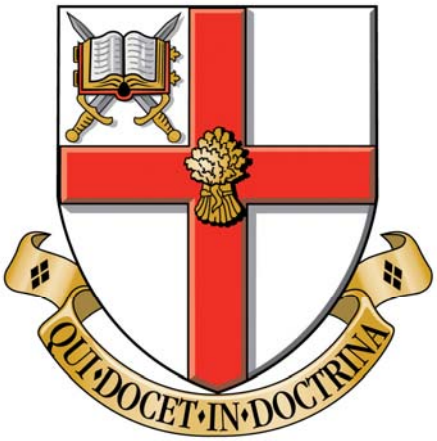

University of Chester

Author(s): Kevin L Lamb; Louise Rogers

Title: A re-appraisal of the reliability of the 20 m multi-stage shuttle run test

Date: 2007

Originally published in: European Journal of Applied Physiology

Example citation: Lamb, K. L. & Rogers, L. (2007). A re-appraisal of the reliability of the 20 m multi-stage shuttle run test. *European Journal of Applied Physiology*, 100, 287-292

A re-appraisal of the reliability of the 20 m multi-stage shuttle run test.

Kevin L. Lamb (✉) **and Louise Rogers**

[1]Department of Sport & Exercise Sciences, University of Chester, Chester, UK

Correspondence

Email: k.lamb@chester.ac.uk

Phone: +44 1244 513425

Fax: +44 1244 511337

Abstract

Most studies that have set out to quantify the test-retest reliability of the Multi-Stage Shuttle Run test (MSSRT) have typically used inappropriate statistics (correlation coefficients) and/or one of two possible calculation methods for deriving predictions of maximal oxygen uptake ($\dot{V}O_{2max}$), and have not considered the impact of habituation on performance in the MSSRT and whether its reliability improves beyond a single repeat trial. In this context, the aim of this study was to assess the reliability of the MSSRT over three trials. Thirty five healthy and active university games players (22 males and 13 females) performed the MSSRT on three occasions, one week apart. Individual total numbers of completed shuttles were recorded and estimated $\dot{V}O_{2max}$ values were derived via established equation and table methods. Analysis revealed that the overall mean $\dot{V}O_{2max}$ score from the equation method (52.5 ± 7.8 ml kg^{-1} min^{-1}) was significantly higher than that for the table (46.9 ± 8.9 ml kg^{-1} min^{-1}), whilst the mean trial 2 and trial 3 scores were significantly higher than that for trial 1, but not different to each other. The Limits of Agreement for the table method were -1.4 ± 5.0 (trial 1-trial 2) and 0.0 ± 5.5 ml kg^{-1} min^{-1} (trial 2-trial 3), and for the equation method -1.1 ± 4.7 (trial 1-trial 2) and 0.0 ± 5.0 ml kg^{-1} min^{-1} (trial 2-trial 3). These results suggest that systematic bias is eliminated after the first trial (due to habituation), but a considerable amount of random error remains, regardless of the type of score calculated. Hence, among our sample, the MSSRT does not appear to be reliable enough for the purpose of monitoring changes in $\dot{V}O_{2max}$ due to non-random reasons.

Keywords Multi-stage shuttle run test, reliability, limits of agreement, habituation

Introduction

The multi-stage 20m shuttle run (Léger and Lambert 1982) is a field-based continuous running test that is used extensively in many sporting and occupational health settings to predict maximal oxygen uptake (Wilkinson et al. 1999). Following a modification to the original protocol (Léger et al. 1988; Ramsbottom et al. 1988) the multi-stage 20m shuttle run test (MSSRT) in its current form requires participants to run between two lines set 20 meters apart at a pace dictated by an audio cassette. It commences at a velocity of 8.5 km hr^{-1} and increases by 0.5 km hr^{-1} every minute thereafter. The MSSRT is individually terminated when a participant cannot maintain the set pace or reaches volitional exhaustion. The velocity corresponding to the last completed stage is used as an independent variable in a regression equation derived by Léger et al. (1988) to predict maximal oxygen uptake ($\dot{V}O_{2max}$). Alternatively, predicted maximal oxygen uptake can be obtained by cross-referencing the final (completed) stage and shuttle number to a table of oxygen uptake values provided by Brewer et al. (1998).

The validity of the MSSRT predictions (against laboratory determined criterion $\dot{V}O_{2max}$ values) has been assessed via correlation analysis among samples across a range of ages and athletic abilities (e.g. Grant et al. 1995; Léger and Gadoury 1989; McNaughton et al. 1998; Paliczka et al. 1987; Wilkinson et al. 1999), and on account of the size of the correlations reported ($r > 0.82 – 0.93$) has generally been reported in favourable terms. Although some reports offer a contrary view, for example O'Gorman et al. 2000 ($r < 0.61$) and St Clair Gibson et al. 1998 ($r < 0.71$), the consensus appears to be that the MSSRT is able to provide an accurate prediction of an individual's $\dot{V}O_{2max}$. However, the matter of the reliability of the MSSRT, that is, the extent to which it can yield consistent scores on a test-retest basis, has yet to be satisfactorily addressed.

Studies that have investigated the test-retest reliability of the MSSRT are less well documented than those dealing with its validity, but by-and-large researchers have maintained (on the basis of test-retest correlation coefficients being in the range 0.87 to 0.97) that it is able to generate repeatable scores, and consequently repeatable $\dot{V}O_{2max}$ predictions. However, the contemporary view of the way in which researchers such as Léger et al. (1988), Mahoney (1992), Sproule et al. (1993), St. Clair Gibson et al. (1998) and Vincent et al. (1999) quantified the reliability of the MSSRT renders such a claim as highly questionable. Since the publication of particular articles (Atkinson 1995; Atkinson and Nevill 1998; Lamb 1998; Hopkins 2000) brought to the attention of sport and exercise scientists the limitations of the correlation approach, it has become increasingly accepted that other statistical techniques, such as the coefficient of variation, the 'Limits of Agreement' (LoA) or the 'Typical Error' should be the principal statistics in reliability analyses. Whilst there has been debate on which technique is the most appropriate, there is a consensus that correlation coefficients are inappropriate and that assessments of the reliability of instruments/tests should be viewed in terms of the amount of measurement that error exists, and whether such error is small enough to allow the instruments/tests to detect or monitor changes between and within groups.

The case for the LoA technique, promoted by Bland and Altman (1986) primarily for its use in method comparison research, is based on the mis-use of the bi-variate Pearson statistic correlation on uni-variate (test-retest) data, and that the size of a correlation, particularly the Pearson, is strongly influenced by the heterogeneity of a sample's scores. In addition, as a correlation simply measures the strength of the linear relationship between two sets of scores (how test and retest scores vary relative to each other) it does not provide information about the absolute differences in scores (within subjects). Moreover, it has often been assumed that a high correlation (say > 0.9) between repeated measurements is indicative of a high level of agreement between scores (that individual test and retest scores were very similar). This is not necessarily the case. Instead, the LoA technique described by Bland and Altman (1986) is superior as it does allow a calculation of

the degree of trial-to-trial variation (error) present in a given sample (in the units of the particular measure), from which a judgement must then be made on the basis of 'analytical goals' (Atkinson and Nevill 1998).

In adopting the above argument, the previous claims for the MSSRT to be a reliable measure have to be challenged if they were based upon the calculation of correlation coefficients. Recently, three studies, one among a small sample of 12-13 year-old girls (Fairbrother et al. 2005) and the other two among male undergraduate students (Cooper et al. 2005; Lamb and Humphreys 2003) have addressed this issue using the LoA technique, but derived opposing conclusions. A reason for this may be due to the method they employed to calculate predicted maximal oxygen uptake, that is, Cooper et al. (2005) used the table of values provided by Brewer et al. (1988), which is based on the data of Ramsbottom et al. (1988), whereas Fairbrother et al. (2005) and Lamb and Humphreys (2003) used the regression equation provided by Léger et al. (1988). In addition, only one previous investigation involving adults has considered that the reliability of the MSSRT might be influenced by a testing effect, possibly due to the process of habituation or varying levels of participant motivation. McVeigh et al. (1995) examined reliability over three trials among adolescents and reported that $\dot{V}O_{2max}$ predictions stabilised between the second and third, suggesting that a practice trial should precede any use of the MSSRT for fitness assessment or monitoring purposes. However, their statistical analyses did not use the contemporary technique. Therefore, the aims of this study were to assess whether the test-retest reliability of the MSSRT is affected by (i) providing a pre-trial and (ii) the method of deriving the predicted $\dot{V}O_{2max}$ scores.

Method

Participants

Twenty-two male (20.9 ± 1.5 years, 79.6 ± 9.3 kg, 1.79 ± 0.07 m) and 13 female (19.6 ± 1.0 years, 67.9 ± 10.8 kg, 1.62 ± 0.07 m) volunteers were recruited from a university population to take part in the study. The participants were active games players (engaged in two or more training/match sessions per week) representing rugby union (14), football (13) and tennis (8), and were free of injury and apparently healthy. After receiving an oral and written explanation of the study, all participants gave their written consent to participate and prior to each testing session, completed a pre-test health status questionnaire. The study was approved by the School of Health and Applied Science's Research Ethics Committee, in accordance with the standards set down by the 1964 Declaration of Helsinki.

Procedures

The study utilised a cross-sectional design involving a single cohort engaged in repeated measures. To facilitate the testing process, participants were assigned to one of three groups according to the sport in which they participated. This strategy enabled measurements to be incorporated into each group's usual training sessions. Each participant was required to perform the 20 m MSSRT on three separate occasions, exactly one week apart, and at the same time of the day. All the sessions were performed indoors on sprung wooden surfaces, in temperatures maintained between 18 to 20 °C. Two out of the three groups performed the three trials in an indoor gymnasium, whereas the other group performed them in an indoor dance studio. Participants were instructed not to eat in the two hours preceding each test and to wear suitable clothing and footwear.

The MSSRT was administered in the manner described by Brewer et al. (1988). This protocol is guided by an audio cassette tape and an accompanying instruction booklet. Prior to each testing session the MSSRT cassette tape was calibrated by timing samples of the 60 s periods signalled by two tones emitted from the tape. On each occasion, the periods sampled were found to be within 0.5

s the time frame used. Following a directed, standardised, light 5-min warm-up, the participants were required to run back and forth between two parallel lines marked 20 m apart, starting at a speed of 8.5 km hr^{-1}. The running pace was regulated by the audiotape which signalled when the participant needed to be at one end of the 20 m course or the other (representing a completed 'shuttle'). Participants were therefore instructed to adjust their pace so that they turned at the line in unison with the emitted tones. The running speed was increased by 0.5 km hr^{-1} every minute, representing a completed stage, or as the audiotape describes it, *level*). Participants were encouraged to complete as many stages of the test as possible (up to the maximum of 19), and each had their test terminated when they either voluntarily withdrew, or when they were judged to be unable to maintain the prescribed pace. That is, he or she was given a warning on the first two occasions that they failed to be within three metres of the line when the signal sounded, and then stopped on the third occasion.

On completion of the MSSRT each participant's final valid stage and number of shuttles score was recorded by the investigator, but withheld from the participant. From this his/her total number of 20 m shuttles completed was derived (via Table 1, below) along with values for predicted $\dot{V}O_{2max}$ values calculated using both the equation provided by Léger et al. (1988) and the table provided by Brewer et al (1998). For the equation method, the speed corresponding to the last valid stage completed by the participant was entered into the regression equation, $Y = 6.0X - 24.4$ (where $Y = \dot{V}O_{2max}$ in ml kg^{-1} min^{-1} and X = speed in km hr^{-1}). For the table method of Brewer et al. (1998), the number of valid stages *and* shuttles achieved by each participant are used to obtain the predicted $\dot{V}O_{2max.}$

[Table 1 about here]

Data Analysis

Mean (± SD) values were calculated for the number of 20 m shuttles completed prior to test termination and the predicted $\dot{V}O_{2max}$ scores over the three trials. Separate two-way repeated measures ANOVAs were used to assess the variability of $\dot{V}O_{2max}$ scores and the total number of shuttles due to the trial and method factors, followed (where appropriate) by multiple Bonferroni-adjusted dependent *t*-tests to identify significant pair-wise differences. The assumption of sphericity was confirmed for the ANOVA effects. The assessment of reliability between repeated trials (trial 1 versus trial 2, trial 2 versus trial 3 and trial 1 versus trial 3) was conducted with the 95% LoA analysis (Bland and Altman 1986). As part of this, the residual errors (test minus retest differences) were inspected for their normality via the Shapiro-Wilk statistic, along with an assessment of their heteroscedasticity (absolute differences against means) via a Pearson correlation. As these random errors (differences) were found to be normally distributed and unrelated ($r < 0.20$, $P > 0.05$) to the magnitude of the scores, the 95% LoA were computed as the mean difference (bias) ±1.96 multiplied by SD of the differences. To enable comparisons with previous reliability studies, Pearson (*r*) *and* intraclass (model 3, 1) correlations were also computed as alternative indicators of the reliability between trials. Except where specified above, the level of significance was set at $P < 0.05$.

Results

The mean (± SD) total number of 20 m shuttles completed by the group was 93.4 (28.8), 98.7 (30.4) and 96.5 (27.7) for trial 1, trial 2, and trial 3, respectively. These equate (approximately) to stage 10/shuttle 10, stage 11/shuttle 5, and stage 11/shuttle 3. The main effect of trial was significant ($F_{2,33} = 6.1$, $P = 0.004$), with post-hoc analyses revealing trial 2 and trial 3 means to be significantly higher than trial 1 ($t = -3.8$, $P = 0.001$ and $t = -2.6$, $P = 0.01$, respectively), but not different to each other ($t = 0.8$, $P = 0.62$).

Mean predicted $\dot{V}O_{2max}$ values calculated via the equation and table methods for each trial are presented in Table 2. The main effects of trials ($F_{2,33}$ = 4.9, P = 0.01) and method ($F_{1,33}$ = 739.9, P<0.0005) were significant, the latter reflecting the higher mean scores obtained from the equation (52.5 ± 7.8 ml kg^{-1} min^{-1}) than the table (46.9 ± 8.9 ml kg^{-1} min^{-1}). The trial × method interaction was not significant ($F_{2,33}$ = 1.2, P = 0.36).

[Table 2 about here]

The trial-to-trial reliability statistics for each method are shown in Table 3. For both methods the correlations are consistently high (>0.92) and the random error component of the LoA statistics is similar (between 4 and 6 units) and does not vary markedly across the three trials. The LoA for the total number of shuttles were -5.3 ± 16.3, -4.3 ± 19.2 and 0.8 ± 18.2 shuttles for trial 1- trial 2, trial 1- trial 3 and trial 2- trial 3, respectively.

[Table 3 about here]

Discussion

The data from this investigation demonstrate that a method-related difference in the magnitude of the predictions of $\dot{V}O_{2max}$, with mean values obtained from the table markedly lower (5.7 ml kg^{-1} min^{-1}) than those obtained from the equation. In addition, irrespective of the method used to calculate $\dot{V}O_{2max}$, the estimates became more reliable after the first trial (whereby systematic bias was eliminated), confirming the findings of McVeigh et al. (1995). However, this 'improvement' (less than 1.4 ml kg^{-1} min^{-1}) is relatively small and its meaningfulness in a practical sense is questionable. Whilst the analysis also revealed high correlations between the second and third trials,

the amounts of random error observed (approximately ± 10% of the mean of the two trials) were not negligible. Moreover, in terms of test performance, based on the total shuttles completed, a participant (worse case scenario) might have completed 18 shuttles more or 18 fewer in trial 3 than in trial 2, reflecting a performance change of ± 18% (approximately 1.5 stages).

Our analyses highlight the inappropriateness of adopting solely a correlation test approach to reliability assessment since the extent of the trial-to-trial (within subjects) variation noted is at odds with the magnitude of the inter- and intra-class correlations. For example, the correlations between trials 2 and 3 for both estimation methods (and total number of shuttles) are high (0.95) and indicative of 'good' reliability, whereas the 95% LoA suggest that a participant with an estimated $\dot{V}O_{2max}$ of say, 55 ml kg^{-1} min^{-1} in trial 2, might have a value as high as 60.5 or as low as 49.5 ml kg^{-1} min^{-1} in trial 3, which in our judgement is less impressive, particularly as alluded to above, that he/she may have completed 18 more shuttles or 18 less.

To place our findings into context, we note that Cooper et al. (2005), who derived their $\dot{V}O_{2max}$ estimates among undergraduate students via the table method, reported test-retest LoA of -0.4 ± 2.7 ml kg^{-1} min^{-1} (almost half that of the current data) which they deemed narrow enough to indicate an acceptable level of reliability (albeit, they acknowledged that due to the heteroscedasticity present in their data, the agreement was poorer amongst the fittest students). Whilst their male-only sample had a higher mean $\dot{V}O_{2max}$ (by about 8 ml kg^{-1} min^{-1}) than our mixed sample, the random error expressed as a proportion of the mean of their two trials was lower at approximately ± 5%. Similarly, LoA of 0.7 ± 3.17 ml kg^{-1} min^{-1} (based on the equation method) for undergraduate males were also interpreted favourably in the study by Lamb and Humphreys (2003), since this equated to an approximate random error of only ± 5.5%. It is hard to elucidate the differences in LoA between the above two studies and the present one, though we might speculate that they are down to

methodological issues, such as the point at which each participant ends the MSSRT, or biological variations occurring between trials, or perhaps that the heterogeneous nature of the current sample and their lower fitness levels are significant factors. Conversely, Fairbrother et al. (2005) concluded from their study of adolescents (12-15 year-olds) that LoA of 0.42 ± 4.46 ml kg^{-1} min^{-1} (derived from the equation method) did not render the MSSRT sufficiently reliable. This interpretation was based on the view that trial-to-trial variability of this size would not enable the detection of genuine changes in cardio-respiratory function that might follow from a training intervention. Such an explanation of the LoA statistics is both meaningful, being based on an analytical goal (Atkinson and Nevill 1998) and necessary, otherwise appraisals of reliability are made arbitrarily. What is deemed "acceptable" or not should be based on a comprehensive understanding of the measure under scrutiny and the *a priori* adoption of what constitutes a tolerable level of measurement error. Given that the participants in the current study were active and quite likely (in the absence of criterion data) to have relatively high $\dot{V}O_{2max}$ values, the degree of within-subject variation observed between trials 2 and 3 might easily mask any meaningful changes in $\dot{V}O_{2max}$ that might occur due to non-random reasons (e.g. training) among such people. That is, if an improvement in $\dot{V}O_{2max}$ of 5 – 25% is feasible among young adults after a suitable aerobic conditioning programme (Foss and Keteyian 1998, p. 320), then the random trial-to-trial variation observed in the current study (± 10%) - in the absence of training – would impact on the ability of the MSSRT to detect such an increase. Accordingly, our analyses do not promote the MSSRT as being a reliable means of predicting maximal aerobic capacity.

The finding that the two estimation methods yielded different mean $\dot{V}O_{2max}$ values is likely to be a consequence of the sample-specific nature of most regression models, the types of performance (predictor) and dependent variables used in the models, the size of the samples and their degree of heterogeneity, and whether they were formulated on data collected following a practise trial. The

table by Brewer et al. (1998) was constructed via the application of linear regression analysis to the criterion $\dot{V}O_{2max}$ data of Ramsbottom et al. (1988) obtained from 74 men and women (aged 19 to 36 years) who were described as being, "involved in physical training on a regular basis" (p. 142). The performance measure was the highest stage achieved, though it was not stated whether these participants were habituated to the MSSRT. The equation established by Léger et al. (1988) was based upon the highest running speed achieved by a sample 87 men and women (aged 20 to 45) of uncertain physical activity habits and fitness levels, who received two repeat trials. Importantly, and surprisingly, the dependent variable in their analysis was only an estimate of $\dot{V}O_{2max}$ (Strickland et al. 2003). Moreover, it was not stated which trial's data were used in the regression analysis.

Since the origin of the MSSRT, the focus of research attention has been on its ability to estimate aerobic fitness. As indicated above, there are problems inherent with this process due to sample selection, the nature of the statistical analyses (error of prediction), and the issue of habituation or practise. Recently, attention has been afforded to addressing the utility of the MSSRT (and other running-based tests) as markers of sport-specific endurance performance, rather than $\dot{V}O_{2max}$ (e.g. Aziz et al. 2005a; Aziz et al. 2005b; Lemmink et al. 2004a; Lemmink et al. 2004b; Sunderland and Nevill 2005). In the manner of the present study, Aziz et al. (2005a) calculated the total number of shuttles completed by their sample of 16 male athletes and games players, referring to it as their 'performance indicator'. Whilst no interpretation was offered, Aziz et al. (2005a) reported the test-retest reliability of this performance in terms that equate to LoA of approximately ± 12 shuttles (1 stage), which compares favourably to the current study's LoA of ± 18 shuttles. In absolute terms, however, the question remains whether such variability is acceptable for monitoring purposes?

Whilst the MSSRT appears to yield estimates that are reliable using a correlation approach, the 95% LoA indicate that the degree of random error is too large for the test to be considered to have sufficient reliability to monitor changes in aerobic capacity in active male and female

undergraduates. The data also indicate that a small learning or habituation effect occurred between the first and second applications of the test, suggesting that, notwithstanding our interpretation above, researchers and practitioners should be minded to provide at least one practice trial before utilising the scores from the MSSRT. Further, exponents of the MSSRT need to be aware that differences exist in the estimates of $\dot{V}O_{2max}$ obtained due to the method by which they are calculated. Given this fact, either consensus is required on which method to adopt, or the validation process should be re-visited. At a more fundamental level, however, since the current study has questioned the reliability of the MSSRT, it follows that unless researchers explore ways in which improvements can be made, there is a case for the test to be abandoned as a credible predictor of aerobic capacity or exercise performance.

References

Atkinson G (1995) A comparison of statistical methods for assessing measurement repeatability in ergonomics research. In: Atkinson G, Reilly T (eds.) Sport, leisure and ergonomics. E & FN Spon, London, pp. 218-222

Atkinson G, Nevill, AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 26: 217-238

Aziz AR, Chia MYH, Teh KC (2005a) Measured maximal oxygen uptake in a multi-stage shuttle test and treadmill-run test in trained athletes. J Sports Med Phys Fitness 45: 306-314

Aziz AR, Tan FHY, Teh KC (2005b) A pilot study comparing two field tests with the treadmill run test in soccer players. J Sports Sci Med 4: 105-102

Bland JM, Altman, DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet i: 307-310

Brewer J, Ramsbottom, R, Williams, C (1998) Multistage fitness test. National Coaching Foundation, Loughborough

Cooper SM, Baker JS, Tong RJ, Roberts E, Hanford M (2005) The repeatability and criterion validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. Br J Sports Med 39: e19

Fairbrother A, Jones MA, Hitchen PJ (2005) Reliability and concurrent validity of the multistage shuttle run test in adolescent athletes. J Sport Sci 23: 196

Foss ML, Keteyian SJ (1998) Fox's physiological basis for exercise and sport. WCB McGraw-Hill, Boston

Grant S, Corbett K, Amjad AM, Wilson J, Aitchison T (1995) A comparison of methods of predicting maximum oxygen uptake. Br J Sports Med 29: 147-152

Hopkins WG (2000) Measures of reliability in sports medicine and science. Sports Med 30: 1-15

Lamb KL (1998) Test-retest reliability analysis in quantitative physical education research: a commentary. Eur Phys Ed Rev 4: 145-152.

Lamb KL, Humphreys L (2003) The test-retest reliability and method comparison validity of the 20m multi-stage shuttle run and the 12-minute Cooper run/walk estimates of maximal aerobic capacity. In: Muller E, Schwameder H, Zallinger G, Fastenbauer V (eds.) Proceedings of the 8th Annual Congress of the European College of Sport Science, Salzburg, 9th - 12th July. Institute of Sport Science, University of Salzburg, p. 449

Léger L, Gadoury C (1989) Validity of the 20 m shuttle run test with 1 min stages to predict V02 max in adults. Can J Sport Sci 14: 21-26

Léger LA, Lambert JA (1982) A maximal multistage 20 meter shuttle run test to predict V02 max. Eur J App Phys 49: 1-12

Léger LA, Mercier D, Gadoury C, Lambert J (1988) The multistage 20 meter shuttle run test for aerobic fitness. J Sports Sci 6: 93-101

Lemmink KAPM, Verheijen R, Visscher C (2004a) The discriminative power of the Interval Shuttle Run Test and the Maximal Multistage Shuttle Run Test for playing level of soccer. J Sport Med Phys Fitness 43: 233-239

Lemmink KA, Visscher C, Lambert MI, Lamberts RP (2004b) The interval shuttle run test for intermittent sport players: evaluation of reliability. J Strength Cond Res18: 821-827

Mahoney C (1992) 20-MST and PWC170 validity in non-Caucasian children in the UK. Br J Sports Med 26: 45-47

McNaughton L, Hall P, Cooley D (1998) Validation of several methods of estimating maximal oxygen uptake in young men. Percept Mot Skills 87: 575-584

McVeigh SK, Payne AC, Scott S (1995) The reliability and validity of the 20-Meter Shuttle Test as a predictor of peak oxygen uptake in Edinburgh school children, age 13 to 14 years. Pediatr Exerc Sci 7: 69-70

O'Gorman D, Hunter A, McDonnacha C, Kirwan JP (2000) Validity of field tests for evaluating endurance capacity in competitive and international level sports participants. J Strength Cond Res 14: 62-67

Paliczka VJ, Nichols AK, Boreham AC (1987) A multi-stage run as a predictor of running performance and maximal oxygen uptake in adults. Br J Sports Med 21: 163-165

Ramsbottom R, Brewer J, Williams C (1988) A progressive shuttle run test to estimate maximal oxygen uptake. Br J Sports Med 22: 141-144

Sproule J, Kunalan C, McNeill M, Wright H (1993) Validity of 20-MST for predicting V02 max of adult Singaporean athletes. Br J Sports Med 27: 202-203

St Clair Gibson A, Lambert MI, Hawley JA, Broomhead S, Noakes TD (1998) Measurement of maximal oxygen uptake from two different laboratory protocols in runners and squash players. Med Sci Sports Exerc 31: 1226-1229

Strickland MK, Petersen SR, Bouffard M (2003) Prediction of maximal aerobic power from the 20-m multi-stage shuttle run. Can J App Physiol 28: 272-282

Sunderland C, Nevill ME (2005) High-intensity intermittent running and field hockey skill performance in the heat. J Sports Sci 23: 531-540

Vincent SD, Barker R, Clarke M, Harrison J (1999) A comparison of peak heart rates elicited by the 1-Mile Run/Walk and the progressive aerobic cardiovascular endurance run. Res Quart Exerc Sport 70: 75-78

Wilkinson DM, Fallowfield JL, Myers SD (1999) A modified incremental shuttle run test for the determination of peak shuttle running speed and the prediction of maximal oxygen uptake. J Sports Sci 17: 413-419

Table 1. MSSRT stages and cumulative number of shuttles

Stage	Number of Shuttles	Cumulative Total
1	7	7
2	8	15
3	8	23
4	9	32
5	9	41
6	10	51
7	10	61
8	11	72
9	11	83
10	11	94
11	12	106
12	12	118
13	13	131
14	13	144
15	13	157
16	14	171
17	14	185
18	15	200
19	15	215

Table 2. Mean (SD) predicted $\dot{V}O_{2max}$ values (ml kg^{-1} min^{-1}) by method

	Method	
Trial	Table	Equation
1	46.3 (9.0)	52.1 (7.8)
2	47.7 (9.4)	53.3 (8.4)
3	47.3 (8.5)	52.9 (7.6)

Table 3. Reliability statistics for predicted $\dot{V}O_{2max}$ scores.

Comparison	Table Method			Equation Method		
	LoA[1]	*r*	IC	LoA[1]	*r*	IC
Trial 1 – Trial 2	-1.4 ± 5.0	0.96*	0.95	-1.1 ± 4.7	0.96*	0.95
Trial 2 – Trial 3	0.0 ± 5.5	0.95*	0.95	0.0 ± 5.0	0.95*	0.95
Trial 1 – Trial 3	-1.3 ± 5.8	0.94*	0.93	-1.1 ± 5.4	0.94*	0.95

[1] in ml kg^{-1} min^{-1}
* $P < 0.01$