Test-retest reliability of physiological parameters in elite junior distance runners following allometric scaling

Richard C Blagrove [a,b]*, Glyn Howatson [b,c], and Philip R Hayes [b]

[a] *Faculty of Health, Education and Life Sciences, School of Health Sciences, Birmingham City University, City South Campus, Westbourne Road, Edgbaston, Birmingham, B15 3TN*

[b] *Faculty of Health and Life Sciences, Division of Sport, Exercise and Rehabilitation, Northumbria University, Northumberland Building, Newcastle-upon-Tyne, NE1 8ST*

[c] *Water Research Group, Northwest University, Potchefstroom, South Africa.*

Richard Blagrove (*corresponding author)
Tel.: 0121 300 4396
Email: richard.blagrove@bcu.ac.uk

Dr Philip Hayes
Tel.: 0191 227 4690
Email: phil.hayes@northumbria.ac.uk

Prof Glyn Howatson
Tel.: 0191 227 3573
Email: glyn.howatson@nothumbria.ac.uk

Disclosure Statement:
The authors report no conflict of interest.

Test-retest reliability of physiological parameters in elite junior distance runners following allometric scaling

Abstract

This study aimed to quantify the intra-individual reliability of a number of physiological variables in a group of national and international young distance runners. Sixteen (8 male, 8 female) participants (16.7±1.4 years) performed a sub-maximal incremental running assessment followed by a maximal running test, on two occasions separated by no more than seven days. Maximal oxygen uptake ($\dot{V}O_{2max}$), speed at $\dot{V}O_{2max}$ (km.h^{-1}), running economy, and speed and heart rate (HR) at fixed blood lactate concentrations were determined. $\dot{V}O_{2max}$ and running economy were scaled for differences in body mass using a power exponent derived from a larger cohort of young runners (*n*=42). Running economy was expressed as oxygen cost and energy cost at the speed associated with lactate turn-point (LTP) and the two speeds prior to LTP. Results of ANOVA revealed an absence of systematic bias between trials. Reliability indices showed a high level of reproducibility across all parameters (typical error (TE) ≤ 2%; intra-class correlation coefficient > 0.8; effect size < 0.6). Expressing running economy as energy cost appears to provide superior reliability than using oxygen cost (TE ~1.5% vs ~2%). Blood lactate and HR were liable to daily fluctuations of 0.14-0.22 mmol·L^{-1} and 4-5 beats·min^{-1} respectively. The minimum detectable change values (95% confidence) for each parameter, are also reported. Exercise physiologists can be confident that measurement of important physiological determinants of distance running performance are highly-reproducible in elite junior runners.

Keywords: aerobic fitness, youth, testing, endurance

Highlights:

- When oxygen uptake is expressed using an appropriate scaling exponent in young distance runners (two-thirds of body mass), $\dot{V}O_{2max}$ and running

economy exhibit high test-retest reliability (TE: 1-2%, ICC: 0.88-0.98)

- Energy cost, which captures the influence of substrate utilization within a measure of running economy, should be preferred over measures of oxygen cost, as this metric demonstrates superior reliability.

- Blood lactate is liable to daily fluctuations of ~0.2 $mmol^.L^{-1}$ and heart rate 4-5 $beats^.min^{-1}$ for the same intensity of running. Speed at fixed concentrations of blood lactate offers an alternative marker that appears to provide a higher degree of day-today stability.

Introduction

Distance running performance is determined by several important physiological qualities including maximal oxygen uptake ($\dot{V}O_{2max}$), running economy and speed at various points on a lactate curve (Sparling, 1984). Well-established treadmill testing protocols are used to evaluate the efficacy of training interventions on these physiological variables (Winter, Jones, Davison, Bromley, & Mercer, 2006). Practitioners also rely on data acquired from physiology trials to individualize athlete training load and make inferences concerning fatigue status and readiness to perform (Halson, 2014). Providing accurate recommendations to athletes and their coaches can be problematic unless the reliability of the measurement tools has been established, which is particularly important when confounding factors such as lifestyle and diet are not well-controlled. To have a reasonable degree of confidence that an observed change in a physiological measure is meaningful following an intervention, it is important to determine the systematic and biological error associated with that measure, and hence what should be considered real change (Atkinson and Nevill, 1998).

The results of reliability studies conducted on moderately- and well-trained adult distance runners generally show a good-level of reproducibility (typical error (TE): <5%) for variables relating to oxygen uptake (Brisswalter & Legros, 1994; Pereira & Freedson, 1997; Midgley, McNaughton, Polman, & Marchant, 2007). Oxygen cost used as a measure of running economy, has a day-to-day TE of 2.4-4.7% despite showing pronounced inter-individuals differences in well-trained runners (Brisswalter & Legros, 1994; Morgan, Martin, Krahenbuhl, & Baldini, 1991; Saunders, Pyne, Telford, & Hawley, 2004). Quantification of running economy using the energy cost method takes account of the respiratory exchange ratio (RER) value,

thus recognizing that the energy expended for any given sub-maximal running speed is influenced by both $\dot{V}O_2$ and substrate utilization (Fletcher, Esau, & MacIntosh, 2009). It has therefore been suggested that energy cost may provide a more valid (Shaw, Ingham, & Folland, 2014) and reliable (Shaw et al., 2013) measure than oxygen cost for running economy, however the reproducibility of this parameter is largely untested. Current guidelines suggest a test-retest TE of ~10% should be expected for measurements of blood lactate at any given speed or work rate (Winter et al., 2006), and heart rate (HR) will fluctuate 2-8 beats·min^{-1} for the same intensity on different days (Achten & Jeukendrup, 2003; Lambert et al., 1998).

It is currently unknown whether test-retest reliability is similar for physiological parameters in elite junior distance runners. Competitive adolescent distance runners are likely to undertake relatively high-volumes of training for their age (Wilson, Raglin, & Harger, 1999) and are potentially exposed to a variety of non-training stressors such as school work and expectations imposed by significant others (Winsley and Matos, 2011). It is likely that this stress will impact physiological status via hormonal and chemical imbalances (Meeusen et al., 2013), therefore determining day-to-day variability in measurements is crucial. The tempo and timing of biological maturation is highly variable in adolescents and periods of accelerated physical development have been associated with disruptions to motor coordination (Beunen & Malina, 1988). Therefore maturational status has the potential to influence the short-term stability of physical performance measures in young athletes compared to adult runners.

High intra-class correlation coefficients (ICC) values (r=0.81-0.97) have previously been reported for $\dot{V}O_{2max}$ scores obtained during test-retest scenarios in children (Paterson, Cunningham, & Donner, 1981; Pivarnik, Dwyer, & Launderdale,

1996) and trained adolescents (Rivera-Brown & Frontera, 1998; Rivera-Brown, Rivera, & Frontera, 1995), however these studies used outdated and questionable methods to scale for differences in body mass and define attainment of $\dot{V}O_{2peak}/\dot{V}O_{2max}$ in participants. In particular, substantial variability in patterns of growth and maturation in young athletes mean that using a conventional ratio scaling approach to partitioning body mass is theoretically and statistically inappropriate (Eisenmann, Pivarnik, & Malina, 2001). Therefore, a factor that may influence the reliability metrics reported in previous studies relates to the method which is employed to scale for variations in body mass amongst participants. Utilizing allometrically adjusted values to determine $\dot{V}O_{2max}$ and running economy for the population under investigation is likely to provide a more valid and accurate assessment of reliability in these measures.

Running economy, speed at fixed blood lactate concentrations (sFBLC) and speed associated with $\dot{V}O_{2max}$ (s$\dot{V}O_{2max}$) have all been shown to be important determinants of performance in adolescent distance runners (Almarwaey, Jones, & Tolfrey, 2003), however reliability of these measures remains unreported in this population. Previous studies have also tended to measure the reliability of various parameters independently, and no studies have reported the repeatability of a number of measures in the same group of athletes. Thus, the purpose of this study was to compare the inter-session reliability of a number of physiological determinants of performance in an elite group of junior distance runners.

Methods

Study design

This study adopted a test-retest design whereby participants were required to attend

the laboratory for two identical trials separated by 3-7 days. Following a standardized warm-up, each trial involved a sub-maximal incremental running assessment followed by a maximal running test. The sub-maximal trial was used to assess physiological responses ($\dot{V}O_2$, $\dot{V}CO_2$, blood lactate and HR) to incremental exercise and the maximal test component measured $\dot{V}O_{2max}$ and allowed determination of $s\dot{V}O_{2max}$.

Participants

Sixteen (8 male, 8 female) high performing young distance runners participated in the test-retest part of this study. All participants had at least two years of competitive racing experience and were of national ($n=12$) or international standard ($n=4$) in their age group. To calculate an appropriate power function for $\dot{V}O_2$ and $\dot{V}O_{2max}$ to scale participants for differences in body size, data was pooled with a larger group ($n=42$) of young distance runners. Characteristics of the participants are shown in Table 1.

The study was conducted in accordance with the Helsinki declaration and received University level ethical approval. Participants were informed of the purpose, procedure and risks of the experiment and thereafter a parent/guardian (or if >18 years the participant themselves) provided signed consent to participate.

****Insert Table 1 near here****

Procedure

Trials were conducted at the same time of day (±1 h) for each participant under standardized environmental conditions (temperature, 16-20 °C; relative humidity, 29-50%; barometric pressure, 751-771 mmHg) in the same laboratory. Participants wore

similar clothes and the same running trainers for each trial and were instructed to follow a similar pattern of exercise and diet in the 48 h prior to each trial.

Expired air was monitored throughout the sub-maximal and maximal tests via an open circuit metabolic cart (Oxycon Pro, Enrich Jaeger GmbH, Germany). The automated system measured breath-by-breath gas exchange thus enabling calculation of pulmonary ventilation, $\dot{V}O_2$, $\dot{V}CO_2$ and RER. Participants breathed through a mask with low-dead space (99 or 125 ml) into a two-way valve with a dual gas sensor. Prior to every test both gas analyzers were calibrated with known concentrations of standard calibration gas (16% O_2; 5% CO_2), and the ventilation measurement unit with a 3 L syringe.

Breath-by-breath data were initially filtered by excluding any breaths which fell outside four standard deviations of the local mean (Lamarra, Whipp, Ward, & Wasserman, 1987). Filtering was conducted to remove any errant breaths, which do not reflect the underlying physiological response.

Submaximal running assessment

Following a 5 min warm-up at a speed 2 $km^.h^{-1}$ slower than the start speed for the assessment, participants completed a discontinuous incremental test involving 5-7 three minute stages on a motorized treadmill (HP Cosmos Pulsar 4.0, Cosmos Sports & Medical GmbH, Germany). Each stage was interspersed with a 30 s rest for extraction of a capillary blood sample. A judgment of the most appropriate speed for the first stage of the test was made based upon the participant's best times and published recommendations (Winter et al., 2006). Thereafter, speed was increased by 1 $km^.h^{-1}$ every stage until lactate turnpoint (LTP; rise of >1 $mmol^.L^{-1}$ compared to

previous stage) had been surpassed. The gradient of the treadmill was kept at a constant 1% to mimic the effects of outdoor running (Jones & Doust, 1996).

Maximal running assessment

Upon completion of the submaximal test, participants dismounted the treadmill and rested for 5 min. Participants then completed a continuous incremental test to determine $\dot{V}O_{2max}$. The treadmill speed was set to their speed at LTP (sLTP) and gradient initially set to 1%. At the end of each minute the gradient increased by 1% until volitional exhaustion was reached (6.6 ± 0.8 min).

Measurements

Anthropometry

At the start of each trial participants body mass was measured digitally to the nearest 0.1 kg (MPMS-230, Marsden Weighing Group, Oxfordshire, UK), and stature with a stadiometer to the nearest 1 cm (SECA GmbH & Co., Hamburg, Germany). Body fat percentage was estimated via sum of four skinfolds (Durnin & Womersley, 1974).

sFBLC and heart rate

A 20 µl sample of blood was drawn from the earlobe at the end of each stage during the submaximal running assessment. The blood sample was then hemolysed in a pre-filled micro test tube and analysed for blood lactate (Biosen C-Line, EKF Diagnostic, Ebendorfer Chaussee 3, Germany). The analyser was calibrated before all trials with

a known concentration of blood lactate and in accordance with the manufacturer's instructions.

Heart rate was measured continuously throughout the tests (Polar RS400, Polar Electro Oy, Finland). Data was averaged for the final one minute of each stage and used in subsequent analysis. Heart rate and sFBLC were determined from the speed-lactate curve for 2, 3 and 4 $mmol^.L^{-1}$ using a published software (Newell et al., 2007).

Running economy

Oxygen uptake, $\dot{V}CO_2$ and RER values were obtained by averaging the final 60 s of each submaximal stage and values for sLTP and the two speeds prior (sLTP-1 $km^.h^{-1}$, sLTP-2 $km^.h^{-1}$), were used in subsequent analysis. A 60 s collection period was deemed the longest duration where participants were operating at a steady-state during each stage. To verify whether a steady-state had been achieved during the final minute of each submaximal stage, the difference between the first 30 s of the final minute and the last 30 s was calculated. A difference smaller than the minimal detectable change (MDC), calculated as standard error of the mean x 1.96 x $\sqrt{2}$, confirmed a plateau had been achieved. Energy cost of running was estimated from updated non-protein quotient equations (Péronnet & Massicotte, 1991) and the RER values. These values were then added and multiplied by 4.182 to determine total energy cost.

Maximal measures

A participants $\dot{V}O_{2max}$ was defined as the highest $\dot{V}O_2$ achieved in a 30 s period on the maximal test. Verification that a plateau in $\dot{V}O_2$ had been achieved was identified using

the procedure described by Midgley, Carroll, Marchant, McNaughton, & Siegler (2009). Briefly, a least squared linear regression line was obtained for the $\dot{V}O_2$ data for the period +2 min after commencement of the test to -2 min prior to exhaustion. This period was selected to ensure only the linear portion of the $\dot{V}O_2$ response was captured, avoiding any non-linearity caused by the $\dot{V}O_2$ kinetic response during early stages of exercise and the plateau associated with late stages of a test to exhaustion. $\dot{V}O_{2max}$ was then predicted from this relationship and a plateau was confirmed if the difference between the predicted and actual $\dot{V}O_{2max}$ values was greater than 50% of the regression gradient (Midgley et al., 2009). Speed at $\dot{V}O_{2max}$ was calculated by extrapolating the $\dot{V}O_2$-speed relationship from the sub-maximal running assessment via linear regression.

Allometric scaling

To account for the confounding influence of body size, power laws have been suggested based upon the mathematical principles of allometry (Welsman & Armstrong, 1996). There is however little agreement as to the most appropriate scaling exponent to employ for homogenous groups of individuals such as high-performing adolescent athletes. To calculate appropriate exponents for variables requiring expression relative to body size in the current cohort, body mass data from the larger group was first linearized via natural log transformation and homogeneity of regression was compared for males and females for each dependent variable via an ANCOVA. Results of the ANCOVA tests revealed that the slopes of the log/log transformations did not differ for males and females. A common power function was therefore calculated via linear regression on the logarithmic transformation of each

data set by the formula $\ln_y = \ln_a + b.\ln_x$ and used for both groups when scaling each variable. Antilogarithms of the adjusted means from the ANCOVA were divided by the antilogarithm of the mean body mass for all participants included in the analysis raised to the allometric exponent to obtain the mean power function ratio standard for males and females. To assess the extent of a residual size correlation, linear regression analysis was performed between body mass values and each physiological variable normalized by the scaling exponent. For exponents to be accurate, the R^2 value should approach zero, indicating that any differences observed in the physiological variable are independent of body mass.

Statistical analyses

For both the large cohort (n=42) and test-retest cohort (n=16) the normality of the distributions was assessed using a Shapiro-Wilks test and visually inspected using Q-Q plots. The homogeneity of the variance was assessed using Levene's test. All variables were found to be normally distributed (*p*>0.05) and satisfied the assumption of homoscedasticity.

To determine whether a systematic bias was present between trial 1 and trial 2, a two-factor (sex x trial) ANOVA with repeated measures was performed. Effect sizes (ES) were also calculated as the change in the mean scores between trials divided by the pooled standard deviation (SD) from both trials. Values were interpreted using modified Cohen (1988) thresholds as: trivial <0.2; small 0.2-0.6; moderate 0.6-1.2; and large >1.2 (Batterham and Hopkins, 2006). The TE value provides an absolute index of reliability that encapsulates both the random and systematic error associated with a measurement (Batterham and George, 2003), and was calculated as the SD of

the difference between trial 1 and trial 2 divided by $\sqrt{2}$.. Two-way random (single measure) ICC were also calculated as an indicator of the relative consistency for each measure (Weir, 2005) including a 95% confidence interval (CI).

MDC can be used for practical interpretation of the change required in measurements to have 95% certainty that real change has occurred. MDC confidence intervals (MDC$_{95}$) were calculated for each variable as TE x 1.96 x $\sqrt{2}$. The ES, TE and MDC were calculated using Microsoft Excel 2013 and all other data were processed with IBM SPSS Statistics (version 22). Data are displayed as mean ± SD and significance was accepted at $p<0.05$.

Results

Plateaus in expired gases

The MDC values for $\dot{V}O_2$ in the final 60 s of each submaximal stage across both trials were <122.2 mL·kg^{-1}. Values for $\dot{V}CO_2$ were of a similar magnitude (97.8-112.9 ml·kg^{-1}). Of the 174 samples analyzed for the larger cohort (n=42), 11 $\dot{V}O_2$ samples and ten $\dot{V}CO_2$ samples failed to meet steady-state criteria.

All participants with the exception of three (two from the test-retest cohort and one from the larger sample) achieved the criteria for a plateau in $\dot{V}O_2$ at maximum. For the participants who achieved a plateau, at least two of the three traditional criteria for achievement of $\dot{V}O_{2max}$ were also met (RER ≥1.1, HR ≥95% of age-predicted maximum, end lactate ≥8 mmol·L^{-1}). The samples that failed to demonstrate a plateau in gas exchange (for $\dot{V}O_2$, $\dot{V}CO_2$ or $\dot{V}O_{2max}$) were excluded from subsequent analysis. Revised participant numbers used to generate scaling exponents and reliability statistics are shown within each table.

Allometric scaling

The characteristics of the participants used to calculate scaling exponent for whole body mass are shown in Table 2a and 2b. No significant differences were found between scaling exponents for males and females across $\dot{V}O_2$ parameters. Log-linear regression revealed scaling exponents which approximated two-thirds for males and females. When this scaling exponent was applied, a very weak relationship was present between body mass and $\dot{V}O_{2max}$, or $\dot{V}O_2$ ($R^2 < 0.17$, $p > 0.05$).

****Insert Table 2a and Table 2b near here****

Reliability measures

Body mass displayed high consistency between trials (mean difference: 0.4±0.3 kg, TE: 0.34%). Although sFBLC demonstrated high reliability between trials (Table 3), variability of this measure across all speeds was far higher (mean difference: 0.18±0.20 mmol·L^{-1}, 95% CI: 0.14-0.22 mmol·L^{-1} TE: 6.24%). Heart rate at FBLC was also highly reliable (Table 3) and displayed similar consistency across all speeds for each participant (mean difference: 4±4 beats·min^{-1}, 95% CI: 3.7-5.2 beats·min^{-1}, TE: 1.60%).

Results of ANOVA tests revealed no systematic bias between trial 1 and trial 2 across any parameter ($p > 0.05$). Similarly, ES for test-retest differences across all measures were small (<0.6) or trivial (<0.2). Within-subject variability for $\dot{V}CO_2$ was low across assessed speeds (TE: 1.16-1.37%). As shown in Table 3, parameters relying

14

on measurement of $\dot{V}O_2$ and $\dot{V}CO_2$ all demonstrated a high degree of reliability between trials (ICC: 0.82-0.98, TE: 1-2%), however it is notable that measures of energy cost produced lower TE values than oxygen cost at the same speeds. The MDC values are also shown in Table 3. When expressed as a percentage, MDC for energy cost was lower than oxygen cost for each speed.

****Insert Table 3 near here***

Discussion

This investigation aimed to establish the reproducibility of a number of physiological variables associated with distance running performance in an elite group of junior athletes. The results provide evidence that physiological parameters can be measured with a high-degree of reliability in this population, and we are the first to present reliability data in young athletes for $\dot{V}O_{2max}$ and running economy following use of an appropriate allometric scaling exponent. In the assessment of running economy, energy cost appears to be more reliable than using oxygen cost values, so should be the preferred measure for the practitioner.

Regardless of whether $\dot{V}O_{2max}$ or $\dot{V}O_{2peak}$ has been used to define the highest $\dot{V}O_2$ achieved, previous studies have found a high test-retest reliability (ICC=0.74-0.97) in adolescent populations (Paterson et al., 1981; Pivarnik et al., 1996; Rivera-Brown & Frontera, 1998; Rivera-Brown et al., 1995), which is in agreement with findings from this study (ICC=0.98). This is the first study in young athletes to assess

reliability of $\dot{V}O_{2max}$ following appropriate scaling for differences in the body mass within the cohort. We also applied a more stringent criteria (Midgley et al., 2009) to identify $\dot{V}O_{2max}$, than the criterion proposed by Taylor, Buskirk, and Henschel (1955) that has been applied in previous studies, and excluded any participants from analysis who failed to achieve a true plateau. These two factors are likely to explain, in-part, the high ICC value observed in this study.

Oxygen cost is an important determinant of distance running performance and reliability indices in the present study (ICC: 0.88-0.90, TE: 1.81-2.00%) are similar to those reported elsewhere for well-trained (1.3%), highly trained (1.8%) and elite (2.4%) endurance runners (Morgan et al., 1991; Pereira and Freedson, 1997; Saunders et al., 2004). The between-day stability of oxygen cost has previously been quantified in six year old children (Keefer et al., 2000), however this is the first study to show high-levels of between-test reliability in adolescent runners who are engaged in intensive training regimens. Training status appears to be an important influence in the degree of variability observed in running economy (Brisswalter and Legros, 1994). The mean age of the participants in this study was <18 years, however they were all of national or international standard in their age-groups and ratio-scaled $\dot{V}O_{2max}$ scores were similar to those attained by highly-trained adult runners (male: 73.3±4.2, female: 64.3±5.8 $ml{\cdot}kg^{-1}{\cdot}min^{-1}$). Therefore despite the young age of the participants, results suggest a high level of stability in oxygen cost of running exists, which in-part may be due to the training status of these athletes.

The measure of energy cost provides a potentially more robust measure of running economy, mitigating against the potentially confounding influence of substrate utilization when dietary and lifestyle factors are not controlled adequately. Similar between-test reliability has previously been shown for oxygen cost (TE: 2.7-

3.3%) and energy cost (TE: 3.1-3.7%) in trained runners (Shaw et al., 2013). This is somewhat different to the findings of this study, which showed energy cost to be more reproducible than oxygen cost across all three speeds assessed. Reliable measurement of energy cost requires $\dot{V}O_2$ and $\dot{V}CO_2$ to be consistent, as RER is derived using both values. It is likely that differences in within-subject variability for $\dot{V}CO_2$ explain the discrepancy between this study and those of Shaw and colleagues (2013), who reported much higher TE (<5.94%) than the variability we measured (1.16-1.37%). As a consequence, this meant that the RER was also more stable in the present study (TE: <4.35% vs <2.02%). Although the male participants in both studies possessed similar $\dot{V}O_{2max}$ values (75.5±5.2 vs 73.3±4.2 ml·kg^{-1}·min^{-1}), it may also be possible that the juniors runners in this investigation employed less intensive training regimens than the senior runners used in the Shaw et al. (2013) study, therefore we speculate that this would reduce the daily variability in substrate metabolism and thus measures of energy cost.

Speed at $\dot{V}O_{2max}$, which combines both running economy and $\dot{V}O_{2max}$ into a single variable, has been shown to explain differences in performance that other determinants cannot (Billat and Koralsztein, 1996). Previous work has shown low TE values (2.3%) for amateur male runners with near identical s$\dot{V}O_{2max}$ values (18.6 km·h^{-1}) to those observed in the present study (Loureno, Martins, Tessutti, Brenzikofer, & Macedo, 2011). Given that s$\dot{V}O_{2max}$ was estimated from the linear regression equation based upon $\dot{V}O_{2max}$ the speed-$\dot{V}O_2$ relationship, which has high reliability at three speeds, it is unsurprising that this parameter demonstrate high reproducibility using this protocol (ICC: 0.82, TE: 1.83%).

Blood lactate measurement during sub-maximal exercise in athletes is commonplace for monitoring and evaluation purposes. Previous studies have shown

reliability coefficients of 0.92-0.95 for velocity at 4 mmol·L^{-1} and 0.81-0.93 for HR at 4 mmol·L^{-1} in moderately fit (Grant et al., 2002) and trained runners (Heitkamp, Holdt, & Scheib, 1991; Weltman et al., 1990). It should be noted that although the correlation coefficients used in these studies describe the relationship between test and retest values, this statistic does not account for systematic bias. The ICC is a univariate statistic and provides a more robust means of determining agreement between two independent measurements. The reliability values for speed at 2-4 mmol·L^{-1} in the present study are also similar to those observed in a previous study on endurance-trained males, which used three trials and an ICC statistic (Pfitzinger and Freedson, 1998). The present study did however observe a far larger variability for blood lactate measurement across absolute speeds (mean difference: 0.18±0.20 mmol·L^{-1}, 95% CI: 0.14-0.22 mmol·L^{-1}, TE: 6.24%), therefore when interpreting within-subject changes in lactate response to sub-maximal running, sFBLC should be the preferred marker for assessing change.

Monitoring of submaximal HR provides athletes and practitioners with a low cost and non-invasive tool to assess changes in training status and may provide an indication of overtraining (Achten and Jeukendrup, 2003; Lambert et al., 1998). Under controlled conditions, HR has shown daily variability of 5-8 beats·min^{-1} at sub-maximal running speeds in physically active adults (Lamberts, Lemmink, Durandt, & Lambert, 2004). The variability across all speeds in the present study is slightly lower (mean difference: 4±4 beats·min^{-1}, 95% CI: 3.7-5.2 beats·min^{-1}, TE: 1.60%). However the data supports the recommendation that caution should be observed when interpreting subtle changes in HR at absolute running speeds, as a difference of <6 beats·min^{-1} can be attributed to normal variability (Achten and Jeukendrup, 2003). The reproducibility of HR at FBLC (ICC: 0.86-0.92, TE: 0.96-1.55%) was higher than

values found in other studies using recreational runners (ICC: 0.47-0.79, TE: 1.6-4.3%), which have used HR at intensities corresponding to fixed physiological thresholds (Lourenco et al., 2011; Peserico et al., 2015). Despite the age of the participants used in our study, this discrepancy is likely due to the training status of the athletes, as similar dietary and lifestyle constraints were applied, and previous work has shown lesser-trained runners show larger daily variability in HR scores (Heitkamp et al., 1991) compared to well-trained runners (Brisswalter and Legros, 1994).

Time of day, footwear and environmental conditions were controlled for in the present study, but dietary and training constraints were not strictly enforced. In an attempt to control for these potentially confounding issues, most participants trials were separated by seven days. From personal communication with parents (and participants), this period represented a typical microcycle of training for these runners. It is therefore likely that a highly similar pattern of training and lifestyle activities preceded each trial. Requests for consistency in lifestyle and diet in the 48 h prior to trials were also enforced by parents/guardians who were present for most of the testing sessions. Familiarity with the speeds used to assess running economy and adequate control of confounding factors are therefore likely to have contributed towards the low within-subject variability observed for measures in this study.

The MDC_{95} uses the TE associated with measurement consistency and the z-score from a 95% confidence interval. The MDC values for each variable (see Table 3) can be utilized by exercise physiologists to determine whether a 'real' change has been observed following an intervention in high-performing junior distance runners.

This study isn't without limitations. The relatively small sample size (n=42) used to quantify a suitable scaling exponent for this population may have generated

imprecise estimates owing to sampling error. A recent meta-analysis on allometric scaling of $\dot{V}O_{2max}$ established exponents of 0.71 for young individuals, 0.71 for athletes and 0.70 for treadmill testing assessment (Lolli, Batterham, Weston, & Atkinson, 2017) . These values are similar to the estimated scaling exponent calculated in the present study (0.66) and fall within the 95% confidence intervals for each moderator (Lolli et al., 2017), therefore any error which exists is likely to be minimal. The test-retest design of the present study provides practitioners with valuable insight into the reproducibility of physiological measures when familiarization sessions are not possible due to time constraints. However the inclusion of a third trial would provide a more accurate impression of the systematic error which may exist in each measure.

Practitioners and coaches can be confident that measurements of physiological parameters in elite junior distance runners are highly reproducible when external factors are appropriately constrained. Energy cost should be the preferred measure of running economy as this parameter accounts for day-to-day variations in substrate utilization and demonstrates higher reliability than traditional oxygen cost measurement. Caution should be observed when interpreting small changes in blood lactate (<0.2 mmol$^.$L^{-1}) and HR (<6 beats$^.$min^{-1}) as part of an athlete monitoring process as it is likely this is normal variability in the criterion measure. The sFBLC appears to provide a more sensitive metric for reliably identifying a change in an athlete's physiology compared to the lactate value at a given speed.

References

Achten, J., & Jeukendrup, A. (2003). Heart rate monitoring: applications and limitations. *Sports Medicine, 33,* 517-538. doi:10.2165/00007256-200333070-00004

Almarwaey, O. A., Jones, A. M., & Tolfrey, K. (2003). Physiological correlates with endurance running performance in trained adolescents. *Medicine and Science in Sports and Exercise, 35,* 480-487. doi:10.1249/01.MSS.0000053723.16531.D0

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26,* 217-238. doi:10.2165/00007256-199826040-00002

Batterham, A. M., & George, K. P. (2003). Reliability in evidence-based clinical practice: a primer for allied health professionals. *Physical Therapy in Sport, 4,* 122-128. doi:10.1054/ptsp.2000.0010

Batterham, A. M., & Hopkins, W. G. (2006). Making meaningful inferences about magnitudes', *International Journal of Sports Physiology and Performance*, *1,* 50-57. doi:10.1123/ijspp.1.1.50

Beunen, G., & Malina, R. M. (1988). Growth and physical performance relative to the timing of the adolescent growth spurt. *Exercise and Sport Science Reviews, 16,* 503-540. PMID:3292266

Billat, L. V., & Koralsztein, J. P. (1996). Significance of the velocity at VO2max and time to exhaustion at this velocity. *Sports Medicine, 22,* 90-108. doi:10.2165/00007256-199622020-00004

Brisswalter, J., & Legros, P. (1994). Daily stability in energy cost of running, respiratory parameters and stride rate among well-trained middle distance runners. *International Journal of Sports Medicine, 15*, 238-241. doi:10.1055/s-2007-1021053

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Durnin, J. V. G. A., & Womersley, J. (1974). Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. *British Journal of Nutrition, 32,* 77-97. doi: 10.1079/BJN19740060

Eisenmann, J. C., Pivarnik, J. M., & Malina, R. M. (2001). Scaling peak VO2 to body mass in young male and female distance runners. *Journal of Applied Physiology, 90,* 2172-2180. Retrived from http://jap.physiology.org/content/90/6/2172

Fletcher, J. R., Esau, S. P., & MacIntosh, B. R. (2009). Economy of running: beyond the measurement of oxygen uptake. *Journal of Applied Physiology, 107,* 1918–1922. doi:10.1152/japplphysiol.00307.2009

Grant, S., McMillan, K., Newell, J., Wood, L., Keatley, S., Simposn, D., Leslie, K., & Fairlie-Clark, S. (2002). Reproducibility of the blood lactate threshold, 4 mmol.l^{-1} marker, heart rate and ratings of perceived exertion during incremental treadmill exercise in humans. *European Journal of Applied Physiology, 87,* 159-166. doi:10.1007/s00421-002-0608-2

Halson, S. (2014). Monitoring training load to understand fatigue in athletes. *Sports Medicine, 44,* 139-147. doi:10.1007/s40279-014-0253-z

Heitkamp, H.-Ch., Holdt, M., & Scheib, K. (1991). The reproducibility of the 4mmol/l lactate threshold in trained and untrained women. *International Journal of Sports Medicine, 12,* 363-368. doi:10.1055/s-2007-1024695

Jones, A. M., & Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *Journal of Sports Sciences, 14*, 321-327. doi: 10.1080/02640419608727717

Keefer, D. J., Tseh, W., Caputo, J. L., Craig, I. S., Martin, P. E., & Morgan, D. W. (2000). Stability of running economy in young children. *International Journal of Sports Medicine, 21,* 583-585. doi:10.1055/s-2000-8486

Lamarra, N., Whipp, B. J., Ward, S. A., & Wasserman K. (1987). Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. *Journal of Applied Physiology, 62,* 2003-2012. Retrieved from http://jap.physiology.org/content/62/5/2003.short

Lamberts, R. P., Lemmink, K. A. P. M., Durandt, J. J., and Lambert, M. I. (2004). Variation in heart rate during submaximal exercise: Implications for monitoring training. *Journal of Strength and Conditioning Research, 18,* 641-645. doi:10.1519/1533-4287(2004)18<641:VIHRDS>2.0.CO;2

Lambert, M. I., Mbambo, Z. H., & St Clair Gibson, A. (1998). Heart rate during training and competition for long distance running. *Journal of Sports Sciences, 16*(Suppl. 1)*,* 85-90. doi:10.1080/026404198366713

Lolli, L., Batterham, A. M., Weston, K. L., Atkinson, G. (2017). Size exponents for scaling maximal oxygen uptake in over 6500 humans: A systematic review and meta-analysis. *Sports Medicine, 47,* 1405-1419. doi:10.1007/s40279-016-0655-

Lourenco, T. F., Martins, L. E. B., Tessutti, L. S., Brenzikofer, R., & Macedo, D. V. (2011). Reproducibility of an incremental treadmill VO2max test with gas exchange analysis for runners. *Journal of Strength and Conditioning Research, 25,* 1994-1999. doi:10.1519/JSC.0b013e3181e501d6

Meeusen, R., Duclos, M., Foster, C., Fry, A., Gleeson, M., Nieman, D., Raglin, J., Rietjens, G., Steinacker, J., & Urhausen, A. (2013). Prevention, diagnosis and treatment of the overtraining syndrome: Joint consensus statement of the European College of Sport Science (ECSS) and the American College of Sports Medicine (ACSM). *European Journal of Sport Science, 13,* 1-24. doi:10.1080/17461391.2012.730061

Midgley, A. W., Carroll, S., Marchant, D., McNaughton, L. R., & Siegler, J. (2009). Evaluation of true maximal oxygen uptake based on a novel set of standardized criteria. *Applied Physiology, Nutrition and Metabolism, 34,* 115-123. doi:10.1139/H08-146

Midgley, A. W., McNaughton, L. R., Polman, R., & Marchant, D. (2007). Criteria for determination of maximal oxygen uptake. *Sports Medicine, 37,* 1019-1028. doi:10.2165/00007256-200737120-00002

Morgan, D. W., Martin, P. E., Krahenbuhl, G. S., & Baldini, F. D. (1991). Variability in running economy and mechanics among trained distance runners. *Medicine and Science in Sports and Exercise, 23,* 378-383. Retrieved from http://journals.lww.com/acsm-msse/Abstract/1991/03000/

Newell, J., Higgins, D., Madden, N., Cruickshank, J., Einbeck, J., McMillan, K., & McDonald, R. (2007). Software for calculating blood lactate endurance markers. *Journal of Sports Sciences, 25,* 1403-1409. doi:10.1080/02640410601128922

Paterson, D. H., Cunningham, D. A., & Donner, A. (1981). The effect of different treadmill speeds on the variability of VO2 max in children. *European Journal of Applied Physiology and Occupational Physiology, 47,* 113-122. doi:10.1007/BF00421663

Pereira, M. A., & Freedson, P. S. (1997). Intraindividual variation of running economy in highly trained and moderately trained males. *International Journal of Sports Medicine, 18,* 184-188. doi:10.1055/s-2007-972606

Péronnet, F., & Massicotte, D. (1991). Table of nonprotein respiratory quotient: an update. *Canadian Journal of Sport Science 16,* 23-29. Retrieved from https://www.researchgate.net/profile/Francois_Peronnet2/publication/

Peserico, C. S., Zagatto, A. M., & Machado, F. A. (2015). Reproducibility of heart rate and rating of perceived exertion values obtained from different incremental treadmill tests. *Science & Sports, 30,* 82-88. doi:10.1016/j.scispo.2014.12.002

Pfitzinger, P., & Freedson, P. S. (1998). The reliability of lactate measurements during exercise. *International Journal of Sports Medicine, 19,* 349–357. doi: 10.1055/s-2007-971929

Pivarnik, J. M., Dwyer, M. C., & Lauderdale, M. A. (1996). The reliability of aerobic capacity (VO_{2max}) testing in adolescent girls. *Research Quarterly for Exercise and Sport, 67,* 345-348. doi:10.1080/02701367.1996.10607962

Rivera-Brown, A. M., & Frontera, W. R. (1998). Achievement of plateau and

reliability of VO2max in trained adolescents tested with different protocols.

Pediatric Exercise Science, 10, 164-175. doi:10.1123/pes.10.2.164

Rivera-Brown, A. M., Rivera, M. A., & Frontera, W. R. (1995). Reliability of

VO2max in adolescent runners: A comparison between plateau achievers and

nonachievers. *Pediatric Exercise Science, 7,* 203-210. doi:10.1123/pes.7.2.203

Saunders, P. U., Pyne, D. B., Telford, R. D., & Hawley, J. A. (2004). Reliability and

variability of running economy in elite distance runners. *Medicine and Science in

Sports and Exercise, 36,* 1972-1976. doi:10.1249/01.MSS.0000145468.17329.9F

Shaw, A. J. S., Ingham, S. A., & Folland, J. P. (2014). The valid measurement of

running economy in runners. *Medicine and Science in Sports and Exercise, 46,*

1968-1973. doi:10.1249/MSS.0000000000000311

Shaw, A. J. S., Ingham, S. A., Fudge, B. W., & Folland, J. P. (2013). The reliability

of running economy expressed as oxygen cost and energy cost in trained distance

runners. *Applied Physiology, Nutrition and Metabolism, 38,* 1268-1272.

doi:10.1139/apnm-2013-0055

Sparling, P. B. (1984). Physiological determinants of distance running performance.

The Physician and Sportsmedicine, *12*, 68-77.

doi:http://dx.doi.org/10.1080/00913847.1984.11701795

Taylor, H. L., Buskirk, E., & Henschel, A. (1955). Maximal oxygen intake as an

objective measure of cardio-respiratory performance. *Journal of Applied Physiology,

8,* 73-80. Retrieved from http://jap.physiology.org/content/jap/8/1/73.full.pdf

Welsman, J. R., & Armstrong, N. (1996). The measurement and interpretation of aerobic fitness in children: Current issues. *Journal of the Royal Society of Medicine, 89,* 281P-285P. Retrieved from

http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC1295792&blobtype=pdf

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19,* 231-240.

Weltman, A., Snead, D. B., Stein, P. K., Seip, R. L., Schurrer, R., Rutt, R., & Weltman, J. Y. (1990). Reliability and validity of a continuous incremental treadmill protocol for the determination of lactate threshold, fixed blood lactate concentrations, and VO2max. *International Journal of Sports Medicine, 11,* 26-32. doi:10.1055/s-2007-1024757

Wilson, G., Raglin, J., & Harger, G. (1999). Training practices, mood state and staleness in adolescent distance runners. *Medicine and Science in Sports and Exercise, 31*(Suppl), 216. Retrieved from http://journals.lww.com/acsm-msse/Citation/1999/05001/

Winsley, R., & Matos, N. (2011). Overtraining and elite young athletes. In N. Armstrong & A. M. McManus (Eds.), *The elite young athlete* (pp. 97-105). Basel, Switzerland: Karger Medical and Scientific Publishers.

Winter, E. M., Jones, A. M., Davison, R. C., Bromley, P. D., & Mercer, T. (2006). *BASES sport and exercise physiology testing guidelines: Vol. 2 – exercise and clinical testing*. London: Routledge.

Measure	Test-retest cohort		Scaling cohort	
	M (n=8)	F (n=8)	M (n=20)	F (n=22)
Age (y)	16.3 ± 1.2	17.0 ± 1.5	17.0 ± 1.4	17.1 ± 1.2
Stature (cm)	176.6 ± 3.4	169.5 ± 6.3	176.9 ± 5.7	168.5 ± 6.1
Body mass (kg)	59.8 ± 7.1	51.8 ± 7.0	62.1 ± 6.6	52.7 ± 5.8
Sum of skinfolds (mm)	23.1 ± 4.2	38.6 ± 16.5	24.4 ± 4.2	38.2 ± 13.5
$\dot{V}O_{2max}$ (ml·kg^{-1}·min^{-1})	73.3 ± 4.2	64.3 ± 5.8	69.9 ± 8.3	60.4 ± 6.3
sLTP (km·h^{-1})	16.0 ± 1.1	14.5 ± 0.5	16.0 ± 1.4	14.1 ± 1.2

Table 1. Descriptive characteristics of study participants (data are mean ± standard deviation).

Parameter (y)	n	Age (years)	Mass (x) (kg)	Mean ±SD ($L \cdot min^{-1}$)	$y = ax^b$ ($ml \cdot min^{-1}$)
$\dot{V}O_{2max}$	20	17.0 ± 1.4	62.1 ± 6.6	4.322 ± 0.577	$286.5\, x^{0.66}$
$\dot{V}O_2$					
sLTP	19	17.0 ± 1.4	61.8 ± 6.7	3.726 ± 0.456	$243.4\, x^{0.66}$
sLTP -1 $km \cdot h^{-1}$	20	17.0 ± 1.4	62.1 ± 6.6	3.515 ± 0.411	$247.7\, x^{0.64}$
sLTP -2 $km \cdot h^{-1}$	15	17.1 ± 1.3	61.7 ± 6.7	3.377 ± 0.368	$212.0\, x^{0.67}$

Table 2a. Characteristics of male participants and scaling exponents for sub-maximal $\dot{V}O_2$ and $\dot{V}O_{2max}$

Parameter (y)	n	Age (years)	Mass (x) (kg)	Mean ±SD ($L \cdot min^{-1}$)	$y = ax^b$ ($ml \cdot min^{-1}$)
$\dot{V}O_{2max}$	20	17.2 ± 1.3	53.0 ± 5.9	3.224 ± 0.417	$237.6\,x^{0.66}$
$\dot{V}O_2$					
sLTP	19	17.2 ± 1.2	53.0 ± 5.9	2.803 ± 0.312	$202.9\,x^{0.66}$
sLTP -1 $km^{.}h^{-1}$	21	17.1 ± 1.2	52.4 ± 5.8	2.623 ± 0.274	$206.5\,x^{0.64}$
sLTP -2 $km^{.}h^{-1}$	22	17.1 ± 1.2	52.7 ± 5.8	2.462 ± 0.241	$172.5\,x^{0.67}$

Table 2b. Characteristics of female participants and scaling exponents and scaling exponents for sub-maximal $\dot{V}O_2$ and $\dot{V}O_{2max}$

Parameter	n	Trial 1	Trial 2	ES	Interpretation	TE	TE (%)	ICC (95% CI)	MDC$_{95}$	MDC (%)
Maximal running										
$\dot{V}O_{2max}$ (ml·kg$^{-0.67}$·min^{-1})	14	267.2 ± 32.5	268.2 ± 29.8	0.17	Trivial	2.72	1.02	0.98 (0.94 – 0.99)	7.5	2.8
s$\dot{V}O_{2max}$ (km·h^{-1})	14	18.6 ± 1.5	18.6 ± 1.1	0.49	Small	0.34	1.83	0.82 (0.52 – 0.94)	0.9	5.1
Oxygen cost (ml·kg$^{-0.67}$·km^{-1})										
sLTP	15	889.6 ± 77.0	891.4 ± 71.2	0.39	Small	16.1	1.81	0.88 (0.68 – 0.96)	44.7	5.0
sLTP-1 km·h^{-1}	16	894.5 ± 74.4	886.9 ± 69.2	0.38	Small	18.3	1.45	0.90 (0.75 – 0.96)	35.9	4.0
sLTP-2 km·h^{-1}	14	896.7 ± 85.3	881.3 ± 68.8	0.37	Small	17.8	2.00	0.89 (0.68 – 0.96)	49.2	5.5
Energy cost (kJ·kg$^{-0.67}$·km^{-1})										
sLTP	13	19.3 ± 1.7	19.5 ± 1.5	0.33	Small	0.233	1.20	0.93 (0.79 – 0.98)	0.6	3.3
sLTP-1 km·h^{-1}	16	19.2 ± 1.6	19.0 ± 1.5	0.34	Small	0.267	1.40	0.92 (0.78 – 0.97)	0.7	3.9
sLTP-2 km·h^{-1}	13	19.1 ± 1.8	18.8 ± 1.5	0.28	Small	0.305	1.61	0.93 (0.79 – 0.98)	0.8	4.5
Speed at FBLC (km·h^{-1})										
2 mmol·L^{-1}	16	14.5 ± 1.4	14.6 ± 1.4	0.12	Trivial	0.14	0.94	0.99 (0.97 – 1.00)	0.4	2.7
3 mmol·L^{-1}	16	15.5 ± 1.3	15.6 ± 1.3	0.09	Trivial	0.11	0.71	0.99 (0.97 – 1.00)	0.3	2.0
4 mmol·L^{-1}	13	16.3 ± 1.3	16.4 ± 1.2	0.11	Trivial	0.12	0.76	0.98 (0.93 – 0.99)	0.3	2.0
HR at FBLC (b·min^{-1})										
2 mmol·L^{-1}	16	179 ± 11	178 ± 11	0.41	Small	2.77	1.55	0.86 (0.65 – 0.95)	8	4.3
3 mmol·L^{-1}	16	186 ± 11	185 ± 11	0.34	Small	1.87	1.01	0.92 (0.78 – 0.97)	5	2.8
4 mmol·L^{-1}	13	189 ± 9	187 ± 10	0.37	Small	1.81	0.96	0.90 (0.70 – 0.97)	5	2.7

Table 3.   Test-retest reliability of physiological variables during maximal and submaximal test.