Test–Retest Reliability of Eurofit Physical Fitness Items for Children with Visual Impairments

Suzanne Houwen, Chris Visscher, Esther Hartman, and Koen A.P.M. Lemmink

The purpose of this study was to examine the test–retest reliability of physical fitness items from the European Test of Physical Fitness (Eurofit) for children with visual impairments. A sample of 21 children, ages 6–12 years, that were recruited from a special school for children with visual impairments participated. Performance on the following physical fitness items was measured on two test sessions with 4 weeks in between: sit-and-reach, standing broad jump, handgrip, sit-ups, bent-arm hang, and 20-m multistage shuttle run. The 10 × 5-m shuttle run was replaced by a 5 × 10-m shuttle run. Intraclass correlations ranged from .63 to .91, indicating moderate-to-excellent reliability. However, systematic differences between test and retest were found for the sit-and-reach, bent-arm hang, and the modified 5 × 10-m shuttle run items. The results indicate that for most items, test–retest reliability was satisfactory, but that improvements need to be made to the test protocols of the sit-and-reach, bent-arm hang, and the 5 × 10-m shuttle run items to ensure test–retest reliability.

Physical fitness plays an important role in daily living and the sports activities of individuals with visual impairments (those who are blind or have low vision) because these activities demand increased energy when performed with impaired vision (5). Several studies, however, have shown that young adolescents with visual impairments have low physical fitness levels, including poor flexibility, cardiorespiratory endurance, muscle endurance, strength, and speed (20,36,37). Relatively little attention has been given to the physical fitness of primary-school-age children with visual impairments.

From the viewpoint of promoting fitness for its positive health outcomes, children with poor physical fitness should be identified and enrolled in training programs. An important question in this connection is whether younger children
are trainable. Although several studies on this topic have been conducted, opinions differ on the trainability of young children because of the confounding factors of the maturation of children and the methodological design of studies (23). For example, whereas some studies report positive effects of training programs on aerobic capacity (4, 32), others report little or no change after training (38). Little information is available on the effects of training in children younger than 8 years of age (23). Nevertheless, it seems that children are responsive to training, but that results are more significant as children age and when their initial physical fitness levels are lower. Therefore, the assessment and monitoring of physical fitness in children is important, especially for children with visual impairments who are at risk of poor physical fitness (20, 37). This means that physical educators and researchers measuring physical fitness in primary-school-age children with visual impairments need to have reliable physical fitness measures at their disposal.

Physical fitness is a multidimensional concept that consists of a set of attributes that are related to the ability to perform physical activity, such as cardiorespiratory endurance, strength, flexibility, speed, and power. Many of these components are related to both performance and health (31). For individuals with visual impairments, several test batteries have been used to measure these different physical fitness components, such as the Project UNIQUE Physical Fitness Test (42), FITNESSGRAM (20), and the Brockport Physical Fitness Test (43). Although the reliability of many physical fitness items has been established in sighted individuals (13, 33), little information is available on the reliability of physical fitness items for individuals with visual impairments (37). Longmuir (22) recommended caution when assessing children with visual impairments, and that specific characteristics of this population be taken into consideration. Problems in testing children with visual impairments include poor motivation or comprehension of the test, anxiety, and movement inefficiency (35, 37).

To our knowledge, no information is available on the reliability of physical fitness items for young children with visual impairments. The reliability of physical fitness items, however, has been established for children and young adolescents with visual impairments ages 10–17 years. Daquila [as cited in Winnick and Short (41)] reported Cronbach’s alpha coefficients for the items of body composition, grip strength, standing broad jump, sit-and-reach, flexed-arm hang, and softball throw. Reliability coefficients over .90 were reported, with the exception of .84 for the flexed-arm hang.

Most of the test batteries have been developed and used in the United States. In Europe, the European Test of Physical Fitness (Eurofit) is currently the most used and recommended test battery for Europeans (16). Reference scales for boys and girls ages 6–12 years have been developed (17), and most of the test items have adequate reliability and validity (7, 10). Therefore, this test might provide a suitable means for measuring the physical fitness of primary-school-age children with visual impairments. According to our knowledge, the Eurofit has been used once in young adolescents with visual impairments (14). No information on reliability was provided.

Accordingly, the purpose of the present study was to examine the test–retest reliability of physical fitness items from the Eurofit for primary-school-age children with visual impairments.
Method

Participants

Twenty-one children with visual impairments (11 boys, 10 girls) ages 6–12 years (mean age = 9.3 years ± 1.7) participated. The children were recruited from a special school for children with visual impairments in the northern Netherlands. In the Netherlands, children are eligible for support if corrected vision is less than 0.30 (20/60; better eye provides vision from a distance of 20 ft that is equal to what a sighted child can see at 60 ft). According to information supplied by the school, nine children had a severe visual impairment (i.e., acuity of less than 20/200), and 12 children had a moderate visual impairment (i.e., acuity of less than 20/60 but > 20/200). School placement is not only related to the visual characteristics of the child with a visual impairment but also depends on the child’s learning capabilities. In the present study, children experienced learning and pedagogic problems notwithstanding an average IQ (above 70).

Signed consent to participate was obtained from participants’ parents. The procedures were in accordance with the ethical standards of the Faculty of Medical Sciences of the University of Groningen.

Instrument

Physical fitness was assessed using selected items from the Eurofit test battery (7). The test items were: sit-and-reach (flexibility), standing broad jump (explosive strength), handgrip strength (static strength), sit-ups (trunk strength and endurance), bent-arm hang (arm and shoulder muscular endurance), and 20-m multistage shuttle run (20-MST; cardiorespiratory endurance). The 10 × 5-m shuttle run (running speed and agility) item of the Eurofit was replaced by a 5 × 10-m shuttle run to limit the number of turns. These items can be divided into those that are performance related (standing broad jump, handgrip strength, and 5 × 10-m shuttle run) and those that are health related (sit-and-reach, sit-ups, bent-arm hang, and 20-MST).

The plate-tapping item for fine motor coordination and speed of limb movement was excluded because two experts (adapted physical activity educators) regarded the plate-tapping item inappropriate for the children in the present study as a result of coordination difficulties that could interfere with the accurate assessment of physical fitness. The flamingo balance item was excluded because it was regarded more a motor skill than a physical fitness aspect and because of practical reasons (i.e., total test-session duration).

Test Accommodations. Two experts (adapted physical activity educators) were consulted about the accommodations required in test materials and execution. On the basis of this information, test accommodations, when considered necessary, were made in the execution and materials of the items. A general accommodation in test execution was that more time was given to the children to understand and master the test items before testing.

Hamstring and lower-back flexibility was assessed using the sit-and-reach test. The children sat down with bare feet against a box and had to remain with their legs fully extended during the test. This position was secured by the test administrator. Although making children extend both legs fully has been questioned as a
debatable procedure (28), other studies have used the same procedure in children (10, 27). The children were asked to reach forward slowly and push the ruler as far as possible with two hands. The best of two trials was recorded. Measurements were read to the nearest half cm. No accommodations were made in test materials or execution.

Explosive strength was assessed with the standing broad jump. Children stood behind a line and were asked to jump forward as far as possible on a foam mat. They were told to swing their arms and flex their knees prior to jumping and to keep both legs close together at take-off and landing. The best of two trials was recorded. The distance jumped was recorded to the nearest centimeter. An accommodation in test materials was the use of bright orange-colored tape to indicate the position to jump from. There was no space between the take-off line and the foam mat. Accommodations in test execution consisted of letting the child feel the movement in order to give him or her an idea of how to perform the test.

Handgrip strength was measured with a handgrip dynamometer. The children were asked to hold the dynamometer to the side of their body, arm fully extended, and palm facing inward. The best of two trials was recorded. Handgrip strength was read to the nearest 0.1 kg. No accommodations in test materials or execution were made.

Trunk strength and endurance were assessed using the number of sit-ups completed in 30 s. Sit-ups were performed with hands held at the side of the head, knees bent at 90°, and feet securely held just above the ankle by the test administrator. A full sit-up was regarded as touching the knees with the elbows and returning the shoulders to the ground. No accommodations in test materials or execution were made.

Arm and shoulder muscular endurance was assessed with the bent-arm hang. Each child was requested to maintain a bent arm position while hanging from a bar for as long as possible, chin above the bar. The test ended when a child’s eyes went below the bar. Measurements were recorded to the nearest 0.1 s. No accommodations in test materials or execution were made.

Speed was measured with a 5 × 10-m shuttle run. Measurements were recorded to the nearest 0.01 s. Two experts (adapted physical activity educators) regarded the 10 × 5-m shuttle run inappropriate because of the many turnarounds. Turning takes a lot of time for children with visual impairments and could lead to coordination problems. In order to limit the number of turnarounds, the test was adapted into the 5 × 10-m shuttle run. The children had to sprint 5 times between two lines placed 10 m apart, instead of sprinting 10 times between two lines 5 m apart. In addition, some accommodations were made in test execution: when the visual impairment of the child was too severe to observe the turning line, the test administrator indicated verbally when to turn (by saying, “3, 2, 1, turn”). Furthermore, when running the children were accompanied by a sighted partner, who corrected them when they started to deflect to the side (19).

Cardiorespiratory endurance was assessed with the 20-MST (18). Children had to run back and forth between two lines set 20 m apart according to an externally paced rhythm. The start of the test was indicated by a beep. The beginning of each successive 1-min interval was indicated by a beep, followed by an indication of the level. The initial velocity was 8.5 km/hr. The running velocity was increased by 0.5 km/hr after every 1-min stage. The test was stopped if the child failed to
reach the end lines before the beep on two successive occasions. Otherwise, the test ended when the child stopped because of fatigue. Each participant was given encouragement to keep running for as long as possible. The performance of a child was expressed as the number of stages completed. The following accommodation in test execution was made: the children were allowed sighted-partner assistance by means of a short rope (19).

**Testing Procedure.** The children were tested twice, with a 4-week interval in between, to investigate test–retest reliability. This time interval was used to fit into the school’s educational program. It was regarded long enough to minimize temporary practice effects and short enough to render systematic changes. The administration of all test items took place at the gym of the children’s school in a fixed sequence. During the test, a protocol was used in which children proceeded in small groups through the fixed sequence. The same test administrator supervised a test item for both the first and second test session. The test administrators were research assistants in human movement sciences who were trained in administering the items of the test battery.

**Data Analysis**

Data analysis was conducted using SPSS® for Windows® 11.0. Descriptive statistics were calculated for the Eurofit scores of the two test sessions. Test–retest reliability was assessed using the intraclass correlation coefficient (ICC) and the Bland and Altman method simultaneously (29). The two methods give complementary information (1,21).

In order for an ICC to be meaningful, the frequency distribution in question should be sufficiently close to normal. Skewness and kurtosis of all items’ scores were calculated by dividing the raw value by the standard error for skewness and kurtosis, respectively. The resultant values were interpreted as Z scores. Data were considered normally distributed if $Z \leq 2.58$ (9). With the exception of the bent-arm hang and modified 5 × 10-m shuttle run items, the data were normally distributed or approached a normal distribution. The bent-arm hang scores and modified 5 × 10-m scores were positively skewed, which led us to perform a logarithmic transformation on these items. Logarithmic transformation of the data did not lead to the assumption of a normal distribution; therefore, the original untransformed data are presented. When presumptions for computing ICCs are violated, the Lin’s concordance correlation coefficient (CCC) is a good alternative measure (34).

The ICC gives a relative index of the ratio of variance among participants to the variance among participants plus error variance. The ICC [model 2,1] used in the present study also takes account of all the test–retest variance [systematic and random (40)]. Ninety-five percent confidence intervals (95% CI) were determined for all the ICCs (29). There are no universally accepted criteria for reliability coefficients, but the criteria recommended by Fleiss (12) are often used. According to these criteria, ICCs of $\geq .75$ are excellent, $<.40$ are poor, and ICCs between these ranges are considered moderate to good.

One advantage of the Bland and Altman method is that it provides reliability in the measurements units. The Bland and Altman method is based on analysis of differences between measurements and consists of several analyses. The mean of
the differences and standard deviation of the differences ($SD_{\text{diff}}$) are calculated. An estimate of the strength of any bias in the measurement is obtained by calculating the 95% CI for the mean difference using the standard error of the mean difference. If zero lay within the 95% CI of the mean difference, we concluded that no bias existed between the measurements (29). The Bland and Altman method also includes the formation of plots. The differences between test session 2 and 1 (2 minus 1) are plotted against the average of test session 1 and 2, the mean. These plots illustrate systematic variations around the zero line, outliers, and heteroscedasticity, which occurs when the difference between test–retest measurements increases as the mean value of the measurements increases. The possibility of heteroscedasticity was addressed by forming the Pearson’s correlation coefficient of the absolute differences (i.e., ignoring any sign) between test session 2 and 1 and the mean of the two test sessions. An outlier was considered to be present when the difference between the two test sessions was outside two standard deviations ($SD$).

Statistical significance was set at $p < .05$.

### Results

Table 1 shows the ICCs, corresponding 95% CIs, and Lin’s CCC. The range of the values for ICC was .63 to .91, and for the CCC was .54 to .90. According to the ICCs, test–retest reliability was excellent (ICC $\geq .75$) for the sit-and-reach, standing broad jump, bent-arm hang, and 20-MST items. None of the items had poor reliability.

Table 2 shows the mean and $SD$s of all test items at the first and second test sessions, mean difference and $SD_{\text{diff}}$, standard error, and 95% CI of the mean difference. The mean differences were low for most of the test items. The negative value of the mean differences for the sit-and-reach, standing broad jump, handgrip strength, and bent-arm-hang items indicated that performance at the second test session was generally lower than at the first. For the 5 × 10-m shuttle run, the negative

### Table 1  Intraclass Correlation Coefficients (ICC), 95% Confidence Intervals (95% CI), and Lin’s Concordance Correlation Coefficient (CCC) for the Eurofit Physical Fitness Items ($N = 21$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>ICC</th>
<th>95% CI</th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sit-and-reach</td>
<td>.83</td>
<td>.64–.93</td>
<td>.79</td>
</tr>
<tr>
<td>Standing broad jump</td>
<td>.91</td>
<td>.79–.96</td>
<td>.90</td>
</tr>
<tr>
<td>Handgrip strength</td>
<td>.69</td>
<td>.38–.86</td>
<td>.69</td>
</tr>
<tr>
<td>Sit-ups</td>
<td>.73</td>
<td>.45–.88</td>
<td>.73</td>
</tr>
<tr>
<td>Bent-arm hang</td>
<td>.76</td>
<td>.50–.90</td>
<td>.83</td>
</tr>
<tr>
<td>Modified 5 × 10-m shuttle run</td>
<td>.63</td>
<td>.28–.83</td>
<td>.54</td>
</tr>
<tr>
<td>20-MST</td>
<td>.90</td>
<td>.77–.96</td>
<td>.89</td>
</tr>
</tbody>
</table>

*Note. 20-MST = 20-m multistage shuttle run.*
<table>
<thead>
<tr>
<th>Item</th>
<th>T1</th>
<th>SD</th>
<th>T2</th>
<th>SD</th>
<th>d</th>
<th>SD</th>
<th>SE of d</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sit-and-reach (cm)</td>
<td>27.50</td>
<td>6.15</td>
<td>25.52</td>
<td>6.32</td>
<td>-1.98</td>
<td>3.59</td>
<td>0.78</td>
<td>-3.61 to -0.34</td>
</tr>
<tr>
<td>Standing broad jump (cm)</td>
<td>115.60</td>
<td>37.57</td>
<td>110.00</td>
<td>35.04</td>
<td>-5.60</td>
<td>15.44</td>
<td>3.37</td>
<td>-12.62 to 1.43</td>
</tr>
<tr>
<td>Handgrip strength (kg)</td>
<td>17.29</td>
<td>4.99</td>
<td>17.14</td>
<td>6.06</td>
<td>-0.14</td>
<td>4.36</td>
<td>0.95</td>
<td>-2.12 to 1.84</td>
</tr>
<tr>
<td>Sit-ups (No. completed)</td>
<td>11.29</td>
<td>6.05</td>
<td>11.52</td>
<td>7.51</td>
<td>0.24</td>
<td>4.98</td>
<td>1.09</td>
<td>-2.09 to 2.50</td>
</tr>
<tr>
<td>Bent-arm hang (s)</td>
<td>5.83</td>
<td>6.76</td>
<td>3.81</td>
<td>4.30</td>
<td>-2.02</td>
<td>3.91</td>
<td>0.85</td>
<td>-3.80 to -0.24</td>
</tr>
<tr>
<td>Modified 5 × 10-m shuttle run (s)</td>
<td>26.53</td>
<td>10.76</td>
<td>20.91</td>
<td>4.38</td>
<td>-5.62</td>
<td>7.11</td>
<td>1.55</td>
<td>-8.85 to -2.32</td>
</tr>
<tr>
<td>20-MST (stages)</td>
<td>2.81</td>
<td>2.05</td>
<td>3.07</td>
<td>2.45</td>
<td>0.26</td>
<td>1.01</td>
<td>0.22</td>
<td>-0.20 to 0.72</td>
</tr>
</tbody>
</table>

*Note.* 20-MST = 20 m multistage shuttle run.
value indicated that children had generally better performance (i.e., faster) at the second test session compared with the first. For the sit-and-reach, bent-arm hang, and 5 × 10-m shuttle-run items, zero was not included in the 95% CI for the mean difference, implying a significantly \((p < .05)\) different performance between the two test sessions.

Figure 1 shows the Bland and Altman plots of the mean of two measurements against the difference of the measurements for the Eurofit items. From these plots, there were generally more values below the zero line than above for the sit-and-reach, bent-arm hang, and 5 × 10-m shuttle run items, illustrating the significantly lower scores (i.e., worse performance for sit-and-reach and bent-arm-hang items; better performance for the 5 × 10-m shuttle-run item) at the second test session. Furthermore, there were indications of a larger variability for higher test values, that is, heteroscedasticity. The Pearson’s correlation coefficient of the absolute difference between test sessions 2 and 1 and the mean of the two test sessions for each participant ranged from .05 to .92 and was significant \((p < .05)\) for two of the seven tests: bent-arm hang and 5 × 10-m shuttle run. When the correlation was recalculated after the exclusion of the outlier, there was no significant relationship between the absolute difference and the mean of the two test sessions for the bent-arm hang, but a significant relationship still existed for the 5 × 10-m shuttle run.

**Discussion**

The purpose of this study was to examine the test–retest reliability of physical fitness items from the Eurofit among primary-school-age children with visual impairments. Most items demonstrated acceptable levels of reliability, indicating that these items can be used in children with visual impairments. It is necessary, however, to make accommodations in test materials and execution. Information provided by experts in the field (e.g., adapted physical activity educators) is important in this perspective for choosing appropriate and reasonably performable test items for children with visual impairments.

The ICC and the Bland and Altman method were used in the present study to examine test–retest reliability of the physical fitness items. In reliability studies, Pearson’s correlation coefficient is often used as a measure of reliability (2). A disadvantage of this statistic measure is that its value is not influenced by differences of the means and variances between the repeated measures. The ICC is more appropriate because it assesses not only strength of correlation, it also takes systematic variability into account. To rely only on the ICC to evaluate reliability, however, might be misleading (40). For example, in the present study the ICC for sit-and-reach was .83. This coefficient indicates that reliability was high. As the results from the Bland and Altman method demonstrate, however, bias in the sit-and-reach measurements occurred. Correlation coefficients do not indicate differences in repeated measures, and they depend on the range of measurements: the wider the range, the better the result (2,25). Therefore, complementing the ICC with the Bland and Altman method is appropriate because the method is not affected by the range of the scores and is presented in the same measurement units as those used in the test (29).

The Bland and Altman method revealed systematic differences for the sit-and-reach, bent-arm hang, and the 5 × 10-m shuttle run items. For the bent-arm hang
Figure 1 — Bland and Altman plots showing the difference between the two test sessions against the mean of the two test sessions for the Eurofit items.

Note. 20-MST = 20-m multistage shuttle run.
and sit-and-reach items, significantly lower scores, indicating poorer performance, were found for the second test session. Furthermore, considerable individual variability between test sessions was observed. A plausible cause for the results of the bent-arm hang might be the level of motivation (35). The bent-arm hang required considerable effort, so if motivation to exert maximal level of performance was not consistently high, variable performance between the two test sessions would be observed. Questionable reliability for the bent-arm hang, as measured with the coefficient of variance, was also found by Fjørtoft (10) in children ages 5–7 years. According to the author this test might be too demanding for young children. Indeed, Tsigilis, Douda, and Tokmakidis (39) indicated that this item might only be useful for adults. Because of the many zero scores (about one-third of the children with visual impairments) and the variability in scores in the children, the bent-arm hang does not appear to be a useful measure of strength for school-age children with visual impairments. If one decides to use this item anyway, because alternative tests for arm and shoulder muscular endurance for primary-school-age children also have limitations (24), a test accommodation would be to provide sufficient stimulation because it is quite difficult for a child to produce maximal levels of performance without external motivation.

For the sit-and-reach item, results might have been influenced by measurement error ensuing from the testing protocol. No warm-up period was provided in the present study as indicated in the Eurofit manual. Children, however, were allowed practice trials that can be interpreted as an unstandardized warm-up period. The reliability of the sit-and-reach item might have been affected by the absence of a standard warm-up period, so before administering the sit-and-reach item, a standard warm-up period should be provided to reduce variation from session to session. The mean difference between the two test sessions, however, was small compared with the mean of the two test sessions, and the 95% CI was narrow, implying that this test effect was small.

For the 5 × 10-m shuttle run, large differences in performance were observed among the children, with a number of children exhibiting substantial differences between the two test sessions. Although not specifically mentioned in the results, the differences between test sessions could generally be attributed to substantial improvement by children who had a more severe visual impairment. In the first test session, the results of most children with a more severe visual impairment were visibly influenced by anxiety. Running at full speed could have been anxiety producing at first, and could have diminished the child’s motivation to participate to the fullest extent (35,37). Furthermore, these children might not be used to performing activities at maximal speed. In the second test session, the children might have been more familiar with the test item, which might have led to reduced anxiety and hence better performance. It seems that a habituation effect occurred. In the future, possible learning effects should be taken into account, and these might be reduced by allowing practice trials before the actual measurement.

The Bland and Altman plots and the statistical analyses revealed heteroscedasticity for the bent-arm hang and 5 × 10-m shuttle run: the higher the test value, the larger the variability between the two test sessions. For the bent-arm hang, an outlier explained this heteroscedasticity. Furthermore, heteroscedasticity can occur as a result of the floor effect of a test item (11). For the 5 × 10-m shuttle run,
the heteroscedasticity was large and could not be explained by the outlier. Heteroscedasticity often occurs because larger values, by their nature, can give rise to larger absolute variability (11). For example, the children who needed the longest time to complete this item at the first test session made the largest improvement in scores.

The results from the 5 × 10-m shuttle run suggest that there might be differences among children with different degrees of visual impairments in the way they act in several test sessions. In this case, it would be interesting to examine the test–retest reliability for children with different degrees of visual impairments separately. In the present study, dividing the total group into two subgroups would lead to very small sample sizes. With such sizes, random change in the mean between two measurements becomes larger, with detrimental effects on reliability (15). We, therefore, decided not to split the group for the results of the reliability analysis; another reason was that the other items did not seem to give cause to expect differences in test behavior between the children with different degrees of visual impairments. Nevertheless, when we inspected the scores, test–retest reliabilities tended to be somewhat lower for children with a severe visual impairment compared with children with a moderate visual impairment. Future research might explore whether there are differences in reliability of the Eurofit test items for children with different degrees of visual impairments.

The present study must be interpreted in the context of the sample size, but it suggests that reliability of most items was excellent according to the criterion of .75, and none had poor reliability (<.40). The reliability results are comparable to those obtained in sighted children, in which the reliability coefficients ranged from .67 to .89 for the same items (30). It has been suggested that judgment of the acceptability of reliability coefficients has to be made by the investigators and experts in the field (8). One could question whether or not it is acceptable to have similar criteria for assessing reliability for children with visual impairments as for sighted children.

In our opinion, with the exception of the running items, the accommodations did not alter the test items substantially, which means that they had minimal impact on the items the test was intended to measure. According to Burns (6), the goal of test accommodations is to remove the effects of impairment yet maintain the test’s underlying validity. In the present study, doubts might be raised about the running items, especially the 5 × 10-m shuttle run. The adaptations in the 5 × 10-m shuttle run were made because problems were foreseen in the turnaround, subjecting mainly those children with a severe visual impairment to coordination problems. It can be argued, however, that in the modified test, the speed component is more emphasized than the agility component and that, because of this, results cannot be compared to the norm scores. Nevertheless, other studies have used similar tests for speed and agility [4 × 10-m shuttle run (3,26)]. The appropriateness of this item as a way to monitor speed and agility must be determined in further studies.

For the 20-MST, children with a severe visual impairment performed this item with a sighted partner who guided the child via a short piece of rope. The sighted partner was explicitly instructed not to push or pull the child, but it remained difficult to determine accurately whether it was the child’s decision to stop or whether this decision was influenced by the sighted partner. Nevertheless, the
applied adaptation was also suggested by Winnick and Short (42). Furthermore, Lieberman (19) reported that children felt very comfortable running with a sighted partner.

**Conclusion**

In conclusion, findings of this study indicate that physical fitness items of the Eurofit are applicable in primary-school-age children with visual impairments, although some items warrant further attention to improve reliability. Possible explanations for the significant differences between the first and second test session are level of motivation (bent-arm hang), anxiety (5 × 10-m shuttle run), and measurement error ensuing from test protocol (sit-and-reach). Improvements need to be made to the test protocol to ensure test–retest reliability. Our recommendations are: a) for the bent-arm-hang item, high motivation should be maintained by providing external feedback; b) for the sit-and-reach item, a standard warm-up period be included; and c) for the 5 × 10-m shuttle run, a habituation session be provided for the final test. The validity of the 5 × 10-m shuttle run should be addressed in further studies. Furthermore, studies on the reliability of physical fitness items for children with visual impairments should be undertaken using larger samples.

**Acknowledgments**

The authors wish to thank Jelly Hoekstra, Robert Kingma, and Sharon Voet for their assistance in this research project, as well as all the teachers and children who collaborated in this study.

**References**