Journal
Journal of Sports Sciences ›
Volume 26, 2008 - Issue 13

Research article

Repeatability of scores on a novel test of endurance running performance

Ian Rollo, Clyde Williams & Alan Nevill

Page 1379-1386 | Accepted 17 Jun 2008, Published online: 31 Oct 2008

Download citation http://dx.doi.org/10.1080/02640410802277452

Abstract

The aim of the present study was to determine the repeatability of a running endurance test using an automated treadmill system that requires no manual input to control running speed. On three separate occasions, 7 days apart, 10 experienced male endurance-trained runners (mean age 32 years, *s* = 10; *[Vdot]*O_{2peak} 61 ml · kg^{-1} · min^{-1}, *s* = 7) completed a treadmill time trial, in which they were instructed to run as far as possible in 60 min. The treadmill was instrumented with an ultrasonic feedback-controlled radar modulator that spontaneously regulated treadmill belt speed corresponding to the changing running speed of each runner. Estimated running intensity was 70%*[Vdot]*O_{2peak} (*s* = 11) and the distance covered 13.5 km (*s* = 2), with no difference in mean performances between trials. The coefficient of variation, estimated using analysis of variance, with participant and trial as main effects, was 1.4%. In summary, the use of an automated treadmill system improved the repeatability of a 60-min treadmill time trial compared with time trials in which speed is controlled manually. The present protocol is a reliable method of assessing endurance performance in endurance-trained runners.

Keywords: Treadmill, repeatability, performance,

Introduction

A review of the studies on the influences of, for example, nutritional interventions on exercise performance, reveals that the term "exercise performance" has many interpretations. A wide range of exercise tests has been reported in the literature under the heading of "exercise performance". Nevertheless, most of these laboratory tests of exercise performance fall into two simple categories – namely, those that assess endurance capacity and those that assess endurance performance. Laboratory tests that use constant-pace exercise to fatigue, as in cycling (Maughan, Fenn, & Leiper, 1989) and treadmill running (Brewer, Williams, & Patton, 1988), assess endurance capacity. In contrast, tests that require the completion of a pre-set amount of external mechanical work (cycling: Widrick *et al*., 1993) or distance (cycling or running: Chryssanthopoulos, Williams, Wilson, Asher, & Hearne, 1994; Jeukendrup, Brouns, Wagenmakers, & Saris, 1997) in as fast a time as possible, or ask athletes to complete as much work as possible in a specified time (Schabort, Hopkins, & Hawley, 1998b), assess endurance performance.

The most common method of replicating running performance under laboratory conditions involves the use of a motorized treadmill that allows the manipulation of speed and/or gradient to control running intensity. For example, researchers have successfully used treadmill running to investigate the physiological responses to 800 m, 1500 m (Sandals, Wood, Draper, & James, 2006), endurance capacity (time to fatigue) (Tsintzas *et al*., 2003; Wee, Williams, Gray, & Horabin, 1999), 30-km time trials (Williams, Brewer, & Walker, 1992), half (Williams & Nute, 1983) and even full marathon distances (Tsintzas, Williams, Singh, Wilson, & Burrin, 1995).

Atkinson and Nevill (1998) define the term "reliability" (or for the purpose of this paper "repeatability") as the consistency of measurements. Alternatively, others have defined reliability as the "absence of measurement error" (Safrit & Wood, 1989), although Atkinson and Nevill (1998) recognize that some error will always be present in continuous measurements.

Endurance performance in cycling has been reported to have a smaller measurement error than the much used time-to-fatigue protocol (Hickey, Costill, McConnell, Widrick, & Tanaka, 1992; Jeukendrup, Saris, Brouns, & Kester, 1996; Schabort, Hawley, Hopkins, Mujika, & Noakes, 1998a; Schabort*et al.*, 1998b). For example, time trials that require cyclists to complete a set amount of external mechanical work as quickly as possible (~1600, 200, and 14 kJ) have been reported to have a coefficient of variation (CV) of approximately 1% (Hickey *et al.*, 1992; Schabort *et al.*, 1998a). Jeukendrup *et al*. (1996) reported time trial protocols to be more reliable for performance evaluation (CV: 3.35–3.49%) in direct comparison with constant load cycle tests (CV: 26.6%).

The repeatability of endurance performance (~1 h duration) in treadmill running has thus far failed to produce coefficients of variation that are equivalent to the most reliable cycling tests (Hickey *et al.*, 1992; Schabort *et al.*, 1998b). Schabort *et al*. (1998b) reported a coefficient of variation of 2.7% when runners were asked to run as far as possible in 60 min. A similar value of 2% was reported by Whitham and McKinney (2007) when runners were asked to run as far as possible in 45 min after an initial run of 15 min at 65% maximal oxygen uptake (*[Vdot]*O_{2max}). Despite the appearance of these low coefficients of variation, Hopkins and Hewson (2001) state that running tests need require a coefficient of variation of 2.5% or less to detect worthwhile differences in performance for half and full marathons, and one of 1.5% or less for races over shorter distances.

Laursen and colleagues (Laursen, Francis, Abbiss, Newton, & Nosaka, 2007) have recently confirmed that performances during treadmill time trials have a greater repeatability than time-to-fatigue running tests. However, the limiting factor in treadmill running methods, identified by Laursen *et al*. (2007) and other treadmill time trial studies (Hickey *et al.*, 1992; Schabort*et al.*, 1998b; Whitham & McKinney, 2007), is the inability of runners to spontaneously alter running speed. Instead, runners must manually press buttons on the treadmill console to change their running speed. This is a comparatively blunt response compared with cycling, where changes in power output can be achieved simply by altering pedal cadence. In an attempt to overcome this limitation, Minetti and colleagues (Minetti, Boldrini, Brusamolin, Zamparo, & McKee, 2003) described an automated treadmill system that allows runners to rapidly and spontaneously alter treadmill speed, removing any need to manually alter the speed.

The aim of this study, therefore, was to attempt to improve the methods used for treadmill time trials by using an automated treadmill system that allows runners to rapidly change their running speed with no manual input. To this end, the study examined the repeatability of a 60-min time trial.

Methods

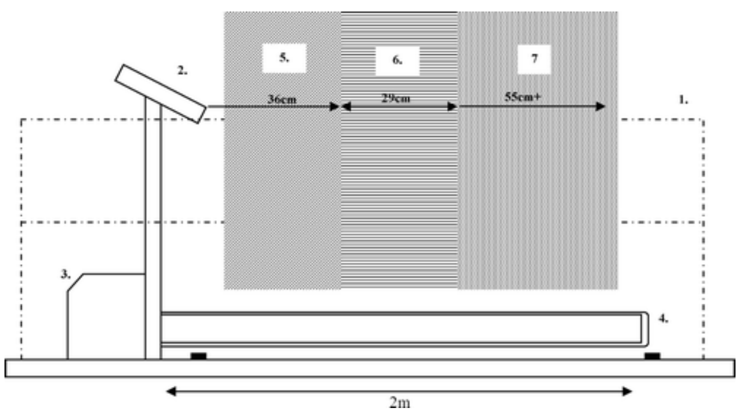
Participants

Ten endurance-trained male athletes (age 32 years, *s* = 10; body mass 72.0 kg, *s* = 6.0; stature 1.78 m, *s* = 0.07; *[Vdot]*O_{2peak} 61.0 ml · kg^{-1} · min^{-1}, *s* = 7) gave their written consent before participating in the study, which was approved by Loughborough University Ethical Advisory Committee. All participants were experienced runners accustomed to training and/or competitions lasting at least one hour. All runners had completed either a half or full marathon distance (36-h ultra-marathon, *n* = 1) within the last year.

Treadmill

All tests were carried out on a motorized treadmill (Runner MT2000, Bianchini and Draghetti, Cavezzo, Italy). The treadmill used in this study had an ultrasonic feedback-controlled radar modulator that spontaneously regulated treadmill belt speed corresponding with the changing position of the runner on the treadmill belt (Minetti *et al.,* 2003). Thus the treadmill speed increased or decreased as the runner moved to the front or the back of the treadmill belt respectively. Changes in speed were therefore achieved without the need for manual input or visual feedback to the participant. More specifically, when the runner moved to the front section of the treadmill (<36 cm from treadmill console), the speed increased (0.8 m · s^{-1}). If the runner remained in the middle (between 36 and 65 cm from the treadmill console), the treadmill speed remained constant. When the runner moved to the rear of the treadmill (>65 cm from the treadmill console), the speed decreased (1.1 m · s^{-1}). Consequently, the runner was always brought back to the centre of the treadmill belt (Figure 1).

Figure 1. Schematic representation of the treadmill (not to proportion for clarity). 1 = support bars; 2 = console; 3 = motor; 4 = 2-m treadmill bed; 5 = acceleration; 6 = constant speed; 7 = deceleration.



PowerPoint slide Original jpg (281.00KB) Display full size

Preliminary tests

After an overnight fast, runners reported to the laboratory and completed a 20-min sub-maximal exercise test to determine oxygen cost and blood lactate concentrations at sub-maximal speeds. Treadmill speed was increased every 4 min, heart rate was recorded at 15-s intervals using short-range telemetry

(Polar Electro, Kempele, Finland), and expired air was collected in the last 60 s of each of the 4-min stages. Expired air was analysed using the Douglas bag method (Williams, Nute, Broadbank & Vinall, 1990) and the coefficient of variation for measures of oxygen uptake at each of the five stages was 6.7%. Fingertip blood samples (20 µl) were taken in duplicate immediately after the expired air collection, deproteinized, frozen, and later analysed for the concentrations of lactate (Maughan, 1982). Following adequate rest, the runners then performed an incremental test to fatigue to determine peak oxygen uptake (*[Vdot]*O_{2peak}) and maximum heart rate. The treadmill speed was kept constant and from an initial gradient of 3%, the gradient was increased by 3% every 3 min until the runner achieved volitional fatigue (Taylor, Buskirk, & Henschel, 1955). The expired air and rating of perceived exertion (Borg, 1982) were collected at the end of each 3-min stage. We have found this method of determining *[Vdot]*O_{2peak} to have a coefficient of variation of approximately 5%.

One-hour run protocol

Runners were fully habituated with the test procedures before the completion of three 1-h running trials. The runners were asked to refrain from heavy exercise and to consume their normal diet in the 48 h before each trial. No caffeine or alcohol was consumed during this period. Each trial was conducted at the same time of day and the trials were separated by 7 days.

Runners arrived at the laboratory following an overnight fast, and emptied their bladder before body mass was recorded. All trials were conducted in a laboratory (temperature 20°C, s = 1; relative humidity 55%) containing only the treadmill and a fan positioned 1 m in front of the runner to provide cooling. Runners were monitored throughout exercise via closed circuit television by an investigator in an adjacent room.

Before each trial, runners were weighed and fitted with a heart rate monitor (Polar Electro, Kempele, Finland) before completing a 5-min warm-up at 60%*[Vdot]*O_{2peak}. During the 5-min warm-up, expired air was collected between 4 and 5 min and then analysed using the Douglas bag method. Ratings of perceived exertion (RPE) were taken at 3 min. On completion of the warm-up, runners were allowed 2 min to prepare for the run and empty their bladder again if required (on these occasions, urine was collected and accounted for in weight loss calculations).

The treadmill display panel and the heart rate monitor were covered so that feedback to the runner was limited to a clock displaying the time remaining throughout the 60-min run. Runners began the trial by standing at the front of the treadmill (1% gradient) and received the following instruction from the same investigator: "This is a running performance test, run as far as you can in 60 minutes".

Runners drank water *ad libitum* during their first trial, the quantity of which was recorded and provided for all subsequent trials. No expired air or blood samples were collected during the 1-h trials. Runners did not receive any feedback from their 1-h run performance until the end of the study.

The runners' stride length was determined by dividing the number of strides by the distance covered between 9–10, 19–20, 29–30, 39–40, and 49–50 min. The relative intensity at which each trial was performed was determined by extrapolating heart rate and oxygen uptake from the preliminary sub-maximal running tests.

Statistics

The agreement (repeatability) between the three 1-h run trials was examined using a repeated-measures analysis of variance (ANOVA) (see Nevill & Atkinson, 1998). The ANOVA estimates the main effects of trial bias and participants and also provides a within-participant measurement error (s_w^2), from which we can estimate the standard deviation of differences between two trial measurements (s) as follows: $s = \sqrt{2 \times s_w^2}$. Provided the residual errors are normally distributed and are not related to the size or level of the measurements, some authors recommend reporting this error as the "95% limits of agreement", defined as ±1.96 × s (Bland & Altman, 1986).

The presence or absence of a relationship between residual errors and the size of measurement can be assessed by plotting absolute residual errors against the predicted measurements. If evidence of heteroscedasticity (evidence of a greater error variation with larger measurements) is detected, a log transformation can be performed to overcome such an effect. Under such circumstances, the analysis described above should be re-applied to the log-transformed measurements. By taking anti-logs of the resulting errors (*s*), we obtain a dimensionless ratio that indicates the measure of unexplained variation, not dissimilar to the concept of a coefficient of variation. From this error ratio we can obtain what Nevill and Atkinson (1997) describe as the 95% ratio limits of agreement, which should contain 95% of the observed ratios (obtained by dividing one trial measurement by a second). All data are reported as means and standard deviations (*s*).

Results

Mean 5-min warm-up speed was 11 km · h^{-1} (s = 2), mean *[Vdot]*O_2 was 36 ml · kg^{-1} · min^{-1} (s = 4) for Trials 1 and 2 and 37 ml · kg^{-1} · min^{-1} (s = 5) Trial 3, equivalent to 59, 59, and 60% *[Vdot]*O_{2peak} for Trials 1–3 respectively. Rating of perceived exertion for the warm-up was 10 (s = 2) and was consistent for the three trials.

Each 60-min time trial was run at an intensity of 70% *[Vdot]*O_{2peak} (s = 11), heart rate was 156 beats · min^{-1} (s = 14), and the runners lost 1.0% (s = 3) of their body mass over the 60 min. Total distance covered by each runner during each 60-min trial, together with the mean distance covered for the three trials, can be seen in Table I.

Table I. Total distance covered (m) by the ten runners during each of the three trials, together with the mean distance covered for the three trials

CSV Display Table

Table 1 of 1

Table I. Total distance covered (m) by the ten runners during each of the three trials, together with the mean distance covered for the three trials

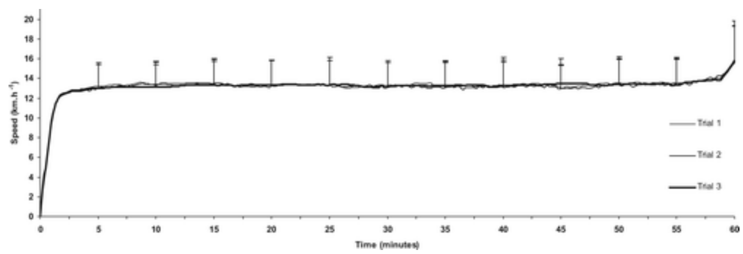
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Mean (m)	*s*
Trial 1 (m)	15,202	15,178	9895	15,875	12,138	14,427	15,583	13,928	9419	10,593	13,224	2493
Trial 2 (m)	15,308	14,791	9825	15,887	12,235	14,228	15,354	14,148	9553	10,453	13,178	2449
Trial 3 (m)	14,866	14,898	9818	15,868	12,329	14,414	15,504	13,944	9606	10,668	13,192	2395
Mean	15,125	14,956	9846	15,877	12,234	14,356	15,480	14,007	9526	10,571	13,198	2446
s	231	200	43	9	96	111	116	123	96	109		

Note: P1–P10 = Participants 1–10.

CSV PDF

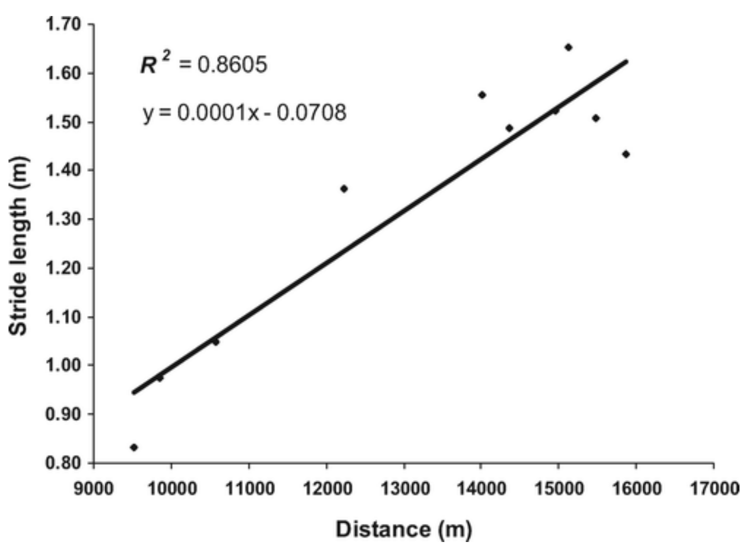
Mean running speed for the three trials was 13.5 km · h^{-1} (s = 2.5) for Trial 1, 13.5 km · h^{-1} (s = 2.4) for Trial 2, and 13.5 km · h^{-1} (s = 2.4) for Trial 3 (Figure 2). Runners increased their running speed from 0 to 2 min until it became stable. Runners maintained a constant pace until approximately 59 min, before increasing their speed to the end of the run.

Figure 2. Treadmill speed (km · h^{-1}) over the 60-min run with the standard deviation (s) shown at 5-min intervals for clarity.



PowerPoint slide Original jpg (42.00KB) Display full size

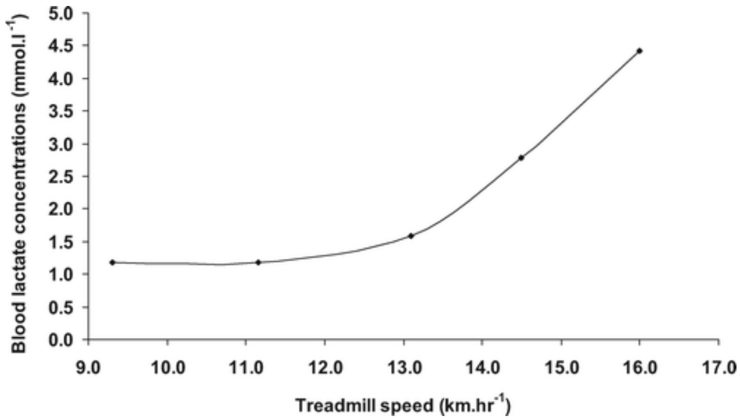
Figure 3. Relationship between mean stride length and total distance covered.



PowerPoint slide Original jpg (91.00KB) Display full size

Total distance covered during the 60-min run was strongly associated with mean stride length (coefficient of determination, R^2 = 0.86) (stride frequency = 180 strides · min^{-1}, s = 48; stride length = 1.34 m, s = 0.28) (Figure 3). Extrapolation from the preliminary sub-maximal running tests suggests that the 60-min runs were performed at a blood lactate concentration of approximately 2 mmol · l^{-1} (Figure 4).

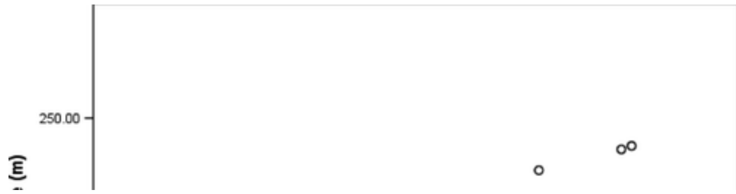
Figure 4. Relationship between blood lactate concentration (mmol · l^{-1}) and running speed.



PowerPoint slide Original jpg (64.00KB) Display full size

Figure 5 shows the association between the residual errors plotted against the predicted distances run in 1 h. When we correlated the absolute residuals against the predicted measurements, we obtained evidence of heteroscedasticity (r = 0.210; P > 0.05). For this reason, we log transformed the dependant variable distance (Figure 6). The resulting within-participant measurement error, obtained from the repeated-measures ANOVA, was (s_w^2) = 0.00000969. Note that the units associated with this error ratio term are dimensionless as explained in the Methods section. The anticipated difference between two trial measurements becomes s = 0.013927. The correlation between the absolute residuals and predicted values for log transformed distance was r = 0.031 (P > 0.05). The unexplained error variation calculated with both participants and trials as main effects results in a coefficient of variation of 1.4%. The equivalent 95% ratio limits of agreement are obtained as an error ratio (*/÷ 1.028) – that is, 95% of the ratios (one trial measurement, divided by a second) should lie between 0.973 and 1.028.

Figure 5. Relationship between the residuals and predicted distances covered in 1 h.



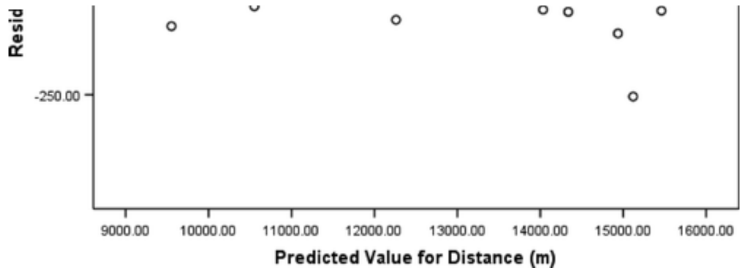
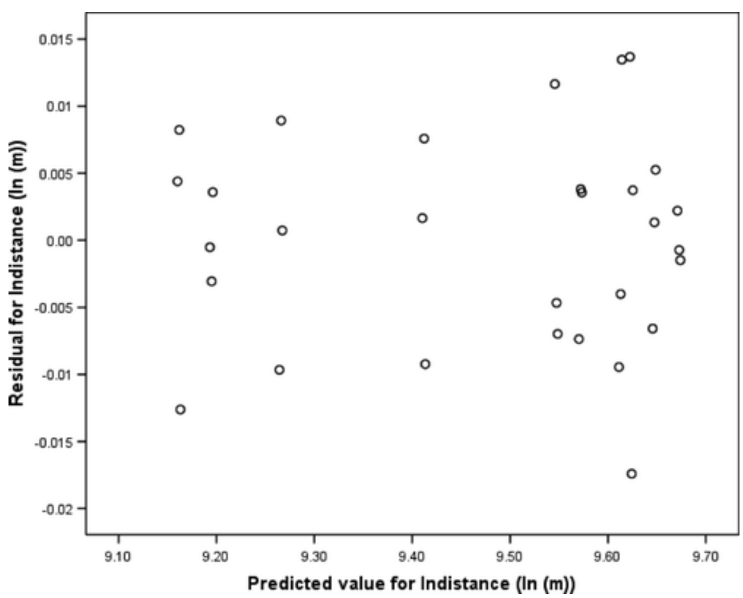


Figure 6. Relationship between the residuals and predicted distances covered in 1 h, all measurements log transformed.



Discussion

Research into the influence of nutrition and training on exercise performance requires simulation of the demands of the event under controlled laboratory conditions. The most common method of achieving this in running involves the use of motorized treadmills. The major limitation with traditional treadmill running, however, is the inability to replicate the free and spontaneous changes in speed that occur during running events. In response, we developed an automated treadmill system that allows runners to freely control their running speed, removing the limitations associated with manual changes in speed.

The main finding of this study was that runners were able to replicate their selection of speed using an automated treadmill. We found a coefficient of variation of 1.4% when runners were asked to run as far as possible in 60 min. Interestingly, the results of the repeated-measures ANOVA identified no significant difference in total distance covered between the three trials. However, the plot of the residuals against the fitted values did identify evidence of heteroscedasticity (a positive correlation between the absolute residuals and the fitted values: $r = 0.210$; $P > 0.05$). This evidence of heteroscedasticity was eliminated by taking logarithms (see Figure 6) and, as such, the unexplained error was reported as a percentage error, in the form of a coefficient of variation equivalent to 1.4%. Hopkins and Hewson (2001) have previously reported that running tests require a coefficient of variation of 1.5% or less to detect small changes in running performance. The coefficient of variation of 1.4% obtained in the present study is therefore sufficiently reliable to detect worthwhile changes in running performance.

The coefficient of variation of 1.4 % is an improvement on the 2.7% reported previously by Schabort *et al.* (1998b), who used the same instruction – that is, to run as far as possible in 60 min – but who used a manually controlled treadmill. Laursen *et al.* (2007) have previously stated that manually changing the treadmill speed by pressing the appropriate console buttons is dependent upon the runner's perception of their ability to run faster or slower. This method of controlling the treadmill speed may not be sufficiently sensitive to detect small differences in performance (Whitham & McKinney, 2007). In the present study, the runners' ability to make spontaneous, accurate adjustments to their speed over the 60 min might be responsible for the improved coefficient of variation with the automated system compared with manually altering the treadmill speed. A smaller coefficient of variation of approximately 1% has been reported in a study on 10-km treadmill time trials (Russell, Redmann, Ravussin, Hunter, & Larson-Meyer, 2004). In this study, however, runners ran for 90 min at 65%*[Vdot]*O_{2max} before completing the 10-km time trial. Although reliable, this test and others that employ prolonged pre-time trial runs (Doyle & Martinez, 1998) might not be appropriate when investigating physiological responses during shorter (~1 h), more intense (~70%*[Vdot]*O_{2peak}) exercise performance (Carter, Jeukendrup, & Jones, 2004; Jeukendrup *et al*., 1997; Whitham & McKinney, 2007).

Although we report a greater reproducibility than previous running time trials (Doyle & Martinez, 1998; Schabort *et al*., 1998b; Whitham & McKinney, 2007) and time-to-fatigue tests (Billat, Renoux, Pinoteau, Petit, & Koralsztein, 1994; Jeukendrup*et al.*, 1996), it still falls short of the 1.0–1.1% coefficine of variation reported for cycling time trials (Hickey *et al.*, 1992; Palmer, Dennis, Noakes, & Hawley, 1996). The exact reason for this is unclear. One possible explanation might be the differences between the automated treadmill and cycle ergometers. Although responsive, the automated treadmill does not have the same sensitivity as cycle ergometers, which allows rapid changes in power output simply by altering pedal cadence.

The time a runner takes to cover a certain distance is determined by stride length and stride frequency (Brandon & Boileau, 1992). Day-to-day stride length and frequency have been shown to be highly reproducible in well-trained runners (Brisswalter & Legros, 1994) and therefore it is not surprising that stride length is strongly correlated with distance covered in 60 min. In the present study, neither stride length nor frequency decreased during the 60-min run, suggesting that the runners did not experience a marked amount of fatigue. From preliminary measurements of blood lactate, it would appear that the chosen speeds were supported almost entirely by aerobic metabolism.

The central governor concept states that athletes have the ability to regulate their metabolic response towards an “anticipated” end point (Noakes, 2007; Rauch, St Clair Gibson, Lambert & Noakes, 2005). Previously, it has been suggested that athletes might be engaging in monitoring processes that allows them to optimize the distribution of their metabolic resources over the duration of the race or exercise task (Foster *et al*., 2003). Consistent with this and previous research in both laboratory (Palmer, Borghouts, Noakes, & Hawley, 1999; Rauch *et al.*, 2005, Weltan, Bosch, Dennis, & Noakes, 1998a, 1998b) and field (Billat, Slawinski, Danel, & Koralsztein, 2001; Sandals *et al.*, 2006) investigations, it would appear that the runners distributed their energetic resources over the 60-min run, so as to be able to sprint during the last minute of exercise. Therefore, knowledge of specific exercise duration appears to be important when repeating endurance performance tests. Finally, we acknowledge the fact that while in the present study the 60-min time trial has been shown to have good repeatability, this may not be the case when using time to complete a fixed distance.

In conclusion, asking runners to cover as much distance as possible in 60 min, using an automated treadmill system that allows runners to control the speed without manual input, is a reliable method of assessing endurance performance in endurance-trained runners.