Reliability of the Health Related Fitness Test for Mainstreamed Educable and Trainable Mentally Handicapped Adolescents

David C. Pizarro
The Wichita State University

This study investigated the reliability and suitability of the Health Related Physical Fitness Test for mainstreamed educable mentally handicapped (EMH) and trainable mentally handicapped (TMH) adolescents. A total of 126 12- to 15-year-old male and female nonhandicapped (NH), EMH, and TMH adolescents were administered the following tests: modified sit-ups, sit and reach, 880-yard run, and skinfold fat measure (triceps only). Reliability coefficients were obtained using an interclass correlation formula. Deviations in test performance were recorded on a checklist. Modified sit-ups, sit and reach, and skinfold fat measurement were determined to be reliable and suitable for use with mainstreamed EMH/TMH adolescents. Reliability scores for the 880-yard run were fair for NH, good for EMH, and excellent for TMH subjects. Procedural deviations in the 880-yard run by TMH adolescents raised questions about the suitability of this test for these subjects. Proper orientation, an allowance for practice, and the development of an appropriate test environment appeared to be important aspects of test preparation for adolescents functioning at a below normal intellectual level.

Appropriate and adequate levels of physical fitness are important for everyone regardless of age or status, and especially for individuals with various physical/mental handicaps (American Alliance for Health, Physical Education and Recreation, 1976; Winnick, 1980). Many handicapped individuals, either because of poor coordination or physical disabilities, need greater levels of fitness in order to perform daily activities. Handicapped individuals therefore need more opportunities than nonhandicapped (NH) individuals to take part in physical activity programs that promote improved levels of fitness (Eichstaedt & Wang, 1980).

In order to establish and monitor fitness levels of handicapped individuals, simple but reliable field tests must be developed. Furthermore, appropriate fitness programs should become an essential part of daily activities for these individuals. Although fitness programs and valid field testing protocols have been established for NH individuals, significantly less attention has been given to the development or modification of fitness related programs and field testing for handicapped individuals (Winnick, 1980).

Request reprints from David C. Pizarro, Physical Education Department, Box 16, The Wichita State University, Wichita, KS 67208.
With the passing of the Education for All Handicapped Children Act of 1975 (PL 94-142), all public schools must provide educational services for every handicapped child under their jurisdiction. Because of PL 94-142, physical educators are expected to evaluate the fitness status of handicapped school-age youth. One tool presently used by many school systems to evaluate fitness of handicapped youth, especially those who have been mainstreamed into a physical education setting, is the Health Related Physical Fitness Test (HRPFT). However, the HRPFT was normed on nonhandicapped public school children ages 5–17. Nowhere in the manual does it support the use of this battery, with or without modification, for assessing mentally handicapped children who have been mainstreamed into physical education settings.

If school systems are to allow mentally handicapped adolescents to participate alongside nonhandicapped children or in parallel fitness programs, then it must be determined whether these test items utilized for nonhandicapped children are administratively feasible and reliable for the mentally handicapped population. Reliable fitness assessment would allow for an expansion of the data base for the mentally handicapped and for possible comparison to their nonhandicapped peers. However, if tests and procedures used with nonhandicapped populations do not prove reliable for mentally handicapped populations, test procedures will need to be modified to obtain the desired data.

The purpose of this study therefore was twofold: first, to determine the reliability of the HRPFT for mainstreamed EMH/TMH adolescents, and second, to determine whether the test procedures need to be modified in order to obtain fitness data about mainstreamed EMH/TMH adolescents.

Methods

The study was divided into two stages. Stage 1 consisted of an initial evaluation of a small group of nonhandicapped (NH) \( n = 11, 6 \text{ males and 5 females} \), educable mentally handicapped (EMH) \( n = 10, 5 \text{ males and 5 females} \), and trainable mentally handicapped (TMH) \( n = 10, 5 \text{ males and 5 females} \) adolescents ranging in age from 12 to 15 years. The objectives of Stage 1 were (a) to determine the appropriateness of the testing procedure for mainstreamed mentally handicapped adolescents, (b) to establish and develop the proper procedures and necessary paperwork for obtaining subjects for the main study (Stage 2), and (c) to become familiar with the equipment used to assess fitness levels.

Stage 2, the main study, involved 45 NH (22 males and 23 females), 44 EMH (25 males and 19 females), and 37 TMH (20 males and 17 females) adolescents ranging in age from 12 to 15 years. Stage 2 consisted of administering the same tests given in Stage 1 on two occasions with the inclusion of administrative and/or procedural modifications developed from the findings in Stage 1.

Subjects

Three populations of adolescents were involved in both stages: NH, EMH, and TMH adolescents. Classification into EMH and TMH was determined by the guidelines set forth by the Georgia Department of Education (1983). EMH adolescents consisted of individuals ranging in IQ between approximately 70 to approximately 55. TMH adolescents consisted of individuals ranging in IQ between approximately 55 to approximately 40. A form of the Stanford Binet test was
used to determine intelligence quotients for each mentally handicapped subject. For both Stages 1 and 2, the subjects were randomly selected from mainstreamed public school settings. A table of random numbers was used to select approximately equal numbers of males and females from each population for testing. No physical handicaps or serious sensory impairments that might preclude participation were noted in any of the subjects.

Test Protocols

Stage I. An orientation to the fitness tests was initially scheduled for each group of subjects. During the orientation period the subjects were familiarized with the procedure for the modified sit-ups, sit and reach, three-site skinfold fat measurements (triceps, subscapular, and abdominal), and 9-minute run/walk for distance.

The modified sit-ups (American Alliance, 1980) were performed by having the subject assume a starting position lying on his/her back on a mat, feet on the floor and heels between 12 and 18 inches from the buttocks. The arms were crossed on the chest with the hands touching the opposite shoulders, and the subject’s feet were held in contact with the mat. To sit up, the subject tightened the abdominal muscles and curled to a sitting position. The arms remained in contact with the chest. When the elbows touched the thighs, the contractive phase of the sit-up was complete. To complete the down phase of the sit-up, the subject reclined until the midback touched the mat. A subject’s score was the number of correctly executed sit-ups performed in 60 seconds.

The sit and reach (American Alliance, 1980) was performed by having the subject remove his/her shoes and sit down at the test apparatus, with knees fully extended and feet slightly apart. The feet were placed flat against the end board so that the measurement board was extended toward the subject, who then placed his/her hands on top of each other and extended the arms forward so that they rested against the slide on the measuring scale. The subject was then instructed to gently push the slide forward along the measuring scale with four distinct movements and to a position of maximum reach on the final movement. This maximum reach determined the subject’s score, which was recorded in centimeters.

The skinfold fat measurement (American Alliance, 1980) was performed by taking a triceps skinfold measure. For measurement the skinfold was firmly grasped between thumb and forefinger and lifted up. The contact surfaces of a Lange skinfold caliper were placed 1 cm below the finger. The measurement was taken three times and the median of the three scores was recorded. Each skinfold measurement was recorded to the nearest .5 mm.

The 9-minute run/walk test (American Alliance, 1980) was performed by having the subjects run/walk as far as possible in 9 minutes on a 220-yard outdoor track. Subjects began running on the signal “ready, start” and continued to run until a whistle was blown at 9 minutes. Walking was permitted, but the objective was to cover as much distance as possible during the 9 minutes.

Orientation consisted of a verbal explanation of the test, a demonstration by the investigator, and a 1-minute practice trial for each subject. During the practice session the investigator provided correction and/or prompting of each subject’s performance to ensure standardization of body position and technique according to test procedures. Approximately 1 week later each group was tested on the modified sit-ups and the skinfold measurement. A week after the sit-up
and skinfold test session, a second session was held during which the sit and reach and the 9-minute run were administered.

Stage 2. This stage involved 45 NH (22 males, 23 females), 44 EMH (25 males, 19 females), and 37 TMH (20 males, 17 females) adolescents ranging in age from 12 to 15 years. Subjects selected for Stage 2 were divided into groups of 11 or less. The first testing session for Stage 2 was an orientation session similar to that of Stage 1. The second session was held approximately 1 week later, during which subjects were administered the four items that comprised the test battery. For Stage 2, the 880-yard run replaced the 9-minute walk/run test and only one skinfold site was measured (triceps), for reasons that will be discussed later. The test items were administered in the following order: modified sit-ups, sit and reach, skinfold measurement, and 880-yard run. During and immediately following the administration of a test item for the second session, the investigator recorded subjective observations of the subjects' performance on a self-made checklist (see Figure 1). The performance checklist was designed to collect subjective data regarding deviations in test procedures for later analysis and interpretation. The data were used to indicate the nature and frequency of procedural deviations for each sample. No subjective performance observations were recorded for the skinfold fat measurement, since it was an anthropometric measurement that did not require an active or strenuous performance by the subject.

Approximately 1 week later during the third session, all subjects were re-administered the four items of the test battery. Procedures and order of test administration were the same as in the second session. All subjects in each population were retested in order to determine reliability of each test item of the HRPFT. The performance checklist was not used during the retest because it was believed that its repeated use would only document learned behaviors that occur under conditions of repeated testing.

Statistical Analysis

For NH, EMH, and TMH adolescent populations in Stage 2, the four test items measured for male and female subjects were analyzed and compared separately. Test-retest reliability coefficients for the four HRPFT items were determined for the three populations. Reliability was determined using the interclass correlation coefficient that estimated the reliability of a single score (Safrit, 1981). The description of a test's degree of reliability (.60 to .69 = poor, .70 to .79 = fair, .80 to .89 = good, and .90 to .99 = excellent) was based on the standards of evaluating reliability coefficients suggested by Mathews (1978).

The subjective data obtained from the test battery performance checklists were compiled for analysis and interpretation. Deviations in test performance were analyzed and are discussed in regard to their frequency and percentage of occurrence.

Results

Stage 1

The results of Stage 1 indicated the need for several changes in the methodology prior to initiating Stage 2. The 9-minute run/walk for distance was modified in favor of a shorter run that involved completing a set distance. This modification was suggested because 10 of 20 of the EMH/TMH subjects failed to complete the 9-minute run/walk for distance. Since 14 of 20 of the EMH/TMH subjects
<table>
<thead>
<tr>
<th>Subject-ID</th>
<th>Group</th>
<th>Sex</th>
<th>Pull-ups</th>
<th>Sit-ups</th>
<th>Sit and reach</th>
<th>Skinfold (triceps)</th>
<th>Distance run</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Distance run**
Performance criteria: Completed run without stopping, lying down, or wandering off course

Performance descriptors
Appeared not to comprehend required task
Wandered off course
Stopped run/walk (sat or lay down)
Poor pacing
Penalized for infractions
Greater % walk than run
Other

**Modified sit-up**
Performance criteria: Completed five or more modified sit-ups within 60-second time limit

Performance descriptors
Appeared not to comprehend required task
Gross change of arm position/penalized
Gross change in leg position
Stopped before fatigued
Other

**Sit and reach**
Performance criteria: Scored a 1 or more on this test of flexibility

Performance descriptors
Appeared not to comprehend required task
Repeatedly complained of being uncomfortable
Repeatedly bent knees
Jumped slide mechanism repeatedly
Gross changes in hand position/warning
Other

**Skinfold fat measurement**  N/A

Figure 1 — Performance checklist for health related physical fitness test.
either quit running/walking or completed less than 1,000 yards in the 9-minute test period, it was felt that the 880-yard run would be a more appropriate distance.

Another modification that was indicated by Stage 1 testing was reducing the three skinfold measures to one, the triceps. Deletion of the subscapular and abdominal skinfold measurements increased test efficiency and prevented unnecessary embarrassment to the subjects. A third modification in the planned protocol for Stage 2 testing was the development of a chart to record subjective observations of the test proceedings. This was done in order to more explicitly define the subjects' performance. The score card was modified to include performance checklists for three of the test items; skinfold measurement was not included.

Stage 2

Test-Retest Reliability. Means, standard deviations, and reliability coefficients for the four items tested in Stage 2 are found in Tables 1 and 2. For the modified sit-up test, EMH adolescents showed good reliability coefficients ($r = .83$)

### Table 1

Means and Standard Deviations for the HRPFT for NH, EMH, and TMH Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 1</th>
<th>Trial 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sit-up</td>
<td></td>
<td>Sit and reach</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH (N = 45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>36.04</td>
<td>37.78</td>
<td>28.09</td>
<td>28.31</td>
</tr>
<tr>
<td>$SD$</td>
<td>6.78</td>
<td>14.09</td>
<td>7.26</td>
<td>7.07</td>
</tr>
<tr>
<td>EMH (N = 44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>27.00</td>
<td>27.75</td>
<td>20.66</td>
<td>20.25</td>
</tr>
<tr>
<td>$SD$</td>
<td>9.20</td>
<td>7.71</td>
<td>7.89</td>
<td>7.54</td>
</tr>
<tr>
<td>TMH (N = 37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>14.38</td>
<td>14.81</td>
<td>23.00</td>
<td>23.00</td>
</tr>
<tr>
<td>$SD$</td>
<td>10.57</td>
<td>9.88</td>
<td>11.84</td>
<td>11.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Skinfold</th>
<th></th>
<th>Distance run</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH (N = 45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>13.50</td>
<td>13.61</td>
<td>270.73</td>
<td>268.42</td>
</tr>
<tr>
<td>$SD$</td>
<td>5.52</td>
<td>5.57</td>
<td>59.10</td>
<td>45.89</td>
</tr>
<tr>
<td>EMH (N = 44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>15.93</td>
<td>16.02</td>
<td>385.41</td>
<td>388.95</td>
</tr>
<tr>
<td>$SD$</td>
<td>7.78</td>
<td>7.98</td>
<td>87.83</td>
<td>86.21</td>
</tr>
<tr>
<td>TMH (N = 37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>15.97</td>
<td>16.30</td>
<td>518.84</td>
<td>520.11</td>
</tr>
<tr>
<td>$SD$</td>
<td>8.01</td>
<td>7.91</td>
<td>158.39</td>
<td>132.59</td>
</tr>
</tbody>
</table>
while TMH adolescents showed excellent reliability coefficients ($r = .94$). A low reliability coefficient was noted for the NH group ($r = .13$). For the sit and reach test, reliability coefficients for the NH, EMH, and TMH were excellent ($r = .94$, .90, and .97, respectively). Reliability coefficients for the skinfold measurements were also excellent for all three sample populations (NH = .96, EMH = .98, and TMH = .99). Mean test reliability coefficients for the 880-yard run ranged from .75 to .90. Reliability coefficients for the NH group were estimated to be fair ($r = .80$), while TMH adolescents showed excellent reliability scores ($r = .90$).

**Checklist Performance Data**

More than 18% (23 out of 126) of the subjects’ scores in the modified sit-up were reduced because of gross changes in arm position. These position infractions occurred with similar frequency across the three populations. Procedural deviations for the sit and reach were noted more often in the EMH/TMH adolescents than in the NH population. EMHs were observed to commit nearly twice as many procedural deviations (11) as their NH peers (6). TMHs committed nearly three times as many infractions (17) as the NH population. Bent knees, jumping off the slide mechanism, and improper position were the deviations most often noted.

The fair-to-excellent reliability coefficients for the mainstreamed EMH/TMH for the 880-yard run are misleading without referring to the data from the performance checklist. Subjective observation indicated more procedural deviations in the distance run than in any other fitness items administered in the test battery. Of the 117 deviations noted, nearly 89% (104 out of 117) were recorded for EMH/TMH subjects. Poor pacing represented the greatest percentage of deviations for EMHs (56.8%). More than 29% of the EMHs were also observed sprinting for short periods of time and walking for increasingly longer periods of time. Subjective data analysis revealed the 880-yard run to be inappropriate for TMH subjects; only 75% of them were able to complete the run/walk without stopping. It was also noted that 51.4% of the TMH group spent a greater percentage of the recorded test time walking than running. Nine TMH subjects either wandered off the course, appeared not to comprehend the task, or required repeated verbal redirection to complete the test.
Discussion

For Stage 2, the modified sit-up, sit and reach, and triceps skinfold measurement demonstrated that mean reliability scores for the EMH/TMH were equal to, and in some instances higher than, those for NH adolescents. Mean reliability coefficients for the modified sit-ups ranged between good to excellent ($r = .83$ to $.94$) for the EMH/TMH samples. Mean reliability coefficients for the sit and reach ($r = .90$ to $.97$) were in the excellent range for EMH/TMH adolescents. Mean reliability coefficients for the triceps skinfold were excellent, ranging from $.98$ to $.99$. However, the validity of this measurement to predict percent body fat needs to be established. The only procedural problem from the triceps skinfold measurement concerned the alleviation of EMH/TMH subjects’ fears related to the pinching effect of the skinfold calipers. Therefore the modified sit-up, sit and reach, and triceps skinfold measurement for the EMH/TMH adolescents demonstrated similar reliability and feasibility when compared to NH adolescents.

The mean reliability coefficients for the 880-yard run, which was substituted for the Pminute run/walk, ranged from good to excellent ($r = .80$ to $.90$) for the mentally handicapped adolescents. These results are similar to those seen for the 600-yard run (Speakman, 1974) and the Cooper 12-minute run (Cressler, Lavay, & Giese, 1988). Although good to excellent reliability coefficients were noted in the sample groups, numerous procedural deviations noted from the checklist data raised many questions concerning the suitability of distance run tests for EMH/TMH adolescents. As was suggested by Speakman (1974), the runs could be too motivational demanding for EMH/TMH individuals.

Indeed, the fact that the mean reliability coefficients of NH adolescents for the 880-yard run were inferior to those of EMH/TMH adolescents raises the question of validity of a field run/walk test when measuring the cardiovascular fitness of adolescents, whether mentally handicapped or not. The only field running test that has been validated for determining cardiovascular fitness involved adult mentally handicapped individuals. In that study, done by Fernhall and Tymeson (1988), mean correlation coefficients for the 300-yard run and 1.5-mile run, when compared to laboratory measured VO$_2$ max, was $-.71$ and $-.88$, respectively. There is a need therefore to validate field running tests for adolescents, both mentally handicapped and nonhandicapped, to determine the separate effects of age and IQ on reliability scores of field running tests.

Conclusion

The results of this study have demonstrated that three of the four items on the HRPFT were found to be reliable and suitable for use with adolescent mainstreamed EMH/TMH individuals. The three tests were the modified sit-up, sit and reach, and triceps skinfold fat measurement. Although these items appeared to be acceptable to administer to mainstreamed mentally handicapped adolescents, test preparation was a concern. Proper orientation to the item, an allowance for practice, and the development of an appropriate test environment appeared to be particularly important aspects of test preparation for adolescents functioning at below normal intelligence.

Furthermore, the 9-minute distance run in the HRPFT was found to be inappropriate for mainstreamed EMH/TMH adolescents during Stage 1. The
880-yard run in Stage 2 was found to be inappropriate not only for EMH/TMH adolescents but also for NH adolescents. There appear to be different reasons why the run may be inappropriate for mentally handicapped and nonhandicapped individuals. The low reliability scores for the NH population, however, imply that the test may not be a valid measure of cardiorespiratory function. The 880-yard run may be inappropriate for mentally handicapped adolescents, as suggested by data from the performance checklists, because of developmental and/or intellectual limitations. This study suggested the need to establish more reliable field test protocols for measuring the cardiovascular fitness of mentally handicapped adolescents. Practitioners who want to use the HRPFT might focus on test preparation with an emphasis on pacing as well as educating mentally handicapped students about feelings of breathlessness and fatigue.

References


Acknowledgments

The author acknowledges Dr. Kenneth Pitetti of Wichita State University and Dr. Barry Lavay of California State University–Long Beach for technical assistance. Also acknowledged are the principals, teachers, and students at Rose Garden School, Nash Middle School, Floyd Middle School, Alpharetta Elementary School, Kimball Bridge Elementary School, Jackson County Elementary School, and Social Circle Middle School for their cooperation and assistance.