Rating of perceived exertion during high-intensity treadmill running

MIKE DOHERTY, PAUL M. SMITH, MICHAEL G. HUGHES, and DAVID COLLINS

Department of Sport and Exercise Science, University of Luton, Luton, Beds LU1 3JU, UNITED KINGDOM; and Department of Physical Education, Sport and Leisure Studies, University of Edinburgh, Edinburgh EH4 6JD, UNITED KINGDOM

ABSTRACT

DOHERTY, M., P. M. SMITH, M. G. HUGHES, and D. COLLINS. Rating of perceived exertion during high-intensity treadmill running. *Med. Sci. Sports Exerc.*, Vol. 33, No. 11, 2001, pp. 1953–1958. **Purpose:** The purpose of this investigation was 1) to evaluate the time course of the rating of perceived exertion (RPE; 6–20 Borg scale) during short-term, high-intensity, constant-load running (ST); and 2) to determine the reproducibility of RPE during ST. **Methods:** Fifteen well-trained males ($\dot{V}O_{2max}$ = 58.0 ± 4.6 $mL \cdot kg^{-1} \cdot min^{-1}$, mean ± SD) performed treadmill running (i.e., between 3 and 4 $m \cdot s^{-1}$ at 10.5% incline) to volitional exhaustion (T_{lim}) at an exercise intensity equivalent to 125% $\dot{V}O_{2max}$. A total of four RPE measurements were taken during each test, one every 30 s during the first 120 s of the exercise. The tests were repeated at the same time of day on three occasions within a 3-wk period. **Results:** T_{lim} for the three tests was 197.6 ± 34.8 s. RPE was linearly related with exercise time (mean ± SD for the three tests: RPE at 30 s = 10.8 ± 2.2; RPE at 60 s = 12.6 ± 1.8; RPE at 90 s = 14.5 ± 1.7; RPE at 120 s = 16.0 ± 1.9; RPE = 9.06 + (0.06 × time (s)); r = 0.71, SEE = 2.0, $P < 0.01$). Repeated ANOVA revealed no systematic bias between the three tests for RPE, and other measures of reliability were also favorable. These included intraclass correlation coefficients ranging from 0.78 to 0.87 and sample coefficients of variation of between 4.4% and 6.0%. The 95% limits of agreement ranged between 0.0 ± 2.3 and 0.0 ± 2.5. **Conclusion:** ST RPE displays a positive linear response during the first 2 min. The measurement of ST RPE appears to be reliable and could thus add a new dimension to ST investigations. **Key Words:** RELIABILITY, ANAEROBIC EXERCISE, PERCEPTUAL RESPONSE, FATIGUE

Perceived exertion has been defined as the act of detecting and interpreting sensations arising from the body during physical exercise (22,p.4). For nearly 40 yr, rating of perceived exertion (RPE) scales have been used as reliable and valid measurements of exercise intensity (6,15,16,22,27,29). Their principal use has involved quantifying subjective feelings of fatigue and exercise tolerance during submaximal exercise (22,pp.215–256) and, increasingly, prescribing exercise intensities for the development and maintenance of cardiovascular fitness for both healthy (12) and "at-risk" populations (17). By comparison, there have not been any systematic attempts to investigate the perceptual response to short-term (2–5 min), high-intensity (i.e., >100% $\dot{V}O_{2max}$) exercise (ST). This probably reflects the belief that subjective estimates of exertion during high-intensity exercise are not viable and lack application to sport and exercise science. However, rather than employ "all-out" ST—where recording of RPE may be impractical—ST investigations may also observe constant-load exercise at an intensity equivalent to, or above, $\dot{V}O_{2max}$ (2,4,14,23). In addition, there are particular modes of ST that could accommodate the measurement of RPE and/or where the velocity of movement could be slow enough for the identification of RPE to be made. Examples of ST exercise modes that might accommodate such measurement include treadmill running against a steep incline (2,14) or isokinetic cycle ergometry (10).

A logical first step in investigating ST RPE would be to establish the reliability of the measurement and to describe the time course of RPE during ST. For example, it would be of interest to determine if ST RPE assumes a positive linear relation, in a manner similar to that shown by RPE and several physiological parameters, including oxygen consumption and heart rate, during incremental submaximal exercise. Alternatively, a nonlinear relation may exist in a fashion similar to that of blood lactate accumulation and pulmonary ventilation during graded submaximal exercise (21).

Thus, the purpose of the present investigation was, 1) to evaluate the time course of the RPE (6–20 Borg scale) during short-term, high-intensity, constant-load running; and 2) to assess the reproducibility of the RPE during ST. We hypothesized that there would be a positive linear response over the first 2 min of ST exercise, and that RPE would be reliable as measured by intraclass correlation coefficients, coefficients of variation, standard errors of measurement, and limits of agreement.

METHODS

Subjects. A power analysis for calculation of sample size for intraclass correlation was performed (NCSS Statistical Software, Kaysville, UT). Using a power of 0.80, an alpha level of 0.05, and an estimated intraclass correlation

0195-9131/01/3311-1953/$3.00/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

Submitted for publication August 2000.
Accepted for publication February 2001.

coefficient of 0.60 with three observations per subject, it was estimated that 11 subjects were required for the investigation. Fifteen male subjects eventually volunteered to take part in the study and were retained to compensate for envisaged subject "dropout." The subjects were physically active students and included sprinters, middle-distance runners, and multiple-sprint sports performers. The mean ± SD for age, height, and body mass of the subjects was 22.4 ± 3.4 yr, 1.76 ± 0.06 m, and 70.9 ± 8.4 kg, respectively. All subjects were familiar with treadmill running and all data collection procedures, including the measurement of RPE.

General procedures. The study was part of an ongoing investigation into the reliability of the maximal accumulated oxygen deficit (MAOD) and run-time to exhaustion (T_{lim}) during ST (14). The nature, purpose, experimental procedures, and possible risks and benefits were outlined for each subject both verbally and in writing. The experimental procedures were in accordance with the policy statements of the American College of Sports Medicine. The subjects were tested on four separate occasions, consisting of one preliminary test and three repeat supramaximal tests. The tests were scheduled at the same time of day for each subject, with a minimum of 2 d separating test sessions. For each subject, all tests were completed within 3 wk. Although the subjects continued to train over the duration of the study, they were instructed to refrain from physical activity and to avoid caffeine in the 24 h before the tests (13), and to present themselves at the laboratory in a 2-h postabsorptive state.

Preliminary test session ($\dot{V}O_{2\mathrm{max}}$ and estimation of 125% $\dot{V}O_{2\mathrm{max}}$). Supramaximal $\dot{V}O_2$ was determined using an extrapolation method adapted from procedure 3 of the methods described by Medbø et al. (19). In order to relate $\dot{V}O_2$ to running velocity, each subject performed three 6-min discontinuous treadmill (Powerjog, Cranlea & Co., Birmingham, U.K.) runs of increasing exercise intensity, with 5-min recovery between runs. All tests were conducted on a 10.5% incline. The preliminary test was adjusted to take account of between-subject differences in fitness levels. The three treadmill speeds were 2.3 ± 0.14, 2.5 ± 0.4, and 2.7 ± 0.4 $m \cdot s^{-1}$. Respiratory gases (Douglas bag, open circuit spirometry), heart rate ($beats \cdot min^{-1}$; Sports tester, Polar Electro, Finland), and RPE (Borg 15-point scale, 6–20) were recorded during the final minute of each 6-min run. Before testing, subjects were reminded of the instructions on the use of the RPE scale (22,pp.77–81). Subjects had to:

1. Understand the definition of RPE and receive an explanation of the nature and use of the scale.
2. "Anchor" the top and bottom perceptual ratings to previously experienced sensations of the easiest and most difficult exercise encountered.
3. Ensure they gave an "all-over," integrated rating, which included both muscular and cardiorespiratory sensations.
4. Understand the subjective nature of the RPE scale; that there are no "right" or "wrong" responses.

Immediately after the third and final run, the treadmill belt velocity was increased by 0.14 $m \cdot s^{-1}$ and thereafter each minute until volitional exhaustion. This took on average an extra 4.2 ± 0.4 min. During these final minutes, both respiratory gas (for subsequent determination of $\dot{V}O_{2\mathrm{max}}$) and heart rate data were collected continuously, with $\dot{V}O_2$ being determined every 30 s.

Attainment criteria for $\dot{V}O_{2\mathrm{max}}$ included any two of the following (7):

1. An increase in $\dot{V}O_2$ of less than 2 $mL \cdot kg^{-1} \cdot min^{-1}$.
2. A respiratory exchange ratio (RER) of > 1.10.
3. A maximum heart rate of $220 - $ age (± 10 $beats \cdot min^{-1}$).

On the basis of the preliminary treadmill test and the relation between $\dot{V}O_2$ and running velocity, an individual linear regression equation was derived for each subject. This was used to calculate the running velocity required to elicit an exercise intensity equivalent to 125% $\dot{V}O_{2\mathrm{max}}$ for the subsequent supramaximal test.

Supramaximal test (T_{lim} at 125% $\dot{V}O_{2\mathrm{max}}$). The supramaximal test was preceded by a 5-min warm-up run on the treadmill at a constant pace that equated to approximately 55% $\dot{V}O_{2\mathrm{max}}$. The subjects then performed a set of stretching exercises, followed by a more test-specific warm-up of 6 × 15-s bouts of running at the predetermined supramaximal velocity, with 15-s recovery periods in between each repetition. The subjects were then given a 5-min recovery period before they began the supramaximal test.

Before the supramaximal test, each subject was instructed to continue running until they could no longer maintain the set pace; at 125% $\dot{V}O_{2\mathrm{max}}$, this was estimated to be approximately 3 min (2,3,13). Subjects were informed that RPE was to be measured throughout the test. Since the purpose of the study was to determine the reliability of the RPE measurements during three repeated runs to exhaustion (T_{lim}), it was important that subjects received no overt cueing of the duration of the test. Subjects thus remained uninformed that the RPE measurements were to be taken every 30 s for the first 2 min of the test. RPE was only taken for the first 2 min of the test because pilot tests revealed that the excessive and disorientating fatigue associated with high-intensity exercise ruled out collection of RPE during the latter part (i.e., approximately the last minute) of the test.

The treadmill belt was adjusted to the predetermined velocity and, when the subject was ready, they lowered themselves onto the moving treadmill belt and a digital stop clock was started to indicate start of exercise. In order to ensure accurate recording of RPE, a printed scale was presented immediately in front of the subject on a large (0.91 × 0.61 m) board. Subjects then pointed to the number or verbal descriptor that adequately represented their RPE. In practice, this took no longer than 1–2 s and thus did not interfere with the exercise test. The treadmill speed (i.e., range, 3.0–4.0 $m \cdot s^{-1}$ at 10.5% incline) was sufficiently slow enough to allow subjects to indicate RPE during the exercise without risk of injury. The supramaximal test was repeated at the same time of day on three occasions within a 3-wk period.

TABLE 1. Mean ± SD RPE for the first four 30 s of the supramaximal run for each test ($N = 15$).

Test	Time (s)			
	30	60	90	120
1	10.7 ± 2.2	12.5 ± 1.9	14.5 ± 1.7	15.9 ± 1.8
2	10.7 ± 2.3	12.5 ± 1.8	14.2 ± 1.9	15.9 ± 2.2
3	10.9 ± 2.5	12.9 ± 2.1	14.7 ± 2.0	16.3 ± 2.1
Mean	10.8 ± 2.2*	12.6 ± 1.8*	14.5 ± 1.7*	16.0 ± 1.9*

* Significantly different from mean RPE at all other time points ($P < 0.01$).

Respiratory gas analysis. Subjects breathed through a Hans Rudolph valve (Kansas City, MO) and were connected to either a series of 200-L Douglas bags (i.e., preliminary assessment) or one 1000-L Douglas bag (supramaximal test) by means of short, lightweight, wide-bore tubing. One liter of air from each Douglas bag was drawn off for the determination of the fractions of oxygen and carbon dioxide using a paramagnetic oxygen analyzer (Servomex 570A; Crowborough, U.K.) and an infrared carbon dioxide analyzer (Servomex PA404), both of which had previously been calibrated against gases of known concentration (15.1% O_2 and 5% CO_2; British Oxygen Co., Wembley, U.K.). The remaining volume of each sample was then measured with a Harvard dry-gas meter (Harvard Apparatus Ltd., Kent, U.K.).

Statistical analyses. As recommended by Atkinson and Nevill (1), we report several measures of "absolute" and "relative" reliability so that different researchers can interpret the one(s) they are most accustomed to. Agreement between the three tests was examined using a two-factor, repeated measures analysis of variance (ANOVA). The design allowed for identification of significant test bias (i.e., test main effect) and estimation of the within-subject measurement error (MSE) (1). The standard deviation (SD) of expected difference between two test measurements was estimated by $SD = \sqrt{2MSE}$ (1,5). Provided the residuals were normally distributed (determined by the Kolmogorov-Smirnov test) and not related to the size of the measurements (determined by interclass correlation coefficients), the 95% limits of agreement (95LoA) were then calculated as ± 1.96 SD (1,5). In addition, a Bland-Altman plot (1,5) was made using the residuals and actual scores. This provides an indication of systematic test bias and random error between tests by examining the direction and magnitude of the scatter around the zero line.

Intraclass correlation coefficient (ICC) was calculated from the ANOVA as $(F - 1)/(F + k - 1)$, where F was the F-ratio for the subject term and k (= 3) was the number of trials (26). The coefficient of variation (CV) for individual subjects was calculated by dividing each subject's SD by their mean for the three tests (1). The unbiased standard error of measurement (SEM) was calculated from $SD\sqrt{1 - ICC}$ (1). Ninety-five percent confidence intervals were computed for all reliability statistics (18). The RPE responses over time were tested for departure from linearity with a regression analysis that adjusted for subject and test (28). An alpha level of 0.05 was chosen to indicate statistical significance. All statistical procedures were performed using SPSS for Windows, Version 9.0 (SPSS, Inc., Chicago, IL).

RESULTS

Preliminary test. The mean ± SD RPE for the three 6-min submaximal treadmill runs were 11.0 ± 1.6, 12.9 ± 1.2, and 14.9 ± 1.7. During the volitional run to exhaustion, 11 of the 15 subjects demonstrated a $\dot{V}O_2$ plateau and the remaining four subjects satisfied both of the secondary criteria for attainment of $\dot{V}O_{2max}$. The mean ± SD $\dot{V}O_{2max}$, maximum heart rate, and maximum RER of the subjects was 58.0 ± 4.6 $mL \cdot kg^{-1} \cdot min^{-1}$, 196 ± 7.4 $beats \cdot min^{-1}$, and 1.10 ± 0.09, respectively.

Supramaximal tests. The mean ± SD T_{lim} for the three tests was 197.6 ± 34.8 s. The reliability of T_{lim} and MAOD are reported elsewhere (14). The RPE responses for the first four 30 s of the run for each of the three tests are given in Table 1. The ANOVA revealed no significant test bias (i.e., test main effect, $P > 0.05$) and no significant interaction of test and time ($P > 0.05$). However, there was a significant main effect for time of RPE measurement ($P < 0.01$) (Table 1). A Bonferroni *post hoc* test revealed that significant differences existed between each of the four RPE time points (i.e., at 30, 60, 90, and 120 s; $P < 0.01$). In comparison with submaximal and graded exercise RPE literature (16,27,29), the traditional measures of reliability were very similar, with mean ICC of between 0.78 and 0.87 and mean CV of between 4.4% and 6.0% (Table 2).

The Kolmogorov-Smirnov test revealed that all of the residuals except for RPE at 30 s were normally distributed ($P = 0.03$, 0.16, 0.08, and 0.41, for RPE at 30 s through RPE at 120, respectively) and ICC of the absolute residual errors against the actual scores showed that the residuals were not related to the size of the measurements (r ranged from 0.07 to −0.33, $P > 0.05$). Thus, the 95LoA were calculated (Table 2). Bland-Altman plots for RPE at 30 s through RPE at 120 s are given in Figure 1. The regression analysis to test for departure from linearity showed $F_{(2,168)} = 0.256$ ($P > 0.05$). Since the repeated ANOVA revealed no differences between tests, a linear regression with data averaged across the three tests was used (Fig. 2).

DISCUSSION

Fatigue has been defined as an acute impairment of exercise performance that includes both an increase in the perceived effort necessary to produce a power output and the eventual inability to maintain that power output (11). Thus, the RPE, with its strong relation to factors indicating fatigue (22,pp.93–104), is an important index in the evaluation of the extent of fatigue. To date, there have not been

TABLE 2. Reliability statistics and 95% confidence intervals (CI) for the first four 30 s of the supramaximal run for each test (N = 15).

Time (s)	ICC (CI)	CV (CI)	SEM (CI)	95LoA (CI)
30	0.87 (0.72–0.95)	5.5 (4.1–7.2)	0.79 (0.49–1.16)	0 ± 2.3* (–0.5–0.5) ± (1.7–3.0)
60	0.81 (0.61–0.92)	6.0 (4.5–7.8)	0.78 (0.50–1.12)	0 ± 2.3 (–0.5–0.5) ± (1.7–3.1)
90	0.78 (0.56–0.91)	6.0 (4.5–7.8)	0.80 (0.51–0.99)	0 ± 2.5 (–0.5–0.5) ± (1.8–3.2)
120	0.84 (0.66–0.94)	4.4 (3.5–6.8)	0.76 (0.47–0.93)	0 ± 2.3 (–0.5–0.5) ± (1.7–2.9)

ICC, intraclass correlation coefficient; CV, coefficient of variation (%); SEM, standard error of measurement, 68 percentile; 95LoA, 95% limits of agreement.
* Residuals not normally distributed (Kolmogorov-Smirnov P = 0.03).

any studies that have investigated RPE during short-term, high-intensity exercise.

Our results suggest that at least every 30 s during ST (at 125% $\dot{V}O_{2max}$), subjects reported an increase in the effort linked to the task. Although only four RPE recordings were made for each of three tests, the results indicate that perceived effort displays a positive linear response to exercise time, with the increase in RPE (Borg 15-point scale, 6–20) amounting to approximately 1.5–2.0 points every 30 s during the first 2 min of ST (Fig. 2). A caveat here is that subjects exercised on average for 1 additional min after the initial 2 min of ST, and it may be that the growth of RPE over this last minute may not be linear. Indeed, since the final (i.e., 2 min) RPE response of 16.0 ± 2.0 (mean ± SD of all three tests) did not approach the high terminal category of 20 in the Borg scale, there is a strong suggestion that the perceptual response during this last minute may well be nonlinear. That is, assuming the terminal category of 20 was

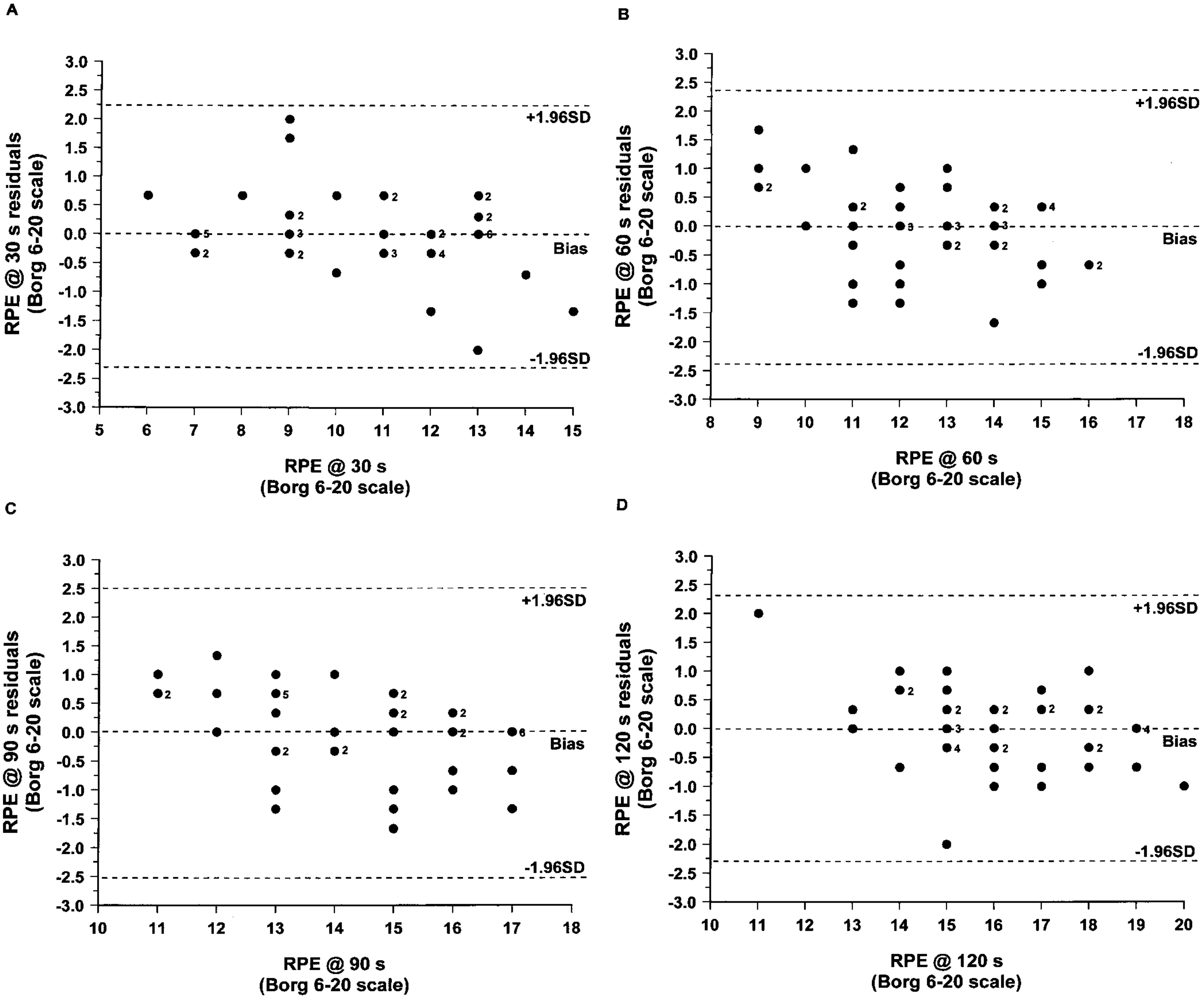

FIGURE 1—Bland-Altman plots for the RPE data at 30 s (A); 60 s (B); 90 s (C); and 120 s (D).

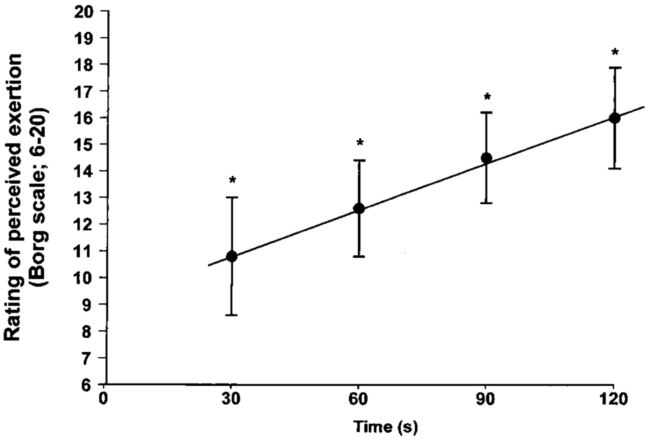

FIGURE 2—Rating of perceived exertion (mean ± SD) during the first 2 min of high-intensity treadmill running at 125% $\dot{V}O_{2max}$ ($N = 15$). The data were averaged across the three tests for each of the four measurement time points. RPE = 9.06 + (0.06 × time (s)); r = 0.71, SEE = 2.0, $P < 0.01$. *Significantly different from RPE at all other time points ($P < 0.01$).

achieved by the end of the test, RPE increased by approximately four RPE categories (cf., 1.5–2.0 RPE categories during every 30 s of the first 2 min) in the final minute. Considering that the subjects were familiar with the concept of RPE, that a perceptual anchoring procedure was followed, and that volitional exhaustion occurred after only approximately 3 min, the RPE values are thus lower than one might have anticipated (Table 1). In fact, the first three ST RPEs are similar to those achieved in the three 6-min steady-state preliminary test runs, where the exercise intensity was equivalent to ~60–80% $\dot{V}O_{2max}$. Interestingly, the initial low RPE values do corroborate subjects' postexercise anecdotal reports that the exercise challenge is "relatively easy" during the first part of the test (i.e., up to 1.5–2 min), after which time fatigue occurs very rapidly (i.e., within ~1 min).

During submaximal exercise using large muscle groups, RPE appears to be mediated to a large extent by the relative metabolic rate (% $\dot{V}O_{2max}$) (22,pp.107–124). Noble and Robertson (22,p.69) suggest that a transience in perceptual intensity may occur from the beginning of exercise until the achievement of a metabolic steady state (i.e., after ~4 min). Thus, in order to achieve a valid steady-state RPE, it is important to record RPE only once a metabolic steady state has been reached. Our results suggest that a (reliable) "transience" in RPE occurs during ST. Thus, a potentially fertile avenue for further research in this area may be to clarify and contrast the initial transient RPE responses during both submaximal and supramaximal exercise. It should be noted that a number of reports suggest that at relatively high submaximal exercise intensities (e.g., above the lactate threshold), perceptual and physiological steady states occur less frequently. In some cases, the magnitude of the increase in RPE is greater than that for the physiological variables (24,25). Our data also suggest a modest uncoupling of exercise intensity and ST RPE, such that the perceptual response is dampened (i.e., at least for the first 2 min of this type of exercise). These findings suggest that the initial setting of perceived exertion is not proportional to the exercise intensity during ST. Thus, some of the physiological cues for ST may be different from submaximal exercise, and/or the perceptual sensitivity or sensory thresholds for the processing of physiological information may somehow be attenuated during ST. During submaximal exercise, it appears that subjects may directly monitor the external manifestations of physiological responses and use them as cues for rating the perceptual response (e.g., shortness of breath, sweating, and "heavy" legs) (22,pp.93–104). If these symptoms are also a prerequisite for assigning RPE during ST, our results might be explained by a delay in the processing of the symptoms. Future work is clearly needed to develop or validate the psychological and physiological/neuromuscular mediators of ST RPE. In comparison with submaximal exercise, the role that pain plays in signaling exertional sensations may be of particular relevance to strenuous exercise (22,p.150).

The reliability of RPE during graded exercise testing or constant submaximal intensity is firmly established (16,27,29). A number of authors have reported reliability coefficients ranging between 0.70 and 0.90 (16,27,29). This investigation displays equally high coefficients (Table 2). Furthermore, the CV and SEM also appear to be acceptable (Table 2). Finally, it is also important to report that there was no systematic bias between the three tests as measured by the repeated ANOVA ($P > 0.01$). This indicates that no significant learning or fatigue occurred over the course of the study. Future ST RPE reliability studies that use similarly trained and highly motivated subjects may therefore find that two rather than three tests are sufficient for data collection purposes.

Thus, when traditional reliability indicators are used, RPE during ST appears to be as reliable as RPE obtained from graded exercise testing and steady-state submaximal exercise (16,27,29). However, it should be borne in mind that reliability coefficients, CV, SEM, and hypothesis tests are measures that have been the subject of continued criticism (1). Although not all sport and exercise scientists fully support the use of the 95LoA (18), a growing body of opinion suggests that it is the most appropriate statistic to report when assessing agreement between two or more trials (1,14). The 95LoA is a relatively new measure of reliability and there are no RPE studies to compare the present results with. However, it can be concluded that the likely difference between two repeat tests should lie within ~0 ± 2 RPE categories at 30, 60, 90, and 120 s. Considering that the 95LoA is a conservative test (18), we believe the limits of agreement are narrow enough for ST RPE to be of practical use in experimental designs where ST RPE may be expected to change.

Even though the subjects were all exercising at the same relative intensity, there is a large intersubject variation in RPE at all four time points, with a SD of ~2 RPE categories (Table 1). Again, as with the intrasubject variation found in this study, the intersubject variation also compares well with RPE measured during graded exercise testing or steady-state submaximal exercise (22,pp.215–255). Apart from the obvious psychobiological variation in sense perception (22,pp.171–195), the ST intersubject variation may also be related to the error involved in estimating the exercise intensity equivalent to 125% $\dot{V}O_{2max}$ (3).

In conclusion, the results from the present study show that RPE displays a reliable and positive linear response during the first 2 min of constant-load, short-term, high-intensity treadmill running. Both the intra- and intersubject RPE variability appear to be similar to the variability found in graded exercise and steady-state, submaximal exercise. Thus, it is not unreasonable to propose that ST RPE could be used to determine subjective, perceptual effort during ST investigations. This could be in the same way that such measurements are made during endurance-based studies (8,9) and in a similar fashion to studies that utilize visual analog scales to investigate the effects of drugs on pain during ischemic skeletal muscle contraction (20). Finally, since one of the first signals that fatigue is imminent is an increased perception of effort at the same absolute exercise intensity (11), RPE during ST might also be used as a marker of overtraining in elite athletes. Inclusion of ST RPE may add a new dimension to ST investigations and contribute to our understanding of the perceptual response to extremely high-intensity exercise.

The authors express their thanks to Neil Willmore for technical assistance and Professor Mike Kenward and Dr. Richard Davison for statistical advice. We are also grateful to the reviewers' and Associate Editors' insightful comments on early drafts of the manuscript.

Address for correspondence: Mike Doherty, Department of Sport and Exercise Science, University of Luton, Park Square, Luton, Beds LU1 3JU, United Kingdom; E-mail: mike.doherty@luton.ac.uk.

REFERENCES

1. Atkinson, G., and A. M. Nevill. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 26:217–238, 1998.
2. Balsom, P. D., S. D. R. Harridge, K. Soderlund, B. Sjodin, and B. Ekblom. Creatine supplementation per se does not enhance endurance exercise performance. *Acta Physiol. Scand.* 149:521–523, 1993.
3. Bangsbo, J. Oxygen deficit: a measure of the anaerobic energy production during intense exercise? *Can. J. Appl. Physiol.* 21: 350–363, 1996.
4. Bangsbo, J., P. D. Gollnick, T. E. Graham, et al. Anaerobic energy production and O_2 deficit-debt relationship during exhaustive exercise in humans. *J. Physiol. (Lond.)* 42:539–559, 1990.
5. Bland, M., and D. G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310, 1986.
6. Borg, G. A. V. *Physical Performance and Perceived Exertion*. Lund, Sweden: Gleerup, 1962, pp. 1–62.
7. British Association of Sport and Exercise Sciences. *Physiological Testing Guidelines*, 3rd Ed. Leeds, U.K.: British Association of Sport and Exercise Sciences, 1997, p. 64.
8. Cole, K. J., D. L. Costill, R. D. Starling, B. H. Goodpastor, S. W. Trappe, and W. J. Fink. Effect of caffeine ingestion on perception of effort and subsequent work production. *Int. J. Sports Nutr.* 6:14–23, 1996.
9. Costill, D. L., G. P. Dalsky, and W. J. Fink. Effects of caffeine ingestion on metabolism and exercise performance. *Med. Sci. Sports* 10:155–158, 1978.
10. Davies, C. T. M., and E. R. Sandstrom. Maximal mechanical power output and capacity of cyclists and young adults. *Eur. J. Appl. Physiol.* 58:838–844, 1989.
11. Davis, J. M., and S. P. Bailey. Possible mechanisms of central nervous system fatigue during exercise. *Med. Sci. Sports Exerc.* 29:45–57, 1996.
12. Dishman, R. K. Prescribing exercise intensity for healthy adults using perceived exertion. *Med. Sci. Sports Exerc.* 26:1087–1094, 1994.
13. Doherty, M. The effects of caffeine on the maximal accumulated oxygen deficit and short-term running performance. *Int. J. Sports Nutr.* 8:95–104, 1998.
14. Doherty, M., P. M. Smith, and K. Schroder. Reproducibility of the maximum accumulated oxygen deficit and run time to exhaustion during short-distance running. *J. Sports Sci.* 18:331–338, 2000.
15. Dunbar, C. C., R. J. Robertson, R. Baun, et al. The validity of regulating exercise intensity by ratings of perceived exertion. *Med. Sci. Sports Exerc.* 24:94–99, 1992.
16. Eston, R. G., and J. G. Williams. Reliability of ratings of perceived effort regulation of exercise intensity. *Br. J. Sports Med.* 22:153–155, 1988.
17. Gutmann, M. C., R. W. Squires, and M. L. Pollock. Perceived exertion-heart rate relationship during exercise testing and training in cardiac patients. *J. Card. Rehabil.* 1:52–59, 1981.
18. Hopkins, W. Measures of reliability in sports medicine and sport science. *Sports Med.* 30:1–15, 2000.
19. Medbø, J. I., A. Mohn, I. Tabata, I. Bahr, and O. Sejersted. Anaerobic capacity determined by the maximal accumulated oxygen deficit. *J. Appl. Physiol.* 64:50–60, 1988.
20. Myers, D. E., Z. Shaikh, and T. G. Zullo. Hypoalgesic effect of caffeine in experimental ischemic muscle contraction pain. *Headache* 37:754–658, 1997.
21. Noble, B. J., G. A. V. Borg, I. Jacobs, R. Ceci, and P. Kaiser. A category-ratio perceived exertion scale: relationship to blood and muscle lactates and heart rate. *Med. Sci. Sports Exerc.* 15:523–528, 1983.
22. Noble, R. J., and R. J. Robertson. *Perceived Exertion*. Champaign, IL: Human Kinetics, 1996.
23. Nummela, A., and H. Rusko. Acclimatization to altitude and normoxic training improve 400-m running performance at sea level. *J. Sports Sci.* 18:411–419, 2000.
24. Pandolf, K. B., E. Caferelli, B. J. Noble, and K. F. Metz. Perceptual responses during prolonged work. *Percept. Mot. Skills.* 35:975–985, 1972.
25. Pandolf, K. B., D. S. Billings, L. L. Drolet, N. A. Pimental, and M. N. Sawka. Differentiated ratings of perceived exertion and various physiological responses during prolonged upper and lower body exercise. *Eur. J. Appl. Physiol.* 53:5–11, 1984.
26. Schabort, E.-J., W. G. Hopkins, and J. A. Hawley. Reproducibility of self-paced treadmill performance of trained endurance runners. *Int. J. Sports Med.* 19:48–51, 1998.
27. Skinner, J. S., R. Hutsler, V. Bergsteinova, and E. R. Buskirk. The validity and reliability of a rating scale of perceived exertion. *Med. Sci. Sports* 5:94–96, 1973.
28. Snedecor, G. W., and W. G. Cochran. *Statistical Methods*, 6th Ed. Ames, IA: Iowa State University Press, 1967, pp. 168–169.
29. Stamford, B. A. Validity and reliability of subjective ratings of perceived exertion during work. *Ergonomics* 19:53–60, 1976.