Validity and Reliability of Predicting Maximum Oxygen Uptake via Field Tests in Children and Adolescents

Michael J. Buono, Julia J. Roby, Frank G. Micale, James F. Sallis, and W. Elizabeth Shepard

The purpose of this study was to determine the validity and reliability of three of the most commonly used field tests to predict maximum oxygen uptake in children and adolescents. VO₂max was directly measured during a maximal treadmill test in 90 children (10-18 yrs). Each subject also performed, in duplicate, a timed distance run (1 mile), a step test, and a submaximal cycle ergometer test. A multiple regression equation was developed with directly measured VO₂max as the dependent variable and mile-run time, gender, skinfold thickness, and body weight as independent variables. The equation had a multiple $R=0.84$ and a standard error of estimate of 9%, or 4.3 ml/kg/min. The results suggest that when the three most commonly used field tests to predict aerobic capacity are compared in the same group of children, the timed distance run is superior in both validity and reliability.

In recent years the assessment of cardiorespiratory fitness in children and adolescents has grown in importance because of findings linking this variable with cardiovascular disease risk factors (11, 13). Maximal oxygen uptake (VO₂max) is generally regarded as the best measure of cardiorespiratory fitness. The direct (i.e., laboratory) measurement of VO₂max, however, is time-consuming, expensive, and requires cumbersome instrumentation. Therefore it is generally not applicable for most public school settings, health promotion programs, or epidemiologic field studies.

Over the years numerous field tests have been developed to predict VO₂max in children and adolescents. The three most commonly used ones are the timed distance run (6, 7), step test (2, 5), and submaximal cycle ergometer test (3, 14). A review of the literature revealed that although numerous studies...
have examined the relationship between any one of the above three predictive tests and \( V_{O2}\text{max} \), no study has examined the potential use of all three predictive tests in the same group of children. In light of the above, the purpose of this study was to determine the validity and reliability of a timed distance run, cycle ergometer test, and step test in a group of children and adolescents.

**Methods**

The subjects for the study were 90 students from a public school setting in Poway, California. Thirty (15 boys, 15 girls) students were recruited from each of three grade school levels (5th, 8th, and 11th). All were volunteers, and written informed consent was obtained from both child and parent prior to testing. In addition, the children were screened by a pediatrician and cleared for maximal exercise. Descriptive and test data for the group are presented in Table 1.

All subjects performed a graded treadmill test to exhaustion, during which expiratory gases were collected each minute of the test to determine \( V_{O2}\text{max} \). The test protocol consisted of 3-min stages at 3, 4, 5, and 6 mph, followed by increases in grade of 2.5% per minute. Similar running protocols have been successfully used in the past to obtain \( V_{O2}\text{max} \) values in children (10). Oxygen uptake was measured via open-circuit spirometry using a calibrated Beckman metabolic cart. Mean \((\pm SD)\) \( V_{O2}\text{max} \) for the group was \( 47.7 \pm 7.7 \text{ ml/kg/min} \). This value is similar in magnitude to what others have reported in the literature (6, 7). Heart rate was measured using a Hewlett-Packard model 1500 electrocardiograph.

In the current study, 94% of the subjects met at least one of three commonly used criteria for maximal effort. Specifically, 91% of them attained 90% of their age-predicted (220 minus age) maximal heart rate, 74% attained a respiratory exchange ratio of 1.0 or greater, and 57% had a difference of 2.1 ml/kg/min or less for the last two stages of their treadmill test (10).

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th>5th</th>
<th>8th</th>
<th>11th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td>10.6 ± 0.5</td>
<td>12.8 ± 0.6</td>
<td>16.1 ± 1.1</td>
</tr>
<tr>
<td>Ht (cm)</td>
<td>146.1 ± 6.4</td>
<td>160.0 ± 9.6</td>
<td>163.8 ± 8.1</td>
</tr>
<tr>
<td>Wt (kg)</td>
<td>38.6 ± 7.1</td>
<td>47.8 ± 8.7</td>
<td>56.1 ± 7.8</td>
</tr>
<tr>
<td>Skinfolds (mm)*</td>
<td>21.8 ± 10.1</td>
<td>23.6 ± 9.4</td>
<td>29.3 ± 10.2</td>
</tr>
<tr>
<td>Measured ( V_{O2}\text{max} ) (ml/kg/min)</td>
<td>47.8 ± 7.1</td>
<td>47.6 ± 8.2</td>
<td>48.0 ± 8.1</td>
</tr>
<tr>
<td>Mile-run time (sec)</td>
<td>583.0 ± 86.0</td>
<td>487.0 ± 87.0</td>
<td>493.0 ± 98.0</td>
</tr>
<tr>
<td>Step test HR (bpm)</td>
<td>156.0 ± 14.6</td>
<td>147.5 ± 18.2</td>
<td>149.6 ± 23.7</td>
</tr>
<tr>
<td>Predicted ( V_{O2}\text{max} ) from cycle ergometry (ml/kg/min)</td>
<td>38.5 ± 7.6</td>
<td>34.9 ± 9.6</td>
<td>41.0 ± 6.3</td>
</tr>
</tbody>
</table>

Values are mean ± standard deviations.

*Sum of tricep and subscapular skinfold.
Each subject also performed two trials of each of three predictive tests (i.e., timed distance run, step test, cycle test). The sequence of the tests was randomly assigned and 1 or 2 days separated each test, depending on the student’s class schedule.

The timed distance run consisted of having the subject run 1 mile on a 1/4-mile oval dirt track. Subjects ran individually and were motivated to give their best effort. Running time and encouragement were given at each 1/4 mile, and mile time was recorded to the nearest second.

The step test was conducted according to the procedure suggested by McArdle, Katch, and Katch (9). Subjects stepped for 3 minutes on a 16-1/4-inch bench. Stepping rate was maintained by a metronome, and heart rate was determined via palpation by a trained investigator for 15 seconds (5-sec post-exercise to 20 sec postexercise) following the test. The 15-sec heart rate was then converted to beats per minute by multiplying the value by 4.

The cycle ergometer test was conducted on a calibrated Monark (model 868) ergometer. A modification of the Siconolfi et al. (12) protocol was employed in which the subject pedaled at 50 revolutions per minute throughout the test. The initial workload was 25 W, and it was increased 25 W every 3 minutes until the subject reached 70% of his or her age-predicted (220 minus age) maximal heart rate. After reaching this point the subjects exercised at the same workload for an additional 3 minutes or until they attained consecutive (1 minute apart) heart rates that differed by 5 bpm or less. Heart rates were obtained via palpation by a trained investigator. VO2max was estimated from the Astrand (1) nomogram using the mean steady-state heart rate from the final 3-min workload.

Tricep and subscapular skinfold thickness was measured, in triplicate, using Lange calipers according to the procedure outlined by Jackson and Pollock (4). Body weight was measured to the nearest 0.1 kg on a balance beam scale.

Validity correlations were determined using Pearson product moment correlations comparing measured VO2max (ml/kg/min) versus run time (sec), step-test heart rate (bpm), and predicted VO2max from the Astrand nomogram. Test-retest reliability was determined via intraclass correlation procedures comparing Trial 1 versus Trial 2 for the three predictive tests. Furthermore, multiple regression analysis was used to develop the best equation for predicting VO2max.

Results and Discussion

Table 2 shows the test-retest reliabilities for the three field tests, both for the entire sample (n=90) and for each of the three grade levels studied. The timed distance run is superior in reliability (R=0.95) as compared to the other two predictive tests at all grade levels. These results would suggest that day-to-day variability in near maximal run performance is significantly less than the submaximal heart rate response to exercise.

Table 3 shows the validity correlations for the three tests. Again, the timed distance run had a higher correlation with VO2max (ml/kg/min) than the other two tests at all grade levels. Furthermore, it was the only one with an r>0.60, which has been used in the past (7, 8) as the lower limit for a useful fitness test.

Such findings agree with the results of Burke (2), who determined the validity of cycle, step, and run tests in college males (ages 17–30 yrs). He re-
Table 2
Test-Retest Reliability for the 3 Predictive Tests

<table>
<thead>
<tr>
<th></th>
<th>Distance run</th>
<th>Step test</th>
<th>Cycle test</th>
</tr>
</thead>
<tbody>
<tr>
<td>5th grade (n=30)</td>
<td>0.91</td>
<td>0.75</td>
<td>0.88</td>
</tr>
<tr>
<td>8th grade (n=30)</td>
<td>0.93</td>
<td>0.87</td>
<td>0.62</td>
</tr>
<tr>
<td>11th grade (n=30)</td>
<td>0.98</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>Entire sample (n=90)</td>
<td>0.95</td>
<td>0.82</td>
<td>0.77</td>
</tr>
</tbody>
</table>

*p<0.01 for all.

Table 3
Validity Correlations for the 3 Predictive Tests

<table>
<thead>
<tr>
<th></th>
<th>Distance run</th>
<th>Step test</th>
<th>Cycle test</th>
</tr>
</thead>
<tbody>
<tr>
<td>5th grade (n=30)</td>
<td>-0.76*</td>
<td>-0.36</td>
<td>0.51*</td>
</tr>
<tr>
<td>8th grade (n=30)</td>
<td>-0.80*</td>
<td>-0.49*</td>
<td>0.34</td>
</tr>
<tr>
<td>11th grade (n=30)</td>
<td>-0.85*</td>
<td>-0.56*</td>
<td>0.80*</td>
</tr>
<tr>
<td>Entire sample (n=90)</td>
<td>-0.73*</td>
<td>-0.48*</td>
<td>0.49*</td>
</tr>
</tbody>
</table>

*p<0.01.

ported correlations of -0.74 for the mile run, 0.62 for the Astrand cycle test, and 0.55 for the Harvard step test. More recently, Zwiren et al. (15) reported validity correlations of 0.79 for the 1.5-mile run, 0.66 for the Astrand cycle test, and 0.55 for the Queens College Step Test in a group of 30- to 39-year-old women. The above findings (2, 15) are similar to the validity correlations of -0.73 (run), 0.49 (cycle), and -0.48 (step) found in the current work. All of these studies suggest that, in general, the performance based field tests had greater shared variance with VO2max than did the submaximal heart rate tests.

At least three factors may be responsible for this trend. First, the different modes of ergometry used may have played a role. Typically, American children and adults are not as accustomed to bench stepping and cycle riding as they are to walking and running. Therefore a field running test may relate closer to treadmill VO2max as compared to either a step test or a cycle test simply because it is the same mode of exercise.

Second, these results seem to indicate the advantages of near-maximal tests in predicting aerobic capacity, both in children and adults. Specifically, in the current study the mean postrun heart rate for the group was 178 bpm, or 89% of their treadmill max heart rate (M=201 bpm). This value is significantly greater
than the mean heart rate obtained during the cycle test ($M = 148$ bpm, or 74\%) and following the step test ($M = 151$ bpm, or 75\%).

Finally, predicting VO$_2$max from submaximal heart rate is limited by one or more of the following assumptions: (a) linearity of the heart rate/oxygen uptake relationship, (b) accuracy of the age-predicted maximal heart rate, (c) constant oxygen cost of external work, and (d) day-to-day variation in heart rate (9).

In defense of the cycle and step tests, they are generally free from the error that may be introduced from the need for motivation on the part of the subject during near-maximal efforts. In the current study, however, motivation and pacing during the run did not seem to pose a problem, as evidenced by the high heart rates obtained immediately after the run ($M = 89$\% of max HR) and the high test-retest reliability ($R = 0.95$).

Multiple regression analysis resulted in the following best equation to predict VO$_2$max in children: 

$$VO_2\text{max} (\text{ml/kg/min}) = 86.1 - 0.04 \text{ (mile-run time in seconds)} - 0.08 (\text{sum of tricep and subscapular skinfolds in mm}) - 4.7 (\text{gender, with male} = 1, \text{female} = 2) - 0.15 (\text{weight in Kg})$$

This equation had a multiple $R = 0.84$ and a standard error of estimate of 4.3 ml/kg/min, or 9.0\%. This equation agrees quite favorably with the results of Massicotte et al. (7). They have previously reported a multiple $R$ of 0.80 and a standard error of estimate of 10–13\% for timed distance runs versus VO$_2$max in 321 children ages 10 to 17 years. Furthermore, the current equation is in agreement with previous studies which have reported that the addition of skinfold data (6) and body weight (7) improves the prediction of VO$_2$max from run time in children.

In conclusion, the current results suggest that when three commonly used field tests to predict aerobic capacity are compared in the same group of children and adolescents, the timed distance run is superior in both validity and reliability. This conclusion was supported in 5th, 8th, and 11th grade samples. We recommend that the timed distance run be used as the measure of cardiovascular fitness in field studies of children and adolescents. Furthermore, with the use of multiple regression analysis, it appears that VO$_2$max can be predicted as accurately in children as in adults.

References


