The Reliability of Four Measures of Cardiovascular Fitness With Mentally Retarded Adults

Marsha Cressler, Barry Lavay, and Mark Giese
Fort Hays State University

The purpose of this investigation was to determine the test-retest reliability of four submaximum oxygen uptake (VO$_2$) test protocols: (a) Modified Physical Working Capacity Cycle Ergometry, (b) Balke Ware Treadmill, (c) Canadian Step Test, and (d) Cooper Twelve-Minute Run/Walk in predicting the cardiovascular fitness of adults who were mentally retarded. The subjects worked in a sheltered workshop setting (N=17, M age=35 years, M IQ=54). The four submaximal VO$_2$ tests were administered over eight sessions with test-retest reliability scores determined by administering each protocol separately and 1 week apart. An intraclass correlation coefficient revealed the following scores on each of the four protocols: Modified PWC Cycle Ergometry, R=.64; Balke Ware Treadmill, R=.93; Canadian Step Test, R=.95; Cooper Twelve-Minute Run/Walk, R=.81. Discussion includes considerations for administering predictor VO$_2$ tests with persons who are mentally retarded. In this particular investigation the Balke Ware Treadmill Test (R=.93) and the Canadian Step Test (R=.95) revealed the highest reliability scores.

The benefits of cardiovascular fitness training with persons who are mentally retarded are well documented in the literature (Campbell, 1973; Moon & Renzaglia, 1982). However, to date few investigations outline assessment procedures that are valid, reliable, and administratively feasible for testing the cardiovascular fitness level of this population. Consequently, when this fitness parameter is tested it is done so with procedures previously established for use with the general population and is administered indiscriminately with persons who are mentally retarded (Reid, Montgomery, & Seidl, 1985). Determining the reliability of commonly used fitness protocols with this population have been extremely limited, and thus insufficiently researched (Bundschuh & Cureton, 1982; Campbell & Shannon, 1970; Reid et al., 1985; Seidl, Reid, & Montgomery, 1987).

Campbell and Shannon (1970) administered two separate trials of the Balke Maximum Work Capacity Treadmill test to six children, ages 12 to 16 years and...
having an IQ range of 10 to 57, who attended a state school for the mentally retarded. Two orientation periods, both 10 minutes in length, were conducted to familiarize the children to the treadmill walking protocol at various speeds and elevations. The two trials of the actual treadmill test were administered 1 week apart, both under identical conditions. Results of the mean resting, exercise, and recovery heart rates between the two trials yielded high reliability correlation coefficients, with the exception of the first minute during the recovery phase. Therefore the investigators concluded that the Balke Maximum Work Capacity test was a reliable means for evaluating the cardiovascular endurance of the children in this investigation.

In addition, Bundschuh and Cureton (1982) evaluated the test-retest reliability of the Physical Working Capacity Cycle Ergometry test (PWC 170) on 14 adolescents who were mentally retarded (age range 11–19; IQ range 41–72). The test was administered twice, a week apart, with both tests being conducted prior to a 16-week intervention program. The general procedures were similar to those used in the Physical Working Capacity of Canadian Children test. A test-retest reliability coefficient of .85 was obtained for PWC 170 (kpm/min). Moreover, the Canadian Standardized Test of Fitness (1981) with slight modifications was administered to 220 adults who were employed in sheltered workshop settings for the mentally retarded in the greater Montreal area (Reid et al., 1985). All testing was performed in the four workshops by trained graduate students in physical education. Reliability coefficients of each test item were determined by administering a retest on 20 of the adults approximately 1 month following initial testing. All reliability coefficient scores were above .85 except the sit-up (.63) and push-up (.62) components.

These testing issues of poor reliability and few tests to choose from leave professionals who are assigned to provide fitness activities to this population with the dilemma of choosing appropriate measures of cardiovascular fitness. Without adequate measurement instruments that are reliable and administratively feasible, proper assessment procedures and, consequently, appropriate programs of physical fitness for persons who are mentally retarded cannot be fully achieved. Therefore the purpose of this study was to expand upon the literature and more specifically ascertain the test-retest reliability of four widely used submaximal oxygen uptake \((\text{VO}_2)\) testing protocols: (a) Modified Physical Working Capacity Cycle Ergometry (MPWC), (b) Balke Ware Treadmill (BWT), (c) Canadian Step Test (CST), and (d) Cooper Twelve-Minute Run/Walk (C12) in order to measure the cardiovascular efficiency of adults who were mentally retarded.

**Methods**

**Subjects**

The investigation consisted of 17 adults (15 men and 2 women) who were mildly to moderately mentally retarded. Their mean chronological age was 35 years (range 25 to 44) and their mean IQ (Weschler Intelligence Scale–Revised) was 54 (range 41 to 72). All subjects worked in a sheltered workshop setting in the rural Midwest (see Table 1). Each adult selected for this study was free from any known physical abnormality that would prevent walking, running, stepping, or bicycling. Initially 21 subjects began the study, but 4 (similar in age and IQ to the other
Table 1
Subject Characteristics and Cardiovascular Fitness Scores

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sex</th>
<th>Age</th>
<th>IQ</th>
<th>PWCT *test</th>
<th>PWCT retest</th>
<th>Balke test</th>
<th>Balke retest</th>
<th>Step test</th>
<th>Step retest</th>
<th>Cooper test</th>
<th>Cooper retest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>44</td>
<td>72</td>
<td>22.50</td>
<td>28.00</td>
<td>19.93</td>
<td>23.17</td>
<td>40.20</td>
<td>40.92</td>
<td>25.50</td>
<td>29.20</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>31</td>
<td>50</td>
<td>28.50</td>
<td>32.50</td>
<td>24.79</td>
<td>24.79</td>
<td>50.31</td>
<td>49.59</td>
<td>26.00</td>
<td>30.00</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>40</td>
<td>54</td>
<td>24.50</td>
<td>23.50</td>
<td>32.89</td>
<td>18.31</td>
<td>31.27</td>
<td>31.49</td>
<td>31.49</td>
<td>30.50</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>35</td>
<td>55</td>
<td>22.50</td>
<td>29.50</td>
<td>32.89</td>
<td>19.93</td>
<td>44.58</td>
<td>47.29</td>
<td>43.89</td>
<td>39.50</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>35</td>
<td>43</td>
<td>33.50</td>
<td>41.50</td>
<td>26.41</td>
<td>28.03</td>
<td>39.27</td>
<td>44.47</td>
<td>16.00</td>
<td>15.00</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>38</td>
<td>54</td>
<td>27.50</td>
<td>34.00</td>
<td>24.79</td>
<td>23.17</td>
<td>34.82</td>
<td>44.47</td>
<td>20.00</td>
<td>22.00</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>35</td>
<td>52</td>
<td>32.50</td>
<td>36.00</td>
<td>18.31</td>
<td>19.93</td>
<td>46.90</td>
<td>48.01</td>
<td>30.00</td>
<td>31.00</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>34</td>
<td>66</td>
<td>27.00</td>
<td>26.00</td>
<td>26.41</td>
<td>26.41</td>
<td>49.16</td>
<td>44.85</td>
<td>27.20</td>
<td>29.80</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>32</td>
<td>54</td>
<td>31.00</td>
<td>28.50</td>
<td>24.79</td>
<td>26.41</td>
<td>36.89</td>
<td>32.50</td>
<td>28.00</td>
<td>33.50</td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>38</td>
<td>61</td>
<td>20.00</td>
<td>24.00</td>
<td>24.79</td>
<td>27.16</td>
<td>38.02</td>
<td>38.97</td>
<td>35.60</td>
<td>37.00</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>25</td>
<td>64</td>
<td>25.00</td>
<td>23.50</td>
<td>36.13</td>
<td>34.51</td>
<td>38.94</td>
<td>36.47</td>
<td>31.00</td>
<td>30.00</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>33</td>
<td>59</td>
<td>25.00</td>
<td>24.00</td>
<td>32.89</td>
<td>26.41</td>
<td>39.04</td>
<td>37.68</td>
<td>17.00</td>
<td>29.00</td>
</tr>
<tr>
<td>13</td>
<td>M</td>
<td>35</td>
<td>54</td>
<td>30.00</td>
<td>25.00</td>
<td>24.00</td>
<td>23.17</td>
<td>35.52</td>
<td>38.16</td>
<td>31.00</td>
<td>27.20</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>41</td>
<td>54</td>
<td>36.00</td>
<td>28.00</td>
<td>19.93</td>
<td>19.93</td>
<td>35.52</td>
<td>38.16</td>
<td>31.00</td>
<td>27.20</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>27</td>
<td>45</td>
<td>24.00</td>
<td>28.00</td>
<td>16.69</td>
<td>16.69</td>
<td>58.23</td>
<td>57.68</td>
<td>30.00</td>
<td>25.00</td>
</tr>
<tr>
<td>16</td>
<td>M</td>
<td>28</td>
<td>54</td>
<td>33.00</td>
<td>24.00</td>
<td>26.41</td>
<td>26.41</td>
<td>55.05</td>
<td>43.97</td>
<td>28.50</td>
<td>29.16</td>
</tr>
<tr>
<td>17</td>
<td>M</td>
<td>41</td>
<td>46</td>
<td>25.00</td>
<td>23.00</td>
<td>21.55</td>
<td>19.93</td>
<td>40.25</td>
<td>27.16</td>
<td>5.79</td>
<td>5.58</td>
</tr>
</tbody>
</table>

*M* Recorded estimated oxygen uptake scores (ml/kg/min).
subjects) were dropped during a 2-week practice session because of their inability to perform the procedures, inconsistency in performance, difficulty in following directions, and/or low interest level. Prior to participation in the testing session, the subjects were screened for any medication that would alter normal heart rate function. They were not involved in any systematic exercise program outside the study, although certain subjects participated in a few Special Olympic activities.

**Test Protocols**

Maximal oxygen consumption ($\dot{V}O_2_{max}$) is defined as the ability of the body to take in, transport, and utilize oxygen during exercise. This is the best single measure to determine physical fitness (Sharkey, 1984). Each of the four experimental protocols for this study has been reported to be a valid and reliable measure of predicting maximal cardiovascular endurance ($\dot{V}O_2_{max}$) with the general population.

The Modified Physical Working Capacity Cycle Ergometry test consisted of three 6-min work bouts at 300 kpm/min, 600 kpm/min, and 750 kpm/min while pedaling at 50 revolutions per minute (DePauw, Hiles, Mowatt, & Goc-Karp, 1985). A Monarch stationary bicycle (Model 868), stopwatch, and tape recorder were necessary for test administration. Subjects were required to pedal continuously for 6 minutes at each of the three work outputs for a total of 18 minutes. A comparison of heart rate data was determined from pulse rates taken manually at the carotid artery by administrative personnel (trained physical education majors) after minutes 10, 11, and 12 of the second work load at 600 kpm/min, and after minutes 16, 17, and 18 of the third work load at 750 kpm/min. Adaptations made to assist the subjects during testing included a tape recording of the proper cadence, a piece of tape placed on the speedometer at the prescribed revolution rate of 18 mph, and a physical prompt of the subject’s thigh from an assistant upon slowing down.

The Balke Ware Treadmill Test is specific in predicting $\dot{V}O_2_{max}$ and simulates uphill walking. Test administration employed a Quinton Model 643 treadmill. The protocol involved maintaining the treadmill at a constant speed of 3.4 mph while increasing the grade from horizontal during the first minute to 2.4% each subsequent minute until the subject reached 90% of his or her age-predicted maximum heart rate or asked to stop. The ending grade and pulse rate were then recorded with oxygen uptake scores calculated from the Cooper Fitness Classification Formula (Cooper, 1970).

A modified step test from the Canadian Standardized Fitness Test (Government of Canada, 1981) is specific to ascending and descending two 8-in. steps (Reid et al., 1985). Administration of the protocol required two 8-in. steps, a tape recorder, and a prerecorded tape of the different stepping cadences. Subjects were required to ascend and descend the steps at the preestablished cadence according to age. The protocol involved the following procedures: Each subject stood on the floor in front of the first step with feet together, then placed the right foot up on the first step, brought the left foot up to the second step, then brought both feet together. The subject started down with the left foot to the first step, lowered the right foot down to ground level, then brought both feet together. The procedure was then repeated. When the music stopped, the subject ceased to step and was required to remain motionless. Postexercise heart rate was deter-
mined for a period of 10 seconds, based on the carotid pulse rate, by the administrative personnel responsible for obtaining the pulse rate. Upon completion of this first 3-min stage, a decision was made whether to stop or continue to the next phase based on each individual’s heart rate. The subject did not continue if his or her heart rate was equal to or exceeded the ceiling postexercise heart rates (Government of Canada, 1981).

If there were no contraindications, subjects completed the second stage and measurements were repeated as in Stage 1. It was determined from taking the carotid pulse rate whether the subject was to continue for a third and final 3-min stage. To ensure that all subjects maintained the proper cadence, the administrator and assistants counted and stepped with each one. When necessary, physical prompts were given by administrative personnel to those subjects who did not maintain proper cadence. An actual stepping tempo was recorded and maximum oxygen uptake was predicted according to the Jette regression equation (Jette, Campbell, Mongeon, & Routhier, 1976).

The Cooper Twelve-Minute Run/Walk test (Cooper, 1970) involved running or walking as many laps as possible (10 laps to a mile) on an indoor track in 12 minutes. A partner (physical education major) was assigned to run with each adult. The number of laps successfully completed to the nearest quarter of a lap were recorded at the end of the 12-min session and submaximal oxygen uptake scores were determined according to the Cooper Fitness Classification Formula.

Procedures

Prior to the actual investigation, a pilot study was conducted with six mentally retarded adults (Lavay, Giese, Bussen, & Dart, 1987). Also, immediately before the actual investigation, two separate practice sessions, each approximately an hour in length and a week apart, were conducted in order to familiarize and accommodate the subjects to the test procedures and equipment used during the actual investigation. Next, the four submaximal VO₂ tests were administered twice over eight sessions in the following order: (a) MPWC during weeks 1 and 2, (b) BWT during weeks 3 and 4, (c) CST during weeks 5 and 6, and (d) C12 during weeks 7 and 8. The same procedures and conditions to collect the pretest scores were used to collect the retest scores.

The particular order of test selection was based on availability of facilities, equipment, and administrative personnel. At the beginning of each of the eight sessions, all subjects participated in a 10–15-min warm-up consisting of slow stretching exercises. After completing the prescribed protocol for that particular week, each subject participated in a 10–15-min cool-down, which consisted of walking and slow stretching exercises. Also, following the cool-down session, for safety purposes all subjects’ recovery pulse rates were carefully monitored.

The tester reliability of the eight administrative personnel—trained undergraduate and graduate students in physical education who were responsible for manually monitoring the pulse rates of the subjects—were demonstrated by comparing their manual scores to those simultaneously taken with an AMF Benchmark Quantum Fitness Monitor. The findings revealed an intratester reliability score of .92 (range of .84-.96) among the administrative personnel. To further strengthen these findings, an analysis of variance was used to determine if any
significant differences existed between the mean scores of the manual pulse rate monitoring and those recorded on the fitness monitor. Analysis of variance revealed no significant difference between the two procedures, $p < .05$, therefore manual pulse rate was justified.

**Results**

Table 1 describes the subject data in terms of individual scores for each of the four separate test protocols. Because each protocol was a repeated-measures design, four separate ANOVA interclass coefficients were determined (Safrit, 1976; Verduci, 1980). These four correlation coefficients ($R$s) determined the magnitude of relationships between the pretests and posttests for each separate protocol in an effort to determine between trial consistency. The reliability coefficients were MPWC, $R = .64$; BWT, $R = .93$; CST, $R = .95$; and C12, $R = .81$.

**Discussion**

The degree of correlation varied greatly among the different protocols. There are several possible reasons for this variation in reliability scores. Reliability scores of .88 have been reported when administering the physical working capacity cycle ergometry test to the general population (DeVries, 1986). In this particular study the MPWC test revealed the lowest reliability score of $R = .64$. This is contrary to the research of Bundschuh and Cureton (1982), who reported a reliability coefficient score of .85 using a similar population but a slightly different test procedure. This may have resulted from pedaling to a prescribed cadence/rhythm which the adults found difficult. Although modifications such as a prerecorded cassette tape, visual cues, and physical prompts were given during the protocol, certain subjects ($n = 5$) found it difficult to successfully maintain the prescribed cadence. Another factor may have been the protocol consisting of three 6-min work loads at 300, 600, and 750 kpm, a protocol modification advocated by DePauw et al. (1985) but perhaps too long an exercise period (total 18 minutes). The reliability may have been affected by subjects' inability to maintain proper cadence throughout the test.

Reliability scores of .80 have been reported between maximal VO$_2$ uptakes and the Balke Ware Treadmill Test with the general population (Glassford, Baycroft, Sedwick, & MacNab, 1965). Results of this study revealed a relatively high reliability score ($R = .93$). This is not surprising, as some investigators believe a motor driven treadmill to be an excellent way to assess cardiovascular fitness with this population (Tomporowski & Jameson, 1985). The treadmill can move at a set speed, allowing for precise control of each person's pace by minimizing periods of acceleration and deceleration during prescribed bouts of exercise. However, previous practice is necessary for this population to feel comfortable while walking/running at different speeds and elevations on the treadmill. A common fault of this population during treadmill walking is to use the handrails for support (Lavay et al., 1987).

Jéte et al. (1976) reported reliability scores of .91 when administering the Canadian Step Test to the general population. This particular test in the study
revealed the highest correlation \((R = .95)\) among the four protocols administered. This not only supports the research of Jette et al. (1976) but also the work of Reid et al. (1985), who reported a reliability correlation coefficient score of \( .84 \) while administering the same protocol to a similar population.

The step test and bicycle ergometer test present a similar administrative problem in maintaining a steady cadence. However, a regression equation was developed by Jette et al. (1976), with modifications made by Reid et al. (1975), for counting the actual ascents and descents and then interpolating the value from Jette in order to predict the oxygen uptake scores. This prediction equation is based on the actual number of steps each subject successfully completed rather than the desired stepping cadence for the prescribed test protocol. This modification may have caused the high test–retest correlation score. For the practitioner who is measuring cardiovascular fitness based on predicted maximum oxygen uptake values of persons who are mentally retarded, the Canadian Step Test is a desirable protocol as it does not require expensive equipment and can be administered in virtually any setting.

In previous investigations with the general population, the Cooper Twelve-Minute Run/Walk test has consistently resulted in high reliability scores to predicted \(VO_2\) maximum oxygen uptake scores (Cooper, 1970). Cooper has reported reliability coefficients of \( .92\) with this test. However, in this particular investigation a correlation coefficient score of \( R = .81\) was reported. One reason for this lower reported reliability group score may have been attributed to the wide variation in individual test–retest scores for Subjects 4 and 14 (Table 1). Also, running with an all-out effort for 12 minutes is a difficult and abstract concept for persons who are mentally retarded. Other issues to consider when conducting the 12-min run/walk with this particular population include inability to cope with breathlessness and fatigue, lack of knowledge about pace, and a lack of motivation and perseverance needed to complete such a run (DePauw et al., 1985; Pizarro, 1982). It should be noted that similar difficulties were found in a pilot study conducted by Lavay et al. (1987). For these reasons a partner ran with each subject during the actual testing (Fernhall & Tymeson, 1986).

Conclusions

In this particular investigation the Balke Ware Treadmill Test \((R = .93)\) and the Canadian Step Test \((R = .95)\) revealed the highest reliability scores and, based solely on this criteria, seemed the most appropriate test to administer to this particular population. The Physical Working Capacity Cycle Ergometry Test with modifications \((R = .64)\) and the Cooper Twelve-Minute Run/Walk \((R = .81)\) revealed much lower reliability scores. It cannot be assumed that tests of cardiovascular fitness that are successfully used with the nonhandicapped population can be used indiscriminately with the mentally retarded population (Reid et al., 1985). Obviously the measurement and administrative issues discussed in this paper warrant further investigation. Finally, quality research conducted in this area will lead to effective cardiovascular fitness testing that can only enhance program application, which in turn will lead to healthy life-style changes in persons who are mentally retarded.
References


Special thanks to Greg Reid and Chris Seidl for their helpful comments on an earlier version of this manuscript.