ORIGINAL ARTICLE

A simple multistage field test for the prediction of anaerobic capacity in female games players

S-M Cooper, J S Baker, Z E Eaton, N Matthews

Br J Sports Med 2004;**38**:784–789. doi: 10.1136/bjsm.2004.012229

See end of article for authors' affiliations

Correspondence to: S-M Cooper, University of Wales Institute Cardiff, School of Sport, PE and Recreation, Cyncoed Campus, Cyncoed Road, Cyncoed, Cardiff CF23 6XD, Wales, UK; smcooper@uwic.ac.uk

Accepted 4 May 2004

Objective: To establish the validity of a 15 m multistage shuttle run test (MSRT) as a predictor of anaerobic capacity (expressed as mean power output (MPO) from the 30 second Wingate anaerobic test (WAnT)) in female university standard games players.

Methods: Data came from three phases using a total of 72 players (mean (SD) age 20.3 (1.5) years, body mass 64.9 (8.8) kg, and stature 1.67 (0.04) m). The repeatability of the MSRT was assessed in phase 1 by applying 95% limits of agreement (LoA) to the test and retest results from a random sample of 20 players. In phase 2, linear relations between MPO and performance on the MSRT were investigated in a random sample of 36 players. As a result, a calibration model ($Y = a + bX$) was developed and cross validated in phase 3, in which the remaining 36 players performed both the WAnT and the MSRT. Time (seconds) to volitional exhaustion/disqualification from the MSRT was substituted into the calibration model from which MPO was predicted. The agreement between MPO predicted and MPO measured from the WAnT was quantified using LoA.

Results: Insignificant bias between repeat applications of the MSRT ($mean_{diff}$ (SD_{diff}) = 1.0 (3.5) seconds (4 (14) m), $t = 1.23$, $p = 0.230$) was found from phase 1. Data were homoscedastic ($r = 0.061$, $p = 0.799$) with LoA $\pm$ 6.9 seconds ($\pm$ 27 m). In phase 2 the strongest correlation was between MPO ($W/kg^{0.67}$) and time to volitional exhaustion/disqualification on the MSRT; $r = 0.715$ ($r^2 = 51.1\%$, $p = 0.0005$). As a result, the calibration model developed was: MPO ($W/kg^{0.67}$) = 12.5 + (0.2 $\times$ time (seconds)) with a standard error of prediction of 2.1 $W/kg^{0.67}$. The cross validation in phase 3 showed insignificant bias between measured and predicted MPO ($mean_{diff}$ (SD_{diff}) = 0.3 (2.8) $W/kg^{0.67}$, $t = 0.75$, $p = 0.460$). Data were homoscedastic ($r = 0.05$, $p = 0.774$) with LoA $\pm$ 5.5 $W/kg^{0.67}$.

Conclusions: The MSRT requires minimal equipment and training of assessors, and it is easy to perform. In the population studied, it provides scores that are repeatable, and anaerobic capacity (MPO) can be successfully predicted from its performance. It would seem therefore to be a useful field based test for use by female games players, their coaches, and support scientists.

In the case of estimating the contribution of the aerobic component to athletic performance, procedures for the direct determination of maximal oxygen uptake, using incremental exercise tests to volitional exhaustion, and non-invasive methods to determine oxygen consumption are well established.[1] These maximal oxygen uptake tests have long been considered the criterion gold standards against which to validate simple field tests of aerobic performance.[2–4]

However, the ability to tolerate high rates of energy expenditure over time—a capacity for intense activity—is one of the most difficult components of athletic performance to objectively quantify.[5 6] In attempting to develop simple field tests of anaerobic capacity, scientists have struggled to agree on a criterion physiological test that will assess the anaerobic contribution to total energy supply. It is now generally considered that, ideally, needle biopsies (muscle metabolites) and arterial and venous cannulation (blood metabolites) should be used to assess the anaerobic energy supplied during short bouts of exercise.[7] Clearly, such techniques are not only invasive but also lack practical applicability because of the sensitive equipment required, the standard of training needed by the testers, and the time to assess each subject.

There is increased agreement that maximal accumulated oxygen deficit[8 9] is an appropriate physiological measure to use as a practical alternative criterion non-invasive physiological test of anaerobic capacity.[7 10] Even so, some authors have questioned the test methodology for determining maximal accumulated oxygen deficit.[11] The technique still requires large investments in terms of equipment, training, and time, and often involves multiple estimations, which can be difficult to obtain.[12]

Some scientists consider that the Wingate anaerobic test[13] is the most sensitive and reliable assessment of anaerobic performance available for many sports performers.[6 14] The test provides two indices of anaerobic performance: peak power output and mean power output. These values are usually associated with maximal rates of ATP splitting (power) and total anaerobic ATP supply (capacity). Although it is doubtful whether the Wingate anaerobic test is the most valid measure of anaerobic performance, it is certainly the most often used. Regardless of its ubiquity, however, it could also be argued that it too lacks practical applicability because of the laboratory based nature of the equipment needed and the standard of training required by the testers. Nevertheless, it is the most popular test used to estimate both anaerobic power and capacity in physically active subjects.[15]

Therefore an easily administered field based test that reliably estimates anaerobic capacity during all out exercise, which requires the minimum of equipment and training of test administrators, would be a very useful addition to the assessment armoury of both sports scientists and sports coaches. The aim of this study was to examine the validity of using a field based, multistage shuttle run test (MSRT) to predict anaerobic capacity as expressed by mean power

Table 1 Data summary for all measured variables and for all phases of the study

Variable	Phase 1 (n = 20)	Phase 2 (n = 36)	Phase 3 (n = 36)
Age (years)	20.4 (1.4)	20.3 (1.1)	20.2 (1.9)
Stature (m)	1.68 (0.05)	1.62 (0.03)	1.70 (0.05)
Mass (kg)	64.3 (10.1)	65.0 (8.4)	65.3 (7.9)
MSRT (shuttles)	18 (3)* 18 (3)†	17 (3)	
MSRT (m)	266 (45)* 263 (46)†	254 (44)	
MSRT (seconds)	70.5 (11.4)* 69.5 (11.5)†	70.9 (10.8)	70.7 (12.7)
WAnT MPO (W)		438 (66)	
WAnT MPO (W/kg$^{0.67}$)		26.7 (3.0)	27.0 (3.6)‡ 26.6 (2.5)¶

Values are mean (SD). Phase 1, Repeatability; phase 2, calibration; phase 3, cross validation.
*Test.
†Retest.
‡Measured data.
¶Predicted data.
MSRT, Multistage shuttle run test; WAnT, Wingate anaerobic test; MPO, mean power output.

output values gathered from the Wingate anaerobic test, in a group of female university standard games players.

METHODS

Subjects

Seventy two female games players (mean (SD) age 20.3 (1.5) years, body mass 64.9 (8.8) kg, and stature 1.67 (0.04) m), all of whom were regular playing members of British Universities Sports Association championship squads (netball, rugby union, and hockey), gave written informed consent and volunteered to act as subjects. Before data collection, the relevant university research ethics subcommittee approved all procedures proposed in the study. All subjects were familiarised with all performance measures before data collection. Each subject performed only one of the tests outlined on any given day.

Data collection procedures

Data were gathered from three separate phases. Phase 1 aimed to establish the repeatability of a MSRT in a group of 20 games players randomly drawn from the 72 volunteers. Each player performed the MSRT twice, on separate days, with a maximum of seven days between the test and retest. The agreement between scores from the test and retest was quantified using the 95% limits of agreement method.[16] Subjects involved in phase 1 were also used as subjects in either phase 2 or phase 3; they were not used as subjects in all three phases.

Linear relations between maximal intensity exercise performances from the Wingate anaerobic test and performances on the MSRT were investigated in phase 2 of the study. Thirty six games players, randomly drawn from the 72 volunteers, performed a 30 second Wingate anaerobic test and the MSRT at about the same time—that is, the two assessments were made on separate days, but within a maximum of seven days. Mean power output results from the Wingate anaerobic test were used as criterion maximal intensity exercise performance indices, and results gathered from the MSRT were used as predictors. From the data generated from phase 2, a calibration model (linear regression[17]) was developed and cross validated in phase 3.

In phase 3, the remaining 36 games players performed both the 30 second Wingate anaerobic test (criterion, Y variable) and the MSRT (predictor, X variable). As a result, the agreement between subjects' measured mean power outputs from the Wingate anaerobic test and predicted mean power outputs from substitution of the most relevant index from the MSRT into the calibration model developed as a result of phase 2, was quantified using the 95% limits of agreement method.

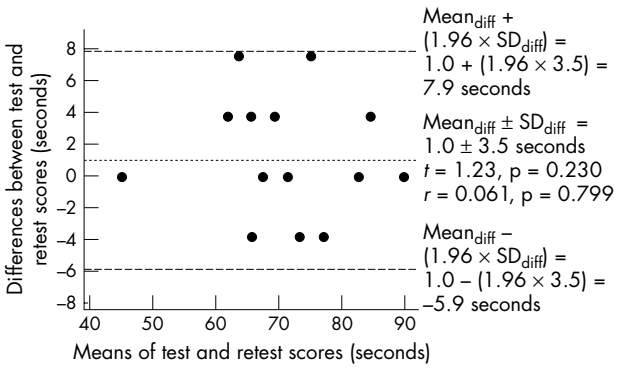

Figure 1 Bland-Altman plot summarising the results from phase 1. Limits of agreement (95%) for repeat applications of the multistage shuttle run test (expressed as time (seconds) to volitional exhaustion/disqualification) have been superimposed on the plot, as have both the bias and the heteroscedasticity summaries. $t_{19}(0.01) = 2.861$, $r_{18}(0.01) = 0.561$ (both two tailed tests).

Criterion maximal intensity exercise performances

Maximal intensity exercise performance was described in terms of mean power output estimated from performing a 30 second Wingate anaerobic test on a friction belt cycle ergometer (Monark 814E), which used a basket weight loading system interfaced to a microcomputer.[13] Mean power output was defined as the arithmetic average of the total work performed during the 30 second test period. The test was conducted in accordance with British Association of Sport and Exercise Sciences (BASES) guidelines[14] and in a laboratory accredited for the purpose by BASES. The external resistive load against which subjects were required to pedal was equivalent to 7.5% of body weight.[18] Subjects were fully habituated to and familiarised with the Wingate anaerobic test procedures on three separate occasions before experimental data collection. Before each administration of the test, subjects were weighed, and then the seat, handlebars, and toe clips of the cycle ergometer were adjusted to the needs of each subject. Each assessment was preceded by a standardised five minute warm up against a 100 W resistance, followed by a five second sprint against the calculated external load. After a five minute recovery, subjects maintained a pedal frequency of 60 rpm before the full braking force was applied. Subjects were required to remain seated throughout the test and were verbally encouraged to pedal maximally. After the test, subjects performed a standardised five minute cool down against a 100 W resistance.

Mean power output was expressed both absolutely (W) and relative to body mass. Relative performance was derived using the surface law exponent: absolute mean power output was divided by body mass raised to the power 0.67 and expressed as W/kg$^{0.67}$.[19–21]

Multistage shuttle run test (MSRT)

The protocol for the MSRT was adapted from the 20 m multistage fitness test described by Brewer *et al*.[22] Modifications to the test included a reduction in running distance from 20 m to 15 m. As in the original test, MSRT running speed cues were indicated by signals emitted from a pre-recorded audiocassette tape. The multistage fitness test audiocassette tape dictates that subjects start running at

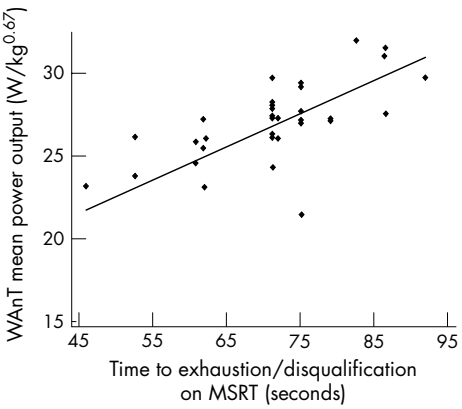

Figure 2 Summary of the results from phase 2. Details of the calibration model developed to predict relative mean power output (W/kg$^{0.67}$) from time to volitional exhaustion/disqualification on the multistage shuttle run test (MSRT) (seconds) are superimposed on the plot of the line of best fit. Mean power output (W/kg$^{0.67}$) = 12.5 + (0.20 × time (seconds)), with a standard error of prediction of 2.1 W/kg$^{0.67}$. r^2 = 51.1%. $r_{34}(0.01)$ = 0.381 (one tailed test).

2.36 m/s which increases by 0.14 m/s each minute. In the MSRT, the tape was modified by doubling the speed so that, at the start of the test, subjects began running at 4.72 m/s, and this was increased by 0.28 m/s every 30 seconds. The changes in running distance and audiocassette tape speed were both proposed as a result of a small pilot study conducted on 15 female university standard games players.

The MSRT was conducted in a sports hall where the test's start point was clearly located on the floor by a line identified with markers. Another similar line, parallel to the start line, was located and marked 15 m away. Before each testing session, the audiocassette tape was checked to ensure that it had not stretched, and the tape player was checked to ensure that it was running at the correct speed. This was achieved by playing the audiocassette tape, at the beginning of which two bleeps were omitted indicating an accurately timed 30 second interval. Accuracy to within ±0.25 second was considered acceptable. Before the MSRT, all subjects performed the same supervised 10 minute warm up, which included prescribed jogging and stretching. Immediately before the test began, subjects were given a familiarisation trial of five low intensity shuttle runs which simulated the test procedures and conditions and in which turning procedures were standardised.

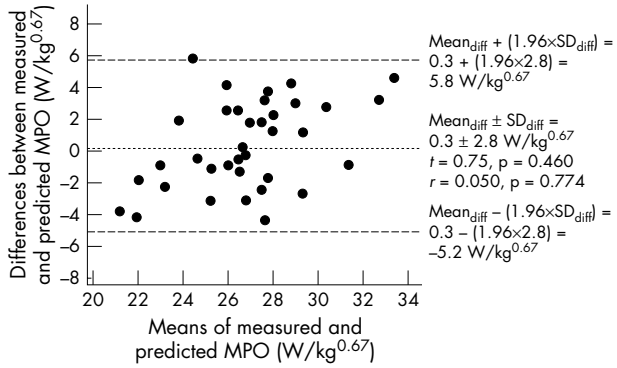

Figure 3 Bland-Altman plot summarising the results from phase 3. Limits of agreement (95%) for measured and predicted mean power output (MPO) have been superimposed on the plot, as have both the bias and heteroscedasticity summaries. $t_{35}(0.01)$ = 2.727, $r_{34}(0.01)$ = 0.418 (both two tailed tests).

Subjects were required to run the test maximally, in groups of five, to add an element of competition and to aid maximal effort. All were verbally encouraged to perform maximally during each test. The protocol consisted of running from the start line to the parallel line, turning, and running back to the start line in time with the signals emitted from the modified audiocassette tape. Subjects continued this pattern of shuttle running until they could run no more (volitional exhaustion) or they failed to make the line in time with the audio signals on two successive occasions which resulted in the assessor disqualifying them from the test. After the test, the same supervised five minute cool down was performed which also included prescribed jogging and stretching.

MSRT results were expressed as: (*a*) the number of shuttles achieved from the start to the point of volitional exhaustion or disqualification; (*b*) the total distance (m) covered during the test; (*c*) the time (seconds) from the start of the test to the point of volitional exhaustion or disqualification. Only fully completed 15 m shuttle runs were recorded.

Statistical analysis

The normality of appropriate data sets was confirmed by the Anderson-Darling normality test (Minitab Inc[23]). It was considered appropriate therefore to test stated hypotheses using parametric statistical methods. As the aim of the study was to validate the MSRT, we were anxious to avoid making type 1 errors. We therefore applied a maximum a priori α level of 0.01 throughout. In phase 1, the degree of agreement between scores gathered from repeat performances on the MSRT (test-retest) was quantified using the 95% limits of agreement method originally described by Bland and Altman.[16] This included plotting a graph (Bland-Altman plot) of the mean of the test and the retest results ((test + retest)/2) for each subject on the x axis, corresponding to the difference (residual errors) between each subject's test and retest results (test − retest) on the y axis. To investigate systematic bias, a dependent *t* test was conducted to test the hypothesis of no difference between the sample mean score for the test and the sample mean score for the retest.

Heteroscedasticity occurs in test data when the amount of random error increases as the measured values increase.[24] Heteroscedasticity was investigated in this study by calculating the zero order correlation coefficient (heteroscedasticity coefficient) between the mean of test and retest scores (indicative of the size of measured values) and the absolute differences between test and retest scores (indicative of random error). If the heteroscedasticity coefficient is close to zero, and residual errors are normally distributed, the 95% limits of agreement can be expressed as ± 1.96 multiplied by the standard deviation of the residual errors—that is, ± 1.96 × SD_{diff}. In such cases, results can be described in the actual units of measurement.[25]

The null hypothesis of no linear relation between the criterion estimates of maximal intensity exercise performance derived from the Wingate anaerobic test and the results generated from the MSRT was analysed in phase 2 using zero order correlation coefficients. Research hypotheses were directionalised, and as a consequence, critical values were established using a one tailed test. Post hoc analysis was conducted using the coefficient of determination (r^2 × 100). From these correlations, linear regression methods were used to develop an equation (a calibration model) to predict a criterion maximal intensity performance variable (Y) from the most appropriate variable generated by the MSRT (X). In deciding on the component variables to use in the final calibration model for cross validation purposes, primary regard was taken of the highest coefficient of determination (adjusted for X) that corresponded to the lowest standard

error of prediction ($\pm SE_{YX}$) resulting from generation of these regression equations.

In the final phase of the study (phase 3), MSRT results were entered into the calibration model developed in phase 2 to predict mean power output for each subject. Cross validation of the model was conducted by quantifying the degree of agreement between subjects' measured mean power output from the Wingate anaerobic test and their predicted mean power output from substitution of the most relevant MSRT values into the calibration model developed. Agreement was once again quantified using the 95% limits of agreement method.

RESULTS

Phase 1: test-retest repeatability of MSRT scores

Two administrations of the MSRT (test-retest) were performed by a group of 20 female games players (mean (SD) age 20.4 (1.4) years, body mass 64.3 (10.1) kg, and stature 1.68 (0.05) m; table 1). The MSRT performances achieved on the test were 70.5 (11.4) seconds, 266 (45) m, 18 (3) shuttles. For the retest they were 69.5 (11.5) seconds, 263 (46) m, 18 (3) shuttles. Figure 1 shows that the dependent *t* test conducted to test the hypothesis of no difference between the mean score for the test versus the mean score for the retest showed no significant bias ($mean_{diff}$ (SD_{diff-}) = 1.0 (3.5) seconds (4 (14) m), $t = 1.23$, $p = 0.230$). The residual errors between scores on the test and retest were normally distributed ($p = 0.06$), and the heteroscedasticity coefficient was $r = 0.061$ ($p = 0.799$). The mean difference (bias) $\pm$ the 95% limits of agreement was 1.0 ± 6.9 seconds (4 ± 27 m).

Phase 2: linear relations between Wingate anaerobic test performances (criterion) and performances on the MSRT (predictors)

Thirty six female games players (age 20.3 (1.1) years, body mass 65.0 (8.4) kg, and stature 1.62 (0.03) m; table 1) performed both the Wingate anaerobic test and the MSRT at about the same time. Mean power output was 26.7 (3.0) $W/kg^{0.67}$ (438 (66) W). MSRT data were 7 (3) shuttles, 70.9 (10.8) seconds, and 254 (44) m. When mean power output was expressed absolutely (W), zero order correlation coefficients were: $r = 0.579$ ($p = 0.0005$) for the relation to distance (m), $r = 0.629$ ($p = 0.0005$) for the relation to the maximum shuttle number achieved, and $r = 0.656$ ($p = 0.0005$) for the relation to time to volitional exhaustion/disqualification. When mean power output data were expressed allometrically relative to body mass ($W/kg^{0.67}$), the values of the correlation coefficients increased. Zero order correlations were: $r = 0.667$ ($p = 0.0005$) for the relation to distance ran, $r = 0.687$ ($p = 0.0005$) for the relation to maximum shuttles achieved, and $r = 0.715$ ($p = 0.0005$) for the relation to time to volitional exhaustion/disqualification.

On the basis of these results and confirmation that the assumption of linearity in these data was met (fig 2), the variables chosen to develop the calibration model to be cross validated in phase 3 were: Y (criterion) = mean power output ($W/kg^{0.67}$) obtained from the laboratory based Wingate anaerobic test (WAnT); X (predictor) = time to volitional exhaustion/disqualification (seconds) from the MSRT. The coefficient of determination between these two variables was 51.1% (adjusted to 49.7%), and the subsequent calibration model developed was: mean power output ($W/kg^{0.67}$) = 12.5 + (0.20 × time (seconds)), with a standard error of prediction of 2.1 $W/kg^{0.67}$.

Table 2 Table of predicted mean power output (MPO) from substitution of multistage shuttle run testing (MSRT) data (time to volitional exhaustion/disqualification) in the calibration model

Shuttle number	Distance ran (m)	Time to exhaustion or disqualification (seconds)	Predicted MPO ($W/kg^{0.67}$)
7	105	30.0	18.5
8	120	33.8	19.3
9	135	37.6	20.0
10	150	41.4	20.8
11	165	45.2	21.5
12	180	49.0	22.3
13	195	52.8	23.1
14	210	56.6	23.8
15	225	60.0	24.5
16	240	63.8	25.3
17	255	67.6	26.0
18	270	71.4	26.8
19	285	75.2	27.5
20	300	79.0	28.3
21	315	82.8	29.1
22	330	86.6	29.8
23	345	90.0	30.5
24	360	93.3	31.2
25	375	96.6	31.8
26	390	99.9	32.5
27	405	103.2	33.1
28	420	106.5	33.8
29	435	109.8	34.5
30	450	113.1	35.1

Additional MSRT indices are also provided for cross reference purposes. Calibration model: MPO ($W/kg^{0.67}$) = 12.5 + (0.2 × time (s)) $\pm$ 2.1 $W/kg^{0.67}$.

Phase 3: cross validation of the calibration model

Both the Wingate anaerobic test and the MSRT were performed by 36 female games players (age 20.2 (1.9) years, body mass 65.3 (7.9) kg, and stature 1.70 (0.05) m; table 1) at about the same time. Table 1 shows that the laboratory determined mean power output for this group was 27.0 (3.6) $W/kg^{0.67}$, and the corresponding predicted mean power output from substitution of time to volitional exhaustion/disqualification from the MSRT in the calibration model developed as a result of phase 2 of the study was 26.6 (2.5) $W/kg^{0.67}$. The dependent *t* test conducted to assess the hypothesis of no difference between mean power output (measured) and mean power output (predicted) showed no significant bias (fig 3, $mean_{diff}$ (SD_{diff}) = 0.3 (2.8) $W/kg^{0.67}$, $t = 0.75$, $p = 0.460$). The heteroscedasticity coefficient was $r = 0.050$ ($p = 0.774$). As the residual errors between measured and predicted mean power output were normally distributed ($p = 0.272$), the mean difference $\pm$ 95% limits of agreement was 0.3 ± 5.5 $W/kg^{0.67}$.

DISCUSSION

Because most previous researchers have estimated test repeatability in terms of correlation coefficients, it was only possible to compare the limits of agreement from phase 1 directly with those from one other study available in the literature. In terms of the correlation, however, the coefficient between the phase 1 test and retest results was encouraging: $r = 0.955$ ($p = 0.0005$). This is considerably higher than the $r = 0.86$ ($p<0.01$) recorded for repeat maximal shuttle runs over 40 m by male undergraduates reported by Baker *et al*.[26] It is also higher than the test-retest correlation ($r = 0.84$; $p<0.01$) recorded by Ramsbottom *et al*,[7] who used the combined results from two high intensity 20 m shuttle runs by 18 male subjects (24.5 (5.0) years) to estimate anaerobic capacity (determined as maximal accumulated oxygen deficit).

The present test-retest correlation illustrates one of the major weaknesses of the coefficient as a measure of repeatability: the calculated value is highly influenced by the range of values of the characteristic being analysed—that is, data heterogeneity.[27–29] Both Baker *et al*[26] and Ramsbottom *et al*[7] used

subject samples of active young men of various abilities in their repeatability studies. In contrast, our subjects were female university games players of similar age and stature, but who were of variable body mass, physical condition, and training status, and who were drawn from three different sports (netball, rugby union, and hockey). As a consequence, even though they were collectively classified as games players, it is likely that the resulting test-retest correlation coefficient is indicative of the heterogeneity of the sample group.

The Bland-Altman plot (fig 1) provides a visual indication of both systematic bias and random error. It can be seen from both the direction and size of the data scatter around the zero line (y axis), that there is some evidence of a positive systematic bias and also some evidence of random error in these data. However, there is no evidence that the size of the residual errors depends on the size of individual mean scores (heteroscedasticity).

In the test-retest data from phase 1, the 95% limits of agreement for the repeatability of the MSRT were given as $\pm$6.9 seconds ($\pm$27 m). Atkinson and Nevill[24] suggested that, when there is no significant systematic bias, as was the case in phase 1, there is a rationale for expressing the limits of agreement as $\pm$ the value of this bias. Consequently, 95% of the differences between test and retest scores should therefore be expected to lie within the limits: 1.0$\pm$6.9 seconds (4$\pm$27 m) or $-$5.9 to 7.9 seconds ($-$23 to 31 m) regardless of the subject's MSRT performance. The examination of heteroscedasticity resulted in a coefficient of r = 0.061 (p = 0.799). The assumption that the limits of agreement remain constant throughout the range of measurements can therefore be accepted.[16]

To put these results into a practical context, if a subject from the study population presented with an estimated MSRT performance of 71.4 seconds (270 m) on the first test, the worst case scenario is that the same subject could score as low as 65.5 seconds (247 m) or as high as 79.3 seconds (301 m) on the retest. Indeed, these results compare favourably with those from the only other study available in the literature that expresses the test-retest repeatability of a high intensity shuttle run test in terms of limits of agreement. Ramsbottom *et al*[7] also modified the 20 m multistage fitness test to predict maximal accumulated oxygen deficit as an indicator of anaerobic capacity in a sample of 18 young men. Their results showed no significant difference between the test and retest (402 (85) m *v* 399 (95) m, p>0.05) with 95% limits of agreement of $-$5 to 5 shuttles (or $-$104 to 97 m). Even accounting for the 5 m difference in distance per shuttle between the MSRT and the test used in the study of Ramsbottom *et al*,[7] it is clear that the limits of agreement from phase 1 of the present study are considerably narrower.

The results from phase 2 showed significant (p<0.01) coefficients in all of the correlations between Wingate anaerobic test scores and the variables derived from performing the MSRT. However, the largest correlation coefficients were found when MSRT indices were correlated with mean power outputs expressed relative to body mass (W/kg$^{0.67}$), rather than when expressed absolutely (W). These results support the work of Tharp *et al*,[30] who identified that, in 10–15 year old boys (n = 56), mean power outputs obtained from a 30 second Wingate anaerobic test were better predictors of sprinting performance (50 yard dash) when expressed relative to body mass (r = $-$0.69, p<0.01) than when expressed absolutely (r = $-$0.53, p<0.05). It is important to note that Tharp *et al*[30] calculated relative mean power output as the commonly applied ratio standard (mean power output (W) divided by absolute body mass (kg) expressed as W/kg) and not as we have done using the surface law exponent. The results do highlight, however, the importance of body mass as a measure of size and muscle mass (indirectly) and as a component in the assessment of anaerobic performance in athletes identified by previous researchers.[31 32]

The significant relations between MSRT indices and mean power output were not altogether surprising. Many other researchers consider mean power output obtained from the 30 second Wingate anaerobic test to reflect anaerobic capacity.[5 6 14] The mean time that the subjects in phase 2 were able to sustain the MSRT was 70.9 (10.8) seconds, a little over double the duration of the Wingate anaerobic test used and therefore probably reflecting mean power output. Indeed, Inbar *et al*[6] were of the opinion that 45 seconds cycling test protocols would elicit more mechanical work and therefore bring subjects closer to their anaerobic capacity. However, 45 second cycling protocols are problematic in motivational terms, making them less suitable when repeat testing is required. The subjects in our study reported that they found the MSRT more conducive to repeated applications because of the dictated nature of its progressively increasing work intensity.

In the data from phase 2, 49.7% (51.1% unadjusted) of the variance in relative mean power output was accounted for by the MSRT when scores were expressed as time to volitional exhaustion/disqualification. The 50.3% variance unaccounted for might be primarily attributed to the differences between the two tests: (*a*) the MSRT is a running test in which subjects support their own body mass, whereas the Wingate anaerobic test is a fixed cycle ergometer test, so the body mass is supported; (*b*) issues related to the running economy and mechanical efficiency of the subjects, particularly towards the end of the MSRT, which are not identified by the Wingate anaerobic test; (*c*) the disparity in test duration between the Wingate anaerobic test (30 seconds) and the MSRT (70.9 (10.8) seconds); (*d*) the Wingate anaerobic test involves 30 seconds of all out effort, whereas exercise intensity in the MSRT is dictated and is progressively increasing; (*e*) the deleterious effect that deceleration at the end of each shuttle might have on relations with cycle ergometry (previously identified by Baker *et al*[26]).

In the development of useful calibration models, the regression equation established from the results from one sample of the chosen population should be cross validated against results provided by another, equivalent sample. Without cross validation to test the accuracy of the predictions, results will always be suspect.[24 33 34] Indeed, Atkinson and Nevill[24] believe that many of the most commonly used field tests of physiological fitness, which provide tables for the prediction of the directly measured physiological variable from indirect measures, lack this key element of validity.

The Bland-Altman plot (fig 3) provides an indication of both systematic bias and random error between predicted and measured mean power outputs in the sample as a result of phase 3. From both the direction and the size of the scatter of these data around the zero line (y axis), there is little evidence of systematic bias, but some evidence of random error. In addition, there is no evidence in fig 3 that these data are heteroscedastic. Indeed, a coefficient of r = 0.050 (p = 0.774) confirmed that these data were homoscedastic and further confirmed that the assumption that the calculated limits of agreement will remain constant throughout the range of measurements can therefore be accepted.[16] About 95% of the differences between measured and predicted mean power output scores for female university games players would therefore be expected to lie within the limits of 0.3$\pm$5.5 W/kg$^{0.67}$—that is, from $-$5.2 to 5.8 W/kg$^{0.67}$ regardless of performance.

In the case of the limits of agreement identified as a result of phase 3, for a subject from the population considered, it

would be expected (a 95% probability) that the variability between laboratory determined mean power output and mean power output predicted from the calibration model developed would lie within the calculated limits of agreement (−5.2 to 5.8 W/kg$^{0.67}$). That is, in the case of a female university games player who presented with a laboratory determined mean power output of 25 W/kg$^{0.67}$ from performing the Wingate anaerobic test, there is a 95% probability that substitution of her time to volitional exhaustion/disqualification on the MSRT into the calibration model would result in a predicted mean power output as low as 19.8 W/kg$^{0.67}$ or as high as 30.8 W/kg$^{0.67}$.

As a consequence of these results, we are prepared to acknowledge that there is some doubt about whether the MSRT is sensitive enough to identify the small changes in performance that might accompany the improved training status of a female games player who already has a highly developed anaerobic capacity. The discriminating ability of the MSRT was not part of the original research design for this study, but we are of the opinion that it should form part of any future developments with the test.

From the results of the study, estimates of anaerobic capacity (mean power output (W/kg$^{0.67}$)) derived from appropriate data gathered from performing the MSRT were prepared for use by the coaches of the university sports academies attended by the subjects, and have been reproduced for information here (table 2).

CONCLUSIONS

The equipment required to perform the MSRT is easy to obtain. Indeed, it should not be beyond the ingenuity of any athlete, coach, or sports scientist to adapt the 20 m multistage fitness test[22] audiocassette tape as we did to replicate that needed to administer the MSRT. In addition, the test itself is easy to perform, and it requires little training for the assessors. From the results of this study, we conclude that the calculated 95% limits of agreement were narrow enough for the MSRT to be considered repeatable when used with female university standard games players. The results also suggest that the anaerobic capacity of this population, reflected by mean power output relative to body mass (W/kg$^{0.67}$), can be successfully predicted from a calibration model, from the time (seconds) to the point of volitional exhaustion/disqualification on the MSRT. When cross validated, the calibration model yielded 95% limits of agreement, which are probably too wide to conclude that the MSRT can be used to monitor the small changes in anaerobic capacity that might result from the improved training status of an already well trained female games player. The test might prove useful, however, in predicting the more substantial effect that could accompany anaerobic training conducted by less well trained female games players. We believe therefore that the MSRT could be used occasionally by female games players to provide them, or their coaches or support scientists, with a snapshot of their anaerobic capacity. Alternatively, it could be used long term as an easily administered field based test for monitoring a player's progress through an anaerobic fitness training programme.

Authors' affiliations
S-M Cooper, Z E Eaton, N Matthews, University of Wales Institute Cardiff, Wales, UK
J S Baker, University of Glamorgan, Wales, UK

Conflict of interest: none declared

REFERENCES

1 **Lange Andersen K**, Shephard R, Denolin H, *et al. Fundamentals of exercise testing*. Geneva: World Health Organisation, 1971.
2 **Léger L**, Lambert J. A maximal multistage 20-m shuttle run test to predict VO_{2max}. *Eur J Appl Physiol* 1982;**49**:1–12.
3 **Léger L**, Mercier D, Gadoury C, *et al*. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sc* 1988;**6**:93–101.
4 **Ramsbottom R**, Brewer J, Williams C. A progressive shuttle run test to estimate maximal oxygen uptake. *Br J Sports Med* 1988;**22**:141–4.
5 **Bouchard C**, Taylor A, Simoneau J-A, *et al*. Testing anaerobic power and capacity. In: MacDougall J, Wenger H, Green H, eds. *Physiological testing of the high-performance athlete*. Champaign, IL: Human Kinetics, 1991:175–221.
6 **Inbar O**, Bar-Or O, Skinner J. *The Wingate anaerobic test*. Champaign IL: Human Kinetics, 1996.
7 **Ramsbottom R**, Nevill M, Nevill A, *et al*. Accumulated oxygen deficit and shuttle run performance in physically active men and women. *J Sports Sci* 1997;**15**:207–14.
8 **Medbø J**, Mohn A-C, Tabata I, *et al*. Anaerobic capacity determined by maximal accumulated O_2 deficit. *J Appl Physiol* 1988;**64**:50–60.
9 **Medbø J**, Tabata I. Relative importance of aerobic and anaerobic energy release during short-lasting exhausting bicycle exercise. *J Appl Physiol* 1989;**67**:1881–6.
10 **Green S**, Dawson B. Measurement of anaerobic capacities in humans: definitions, limitations and unsolved problems. *Sports Med* 1993;**15**:312–27.
11 **Bangsbø J**. Is the O_2 deficit an accurate quantitative measure of anaerobic energy production during intense exercise? *J Appl Physiol* 1992;**73**:1207–8.
12 **Maxwell N**, Nimmo M. Anaerobic capacity: a maximal anaerobic running test versus the maximal accumulated oxygen deficit. *Can J Appl Physiol* 1996;**21**:35–47.
13 **Bar-Or O**. The Wingate anaerobic test: an update of methodology, reliability and validity. *Sports Med* 1987;**4**:381–94.
14 In: Bird S, Davidson R, eds. *Guidelines for the physiological testing of athletes*, 3rd ed. Leeds: British Association of Sport & Exercise Sciences, 1997.
15 **Scott C**, Roby F, Lohman T, *et al*. The maximally accumulated oxygen deficit as an indicator of anaerobic capacity. *Med Sci Sports Exerc* 1991;**23**:618–24.
16 **Bland J**, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**i**:307–10.
17 **Atkinson G**, Nevill A. Selected issues in the design and analysis of sport performance research. *J Sports Sci* 2001;**19**:811–27.
18 **Winter E**, MacLaren D. Assessment of maximal-intensity exercise. In: Eston R, Reilly T, eds. *Kinanthropometry and exercise physiology laboratory manual: volume 2: exercise physiology*. 2nd ed. London: E & FN Spon, 2001:263–88.
19 **Schmidt-Nielsen K**. *Scaling: why is animal size so important?* Cambridge: Cambridge University Press, 1984.
20 **Nevill A**, Ramsbottom R, Williams C, *et al*. Scaling individuals of different body size. *J Sports Sci* 1992;**9**:427–8.
21 **Winter E**, Nevill A. Scaling: adjusting for differences in body size. In: Eston R, Reilly T, eds. *Kinanthropometry and exercise physiology laboratory manual: volume 2: exercise physiology*. 2nd ed. London: E & FN Spon, 2001:275–93.
22 **Brewer J**, Ramsbottom R, Williams C. *Multistage fitness test*. Leeds: National Coaching Foundation, 1988.
23 **Minitab Inc**. M*initab Reference Manual*. Philadelphia: Minitab Inc, 1995.
24 **Atkinson G**, Nevill A. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998;**26**:217–38.
25 **Nevill A**, Atkinson G. Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. *Br J Sports Med* 1997;**31**:314–18.
26 **Baker J**, Ramsbottom R, Hazeldine R. Maximal shuttle running over 40 m as a measure of maximal exercise performance. *Br J Sports Med* 1993;**27**:228–32.
27 **Bland J**, Altman D. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983;**32**:307–17.
28 **Atkinson G**. A comparison of statistical methods for assessing measurement repeatability in ergonomics research. In: Atkinson G, Reilly T, eds. *Sport, leisure and ergonomics*. London: E & FN Spon, 1995:218–22.
29 **Bates B**, Zhang S, Dufek J. The effects of sample size and variability on the correlation coefficient. *Med Sci Sports Exerc* 1996;**28**:386–91.
30 **Tharp G**, Newhouse R, Uffleman L, *et al*. Comparison of sprint and run times with performance on the Wingate Anaerobic Test. *Res Q* 1984;**56**:73–6.
31 **Patton J**, Kraemer W, Knuttgen H, *et al*. Factors in maximal power production and in exercise endurance relative to maximal power. *Eur J Appl Physiol* 1990;**60**:222–7.
32 **Davis J**, Brewer J. Applied physiology of female soccer players. *Sports Med* 1993;**16**:180–9.
33 **Nevill A**. Validity and measurement agreement in sports performance [editorial]. *J Sports Sci* 1996;**14**:199.
34 **Vincent W**. *Statistics in kinesiology*, 2nd ed. Champaign, IL: Human Kinetics, 1999.