Published in final edited form as:
Res Q Exerc Sport. 2013 June ; 84(2): 223–231.

Test–Retest Reliability and Minimum Detectable Change Using the K4b²: Oxygen Consumption, Gait Efficiency, and Heart Rate for Healthy Adults During Submaximal Walking

Benjamin J. Darter,
Virginia Commonwealth University, Brooke Army Medical Center

Kelly M. Rodriguez, and
Brooke Army Medical Center

Jason M. Wilken
Brooke Army Medical Center

Abstract

Purpose—Oxygen consumption (VO_2; mLO_2/kg/min), gait efficiency (GE; mlO_2/kg/m) and heart rate (HR; beats per minute) are measures of physiological gait performance. However, the collection device, procedures for data normalization, and biological factors can affect measurement variability. The purpose of this study was to determine the test–retest reliability and minimum detectable change (MDC) for VO_2, GE, and HR with the $K4b^2$ at submaximal walking speeds in healthy young adults. A second purpose was to determine if net measures improved reproducibility.

Method—Twenty-two participants completed 2 identical treadmill tests on separate days at submaximal walking speeds from 0.71 m/s to 1.65 m/s.

Results—Intraclass correlation coefficient (ICC) values for gross VO_2, gross GE, and HR were greater than .85 for all walking speeds. Associated MDC values were approximately 7% to 10% for gross VO_2 and GE, and approximately 9% to 12% for HR. ICC values for resting VO_2 were lower, with MDC values approaching 25%. Subtracting out resting values to derive net VO_2 and GE values produced ICC values below.76 for the 2 slowest speeds but ICC values greater than .83 for the faster speeds. MDC values for net VO_2 and GE were up to 20% for the slowest speeds.

Conclusions—The results demonstrate metabolic cost can be assessed reliably using the $K4b^2$ during submaximal walking and that gross measures are more reliable than net measures. Furthermore, changes at self-selected speeds exceeding 1.0 mLO_2/kg/min in gross VO_2 and 0.01 mLO_2/kg/m in gross GE can be considered a true change in walking performance.

Correspondence should be addressed to Benjamin J. Darter, Department of Physical Therapy, Virginia Commonwealth University, 1200 East Broad Street, P. O. Box 980224, Richmond, VA 23298. bjdarter@vcu.edu.

The view(s) expressed herein are those of the author(s) and do not reflect the official policy or position of Brooke Army Medical Center, the U.S. Army Medical Department, the U.S. Army Office of the Surgeon General, the Department of the Army, Department of Defense, the U.S. government, or the National Institutes of Health.

We, the authors, affirm that we have no financial affiliation (including research funding) or involvement with any commercial organization that has a direct financial interest in any matter included in this manuscript.

Keywords

exercise test; gait; reproducibility of results

Oxygen consumption (VO_2; mLO_2/kg/min), gait efficiency (GE; mLO_2/kg/m), and heart rate (HR; beats per minute [bpm]) are commonly used measures to evaluate physiological effort. In recent years, many physiological variables, including VO_2, GE, and HR, have been assessed for measurement accuracy using the COSMED $K4b^2$ (Rome, Italy) and other current metabolic testing systems (Brehm, Harlaar, & Groepenhof, 2004; Carter & Jeukendrup, 2002; Crouter, Antczak, Hudak, DellaValle, & Haas, 2006; Duffield, Dawson, Pinnington, & Wong, 2004; Eisenmann, Brisko, Shadrick, & Welsh, 2003; Larsson, Wadell, Jakobsson, Burlin, & Henriksson-Larsen, 2004; Laurent et al., 2008; McLaughlin, King, Howley, Bassett, & Ainsworth, 2001; McNaughton, Sherman, Roberts, & Bentley, 2005; Meyer, Georg, Becker, & Kindermann, 2001; Perret & Mueller, 2006; Pinnington, Wong, Tay, Green, & Dawson, 2001; Prieur, Castells, & Denis, 2003; Schrack, Simonsick, & Ferrucci, 2010). Comparisons of measures from the $K4b^2$ with Douglas bags (McLaughlin et al., 2001), metabolic carts (Duffield et al., 2004; Eisenmann et al., 2003; Pinnington et al., 2001; Schrack et al., 2010), and a mass spectrometer (McNaughton et al., 2005) demonstrate acceptable concurrent validity for the $K4b^2$. Reliability, conceptualized as the reproducibility or dependability of the measurement, is less understood for many current metabolic testing systems including the $K4b^2$ (Macfarlane, 2001). In human testing, measurement reliability is influenced by error associated with the device as well as biological factors such as postabsorptive status, body position, and time of day (Compher, Frankenfield, Keim, & Roth-Yousey, 2006). Other factors such as the walking speed used and the reporting of net measures (gross-rest) may also affect measurement reliability, but the evidence is equivocal (Amorim, Byrne, & Hills, 2009; Brehm, Becher, & Harlaar, 2007). Despite potential sources of increased variability, it is important for clinical and research purposes that the physiological assessments are reliable and useful in determining if changes over time are in all probable likelihood a true change in the measure.

Few studies have assessed test–retest reliability using the $K4b^2$. Duffield et al. (2004) reported satisfactory reliability with approximately 4% variability in absolute VO_2 between repeated bouts of running. More recently, Thomas et al. (2009) reported intraclass correlation coefficients (ICCs) between .87 and .96 and coefficient of variation (CV) values of 11% to 13% for self-selected walking speeds in children with cerebral palsy. However, walking in healthy adults should be examined separately. Evidence from other commercially available systems suggests reliability is affected by biological factors such as the activity tested (walking vs. running; Blessinger et al., 2009) and the presence of walking pathology (Brehm, Nollet, & Harlaar, 2006). Acceptable reliability has been reported for walking with healthy adults using comparable systems (Blessinger et al., 2009), but generalizing the measurement reliability reported for other commercially available systems to the $K4b^2$ could be problematic because of systematic and unsystematic differences between the devices (McNaughton et al., 2005). Unfortunately, to this point, no published study has reported the reliability of physiological gait performance measures with the $K4b^2$ in healthy adults.

Research specifically examining the reliability of measures using K4b^2 in healthy adults is warranted given these limitations.

Regardless of the device and biological factors, the decision to report data in terms of gross or net values and the selection of walking intensity are important as they may also affect measurement reliability. Several studies recommended the use of net measures to decrease variability (Baker, Hausch, & McDowell, 2001; Schwartz, Koop, Bourke, & Baker, 2006), especially when metabolic changes are expected over time such as occurs with growth in adolescents. Others reported no added benefit to the use of net measures (Brehm et al., 2007; De Groot et al., 2010; de Mendonca & Pereira, 2008). At this time, both gross and net measures are routinely reported because it remains unclear if either provides higher measurement reliability. Likewise, reliability may improve as walking intensity is increased (de Mendonca & Pereira, 2008). In their study, de Mendonca and Pereira (2008) used treadmill grades up to 10% to increase walking intensity and reported CV percentages that decreased with higher intensities (gross VO_2 = 7.3% to 3.6%) and ICC values (> .89) that were higher than previous studies examining a single slow-to-moderate walking speed at 0% grade. Elsewhere, multiple walking speed protocols were used to increase intensity. The results from Amorim et al. (2009) partially support de Mendonca and Pereira's conclusion, with decreased CV values for VO_2 (from 7.9% to 4.4%) and GE (from 7.9% to 4.7%) as speed was increased to the adolescent's self-selected speed, but the results revealed no improvement at speeds > 0.22 m/s higher than self-selected. Conversely, data from Blessinger et al. (2009) indicate negligible differences in healthy adults for walking speeds between 0.89 m/s and 1.78 m/s (CV values, 7.47% to 6.56%). Overall, it remains unclear from ICC and CV data if the reliability of physiological measures obtained using the K4b^2 would reflect improvements across a range of submaximal walking speeds.

A shortcoming of common measures like the ICC is the statistic only provides a relative measure of reliability, which offers little assistance in interpreting changes between repeat tests. The "minimum detectable change" (MDC), also referred to as the "smallest detectable difference," is an absolute measure of reliability (measurement error), which accounts for various sources of variability in defining a confidence interval in units of the measure. MDC values are being increasingly used to assist in interpreting results and determining whether a change between repeated tests is random variation or a true change in performance (Haley & Fragala-Pinkham, 2006). In practical terms, MDC values can facilitate researchers and clinicians in identifying when circumstances such as experimental conditions, natural changes (aging), or surgical or rehabilitative interventions have truly altered a person's physiological gait performance. It is important to note the MDC differs from the "minimum clinically important difference," which is a change in performance judged to be relevant through comparison to some statistical or expert-based external criterion (Haley & Fragala-Pinkham, 2006). MDC values have only recently been reported for measures related to VO_2 during walking (Brehm et al., 2006, 2007; da Cunha-Filho, Henson, & Protas, 2007; De Groot et al., 2010; Thomas et al., 2009). In healthy adults, a 6% MDC value was reported for walking at a self-selected speed (Brehm et al., 2006). Another study revealed an MDC of approximately 20% for total energy expended during a 5-min walk test (da Cunha-Filho et al., 2007). Unfortunately, MDC values while using the K4b^2 are limited to a single study reporting an MDC of more than 30% for GE in children with cerebral palsy (Thomas et al., 2009). Furthermore, no study in healthy or clinical populations has examined the effect of walking speed on MDC values. As is the case with other measures of reliability, MDC values for measurements obtained with other devices or in populations with gait dysfunctions should not be applied universally. Research is especially needed to define MDC values for measures collected with the $K4b^2$ for walking in healthy adults because of how frequently healthy adults are tested in research studies examining gait.

The purpose of this study was to determine the test– retest reliability and MDC of VO_2, GE, and HR measures collected using the $K4b^2$ at multiple submaximal walking speeds in healthy young adults. A second purpose was to determine if the use of net measures for VO_2 and GE improved reproducibility.

METHODS

Participants

A total of 22 volunteers (11 men and 11 women; $M_{\text{age}} = 24.7 \pm 6.1$ years; $M_{\text{height}} = 1.71 \pm 0.10$ m; $M_{\text{weight}} = 70.2 \pm 12.6$ kg; $M_{\text{bodymassindex}} = 23.7 \pm 2.5$ kg/m^2; $M_{\text{leglength}} = 0.88 \pm 0.06$ m) completed the study. All participants recruited for the study were active, healthy young adults from the local community who were screened to exclude those with a history of lower-extremity injury requiring surgery, known balance impairment, neurological disorder, or chronic musculoskeletal or cardiopulmonary condition that limited or precluded involvement in physical activity. This study was approved by the institutional review board at Brooke Army Medical Center, and all participants provided written informed consent prior to participation.

Procedure

Each participant completed two identical multiple-speed treadmill walking tests on nonconsecutive days (5.3 ± 1.8 days apart). Demographic data were collected during the first session and included the participant’s weight and height on a standard digital scale and height bar (Detecto, Webb City, MO) and a leg-length measure from the greater trochanter of the femur to the floor, which was used to determine leg-length normalized walking speeds (England & Granata, 2007). Participants were tested at approximately the same time of day in a postabsorptive state of at least 2 hr, without caffeine ingestion or exercise on the same day as testing to minimize biological variability. Furthermore, feedback on test performance was provided only after the conclusion of the second test to minimize the risk for the feedback from the first test influencing performance on the second test.

VO_2 and HR were continuously monitored during the 0% grade treadmill test, which consisted of walking at five leg-length normalized speeds (0.71 [0.03] m/s, 0.94 [0.03] m/s, 1.18 [0.04] m/s, 1.41 [0.05] m/s, 1.65 [0.06] m/s) and a speed matching an overground self-selected walking speed. The controlled velocities were used to allow for dynamically equivalent comparisons across participants based on leg length (l): $v = 0.40 \times \sqrt{(g \times l)}$ (England & Granata, 2007). A nominal speed was calculated using the equation, with the other controlled speeds being 20% and 40% faster and slower. The self-selected walking speed was determined prior to the treadmill test by averaging the velocity of a retroreflective marker placed on the participant while walking approximately 10 laps on a 25-m oval walkway at a self-selected pace. Each stage lasted 5 min with steady-state data from the final 2 min used in determining the gross VO_2 and GE. Resting values, determined using steady-state values collected prior to walking while seated comfortably for a minimum of 10 min, were subtracted from the gross measures to calculate net VO_2 and GE. A $K4b^2$ breath-by-breath pulmonary gas exchange system was used for all collections and was calibrated prior to each test to manufacturer recommendations with known volume and gas concentrations. A detailed description of the device can be found elsewhere (Pinnington et al., 2001). In addition, HR, expressed in bpm, was collected via an interfaced Polar T31 HR monitor (Lake Success, NY) placed around the thorax.

Statistical Analysis

Each measure was evaluated for normality with the Shapiro-Wilks Test and homoscedasticity with Levene's Test (Portney & Watkins, 2000). After confirming the data met assumptions for analysis of variance (ANOVA) testing, a two-factor (test, speed) repeated-measures ANOVA was used for each measure to evaluate for systematic differences between test sessions, with statistical significance set at $p < .05$. An ICC (model 2,k) was calculated for each walking speed to evaluate the reliability of the physiological gait performance measures, with coefficients of .75 or higher considered to be good and .75 or lower to be poor-to-moderate (Portney & Watkins, 2000). The use of CV values has been the most common approach previously to examine variability between tests, and in the current study, a CV for method error was calculated as follows: $CV = 100 \times (2 \times (SD_d / \sqrt{2})/(X_1 + X_2)$ (Portney & Watkins, 2000). SD_d represents the standard deviation of the differences between the two tests, and X_1 and X_2 represent the two test means, respectively. The standard error of measurement (SEM) also provides a measure of variability but was primarily used for calculating the MDC. SEM values were calculated as follows: $SEM = SD \times \sqrt{(1 - ICC)}$, with SD representing the standard deviation of the measure. MDC values, which reflect the magnitude of change necessary to provide confidence that a change is not be the result of random variation or measurement error, were calculated as follows: $MDC = z\text{-score } (95\% \text{ CI}) \times SEM \times \sqrt{2}$ (Haley & Fragala-Pinkham, 2006). All statistical calculations were completed using the Statistical Package for the Social Sciences Version 16 (Chicago, IL) and Microsoft Excel 2007 (Redmond, WA).

RESULTS

Data from the 22 participants were used in the final analyses of VO_2 and GE. Data for 1 participant were not included in the final analyses of HR due to a malfunction of the HR-monitoring equipment during the retest. Tables 1 through 3 show the test means, ICC, CV, SEM, and MDC values at each walking speed for VO_2, GE, and HR, respectively. Repeated-measures ANOVAs demonstrated no systematic differences between Tests 1 and 2 across the tested walking speeds for any variable: gross VO_2, $F(1, 21) = 0.04$, $p = .84$; net VO_2, $F(1, 21) = 2.07$, $p = .16$; gross GE, $F(1, 21) = 0.2$, $p = .87$; net GE, $F(1, 21) = 1.79$, $p = .19$; HR, $F(1, 20) = 0.22$, $p = .64$.

VO_2

As listed in Table 1, ICC values for gross VO_2 were excellent for all submaximal walking speeds and ranged from .85 to .96. VO_2 data collected at rest was less reliable, with an ICC value of .43. Calculated between-test variability (CV) for gross VO_2 was 7.3% at rest but markedly improved for walking speeds at 2.0% to 2.6%. The MDC values for VO_2 ranged from 0.75 mLO_2/kg/min to 1.27 mLO_2/kg/min, which represented 6.9% to 9.6% of the first test means, but 25.5% of the resting value. Subtracting out resting values to derive net VO_2 produced ICC values that were less than .74 for the two slowest speeds but were .83 to .93 for the other speeds. In all cases, the ICC for net VO_2 was lower than the equivalent ICC for gross VO_2. CV values for net VO_2 ranged from 3.0% to 7.3% and progressively decreased as speed increased unlike gross measures, which changed negligibly as speed increased. Furthermore, MDC values for net VO_2 were higher than gross VO_2 values despite smaller test means at 1.11 mLO_2/ kg/min to 1.49 mLO_2/kg/min and decreased from 20.3% to 9.8% as speed increased.

GE

Table 2 shows ICC values for gross GE were excellent for all submaximal walking speeds and ranged from .84 to .93. Gross GE CV values were nearly identical to gross VO_2 (CV range = 2.0%–2.6%), and the MDC values were 0.01 mLO_2/kg/m to 0.02 mLO_2/kg/m (MDC% = 6.8%–9.8%). In terms of net GE, ICC values were less than .76 for the slowest two speeds but were .85 to .90 for the faster walking speeds. As with VO_2, net GE ICC values were lower, and CV values (3.0%–7.4%) were higher for each corresponding speed. Likewise, MDC values were higher for net GE than for gross GE at 0.02 mLO_2/kg/m to 0.03 mLO_2/kg/m and decreased from 20.3% to 10.0% as speed increased.

HR

Table 3 presents excellent ICC values for HR (ICC range = .81–.96). For HR, the CV values were 6.5% at rest and 3.2%to 3.9% for walking. The MDC values for HR were 9.2 bpm to 14.3 bpm, which represented 8.9% to 11.7% while walking but 22.0% at rest.

DISCUSSION

Technological advances allowing the transition from more time-consuming techniques to modern portable devices such as the $K4b^2$ have played a significant role in increasing the use of indirect calorimetry for measuring physiological effort. Studies on measurement accuracy for many currently used devices can be found in the literature, but reliability data are often lacking. The purpose of this study was to evaluate the test–retest reliability and calculate MDC values for level-grade walking at multiple submaximal speeds in healthy adults. Overall, the results suggest excellent reliability, and acceptable measurement errors were obtained for VO_2, GE, and HR when collected using the $K4b^2$. However, it is recommended to report gross measures, which were more reproducible than net measures especially at slower speeds.

The test–retest results compare favorably with previously published reliability data obtained using other devices at walking speeds in healthy adults (Table 4). In the Blessinger et al.

(2009) study, which most closely matches the design of the current study, participants walked on a treadmill at 0.89 m/s, 1.34 m/s, and 1.78 m/s and ran at 2.68 m/s. ICC values between .77 and .85 for walking and CV values between approximately 6% and 7% suggested good reliability with the VmaxST device (VIASYS, Yorba Linda, CA) used in VO_2 measures. The results of the current study using the $K4b^2$ during submaximal walking revealed consistently higher ICC values (ICC range = .84–.96) and CV values less than half of those presented by Blessinger et al. The fact that Blessinger et al. reported absolute VO_2 values should be of minimal consequence as only a nominal difference was found previously in a study reporting ICC values for both absolute and relative VO_2 (Dobrovolny, Ivey, Rogers, Sorkin, & Macko, 2003).

Increasing the walking intensity by increasing the treadmill speed did not appreciably change variability for gross or net measures of VO_2 and GE. Improvements were noted for net measures, but only for speeds below the nominal walking speed (1.18 m/s). These results are in contrast to the work of de Mendonca and Pereira (2008), who found lower variability with increasing walking grades for both gross and net measurements. Similar mean HRs between the current study and the study by de Mendonca and Pereira suggest that intensity itself was not the determining factor. Rather, some other factor related to walking at increasing grades likely led to progressive decreases for within-participant variability. Instead, the findings agree with the results from Blessinger et al. (2009), who reported no change in measures across multiple speeds at 0% grade in healthy adults. The results also partially agree with those of Amorim et al. (2009), who reported improvement in gross VO_2 and GE measures up to self-selected walking speeds. However, although lower variability was observed at the lowest speeds in the current study, the improvement was limited to net measures only and not gross measures as was reported by Amorim et al.

Similar conclusions regarding overall reproducibility and the effect of walking intensity can be drawn from the MDC values for VO_2 and GE measures obtained using the $K4b^2$. The MDC values calculated in the current study are acceptable and similar to that of Brehm et al. (2006) at approximately 8% for gross VO_2 and GE for a self-selected walking speed. In addition, finding MDC values in the current study that were larger for net measures than gross measures also agrees with previous results from studies in healthy children (Brehm et al., 2007), children with cerebral palsy (Brehm et al., 2007), and children with spina bifida (De Groot et al., 2010). Regarding the effect of walking speed, for gross measures, it was minimal, with all MDC values within approximately 3% and no clear trend of decreasing values as speed increased. As might be expected from the preceding results, MDC values for net measures did not improve meaningfully except for below the nominal walking speed.

Healthy adults are frequently used in studies examining experimental conditions such as walking with added loads (Grabowski, Farley, & Kram, 2005; Puthoff, Darter, Nielsen, & Yack, 2006). Changes in steady-state VO_2 of less than 10% have been cited as being clinically insignificant (McLaughlin et al., 2001). Although the exact relevant value may be debated, for a device to be truly useful, it must provide test–retest MDC values that are sufficiently small to detect meaningful change. The current findings suggest that only gross measures collected using the $K4b^2$ provide sufficient resolution to detect changes less than 10% across the range of speeds tested. We have found that changes in gross VO_2 and GE of approximately 1.0 mLO_2/kg/min and 0.01 mLO_2/kg/m, respectively, can serve as benchmarks of true change in healthy adults at self-selected walking speeds. The MDC values may also be of use in interpreting changes in clinical populations who have not otherwise had test–retest reliability established, as it is unlikely that patient groups will demonstrate MDC values lower than those of healthy adults. However, the benchmarks established in this manuscript are most appropriate for use in interpreting data collected during treadmill walking because physiological gait performance for over-ground and treadmill walking may not be equivalent (Berryman et al., 2012).

The results indicate that resting VO_2 measurements have the potential to be highly variable and can negatively influence the reliability of net measures. This detrimental effect was especially true at lower speeds, where resting values represent a greater portion of the VO_2 consumed during steady-state walking (Brehm et al., 2007; De Groot et al., 2010). Although it is impossible to determine what exactly produced the variable measures at rest, two possible contributing factors are suggested. First, the $K4b^2$ device may have been inherently more variable for resting measures and produced large measurement error (Littlewood et al., 2002). Second, the procedures used to obtain resting values may not have been sufficient to ensure reliable collection of a true resting value. Although the resting procedures were consistent with previously published methods involving exercise testing (Brehm et al., 2006, 2007), more restrictive guidelines (such as a longer fasting period, testing in a reclined position, and controlling for the menstrual cycle) have been recommended in studies specifically measuring resting values (Compher et al., 2006; Haugen, Chan, & Li, 2007). Improved reliability at rest may have been possible through stricter adherence to those published guidelines.

A potential limitation of the current study is the relatively small sample size. Previous research on measurement reliability suggests a minimum of 50 participants (Terwee et al., 2007). However, post-hoc power analysis using study data indicated that increasing the sample from 22 to 50 would have had a negligible effect. An unrealistically large sample (> 1,000 participants) would have been necessary to find statistical significance given the data's proportionally small mean differences and variances.

CONCLUSIONS

In conclusion, several common measures of physiological effort obtained using the $K4b^2$ are reliable for healthy adults across a range of functional walking speeds. MDC values for grossVO_2 and GE measures range from7%to 10%, providing further evidence that a change in gross values of more than 10% between tests should be interpreted as a true change in walking performance. Finally, it is recommended to restrict the reporting of net measures to the studies that closely follow recommended guidelines for determining resting values.

WHAT DOES THIS ARTICLE ADD?

Measures of physiological gait performance are affected by a variety of factors that increase data variability. This variability can decrease the measurement's clinical utility. The current study is the first to identify the reliability and measurement error for several common physiological parameters (VO_2, GE, and HR) collected using the COSMED $K4b^2$ during submaximal walking in healthy adults. Principally, we concluded from the results that the reliability and measurement error are acceptable for VO_2 and GE when expressed as gross values, but subtracting the resting values to derive a net value may be problematic. We also identify benchmarks for VO_2 and GE that can be used when interpreting if a change is greater than random variation. This study ultimately provides additional information regarding how physiological data can be used to describe gait performance.

Future research can build upon this study in several ways. Firstly, the $K4b^2$ is one of the few available portable devices. Portability enhances the ability to quantify “real-world” gait performance. However, those walking conditions would likely exhibit greater natural variability with negative consequences for reliability and measurement error. Investigating reliability and measurement error with the device under non-laboratory-based conditions is warranted. Secondly, the current study only examined the $K4b^2$ with healthy adults; future studies should extend this research into clinical populations with gait dysfunctions. For example, in persons with a lower-extremity amputation, population-specific MDC values would be useful for judging the efficacy of different prosthetic devices or the effectiveness of novel rehabilitation interventions.

Acknowledgments

The participants were enrolled in studies approved by the institutional review board at Brooke Army Medical Center in Ft. Sam Houston, TX (C.2008.174).

Support for this project was provided by the Military Amputee Research Program through grant awards to the authors (BJD, JMW) and by Clinical and Translational Science Award No. KL2TR000057 from the National Center for Advancing Translational Sciences (BD).

References

Amorim PRS, Byrne NM, Hills AP. Within- and between-day repeatability and variability in children’s physiological responses during submaximal treadmill exercise. Research Quarterly for Exercise and Sport. 2009; 80:575–582. [PubMed: 19791644]

Baker R, Hausch A, McDowell B. Reducing the variability of oxygen consumption measurements. Gait & Posture. 2001; 13:202–209. [PubMed: 11323226]

Berryman N, Gayda M, Nigam A, Juneau M, Bherer L, Bosquet L. Comparison of the metabolic energy cost of overground and treadmill walking in older adults. European Journal of Applied Physiology and Occupational Physiology. 2012; 112:1613–1620. [PubMed: 21863296]

Blessinger J, Sawyer B, Davis C, Irving BA, Weltman A, Gaesser G. Reliability of the VmaxST portable metabolic measurement system. International Journal of Sports Medicine. 2009; 30:22–26. [PubMed: 18651368]

Brehm MA, Becher J, Harlaar J. Reproducibility evaluation of gross and net walking efficiency in children with cerebral palsy. Developmental Medicine & Child Neurology. 2007; 49:45–48. [PubMed: 17209976]

Brehm MA, Harlaar J, Groepenhof H. Validation of the portable VmaxST system for oxygen-uptake measurement. Gait & Posture. 2004; 20:67–73. [PubMed: 15196523]

Brehm MA, Nollet F, Harlaar J. Energy demands of walking in persons with postpoliomyelitis syndrome: Relationship with muscle strength and reproducibility. Archives of Physical Medicine and Rehabilitation. 2006; 87:136–140. [PubMed: 16401452]

Carter J, Jeukendrup AE. Validity and reliability of three commercially available breath-by-breath respiratory systems. European Journal of Applied Physiology. 2002; 86:435–441. [PubMed: 11882930]

Compher C, Frankenfield D, Keim N, Roth-Yousey L. Best practice methods to apply to measurement of resting metabolic rate in adults: A systematic review. Journal of the American Dietetic Association. 2006; 106:881–903. [PubMed: 16720129]

Crouter SE, Antczak A, Hudak JR, DellaValle DM, Haas JD. Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems. European Journal of Applied Physiology. 2006; 98:139–151. [PubMed: 16896734]

da Cunha-Filho IT, Henson H, Protas EJ. Reliability of a portable gas analyzer during a 5-min walk test. American Journal of Physical Medicine & Rehabilitation. 2007; 86:469–473. [PubMed: 17515686]

Dawes H, Collet J, Ramsbottom R, Howells K, Sackley C, Wade D. Measuring oxygen cost during level walking in individuals with acquired brain injury in the clinical setting. Journal of Sport Science and Medicine. 2004; 3:76–82.

De Groot JF, Takken T, Schoenmakers MA, Tummers L, Vanhees L, Helders PJ. Reproducibility of energy cost of locomotion in ambulatory children with spina bifida. Gait & Posture. 2010; 31:159–163. [PubMed: 19875289]

de Mendonca GV, Pereira FD. Between-day variability of net and gross oxygen uptake during graded treadmill walking: Effects of different walking intensities on the reliability of locomotion economy. Applied Physiology Nutrition and Metabolism. 2008; 33:1199–1206.

Dobrovolny CL, Ivey FM, Rogers MA, Sorkin JD, Macko RF. Reliability of treadmill exercise testing in older patients with chronic hemiparetic stroke. Archives of Physical Medicine & Rehabilitation. 2003; 84:1308–1312. [PubMed: 13680566]

Duffield R, Dawson B, Pinnington HC, Wong P. Accuracy and reliability of a Cosmed K4b2 portable gas analysis system. Journal of Science & Medicine in Sport. 2004; 7:11–22. [PubMed: 15139160]

Eisenmann JC, Brisko N, Shadrick D, Welsh S. Comparative analysis of the Cosmed Quark b2 and K4b2 gas analysis systems during submaximal exercise. Journal of Sports Medicine & Physical Fitness. 2003; 43:150–155. [PubMed: 12853896]

England SA, Granata KP. The influence of gait speed on local dynamic stability of walking. Gait & Posture. 2007; 25:172–178. [PubMed: 16621565]

Grabowski A, Farley CT, Kram R. Independent metabolic costs of supporting body weight and accelerating body mass during walking. Journal of Applied Physiology. 2005; 98:579–583. [PubMed: 15649878]

Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Physical Therapy. 2006; 86:735–743. [PubMed: 16649896]

Haugen HA, Chan LN, Li F. Indirect calorimetry: A practical guide for clinicians. Nutrition in Clinical Practice. 2007; 22:377–388. [PubMed: 17644692]

Larsson PU, Wadell KM, Jakobsson EJ, Burlin LU, Henriksson-Larsen KB. Validation of the MetaMax II portable metabolic measurement system. International Journal of Sports Medicine. 2004; 25:115–123. [PubMed: 14986194]

Laurent CM, Meyers MC, Robinson CA, Strong LR, Chase C, Goodwin B. Validity of the VmaxST portable metabolic measurement system. Journal of Sports Sciences. 2008; 26:709–716. [PubMed: 18409102]

Littlewood RA, White MS, Bell KL, Davies PS, Cleghorn GJ, Grote R. Comparison of the Cosmed K4 b(2) and the Deltatrac II metabolic cart in measuring resting energy expenditure in adults. Clinical Nutrition. 2002; 21:491–497. [PubMed: 12468369]

Macfarlane DJ. Automated metabolic gas analysis systems: A review. Sports Medicine. 2001; 31:841–861. [PubMed: 11665912]

McLaughlin JE, King GA, Howley ET, Bassett DR Jr, Ainsworth BE. Validation of the COSMED K4 b2 portable metabolic system. International Journal of Sports Medicine. 2001; 22:280–284. [PubMed: 11414671]

McNaughton LR, Sherman R, Roberts S, Bentley DJ. Portable gas analyser Cosmed K4b(2) compared to a laboratory based mass spectrometer system. Journal of Sports Medicine and Physical Fitness. 2005; 45:315–323. [PubMed: 16230983]

Meyer T, Georg T, Becker C, Kindermann W. Reliability of gas exchange measurements from two different spiroergometry systems. International Journal of Sports Medicine. 2001; 22:593–597. [PubMed: 11719895]

Perret C, Mueller G. Validation of a new portable ergospirometric device (Oxycon Mobile) during exercise. International Journal of Sports Medicine. 2006; 27:363–367. [PubMed: 16729377]

Pinnington HC, Wong P, Tay J, Green D, Dawson B. The level of accuracy and agreement in measures of FEO2, FECO2 and VE between the Cosmed K4b2 portable, respiratory gas analysis system and a metabolic cart. Journal of Science & Medicine in Sport. 2001; 4:324–335. [PubMed: 11702919]

Portney, LG.; Watkins, MP. Statistical measures of reliability. In: Portney, LG.; Watkins, MP., editors. Foundations of clinical research: Applications to practice. 2. Upper Saddle River, NJ: Prentice Hall; 2000. p. 565p. 579-580.

Prieur F, Castells J, Denis C. A methodology to assess the accuracy of a portable metabolic system (VmaxST). Medicine & Science in Sports & Exercise. 2003; 35:879–885. [PubMed: 12750601]

Puthoff ML, Darter BJ, Nielsen DH, Yack HJ. The effect of weighted vest walking on metabolic responses and ground reaction forces. Medicine & Science in Sports & Exercise. 2006; 38:746–752. [PubMed: 16679992]

Schrack JA, Simonsick EM, Ferrucci L. Comparison of the Cosmed K4b(2) portable metabolic system in measuring steady-state walking energy expenditure. PLoS One. 2010; 5:e9292. [PubMed: 20174583]

Schwartz MH, Koop SE, Bourke JL, Baker R. A nondimensional normalization scheme for oxygen utilization data. Gait & Posture. 2006; 24:14–22. [PubMed: 16112865]

Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology. 2007; 60:34–42. [PubMed: 17161752]

Thomas SS, Buckon CE, Schwartz MH, Russman BS, Sussman MD, Aiona MD. Variability and minimum detectable change for walking energy efficiency variables in children with cerebral palsy. Developmental Medicine and Child Neurology. 2009; 51:615–621. [PubMed: 19627334]

TABLE 1

Reproducibility and Minimal Detectable Change for Oxygen Consumption (VO_2) at Submaximal Walking Speeds

	Rest	40% Below Nominal Speed 0.71 (0.02) m/s	20% Below Nominal Speed 0.94 (0.03) m/s	Nominal Speed 1.18 (0.04) m/s	20% More Than Nominal Speed 1.41 (0.05) m/s	40% More Than Nominal Speed 1.65 (0.06) m/s	Self-Selected Walking Speed 1.28 (0.14) m/s
Gross VO_2 (ml/kg/min)							
Test 1 (SD)	3.46 (0.41)	9.22 (1.05)	10.51 (1.04)	12.24 (1.06)	14.98 (1.42)	18.56 (1.64)	13.45 (1.95)
Test 2 (SD)	3.49 (0.44)	9.19 (0.86)	10.47 (0.84)	12.23 (0.98)	14.95 (1.15)	18.47 (1.43)	13.46 (1.70)
ICC [CI]	.43 [−.41, .77]	.92 [.81, .97]	.85 [.63, .94]	.89 [.74, .96]	.89 [.73, .95]	.91 [.78, .96]	.96 [.89, .98]
CV%	7.3	2.6	2.6	2.0	2.3	2.0	2.5
SEM	0.32	0.27	0.36	0.34	0.43	0.46	0.37
MDC (%M)	0.88 (25.5)	0.75 (8.1)	1.01 (9.6)	.94 (7.6)	1.18 (7.9)	1.27 (6.9)	1.01 (7.5)
Net VO_2 (ml/kg/min)							
Test 1 (SD)		5.92 (1.02)	7.22 (1.08)	8.96 (1.11)	11.68 (1.43)	15.26 (1.73)	10.19 (1.79)
Test 2 (SD)		5.70 (0.68)	6.98 (0.69)	8.74 (0.84)	11.46 (1.00)	14.97 (1.37)	9.96 (1.55)
ICC [CI]		.74 [.40, .89]	.66 [.21, .86]	.83 [.59, .93]	.83 [.60, .93]	.88 [.71, .95]	.93 [.84, .97]
CV%		7.3	6.4	4.1	3.7	3.0	4.0
SEM		0.43	0.52	0.40	0.50	0.54	0.44
MDC (%M)		1.20 (20.3)	1.43 (19.8)	1.11 (12.4)	1.39 (11.9)	1.49 (9.8)	1.23 (12.0)

Note. *SD* = standard deviation; ICC = intraclass correlation coefficient; CI = 95% confidence interval; CV = coefficient of variation; SEM = standard error of measurement; MDC = minimum detectable change; %M = MDC expressed as a percent of measurement mean.

TABLE 2

Reproducibility and Minimal Detectable Change for Gait Efficiency (GE) at Submaximal Walking Speeds

	40% Below Nominal Speed 0.71 (0.02) m/s	**20% Below Nominal Speed 0.94 (0.03) m/s**	**Nominal Speed 1.18 (0.04) m/s**	**20% More Than Nominal Speed 1.41 (0.05) m/s**	**40% More Than Nominal Speed 1.65 (0.06) m/s**	**Self-Selected Walking Speed – 1.28 (0.14) m/s**
Gross GE (ml/kg/m)						
Test 1 (SD)	0.22 (0.02)	0.19 (0.02)	0.17 (0.01)	0.18 (0.02)	0.19 (0.02)	0.17 (0.02)
Test 2 (SD)	0.22 (0.02)	0.19 (0.02)	0.17 (0.01)	0.18 (0.02)	0.19 (0.02)	0.17 (0.02)
ICC [CI]	.93 [.82, .97]	.84 [.62, .93]	.89 [.75, .96]	.90 [.75, .96]	.93 [.83, .97]	.89 [.74, .96]
CV%	2.6	2.6	2.0	2.3	2.0	2.5
SEM	0.01	0.01	0.01	0.01	0.01	.01
MDC (%M)	0.02 (8.4)	0.02 (9.8)	0.01 (7.8)	0.01 (7.9)	0.01 (6.8)	0.02 (8.4)
Net GE (ml/kg/m)						
Test 1 (SD)	0.14 (0.02)	0.13 (0.02)	0.13 (0.02)	0.14 (0.02)	0.15 (0.02)	0.13 (0.02)
Test 2 (SD)	0.13 (0.02)	0.12 (0.01)	0.12 (0.01)	0.14 (0.01)	0.15 (0.02)	0.13 (0.01)
ICC [CI]	.76 [.44, .90]	.69 [.27, .87]	.85 [.64, .94]	.85 [.64, .94]	.90 [.77, .96]	.87 [.68, .94]
CV%	7.4	6.4	4.0	3.8	3.0	4.2
SEM	0.01	0.01	0.01	0.01	0.01	.01
MDC (%M)	0.03 (20.3)	0.03 (19.4)	0.02 (12.1)	0.02 (12.0)	0.02 (10.0)	0.02 (11.8)

Note. *SD* = standard deviation; ICC = intraclass correlation coefficient; CI = 95% confidence interval; CV = coefficient of variation; SEM = standard error of measurement; MDC = minimum detectable change; %M = MDC expressed as a percent of measurement mean.

TABLE 3

Reproducibility and Minimal Detectable Change for Heart Rate (HR) at Submaximal Walking Speeds

	Rest	**40% Below Nominal Speed 0.71 (0.02) m/s**	**20% Below Nominal Speed 0.94 (0.03) m/s**	**Nominal Speed 1.18 (0.04) m/s**	**20% More Than Nominal Speed 1.41 (0.05) m/s**	**40% More Than Nominal Speed 1.65 (0.06) m/s**	**Self-Selected Walking Speed 1.28 (0.14) m/s**
Test 1 (SD)	65.2 (11.4)	83.4 (13.8)	88.3 (14.1)	94.3 (12.2)	103.6 (16.3)	115.9 (19.4)	98.1 (14.5)
Test 2 (SD)	66.7 (12.3)	83.4 (13.4)	87.7 (14.2)	93.5 (15.2)	102.2 (17.0)	113.2 (20.0)	97.2 (15.4)
ICC [CI]	.81 [.53, .92]	.93 [.84, .97]	.93 [.82, .97]	.94 [.84, .97]	.96 [.89, .98]	.96 [.90, .99]	.93 [.82, .97]
CV%	6.5	3.9	3.9	3.5	3.3	3.2	3.5
SEM	5.2	3.6	4.0	3.9	3.7	3.9	4.2
MDC (%M)	14.3 (22.0)	10.0 (12.0)	10.4 (11.7)	10.0 (10.6)	9.2 (8.9)	10.9 (9.4)	11.0 (11.2)

Note. *SD* = standard deviation; ICC = intraclass correlation coefficient; CI = 95% confidence interval; CV = coefficient of variation; SEM = standard error of measurement; MDC = minimum detectable change; %M = MDC expressed as a percent of measurement mean.

TABLE 4

Summary of the Reported Reliability for Walking in Healthy Adults at 0% Grade

Author	**Measures**	**ICC**	**CV**
Dawes et al. (2004) *n* = 21	VO_2 (ml/kg/min)	.55	
	net VO_2 (ml/kg/min)	.68	
	GE (ml/kg/m)	.42	
	net GE (ml/kg/m)	.61	
Brehm et al. (2006) *n* = 14	VO_2 (J/kg/min)	.98	
	GE (J/kg/m)	.93	
da Cunha-Filho et al. (2007) *n* = 40	total VO_2 (ml/kg/min)	.99	7.4
	GE (ml/kg/m)	.67	9.0
de Mendonca and Pereira (2008) *n* = 16	VO_2 (ml/min)	.89	7.3
	net VO_2 (ml/min)	.87	8.8
	HR (bpm)	.93	7.5
Blessinger et al. (2009) *n* = 45	VO_2 (L/min)	.77–.85	6.6–7.6

Note. VO_2 = oxygen consumption; GE = gait efficiency; HR = heart rate; ICC = intraclass correlation coefficient; CV = coefficient of variation.