Reliability of Eurofit Physical Fitness Items for Adolescent Males With and Without Mental Retardation

Ciaran Mac Donncha and Anthony W.S. Watson
University of Limerick

Terence McSweeney
Cope Foundation

Daniel J. O’Donovan
University College Galway

The purpose was to examine the reliability of physical fitness items from the Eurofit Test Battery for adolescent males with mild mental retardation (MMR; $n = 63$, mean IQ = $63.0 \pm 11.5$, mean age = $15.5 \pm 1.2$) and those without ($n = 22$, mean age = $15.6 \pm 0.6$). Males with MMR scored significantly lower ($p \leq .005$) than those without on all items except sum of skinfolds, height, and weight. Intraclass correlations (ICCs) ranged from .94 to .99 for males with MMR and .85 to .99 for those without. Percentage error of the mean for all items ranged from 0.5 to 47.5% for participants with MMR and 0.4 to 32.2% for those without. ICCs indicated that Eurofit physical fitness test items are reliable measures for males with and without MMR. However, the percentage error of the mean is quite large for sit-and-reach and 20-m shuttle test (20-MST) items.

Little information is available on the reliability of physical fitness items for individuals with mild mental retardation (MMR). Many studies report only interclass correlation coefficients calculated by the Pearson product moment method (e.g., Rarick, Dobbins, & Broadhead, 1976; Pizzaro, 1990), although intraclass correlation techniques are now favored for test-retest data (Thomas & Nelson, 1996). In particular, little research has focused on physical fitness field test items for older adolescents (> 15 years) with MMR.
Reliability of Fitness Items

Rarick et al.'s (1976) study was one of the first to address reliability of fitness items for individuals with mild and moderate MR and to provide test-retest coefficients for six subgroups of boys and girls, ages 6–9 and 10–13, with and without MR. Test-retest Pearson product moment coefficients with Spearman-Brown corrections for standing long jump and handgrip strength were $r = .94–.98$ for individuals with and without MMR. Other items in the Rarick et al. study were not similar to those examined by the present investigators.

Pizarro (1990) concluded that the modified sit-up, sit-and-reach test, 880-yd run, and triceps skinfold measure were reliable for use with male and female adolescents ($n = 126$, ages 12–15) with mild to moderate mental retardation (MR). Pizarro stated that the apparent reliability of the 880-yd run is misleading because a large number of procedural difficulties were recorded during test administration (e.g., judgment of pacing and lack of task comprehension).

One of Pizarro’s (1990) most valuable contribution was recording all procedural deviations during the testing process. For example, during the sit-and-reach test, individuals with mild ($n = 10$) and moderate ($n = 10$) MR committed almost twice as many procedural deviations as participants without MR ($n = 11$). Bending the knees, jumping of the slide mechanism, and improper positioning were most commonly noted during the sit-and-reach test. Of 117 noted deviations during the 880-yd run, 89% were recorded for individuals with mild and moderate MR. Individuals without MR accounted for the remainder. Numerous other researchers corroborate these findings (Dishman, 1978; Koh & Watkinson, 1988; Lavay, Zody, Solko, & Era, 1990; Reid, Seidl, & Montgomery, 1989). Test preparation, familiarization, participant-tester confidence, and emphasis on pacing are essential components of reliable physical fitness testing for individuals with MR (Pizarro, 1990). More research is needed that attends to these essential components.

Extensive research exists regarding the reliability and validity of cardiorespiratory endurance tests for individuals with MR (Baumgartner & Horvat, 1991; Cressler, Lavay, & Giese, 1988; Fernhall & Tymeson, 1988; Rintala, Dunn, McCubbin, & Quinn, 1992), but investigators have not reached a consensus about the most suitable test. Cressler et al. (1988) reported test-retest reliability of four measures of cardiorespiratory fitness for individuals with mild to moderate MR ($n = 17$, mean age = 35, mean IQ = 54). Results showed intraclass correlations (ICCs) of $R = .64$ for the modified PWC170 (Wahlund, 1948), .81 for the 12-min run/walk (Cooper, 1970), .93 for the Balke Ware treadmill test (Balke & Ware, 1959), and .95 for the Canadian step test (Jette, Campbell, Monegon, & Routhier, 1976). Liu, Plowman, and Looney (1992) reported $R = .91$ for the 20-m shuttle test (20-MST) of Leger, Mercier, Gadoury, and Lambert (1988) performed by males without MR ($n = 12$, ages 12–15). Fernhall et al. (1996) reported a correlation of .95 for the 20-MST when performed by individuals with mild and moderate MR (22 boys, 12 girls; mean age = 14.3). Fernhall et al. (1996) did not indicate which type of correlation was used.

Reliable data are required to determine the need for intervention and to evaluate the impact of intervention programs aimed at enhancing the physical fitness of individuals with MMR. Several studies have indicated the reliability of cardiorespiratory fitness tests for individuals with mild and moderate MR. Few studies, however, have addressed the reliability of physical fitness items for adolescents with MMR. Thus, the purpose of this investigation was to examine the reliability of items selected from the Eurofit Test Battery (Council of Europe, 1988) for adolescent males with and without MMR.
Method

Participants

Written informed consent to participate in the study was obtained from parents of all males (ages 14–18) attending two schools, one for individuals with MMR and another for those without. Both schools were located in Cork City, Ireland, and were within a 1-mile radius of each other. A random sample was drawn from all males whose parents gave written consent. Participants were 63 males with MMR (mean IQ = 63.0 ±11.5, mean age = 15.5 ± 1.2) and 22 males (mean age = 15.6 ± 0.6 years) with average or better IQ. MMR classification was based on two measures: IQ between 50 and 70, using the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1991), and a measure of social adaptation, based on the Vineland Adaptive Behaviour Scales (Sparrow, Balla, & Cicchetti, 1984). Results of both tests were combined for general MMR classification by school personnel.

In Ireland, the term "learning difficulty" is preferred over "mental retardation," and no distinction is made between these terms. Learning difficulty is further defined by specific and general learning difficulty. Specific learning difficulty refers to a reading, arithmetic, or perceptual problem with no MR. General learning difficulty indicates the existence of MR. When both specific and general learning difficulty are present, the condition is equivalent to MR as used in North America. Although participants in this study were diagnosed as having a general learning difficulty in Ireland, MMR is used in this article to prevent confusion between learning difficulty and disability. The ethical committee of the Cope Foundation sanctioned the study.

Procedure

The battery of selected tests was based on those described in the Eurofit handbook (Council of Europe, 1988) and subsequently used in the Northern Ireland Fitness Survey (Riddoch, 1990). This survey examined the physical fitness of 3,211 nondisabled adolescents and is valuable as a normative resource. The Eurofit handbook defines physical fitness by the following dimensions: cardiorespiratory endurance, strength, muscular endurance, speed, flexibility, and balance. Anthropometric measures such as height, weight, and skinfold thickness are also included in the Eurofit Test Battery. Tests that represented each dimension of physical fitness were examined, with the exception of balance, which was excluded for practical reasons (i.e., test organization and duration). Balance was not assessed in the Northern Ireland Fitness Survey.

To estimate reliability, tests were administered twice, 1 week apart. All participants had been tested previously and were familiar with the equipment. Test-retest conditions, such as testing order, day, and time, were standardized for each participant. When individuals with MMR demonstrated a lack of understanding for the testing procedures, practice trials were permitted until test procedures were mastered. For some tests (e.g., sit-ups), in which repeated trials resulted in fatigue, participants were permitted to take the test on another day.

The following physical fitness tests were administered: height (mm), weight (kg), sit-ups (number completed in 30 s), standing long jump (cm), sit-and-reach (cm), handgrip strength (kg), agility as predicted by a 10- by 5-m shuttle run (s), skinfold thickness (mm), and cardiorespiratory endurance as predicted by the
20-MST (measured by the number of laps completed). Alternate tests, which represented similar physical fitness dimensions as outlined in the Eurofit handbook, were not examined. These included a bicycle ergometer test for assessing cardiorespiratory endurance, a bent-arm hang for muscular endurance, and a plate tapping test for fine motor coordination or speed.

Height was measured using a stadiometer (Seca). Measurements were read to the nearest millimeter. Weight was measured using an electronic weighing scale (Seca). Measurements were read to the nearest 0.1 kg. Participants wore only shorts for height and weight measurements.

Abdominal muscle endurance was measured as the number of sit-ups completed in 30 s. Sit-ups were performed with hands placed at the side of the head, knees bent at 90°, and the feet securely held by a partner. A full sit-up was defined as touching the knees with the elbows and returning the shoulders to the ground.

Explosive strength was assessed by a standing long jump, using a tape measure attached to foam mats. Participants were asked to stand behind a line drawn perpendicular to the tape measure and jump forward as far as possible. They were advised to swing their arms and flex their knees prior to jumping. They were also advised to keep both legs close together at take-off and landing and not to fall backward. The distance jumped was recorded from the take-off line to the farthest point backward of the participant. The best of two trials was recorded. The distance jumped was recorded to the nearest 0.5 cm.

Flexibility was assessed using the sit-and-reach test, which indicates range of motion in the hamstrings and lower back regions. A standard sit-and-reach box was used. Participants sat down with their shoes removed and placed their feet against the front of the box. The knees remained fully extended throughout the test. The legs were held in position by the investigator, and participants were asked to reach forward slowly and push the ruler forward as far as possible. A smooth stretching performance was encouraged. The best of three trials was recorded. Measurements were read to the nearest 0.5 cm.

Handgrip strength was measured using an electronic handgrip dynamometer. Dominant handgrip measures were read to the nearest 0.1 kg. The dynamometer was held by the side, with the arm fully extended and palm facing inward. The dynamometer was compressed as hard as possible for 2–3 s. The best of two trials was recorded.

Agility was measured using a 10- by 5-m shuttle run. Each participant was required to sprint 10 times between two lines placed 5 m apart. The track was 1.3 m wide. Measurements were recorded to the nearest 0.01 s.

Cardiorespiratory endurance was estimated using the 20-MST (Leger et al., 1988), which requires participants to run back and forth between two lines set 20 m apart. Running pace is determined by an audio signal. The initial velocity was 8.5 km · hr⁻¹ (5.1 mph) and increased by 0.5 km · hr⁻¹ (0.3 mph) every minute. The test is terminated when the participant cannot reach the end lines in time with two consecutive audio signals. Distance covered was recorded in laps (one lap = 20 m). Approximately 10 participants were grouped together for this test. A staff member initially joined participants to set the correct running pace and only withdrew when everyone had achieved this pace. A similar procedure was used by Boreham, Paliczka, and Nichols (1990), Riddoch (1990), and Liu et al. (1992). This test is similar to the Progressive Aerobic Cardiovascular Endurance Run (PACER) used in North America (Cooper Institute for Aerobics Research, 1994; Seaman, 1995).
Skinfold thickness was measured using a Harpenden skinfold caliper. Four sites were used: midbiceps, midtriceps, subscapular, and suprailiac. All were measured on the right side of the body. Measurements were read to the nearest 0.2 mm.

Statistical Analysis

Statistical analysis was conducted using SPSS for Windows 6.0. Descriptive statistics were computed, and test-retest reliability was established using ICCs. The 95% confidence intervals (the standard deviation of the mean difference between test and retest multiplied by 2.0) was calculated for each measurement procedure. The percentage error of the mean was also calculated by dividing the 95% confidence interval by the maximum mean score. Planned group comparisons were conducted using independent student $t$ tests. A Bonferroni adjustment (alpha = .05 divided by number of comparisons = 9) was applied to the alpha level because multiple $t$ tests result in a greater probability of a Type I error. Alpha was calculated as $p \leq .005$. Skewness of test data was calculated by dividing the raw value by the standard error for skewness. The resultant values were interpreted as $Z$ scores (i.e., values $\pm 1.96$ exceed $p \leq .05$, and values $\pm 2.58$ exceed $p \leq .01$). Data were considered normally distributed if $Z \leq 2.0$ (Vincent, 1995).

Results

ICCs, corresponding 95% confidence intervals (95% CIs), and percentage error of the maximum mean for individuals with and without MMR are presented in Table 1. All test data were normally distributed, except for sum of skinfolds in both groups.

ICCs for physical fitness items were $R = .94-.99$ for males with MMR and .85-.99 for those without. ICCs provide unitless estimates of measurement reliability. To provide a measure with the original measurement units, 95% CI was calculated. In an intervention study, if an individual does not change more than the 95% CI between pre- and posttest scores, no change can be claimed, as this may be due only to the measurement error inherent in the test and testing procedures. Also, correlation coefficients are influenced by the dispersion of the scores. In the current study, the standard deviations of the individuals with MMR are generally greater than those without. Largely dispersed scores usually result in higher correlation coefficients (Baumgartner & Horvat, 1991; Bland & Altman, 1990; Sale, 1991; Watson, 1995) The use of 95% CIs thus offers a more accurate and applicable estimate of reliability.

Independent student $t$ test group comparisons (see Table 2) indicated that male adolescents with MMR scored significantly poorer ($p \leq .005$) on all variables except for height and weight. No significant difference existed between groups for the sum of skinfolds.

Discussion

ICCs for physical fitness items were $R = .94-.99$. for males with MMR and .85-.99 for those without. These correlations suggest that the reliability of items in this study were high for both groups. ICCs for sit-and-reach ($R = .97$), standing long
Table 1  Intraclass Correlations (ICCs), 95% Confidence Intervals (95% CIs), and Percentage Error of the Mean for Males With \((n = 63)\) and Without Mental Retardation \((n = 22)\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>With mental retardation</th>
<th>Without mental retardation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>95% CI of mean</td>
</tr>
<tr>
<td>Height (mm)</td>
<td>.99</td>
<td>8.4</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>.99</td>
<td>1.6</td>
</tr>
<tr>
<td>Sit-and-reach (cm)</td>
<td>.97</td>
<td>4.9</td>
</tr>
<tr>
<td>Standing long jump (cm)</td>
<td>.98</td>
<td>19.6</td>
</tr>
<tr>
<td>Sit-ups (reps)</td>
<td>.95</td>
<td>4.7</td>
</tr>
<tr>
<td>Handgrip strength (kg)</td>
<td>.95</td>
<td>6.9</td>
</tr>
<tr>
<td>Shuttle run (s)</td>
<td>.97</td>
<td>2.0</td>
</tr>
<tr>
<td>20-MST (laps)</td>
<td>.94</td>
<td>21.6</td>
</tr>
<tr>
<td>Sum of skinfolds (mm)*</td>
<td>.99</td>
<td>2.4</td>
</tr>
</tbody>
</table>

*Sample sizes for sum of skinfolds were 20 and 21 for males with and without MMR, respectively.

Table 2  Comparison of Male Adolescents With \((n = 63)\) and Without \((n = 22)\) Mental Retardation on Fitness Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>With mental retardation</th>
<th>Without mental retardation</th>
<th>(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Height (mm)</td>
<td>1,680.0</td>
<td>103.5</td>
<td>1,742.0</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>58.1</td>
<td>11.9</td>
<td>64.0</td>
</tr>
<tr>
<td>Sit-and-reach (cm)</td>
<td>10.1</td>
<td>7.9</td>
<td>17.7</td>
</tr>
<tr>
<td>Standing long jump (cm)</td>
<td>144.9</td>
<td>34.5</td>
<td>198.6</td>
</tr>
<tr>
<td>Sit-ups (reps)</td>
<td>17.0</td>
<td>5.2</td>
<td>22.9</td>
</tr>
<tr>
<td>Handgrip strength (kg)</td>
<td>30.2</td>
<td>8.1</td>
<td>39.1</td>
</tr>
<tr>
<td>Shuttle run (s)</td>
<td>20.4</td>
<td>2.4</td>
<td>17.1</td>
</tr>
<tr>
<td>20-MST (laps)</td>
<td>59.1</td>
<td>24.6</td>
<td>84.6</td>
</tr>
<tr>
<td>Sum of skinfolds (mm)*</td>
<td>35.8</td>
<td>24.0</td>
<td>36.8</td>
</tr>
</tbody>
</table>

*Sample sizes for sum of skinfolds were 20 and 21 for males with and without MMR, respectively.

*\(p \leq .05\). **\(p \leq .005\).
jump \( (R = .98) \), sit-ups \( (R = .95) \), and handgrip strength \( (R = .95) \) were higher for males with MMR. ICCs for males without MMR were \( R = .91, .93, .85, \) and \(.91 \), respectively, for the measures in question. This difference may be explained by the greater dispersion of scores as reflected by the standard deviations of males with MMR. ICCs were higher than the Pearson correlations reported by Pizarro (1990) for individuals (ages = 12–15) with MR \( (n = 44) \), where \( r = .83 \) for modified sit-ups and .90 for sit-and-reach. Pizarro’s (1990) test items were similar to those in the present study. However, Pizarro (1990) used interclass correlations (Pearson product moment) to estimate reliability. Interclass correlations tend to be smaller than ICCs when based on the test-retest scores of the same population (Erbaugh, 1990).

Safrit (1990) reported correlation coefficients for 11- to 14-year-old boys without MR ranging from .94 to .97 and .62 to .93 for the sit-and-reach and sit-up tests, respectively. The type of correlation used was not stated. In the present study, individuals with MMR had ICCs equal to the upper level of the correlation ranges described by Safrit (1990), but participants were older than those in Safrit’s study.

Rarick et al. (1976) described interclass correlation coefficients greater than \( r = .90 \) for boys with MMR (ages 10–13) for the following tests: right and left grip strength, standing long jump, and forward spinal flexion (toe touch). Balogun, Adenlola, and Akinloye (1991) reported a reliability coefficient (Pearson) greater than \( r = .90 \) for handgrip strength of college male students. The present study supports these researchers’ findings.

Reliability coefficients for the shuttle run and the 20-MST in the present study were similar for males with and without MR. Correlation coefficients for the 20-MST in the present study \( (R = .94 \) and .95 for individuals with and without MMR, respectively) are greater than those reported by Liu et al. (1992) for participants without MMR and similar to those of Fernhall et al. (1996) for participants with MMR. Based on correlation evidence, the 20-MST shows promise as a reliable test of aerobic endurance for individuals with MMR. Very few procedural difficulties were noted during this test. Participants adapted very quickly to test requirements. This test seems suitable for individuals with MMR as the intensity level is controlled by the test, unlike distance runs. Also, in the earlier stages, the participant can adapt to test demands, as the level of intensity is low. A potential difficulty with the 20-MST is that participants are generally measured in groups, and an individual’s score may be influenced by others dropping out or persevering. In the current study, 20-MST reliability for individuals with MMR was not as high as that reported by Cressler, Lavay, and Giese (1988) for the Balke Ware treadmill test and the Canadian step test. In the present study, participants with and without MMR scored 59.1 ± 24.6 and 85 ± 20.0 laps, respectively. Boreham et al. (1990) reported a score of 81.7 ± 15.9 laps for 15-year-old males without MMR. Riddoch (1990) reported a similar score (86 ± 20 laps) for same-age peers. Liu et al. (1992) found that 12- to 15-year-old males without MMR scored 71.8 ± 27.0 laps on the 20-MST.

Research into the validity of 20-MST as a measure of cardiorespiratory endurance for individuals with MMR is required. Montgomery, Reid, and Koziris (1992) reported a validity correlation of \( r = .78 \) between a modified shuttle run and the \( V_{O, max} \) score of 18 individuals with mild to moderate MR (ages 20–35). However, the shuttle run underestimated \( V_{O, max} \) values by 28%. Fernhall et al. (1996) reported \( r = .88 \) between 20-MST and \( V_{O, max} \) scores of individuals with mild and moderate MR (22 boys, 12 girls, mean age = 14.3).
The reliability of nonperformance tests, such as height, weight, and skinfolds, was high for both groups in the present study. To achieve a reproducible score, test procedures must be standardized. A difference in the 95% CI and the percentage error for the mean for the sum of skinfolds for both groups are presented in Table 1. The greater 95% confidence interval for male adolescents without MMR was accounted for by variability in one individual’s test-retest score.

The magnitude of the 95% CI and percentage error of the mean of some tests in the present study do not support the corresponding ICCs. A high correlation should suggest a low 95% CI and low corresponding percentage error of the mean. The magnitude of the 95% CI and percentage error for the sit-and-reach and 20-MST do not agree with the strength of the corresponding ICCs in both groups (see Table 1). For example, the ICC for the sit-and-reach test for male adolescents with MMR is R = .97, the corresponding 95% CI is 4.9 cm, and the percentage error of the mean is 47.6%. This lack of agreement may be explained by the effect that the dispersion of scores has on correlations (Baumgartner & Horvat, 1991; Bland & Altman, 1990; Sale, 1990; Watson, 1995). Thus, investigators should carefully consider which analysis method is used to establish reliability.

Procedural deviations similar to those recorded by Pizarro (1990) were noted in the current study, but these were overcome by insisting that participants become adept at each test and by administering each item so as to meet performance criteria (e.g., knees were held down by an assistant in the sit-and-reach test, jerking forward was discouraged, sit-ups were not counted unless elbows touched the knees, and multiple trials were permitted so that participants could learn a test item). Procedural difficulties do exist when testing individuals with MMR, but reliable results can still be obtained, as demonstrated by several studies.

**Conclusion**

Findings of this study indicate that physical fitness items as outlined by the Eurofit handbook (Council of Europe, 1988) are appropriate for individuals with MMR. Moreover, ICC as a reliability estimate indicates that the physical fitness items in this study are reliable for male adolescents with and without MMR.

Using a 95% CI may be a more appropriate method for establishing test item reliability because it is not affected by the dispersion of the scores, and it is presented in the same measurement units as used in the test. Future studies should indicate whether interclass or intraclass correlations were used to estimate test-retest reliability and should also present a 95% confidence interval for each variable examined.

**References**


**Author’s Note**

This study was supported in part by the National Rehabilitation Board and the Cope Foundation, Cork, Ireland. Requests for reprints should be sent to C. Mac Donncha, Physical Education and Sports Science Department, University of Limerick, Ireland.