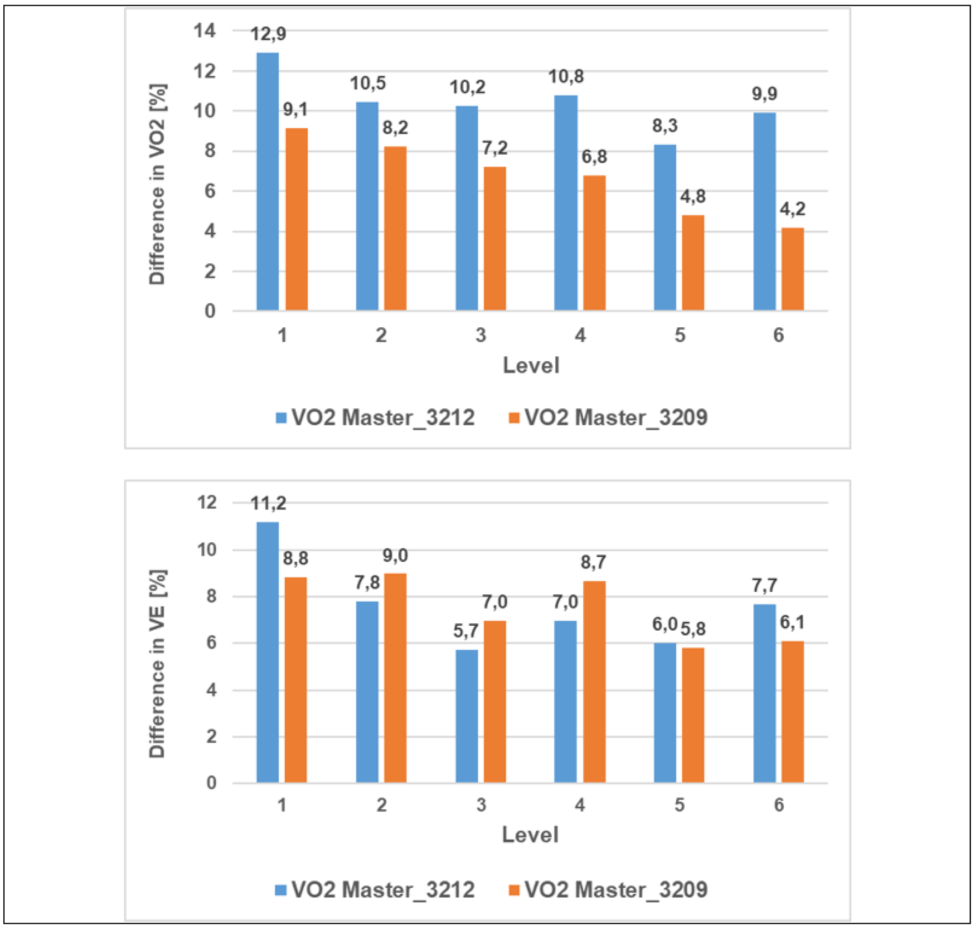


Table 3 shows the statistical results of the validity test of the VO2 Masters 3212 and 3209 in comparison to the VacuMed.

Table 3. **Measures of concurrent validity of VO_2 Masters 3212 and 3209 compared to VacuMed**

Parameter	Intercept	Slope	SDD	lLoA	uLoA	CCC
	(95% CI)					
VO_2 Master 3212						
$\dot{V}O_2$	– 37.45	0.92*	0.24#	–56.79#	57.27#	0.97
	(–88.23; 13.33)	(0.89; 0.95)	(–30.29; 30.78)	(–112.30; –1.28)	(1.76; 112.79)	(0.87; 0.99)
$\dot{V}E$	–1.99	0.96*	–0.04#	–1.53#	1.44#	0.98
	(–4.47; 0.48)	(0.93; 0.98)	(–0.84; 0.75)	(–2.97; –0.08)	(0.00; 2.89)	(0.92; 1.00)
VO_2 Master 3209						
$\dot{V}O2$	–74.28	0.97	1.34	–74.45#	77.12#	0.99
	(–233.83; 85.27)	(0.92; 1.02)	(–39.24; 41.92)	(–148.21; –0.68)	(3.35; 150.89)	(0.95; 1.00)
$\dot{V}E$	–1.11	0.94*	0.05	–2.33#	2.42#	0.98
	(–4.02; 1.80)	(0.91; 0.98)	(–1.23; 1.32)	(–4.64; –0.02)	(0.11; 4.74)	(0.91; 0.99)

Note: ($\dot{V}O_2$) oxygen uptake; ($\dot{V}E$) ventilation; (SDD) standard deviation of differences; (lLoA) lower limits of agreement; (uLoA) upper limits of agreement; (CCC) concordance correlation coefficient; (CI) confidence interval; * slope significantly different from 1; # calibrated value.

The Deming regression shows a significant proportional bias between the VacuMed and the VO2 Master 3212 for $\dot{V}O_2$ and $\dot{V}E$ measurement (Table 3). Due to the significant proportional bias, the VO2 Master 3212 data for $\dot{V}O_2$ and $\dot{V}E$ were recalibrated in relation to the VacuMed using the coefficients of the Deming regression via equations three and four (VO2 Master 3212_{cali}). Bland–Altman analysis was then carried out using the VO2 Master 3212_{cali} data.

$\dot{V}O_2$

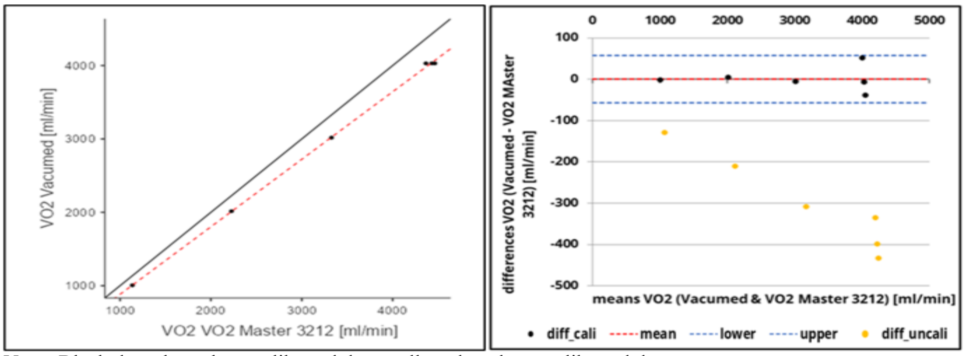
$$VO2\ \ Master\ \ 3212_{cali} = -37.45 + VO2\ \ Master\ \ 3212 \times 0.92$$

$\dot{V}E$

$$VO2\ \ Master\ \ 3212_{cali} = -1.99 + VO2\ \ Master\ \ 3212 \times 0.96$$

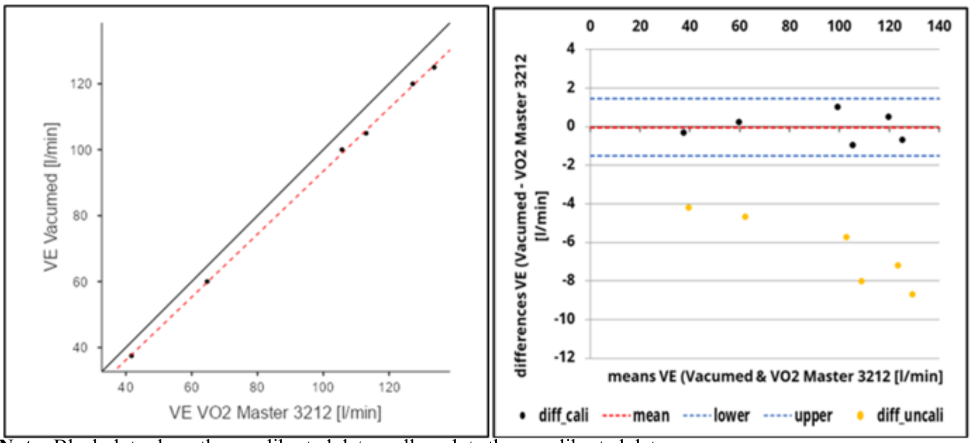
The Bland–Altman analysis shows that for the comparison between VO2 Master 3212 and VacuMed, the random error for $\dot{V}O_2$ with the recalibrated data is 0.24 ml/min with a LoA between –56.79 ml/min and 57.27 ml/min. For $\dot{V}E$, the random error is –0.04 l/min with a LoA between –1.53 l/min and 1.44 l/min (Figures 4 and 5 [right]).

Figure 4. **Graphical presentation of the Deming regression [left] and the Bland–Altman analysis [right] for the comparison of $\dot{V}O_2$ measured with the VacuMed and the VO_2 Master 3212**



Note: Black dots show the recalibrated data, yellow dots the uncalibrated data.

Figure 5. **Graphical presentation of the Deming regression [left] and the Bland–Altman analysis [right] for the comparison of $\dot{V}E$ measured with the VacuMed and the VO_2 Master 3212**



Note: Black dots show the recalibrated data, yellow dots the uncalibrated data.

For the VO2 Master 3209, the Deming regression shows no significant proportional bias for $\dot{V}O_2$, but there is a significant one for $\dot{V}E$. Due to the small number of measurement points ($N = 6$) and the visual inspection of the Line of Identity (Figure 6 [left]), the data from the VO2 Master 3209 were also recalibrated for $\dot{V}O_2$ and $\dot{V}E$ in relation to the VacuMed using the coefficients of the Deming regression via equations 5 and 6 (VO2 Master 3209_{cali}).

$\dot{V}O_2$

$$VO2\ \ Master\ \ 3209_{cali} = -74.28 + VO2\ \ Master\ \ 3209 \times 0.97$$

$\dot{V}E$

$$VO2\ \ Master\ \ 3209_{cali} = -1.11 + VO2\ \ Master\ \ 3209 \times 0.94$$

The Bland–Altman analysis shows that for the comparison between VO2 Master 3209 and VacuMed, the random error for $\dot{V}O_2$ with the recalibrated data is 1.34 ml/min with a LoA between –74.45 ml/min and 77.12 ml/min. For $\dot{V}E$, the random error is –2.33 l/min with a LoA between –2.33 l/min and 2.42 l/min (Figures 6 and 7 [right]).

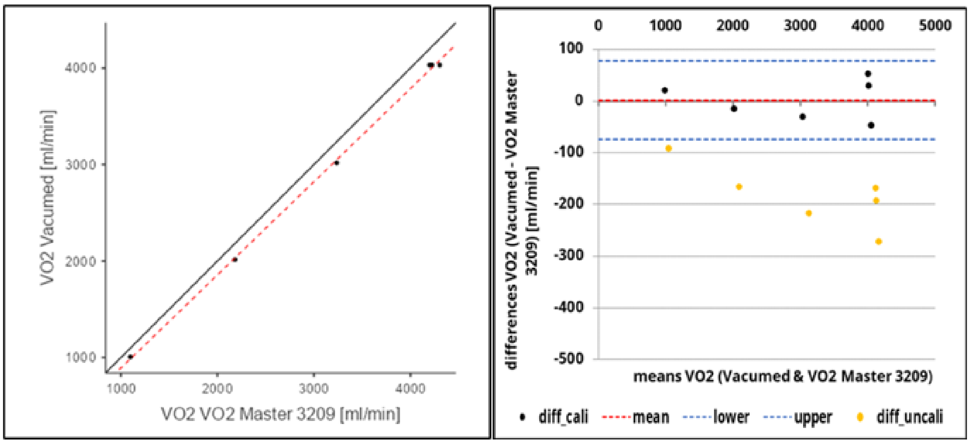


Figure 6. **Graphical presentation of the Deming regression [left] and the Bland–Altman analysis [right] for the comparison of $\dot{V}O_2$ measured with the VacuMed and the VO_2 Master 3209**

Note: Black dots show the recalibrated data, yellow dots the uncalibrated data.

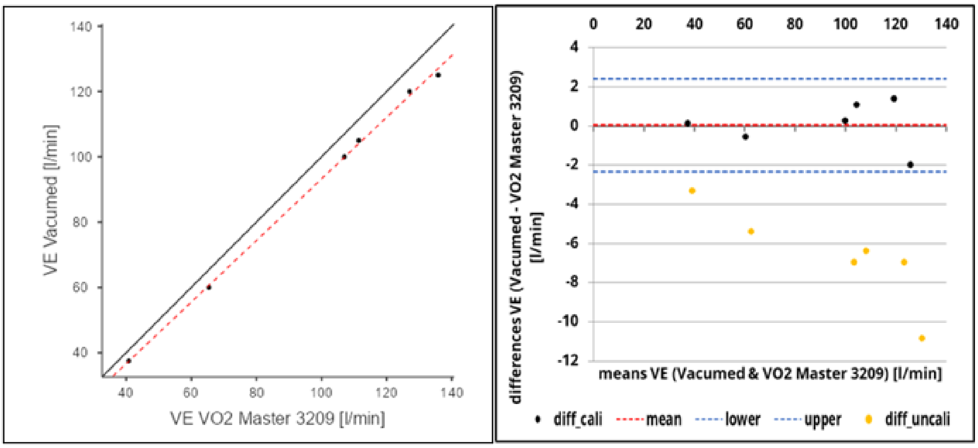


Figure 7. **Graphical presentation of the Deming regression [left] and the Bland–Altman analysis [right] for the comparison of $\dot{V}E$ measured with the VacuMed and the VO_2 Master 3209**

Note: Black dots show the recalibrated data, yellow dots the uncalibrated data.

3.2. Reliability

3.2.1. Circle Test

In the first test, the gymnasts achieved a Δ lactate of 5.25 ± 1.31 mmol, a Δ RPE of 8.00 ± 1.80, and a peak HR (HRpeak) of 170.67 ± 10.70 bpm. In the retest, Δ lactate was 5.05 ± 1.08 mmol/l, Δ RPE was 8.00 ± 1.94, and HRpeak was 170.44 ± 8.31 bpm. All stress parameters showed no significant difference between test and retest in the paired t-tests (Table 4). The effect sizes are in the range of trivial to moderate effects. Thus, the physiological stress levels between test and retest are comparable.

The parameters of the $\dot{V}O_2$ measurement ($\dot{V}O_2$mean: $t = –0.46$; $p = 0.66$; $d = 0.15$, $\dot{V}O_2$peak: $t = –0.07$; $p = 0.95$; $d = 0.02$, O_2 deficit: $t = 2.07$; $p = 0.07$; $d = 0.69$, O_2 integral: $t = –0.11$; $p = 0.92$; $d = 0.04$) also show no significant difference between test and retest in the paired t-tests. The effect sizes for $\dot{V}O_2$mean, $\dot{V}O_2$peak, and O_2 integral are in the range of trivial effects, while the parameter O_2 deficit shows a large effect.

Table 4. **Data of stress parameters and $\dot{V}O_2$ values in the test and retest in the circle test**

Parameter	Test Mean ± SD	Retest Mean ± SD	t-test (p)	Cohen's d
Δ lactate [mmol/l]	5.25 ± 1.31	5.05 ± 1.08	0.23	0.43
Δ RPE	8.00 ± 1.80	8.00 ± 1.94	1.00	0.00
HRpeak [bpm]	170.67 ± 10.70	170.44 ± 8.31	0.91	0.04
$\dot{V}$O2mean [ml/min/kg]	26.96 ± 2.81	27.19 ± 3.20	0.66	0.153
$\dot{V}$O2peak [ml/min/kg]	34.36 ± 3.10	34.41 ± 3.18	0.95	0.02
O2 deficit [ml/kg]	6.22 ± 0.86	5.89 ± 0.58	0.07	0.69
O2 integral [ml/kg]	14.17 ± 2.18	14.22 ± 2.21	0.92	0.04

Note: (RPE) Rating of perceived exertion; (HR) Heart rate; ($\dot{V}$O2) Oxygen uptake.

The $\dot{V}O_2$ parameters in the CT show ICCs ranging from moderate to excellent. The percentage SEMs range from 3.84 to 7.71%, which indicates good to moderate reliability (Table 5).

Table 5. **Statistics of the reliability test of the stress parameters and the $\dot{V}O_2$ values using VO_2 Master 3212 in the circle test (Mean ± SD)**

Parameter	Absolute Error	%Error	SEM	%SEM	ICC (95% CI)
Δ lactate [mmol/l]	0.31 ± 0.36	6.41 ± 7.01	0.31	6.07	0.93 0.73–0.98
Δ RPE	0.67 ± 0.71	5.36 ± 6.51	0.71	8.84	0.94 0.74–0.99
HRpeak [bpm]	4.89 ± 2.20	2.84 ± 1.25	3.98	2.33	0.84 0.44–0.96
$\dot{V}O_2$mean [ml/min/kg]					0.91 0.63–0.98
$\dot{V}O_2$peak [ml/min/kg]	1.04 ± 1.00	3.97 ± 3.79	1.04	3.84	0.75 0.18–0.94
O_2 deficit [ml/kg]	1.88 ± 1.23	5.56 ± 3.64	1.66	4.81	0.74 0.17–0.94
O_2 integral [ml/kg]	0.40 ± 0.40	6.21 ± 6.06	0.33	5.48	0.80 0.32–0.95

Note: (RPE) Rating of perceived exertion; (HR) Heart rate; ($\dot{V}O_2$) Oxygen uptake; (SEM) Standard error of measurement; (ICC) Intra-class correlation coefficient.

The Bland–Altman analysis show no significant bias in CT for the four parameters examined (Figure 8). The random error for $\dot{V}O_2$mean is –0.23 ml/min/kg with a LoA between –3.11 and 2.66 ml/min/kg, for $\dot{V}O_2$peak it is –0.05 ml/min/kg with a LoA between –4.64 and 4.54, for the O_2 deficit 0.33 ml/kg with a LoA between –0.60 and 1.24 ml/kg, and for the O_2 integral 0.05 ml/kg with a LoA between –2.89 and 2.79.

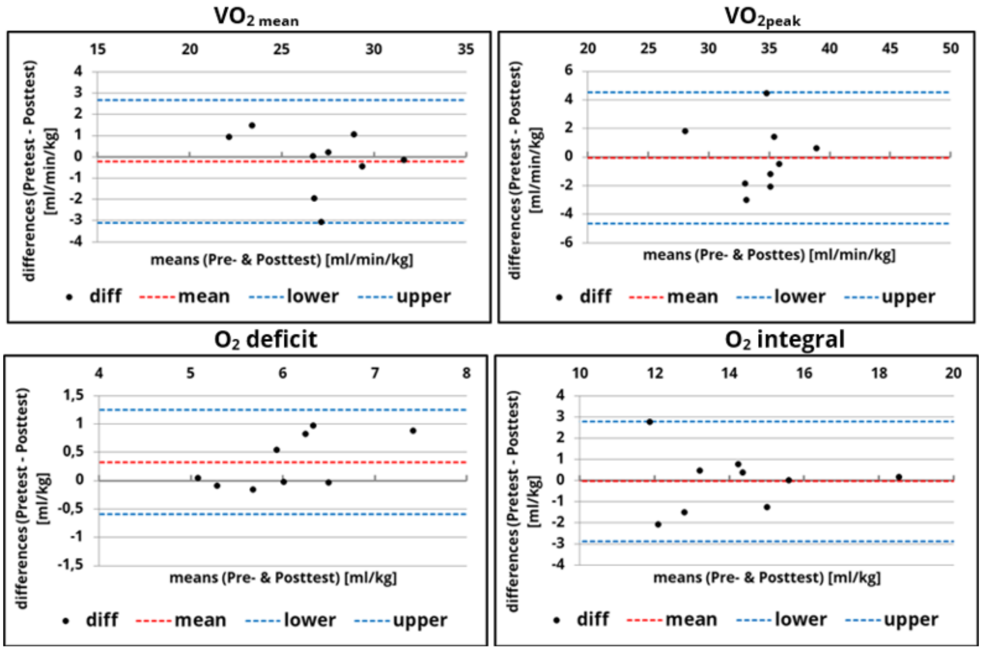


Figure 8. **Graphical presentation of Bland–Altman analysis of the parameters ($\dot{V}O_2$ mean, $\dot{V}O_2$ peak, O_2 deficits, and O_2 integral) in the circle test**

3.2.2. Floor-Specific Test

In the first test of the FST, the subjects achieved a Δ lactate of 6.46 ± 0.91 mmol, a Δ RPE of 8.43 ± 1.27, and a HRpeak of 162.57 ± 10.78 bpm. In the retest, Δ lactate was 5.54 ± 0.98 mmol/l, Δ RPE was 8.29 ± 1.11, and HRpeak was 162.43 ± 13.56 bpm. As in the CT, all stress parameters show no significant difference between test and retest in the paired t-tests (Table 6). The effect sizes are in the trivial range. This means that the physiological stress levels between test and retest are comparable.

The parameters of the $\dot{V}O_2$ measurement ($\dot{V}O_2$mean: $t = –0.09$; $p = 0.93$; $d = 0.03$, $\dot{V}O_2$peak: $t = –0.50$; $p = 0.64$; $d = 0.18$, O_2 deficit: $t = –2.30$; $p = 0.05$; $d = 0.77$, O_2 integral: $t = 0.13$; $p = 0.90$; $d = 0.04$) also show no significant difference between test and retest in the paired t-tests. As with the CT, the effect sizes for $\dot{V}O_2$mean, $\dot{V}O_2$peak, and O_2 integral are trivial, while the parameter O_2 deficit shows a large effect.

Table 6. **Data on the stress parameters and $\dot{V}O_2$ values in the test and retest in the floor-specific test**

Parameter	Test Mean ± SD	Retest Mean ± SD	t-test (p)	Cohen's d
Δ lactate [mmol/l]	5.55 ± 0.82	5.54 ± 0.98	0.98	0.01
Δ RPE	8.43 ± 1.27	8.29 ± 1.11	0.69	0.16
HRpeak [bpm]	162.57 ± 10.78	162.43 ± 13.56	0.93	0.04
$\dot{V}O_2$mean [ml/min/kg]	36.21 ± 4.75	36.11 ± 5.37	0.93	0.03
$\dot{V}O_2$peak [ml/min/kg]	52.53 ± 8.53	51.87 ± 6.99	0.64	0.18
O_2 deficit [ml/kg]	17.56 ± 3.36	18.31 ± 3.33	0.05	0.77
O_2 integral [ml/kg]	28.98 ± 4.81	28.90 ± 4.68	0.90	0.04

Note: (RPE) Rating of perceived exertion; (HR) Heart rate; ($\dot{V}O_2$) Oxygen uptake.

In FST, the $\dot{V}O_2$ parameters show ICCs in the range of good to excellent (Table 7). The percentage SEMs are in the range of 3.89–7.23%, which is classified as good to moderate.

Table 7. **Statistics of the reliability test of the stress parameters and the $\dot{V}O_2$ values using VO_2 Master 3212 in the floor-specific test (Mean ± SD)**

Parameter	Absolute Error	%Error	SEM	%SEM	ICC (95% CI)
Δ lactate [mmol/l]	0.41 ± 0.37	7.88 ± 8.31	0.41	7.31	0.82 0.36–0.96
Δ RPE	0.71 ± 0.49	8.79 ± 6.24	0.64	7.61	0.74 0.02–0.95
HRpeak [bpm]	2.71 ± 2.50	1.73 ± 1.73	2.72	1.67	0.96 0.77–0.99
$\dot{V}O_2$mean [ml/min/kg]	3.24 ± 1.37	9.01 ± 3.71	2.61	7.23	0.76 0.21–0.94
$\dot{V}O_2$peak [ml/min/kg]	3.26 ± 1.81	6.08 ± 3.01	2.76	5.18	0.89 0.59–0.98
O_2 deficit [ml/kg]	1.00 ± 0.70	5.53 ± 4.05	0.70	3.89	0.94 0.66–0.99
O_2 integral [ml/kg]	1.44 ± 0.82	5.20 ± 2.98	1.22	4.23	0.94 0.76–0.99

Note: (RPE) Rating of perceived exertion; (HR) Heart rate; ($\dot{V}O_2$) Oxygen uptake; (SEM) Standard error of measurement; (ICC) Intra-class correlation coefficient.

4. DISCUSSION

The VO2 Master is a suitable metabolic analyser for use in acrobatic sports such as artistic gymnastics. However, the validity and, especially, the reliability of this measurement system in gymnastic-specific use must be examined in more detail before use. Therefore, the aim of the study was to assess validity and gymnastic-specific reliability of the VO2 Master. Concurring with the first hypotheses, we found acceptable validity of our VO2 Master analyser under laboratory conditions. The MPE between the VacuMed and the VO2 Masters for $\dot{V}O_2$ is 10.44 ± 1.48% (model 3212) and 7.56 ± 1.97% (model 3209). Both VO2 Masters overestimate the $\dot{V}O_2$ at each level, with the percentage error decreasing with higher $\dot{V}O_2$. The CCCs of 0.97 and 0.99 for $\dot{V}O_2$ indicate a substantial to almost perfect agreement between VacuMed and VO2 Masters. For the $\dot{V}E$, the MPE is 7.56 ± 1.97% (model 3212) and 7.56 ± 1.45% (model 3209). The CCCs of 0.98 for both models indicate a substantial agreement between VacuMed and VO2 Masters for the $\dot{V}E$. Thus, our results confirm the results of earlier studies, which also showed that the VO2 Master has an acceptable validity (Montoye et al., 2020). The MPEs found in our study are, however, slightly higher than the MPE (8.2 ± 6.0%) determined in the study by Montoye et al. (2020). The differences could be due to the use of a different reference system since Montoye et al. (2020) used the metabolic analyser Parvo Medics' TrueOne 2400 and not the metabolic simulator VacuMed. Further studies examining the validity of the VO2 Master compared to a metabolic simulator show contradictory results. While Thiessen et al. (2025) demonstrated good validity of the VO2 Master, almost similar to the manufacturer's specifications, Van Hooren et al. (2024) reported poor validity of the VO2 Master with an average percentage error of −17.9 ± 13.3% in $\dot{V}O_2$ measurement. In contrast, the $\dot{V}E$ measurement shows, in line with Montoye et al. (2020) and the present study, acceptable validity with a mean relative percentage error of −5.95 ± 7.03% (Van Hooren et al., 2024). Thiessen et al. (2025) suspect that the large mean relative error of the VO2 Master reported by Van Hooren et al. (2024) is due to a methodological error in the conversion of ATPS to STPD, which has not yet been confirmed. However, based on the results of other studies, it can be assumed that the VO2 Master has acceptable validity.

Comparisons with other brands of metabolic analysers (K5 or MetaMax 3B) that also use a metabolic simulator as a reference system are informative and allow indirect comparisons of device errors between brands. The K5 with MPEs of −0.04% (−2.55 min to 3.52 max) for $\dot{V}O_2$ and 0.50% (−2.79 min to 0.66 max) for $\dot{V}E$ and the MetaMax 3B with MPEs of 4.2% (−3.0 min to 7.8 max) for $\dot{V}O_2$ and 3.2% (2.6 min to 4.0 max) for $\dot{V}E$ show better validity than the VO2 Master (Guidetti et al., 2018; Vogler et al., 2010). Due to their design, however, these systems have only limited use in artistic gymnastics. In this way, these systems could be used for high-level still ring routines, but not for high-level pommel horse or floor routines due to the increased risk of injury and balance restrictions (Seemann-Sinn et al., 2023; Seemann-Sinn et al., 2024). However, if the aim is to analyse $\dot{V}O_2$ during high-level pommel horse or floor routines, the validity of the measurement system has to be compromised to ensure practicability.

In terms of reliability, the results of the present study show that the VO2 Master has excellent to moderate relative and absolute test–rest reliability for gymnastic-specific short-term exercise. This confirms our second hypothesis (H2). In CT, $\dot{V}O_2$mean, $\dot{V}O_2$peak, O_2 deficit, and O_2 integral showed excellent to moderate relative reliability with ICCs between 0.74 and 0.95. The absolute reliability, measured as %SEM, was between good and moderate in CT with values of ≤10% and ≤5%. In FST, $\dot{V}O_2$mean, $\dot{V}O_2$peak, O_2 deficit, and O_2 integral showed good to excellent relative reliability (ICCs between 0.76 and 0.94). The absolute reliability in FST is also good to moderate, with values of ≤10% and ≤5% for the SEM. Montoye et al. (2020) also report acceptable test–retest reliability for the VO2 Master with a mean absolute percent error of 10.9 ± 9.4% for $\dot{V}O_2$ and 8.9 ± 7.3% for $\dot{V}E$. A direct comparison with the above-mentioned study regarding reliability is not advisable because the present study examined short-term exercises (50–70 s), while Montoye et al. (2020) examined long-term exercise (60 minutes). For this reason, the study was conducted to obtain data on the reliability of the VO2 Master during short-term exercise. Studies of other mobile metabolic analysers for $\dot{V}O_2$ measurement during short-term exercise could not be found. Thiessen et al. (2025) and Van Hooren et al. (2024) investigated the reliability of the VO2 Master using a metabolic simulator and are therefore not directly comparable with the results of the present study and that of Montoye et al. (2020), as in this case the reliability was investigated in human experiments. While Thiessen et al. (2025) demonstrated consistent results with a deviation of generally less than ±1% when measuring $\dot{V}O_2$ and $\dot{V}E$ with the VO2 Master over several days, the deviation in the study by Van Hooren et al. (2024) was >20% for the measured $\dot{V}O_2$ values. From our point of view, the large deviations observed by Van Hooren et al. (2024) cannot be explained, while on the other hand, the smaller deviations observed by Thiessen et al. (2025) compared to the present study are probably due to the difference between simulated experiments and human experiments.

This study shows strengths and limitations that should be considered when interpreting the results. The strengths of this study are that it examines reliability in gymnastic-specific exercises and thus provides important insights into how the VO2 Master works in this specific load situation. Furthermore, HR, lactate, and RPE data were included in the study to confirm similar physiological workloads across tests. On the other hand, the following limitations of the study must also be taken into account. While our sample size was comparable to previous studies, it was still small. Future studies should use a larger sample size. Since the tests were carried out on different days, day-to-day variability cannot be completely eliminated. However, the HR, lactate, and RPE data confirm comparable physiological stress levels during the tests. For the validation, only six different metabolic rates were used in this study, which could be problematic as variable error patterns across different intensities can limit extrapolation to higher $\dot{V}O_2$ values, which could be relevant for elite gymnasts (Van Hooren et al., 2024). In future studies, the number of metabolic rates should be increased to improve the consistency of the results. Furthermore, further validation with human data, as proposed by Van Hooren et al. (2024) was not carried out in this study. However, it has been shown that the VO2 Master has consistent validation data between human and simulated data using a metabolic simulator (Thiessen et al., 2025), suggesting that consistency can be assumed.

5. CONCLUSIONS

In conclusion, our study found acceptable validity and moderate test–retest reliability of the VO2 Master for the assessment of $\dot{V}O_2$ and $\dot{V}E$ during gymnastic-specific short-term (50–70 s) exercises. While further work is needed to confirm these findings, our results suggest that the VO2 Master is a practicable, reliable, and valid option to measure $\dot{V}O_2$ and $\dot{V}E$ during short-term exercises like artistic gymnastic routines. Although the VO2 Master shows certain deviations, it is the only metabolic analyser that can be practically used for high-level artistic gymnastics routines due to its design. In this way, the VO2 Master can be used for $\dot{V}O_2$ measurements during acrobatic and combat sports, providing insights into the energy demands of such sports. The study also provides important findings for assessing the validity and reliability of future

$\dot{V}O_2$ measurements in high-level artistic gymnastics routines.

Author Contributions: Conceptualisation, A.S.-S. and P.R.; methodology, A.S.-S., P.R., and I.S.; software, A.S.-S. and P.R.; formal analysis, A.S.-S., F.N., and I.S.; investigation, A.S.-S. and P.R.; data curation, A.S.-S. and P.R.; writing—original draft preparation, A.S.-S.; writing—review and editing, P.R., F.N., and I.S.; visualisation, A.S.-S.; supervision, F.N. and I.S.; project administration, A.S.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry of the Interior and Community and was supported by a decision of the German Bundestag.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study (written informed consent or cooperation agreement between the authors' research institute and the German Gymnastics Federation).

Acknowledgements: The authors thank the athletes, their coaches, and the authors' colleagues for participating in the study.

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

Preprint Declaration: This article is based on a preprint of Alexander Seemann-Sinn, Peter Rüdrich, Falk Naundorf, and Ingo Sandau previously posted on OSF Preprints under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence. The preprint version is available at https://doi.org/10.31219/osf.io/f3vjr

REFERENCES

Bearden, S., & Moffatt, R. (2001). V̇O2 and heart rate kinetics in cycling: Transitions from an elevated baseline. *Journal of Applied Physiology*, *90*(6), 2081–2087. https://doi.org/10.1152/jappl.2001.90.6.2081

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, *14*, 377–381.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Davis, P. (2024). The reliability of the VO2 Master Pro metabolic analyzer and comparison with the Cosmed Quark. *Science & Sports*, *39*(8), 706–709. https://doi.org/10.1016/j.scispo.2024.03.002

Fédération Internationale de Gymnastique. (2020). *2022–2024 Code of points: Men's artistic gymnastics.*

Gómez-Carmona, C. D., Pino-Ortega, J., Rico-González, M., & Bastida-Castillo, A. (2021). Global navigation satellite systems. In J. Pino-Ortega & M. Rico-González (Eds.), *The use of applied technology in team sport* (pp. 20–38). Routledge.

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

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

McBride, G. (2005). A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. *NIWA Client Report: HAM2005-062*, *45*, 307–310.

Montoye, A. H., Vondrasek, J. D., & Hancock II, B. J. (2020). Validity and reliability of the VO2 Master Pro for oxygen consumption and ventilation assessment. *International Journal of Exercise Science*, *13*(4), 1382–1401. https://doi.org/10.70252/THJT1177

Payne, R. (1997). Method comparison: Evaluation of least squares, Deming and Passing/Bablok regression procedures using computer simulation. *Annals of Clinical Biochemistry*, *34*(Pt 3), 319–320. https://doi.org/10.1177/000456329703400317

Perret, C., & Mueller, G. (2006). Validation of a new portable ergospirometric device (Oxycon Mobile) during exercise. *International Journal of Sports Medicine*, *27*(5), 363–367. https://doi.org/10.1055/s-2005-865666

Rodriguez, F. A., Latt, E., Jurimae, J., Maestu, J., Purge, P., Ramson, R., Haljaste, K., Keskinen, K. L., & Jurimae, T. (2016). VO2 kinetics in all-out arm stroke, leg kick and whole stroke front crawl 100-m swimming. *International Journal of Sports Medicine*, *37*(3), 191–196. https://doi.org/10.1055/s-0035-1554695

Rüdrich, P., Samson, A. J., & Wolfarth, B. (2022). Möglichkeiten der qualitätssicherung in der spiroergometrie im spitzen- und hochleistungssport ((Possibilities of quality management in spiroergometry in elite and high-performance sports) [Congress Proceedings]. *Frühjahrsschule „Technologien im Leistungssport", Leipzig IAT*, *4*(20).

Seemann-Sinn, A., Brehmer, S., Naundorf, F., & Wolfarth, B. (2021). Development of the routine duration in artistic gymnastics from 1997 to 2019. *International Journal of Performance Analysis in Sport*, *21*(2), 250–262. https://doi.org/10.1080/24748668.2021.1877601

Seemann-Sinn, A., Rudrich, P., Gorges, T., Naundorf, F., & Wolfarth, B. (2023). Physiological and energetic demands during still-rings routines of elite artistic gymnasts. *International Journal of Sports Physiology and Performance*, *18*(7), 704–710. https://doi.org/10.1123/ijspp.2022-0380

Seemann-Sinn, A., Rudrich, P., Gorges, T., Sandau, I., Naundorf, F., & Wolfarth, B. (2024). Determining physiological and energetic demands during high-level pommel horse routines using a modified method based on heart rate-oxygen uptake functions. *Sports*, *12*(1). https://doi.org/10.3390/sports12010027

Thiessen, J. S., Guluzade, N. A., Faricier, R., & Keir, D. A. (2025). (Re)assessment of the COSMED Quark CPET and VO2 Master Pro systems for measuring pulmonary gas exchange and ventilation. *Scandinavian Journal of*

Medicine & Science in Sports, *35*(2), 1–12. https://doi.org/10.1111/sms.70019

Ungerer, J. P. J., & Pretorius, C. J. (2017). Method comparison – A practical approach based on error identification. *Clinical Chemistry and Laboratory Medicine*, *56*(1), 1–4. https://doi.org/10.1515/cclm-2017-0842

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), 1–21. https://doi.org/DOI:10.1111/sms.14490

Vogler, A. J., Rice, A. J., & Gore, C. J. (2010). Validity and reliability of the Cortex MetaMax3B portable metabolic system. *Journal of Sports Sciences*, *28*(7), 733–742. https://doi.org/10.1080/02640410903582776

Webber, J. (2019). *Validity and reliability of the VM Pro*. https://V̇O2master.com/blog/validity-and-reliability-of-the-V̇O2-master/

Winkert, K., Kamnig, R., Kirsten, J., Steinacker, J. M., & Treff, G. (2020). Inter- and intra-unit reliability of the Cosmed K5: Implications for multicentric and longitudinal testing. *PLoS ONE*, *15*(10), e0241079. https://doi.org/10.1371/journal.pone.0241079

Received on December 10, 2025
Accepted on January 27, 2025