Reliability of Fitness Measures in 3- to 5-Year-Old Children

Thanh Nguyen, Joyce Obeid, and Brian W. Timmons
Children's Exercise and Nutrition Centre McMaster University

The purpose of this study was to determine the reliability of short-term power output, heart rate (HR) response during and after a treadmill test, and time to complete a 25-m dash in healthy preschool children. Thirty-two 3- to 5-year-old boys and girls completed two sessions approximately one week apart. Intraclass correlation coefficient (ICC) and coefficient of variation (CV) were calculated to evaluate reliability. Power output was found to be reliable with ICCs ranging from 0.83 to 0.93 and CVs from 8.1 to 9.7%. Time to complete the 25-m dash was highly reliable (ICC = 0.91, CV = 3.7%). Reliability for HR at submaximal exercise (ICC = 0.28, CV = 18.8%) and HR recovery (ICC = 0.42, CV = 14.0%) was not as strong. These findings should assist in determining appropriate fitness tests for preschoolers.

The preschool years are clearly an important developmental life stage that involves acquiring movement skills and active living behaviors. The association between physical activity and health in preschool-age children (defined here as 3–5 years of age), however, is not as clear as in school-age children (27). Even though preschoolers are not immune to recent trends in obesity (7), we know very little about physical activity and fitness related aspects of health during early childhood. Given the importance of health-related fitness in school-age children for later life health (22), it is important to understand fitness levels at an early age.

Unfortunately, there is a paucity of research devoted to health-related fitness during the early years. This is probably attributable to an assumption that preschoolers are, by nature, physically active and, therefore, physically fit, and to a lack of appropriate measures to assess fitness in young children. Tests of anaerobic and aerobic fitness are common among school-age children and could provide important information regarding the health of preschoolers.

Assessing anaerobic fitness is less common than aerobic fitness in school-age children and also in preschoolers. However, the determination of short-term muscle power seems indicated in preschoolers given that the natural physical activity patterns of children are characterized by short bouts of intense movement (2,3). In contrast to limited information on anaerobic testing, the submaximal heart rate response during treadmill walking and running has been used as a surrogate.
of aerobic fitness in young children (24). Similarly, heart rate recovery (HRR) following exercise is an accepted proxy of fitness (19), as it has been correlated with maximal oxygen consumption (14,30) and trained people tend to have a faster HRR (30,25,9,15). Submaximal heart rate is a common measure of fitness in children (1,24). We therefore sought to determine whether this measure is reliable in preschoolers.

Given the emerging necessity to assess physical activity and fitness at an early age, there is a parallel need to use reliable measures to assess health-related fitness. Because the reliability of common tests of aerobic and anaerobic fitness is not readily available in the literature, we aimed to fill this gap in pediatric research. Since gross motor skills are being developed during the preschool years, we also determined the reliability of a 25-m running test, as a measure of a gross motor skill. The results of this study should help other investigators choose reliable tests of health-related fitness to better understand the role of health-related fitness in healthy growth and development during the early years.

Methods

Participants

Thirty-two 3- to 5-year-old boys (n = 20) and girls (n = 12) from the Hamilton, Ontario region participated in this study approved by the Hamilton Health Sciences/Faculty of Health Sciences Research Ethics Board. Twenty-eight participants were of Caucasian decent, while 4 were non-Caucasian. All participants were prescreened for medical conditions, medical contraindications to exercise, and medications. Participant characteristics are provided in Table 1. None of the children were known to have any chronic diseases (e.g., asthma, diabetes, etc) or disabilities (e.g., cerebral palsy). Informed consent was obtained from parents before their child was included in the study.

Study Protocol

Participants attended two visits, which were separated by 6–7 days and scheduled at the same time of day. However, for 6 children, the time between visits ranged from 2 to 4 weeks due to scheduling difficulties with the families, and for one child, the anaerobic fitness testing was separated by 5 weeks, due to a malfunction of the ergometer combined with difficulty rescheduling the family. During each visit, participants first performed a modified 10-s Wingate cycling test (anaerobic

Table 1 Participant Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>3-year-olds</th>
<th>4-year-olds</th>
<th>5-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height, cm</td>
<td>106.3 ± 6.7 (32)</td>
<td>100.4 ± 4.3 (13)a</td>
<td>107.5 ± 3.6 (8)b</td>
<td>112.5 ± 4.4 (11)c</td>
</tr>
<tr>
<td>Body mass, kg</td>
<td>18.7 ± 3.2 (32)</td>
<td>16.7 ± 1.9 (13)a</td>
<td>18.0 ± 1.7 (8)a</td>
<td>21.5 ± 3.4 (11)b</td>
</tr>
<tr>
<td>%Body fat</td>
<td>22.9 ± 5.6 (31)</td>
<td>22.5 ± 5.0 (12)a</td>
<td>21.5 ± 6.6 (8)a</td>
<td>24.4 ± 5.5 (11)a</td>
</tr>
</tbody>
</table>

Values are mean ± SD with the number of participants in parentheses. Values with different superscript letters are significantly different from each other, p < .05.
fitness) followed by a submaximal treadmill test (aerobic fitness) followed by a 25-m dash (gross motor skill). At least 10 min were provided between each test for rest and recovery. Researchers and children were not acquainted before participants’ induction into the study, except for two participants.

**Anthropometry and Body Composition**

We performed standard measures of anthropometry, including standing height measured without shoes to the nearest 0.1 cm using a calibrated stadiometer and body mass measured without shoes and in light clothing to the nearest 0.1 kg using a digital scale (BWB-800, Tanita Corporation, Japan). A bioelectric impedance analyzer (BIA 101A RJL system, Miami U.S.A) was used to assess body composition, after children rested in the supine position for 5 min. Fat free mass (FFM) was calculated using an age-specific BIA equation (12), which has been validated in 4- to 6-year–old children using doubly-labeled water. Percent body fat was calculated as [(Body weight—FFM) / body weight] x 100.

**Anaerobic Fitness**

Because the majority of bouts of vigorous physical activity in young children last less than 15 s (2,3), we developed a modified 10-s Wingate test to assess anaerobic fitness in preschoolers. The test was performed on a calibrated mechanically braked cycle ergometer (Fleisch-Metabo, Geneva, Switzerland) following a similar protocol as previously described (5). Following a 1–2 min of warm up, maximal pedaling speed was first determined for each child during 20-s of unloaded, all-out cycling. The 10-s Wingate was then performed using a braking force of 4.5% body mass; the braking force was applied when the child reached 80% of their predetermined maximal pedaling speed. Peak power (PP), the highest 3-s power output, and mean power (MP), the average power output over the 10 s test, were calculated and expressed in absolute terms and per kg body mass.

**Aerobic Fitness**

Previous studies involving preschoolers have used treadmill tests to assess aerobic fitness (1,24,21), although none have reported the reliability of this approach. We considered heart rate (HR) at submaximal exercise intensities and the rate of HRR recovery following exercise as markers of aerobic fitness. A Polar HR monitor (Polar Electro Oy, Kempele, Finland) was fastened to each participant using an adjustable band. Before the test children were allowed to warm-up for 1–2 min at speeds that they found comfortable. For the experimental protocol, each participant walked on a motorized treadmill (Free Spirit Treadmill, DF HEBB Industries, Texas, U.S.A) for at least three continuous 2-min stages at (4, 5 and 6 kph) and at 0% grade. Children who were not accustomed to walking on a treadmill as well as those that found this task difficult performed the test at lower speeds. For these participants, speed was increased by either 0.5 or 1.0 kph for every 2-min stage. HR was recorded every minute and averaged for each stage. When only one minute of the last stage was completed we took that value for analysis. Linear regression analysis was performed to interpolate the speed corresponding to a HR of 150 beats per min (Speed\textsubscript{150}). This analysis required at least 3 data points for each child on
both days. Children who did not perform 3 stages on both days were excluded from analysis. To motivate and maintain participant’s attention during the treadmill test, we strategically placed a parent in front of the treadmill for the child to focus on. In addition to verbal encouragement from the researchers, parents provided additional encouragement. Following the final stage, children were immediately seated in a chair beside the treadmill and asked to remain as quiet and still as possible. HR was then recorded every 30 s for 2 min. HRR was calculated as beats per min (bpm) according to the equation:

\[ HRR = HR_{\text{peak}} - HR_{1\text{min}} \]

Where \( HR_{\text{peak}} \) is the HR achieved during the last minute of the last treadmill stage, and \( HR_{1\text{min}} \) is the HR at 1 min of seated recovery.

**Gross Motor Skill**

Participants performed a 25-m dash without shoes to standardize participants’ footwear. The test was conducted in a hallway with the start line and finish line clearly marked. An investigator positioned at the finish line used a hand-held stopwatch to time the test that started with a ready-set-go command. The dash was performed twice, with the second dash done immediately after the first. The fastest time (sec) was used in subsequent analysis.

**Statistical Analysis**

The intraclass correlation coefficient (ICC) was used to assess reliability. ICCs were calculated in SPSS version 17.0 for Windows. The differences between sessions were considered in the calculation of ICC through the “Absolute Agreement” option provided by SPSS. Coefficient of variation (CV) was calculated to evaluate method error, and one-way ANOVAs were performed to compare fitness values between age groups. ANOVAs were also used to compare CVs and whether method error was influenced by age. In both cases, Tukey’s post hoc test was used when appropriate. Data are reported as mean ± SD and statistical significance was set at \( p \leq .05 \).

**Results**

**Anaerobic Fitness**

Thirty-two children completed the modified 10-s Wingate test. One child’s results were lost due to a technical problem with the computer, and this child was, therefore, excluded from analysis. Three children were unable to attain 80% of their predetermined maximal pedaling speed, and braking force was applied at 60% of maximal pedaling speed. These children were also excluded from analysis; 28 sets of data were analyzed. PP and MP values for each age group are shown in Figure 1. One-way ANOVAs revealed that 5-year-olds had significantly higher absolute PP and MP compared with 3 and 4 year olds. In addition, 5-year-olds had a higher relative PP and MP compared with 3-year-olds, but not 4-year-olds.

ICC and CV results for absolute and relative PP and MP are presented in Table 2. Based on ANOVA, there were no intergroup differences in CVs for any measure of anaerobic fitness.
Table 2 Peak and Mean Power Output From the Modified 10-s Wingate Cycling Test

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Day 1</th>
<th>Day 2</th>
<th>ICC</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak power, Watts</td>
<td>28</td>
<td>103.0 ± 40.1</td>
<td>102.6 ± 42.3</td>
<td>0.93***</td>
<td>8.2%</td>
</tr>
<tr>
<td>Peak power, Watts/kg</td>
<td>28</td>
<td>5.4 ± 1.5</td>
<td>5.3 ± 1.4</td>
<td>0.84***</td>
<td>8.1%</td>
</tr>
<tr>
<td>Mean power, Watts</td>
<td>28</td>
<td>81.8 ± 31.6</td>
<td>79.1 ± 33.7</td>
<td>0.91***</td>
<td>9.6%</td>
</tr>
<tr>
<td>Mean power, Watts/kg</td>
<td>28</td>
<td>4.3 ± 1.3</td>
<td>4.1 ± 1.2</td>
<td>0.83***</td>
<td>9.7%</td>
</tr>
</tbody>
</table>

Values are mean ± SD. ICC, intraclass correlation coefficient; CV, coefficient of variation. ***, p < .001.

Figure 1 — Peak and mean power output in 3- (n = 10), 4- (n = 8), and 5- (n = 10) year-old children in absolute (A) and per kg body mass (B) units. Values are mean ± SD from both testing days. *, p < .05, **, p < .01, ***, p < .001.
Aerobic Fitness

Twenty-six children completed at least three stages of the treadmill test on both days. Eighteen children performed the test at 4, 5 and 6 kph, with one child who did one min of the last stage on both days. For the other children, one performed the test at 2, 3 and 4 kph, two at 3, 4 and 5 kph, one at 3.5, 4.0 and 4.5 kph, one at 2, 3, 4 and 5 kph (this child did 1 min of the last stage on both days) and one child did 3, 4, 5, and 6 kph (this child also did 1 min of the last stage on both days). Two of our participants performed the treadmill test at 3, 4, 5, and 6 kph on day one and 4, 5, and 6 kph on day two. These last two participants were included in the analysis because an extrapolation could still be made to Speed150.

Data from 23 children were analyzed for HRR, as only children who performed the same number of stages at the same speeds on both days were included in the analysis. This is because we wanted to ensure children did the same workload on both days, as different workloads may affect HRR and thus reliability. One child’s HRR was not recorded because he needed to leave the laboratory to empty his bladder.

ICC and CV results for Speed150 and HRR are shown in Table 3. The ICC was not significant for Speed150 and the CV was high. We also found no significant differences (p ≥ .05) in speed150 or slopes between children who did the standard protocol and children who performed the alternate protocol. Furthermore, we determined the ICC of speed150 using data from only children who did the standard protocol to be unreliable (p ≥ .05) with a high CV (14.3%). The ICC was significant for HRR and the CV was high. There were no significant differences in the values for Speed150, HRR, or the CVs for these variables between age groups.

Gross Motor Skill

Thirty-two children completed the 25-m dash on both days. The fastest time to complete the test was 8.2 ± 1.3 s on Day 1 and 8.1 ± 1.2 s on Day 2. The ICC was 0.91 (p < .001) with a CV of 3.7%. Average times for each age group are presented in Figure 2. There were significant time differences in 3-year-olds compared with both 4 and 5-year-olds, with 3 year olds having slower times. One-way ANOVA revealed no differences in CVs between age groups.

Discussion

Physical activity during the early years is becoming an important public health issue. Realizing that preschoolers are not immune to growing trends in obesity and the necessity to understand physical activity and fitness at an early age, there

Table 3 HR Results From the Treadmill Test

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Day 1</th>
<th>Day 2</th>
<th>ICC</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed150, km/h</td>
<td>26</td>
<td>4.4 ± 1.7</td>
<td>5.1 ± 0.9</td>
<td>0.28</td>
<td>18.8%</td>
</tr>
<tr>
<td>HRR, bpm</td>
<td>23</td>
<td>54 ± 11</td>
<td>55 ± 16</td>
<td>0.42*</td>
<td>14.0%</td>
</tr>
</tbody>
</table>

Values are mean ± SD. ICC, intraclass correlation coefficient; CV, coefficient of variation; Speed150, interpolated treadmill speed corresponding to a heart rate of 150 bpm; HRR, heart rate recovery. *, p < .05.
is a parallel need to identify appropriate tests to assess health-related fitness. Thus, the purpose of this study was to determine the reliability of tests of health-related fitness in preschoolers.

We investigated the reliability of a modified 10-s Wingate cycling test developed for use with preschool-age children. Our results demonstrate high reliability in values for PP and MP. The reliability of the original 30-s Wingate cycling test is well established for school-age children (4). We modified this test for preschoolers in light of the literature indicating that vigorous physical activity in young children is most often accumulated in bouts of less than 15 s (3,2). Using accelerometry, we confirmed this pattern of vigorous physical activity in 3- to 5-year-old children (20). We therefore reasoned that a test of short-term muscle power is appropriate for preschoolers.

Average values from both visits for PP and MP normalized to body mass were 5.3 ± 1.3 and 4.2 ± 1.2 W/kg, respectively. These results are comparable to the values provided for 24 5- to 7-yr-old healthy children who performed a full 30-s Wingate on a Fleisch-Metabo ergometer (17). Although maturation-related increases in PP and MP are known to occur from age 10 until early adulthood (5), how short-term muscle power output changes during the early years requires additional study. Our results suggest that a modified Wingate cycling test is reliable in children as young as 3 years and provides an option for future investigations in young children.

Given the popularity of submaximal HR during treadmill exercise as a surrogate of aerobic fitness, we deemed it important to assess the reliability of this procedure in preschoolers. Based on our results, the reliability of the submaximal HR response was poor in our sample of children. The Speed150 on Day 2 tended to be higher than on Day 1, suggesting that children may have been less anxious during the second treadmill test. It is possible that at least 3 sessions need to be provided before one might consider the results reliable, although we did not test this possibility in the current study. This extra testing burden, however, is probably not acceptable in most research studies, especially those with a population health focus.

Our treadmill protocol could not be completed by all of our participants, suggesting that not all 3- to 5-year-olds will be able to perform this test. For the
Reliability of Fitness Measures in Children

children who were able to complete the test, submaximal heart rate and HRR were not reliable. The present study demonstrates that HR during this particular submaximal treadmill exercise is not a good indicator of cardiovascular fitness. However, maximal treadmill and HR testing during exercise with 4- to 10-year-olds have been reported to be reliable, with the best reliability found when children were most exhausted (29). Maximal exercise testing may therefore be advantageous to submaximal testing in this age group. Although maximal treadmill testing is not impossible for younger children, it also carries its own limitations. For example, there could be unfamiliarity and safety issues for very young children running at high speeds. Moreover, the need to meet certain criteria to establish a “maximal” test that may or may not be valid in young children could also be questionable.

A previous study testing the reliability of the HR response to submaximal treadmill exercise in older children found it to be reliable in 58 prepubertal girls aged 4.8–10.3 years (10). Girls performed two 10–15 min tests of maximal oxygen uptake 6 weeks apart, in the morning in a fasted state. Researchers reported a CV of 5.1% and nonsignificant t tests between the two sessions, but no ICC analysis was done. These results are in contrast to ours in which we report a high CV and may be due to the age differences or lower number of participants in our study.

HRR at 1 min (8) and 2 min have the capacity to predict mortality in adults, with 2 min being a better predictor (13). In our study, we recorded HRR for up to 2 min, but performed reliability analysis using the 1 min time point, because the children could not usually sit still after the first minute of recovery. Although we found a significant ICC for HRR, the CV (14%) was relatively high for a physiological test. Thus, further research is needed to confidently assess the reliability of HRR after a submaximal graded treadmill test. HRR was recorded with children seated, whereas having participants in the supine position may have improved the results. HRR was recorded with children seated, whereas having participants in the supine position may have improved results because it would have minimized the possibility for movement within the chair.

Reliable measurement of HRR in young children could be a useful tool in assessing fitness and health. Aside from being a predictor of mortality in adults, studies have also found an association between a slower HRR to increased inflammatory markers related to cardiovascular disease (28,16). Furthermore, a reduced HRR at 1 min after a maximal treadmill test has been reported in children with poor exercise endurance, and in overweight children (26).

In comparing our HRR results with other studies, we find that our average HRR over the two days (55 bpm) is greater than that of 4.5- to 5.9-year-olds (34 bpm) who completed a 10-min treadmill test according to the Bruce protocol (18). This is not surprising since it has been reported in children that the greater the intensity of exercise the longer the HRR period (6). In the study by Mimura and colleagues (18), the last treadmill stage was 5.5km/h at a 14% incline and their participants reached an average HRpeak of 185 bpm. In our study, the average HRpeak was 166 bpm. Hence, HR was greater in the study by Mimura and colleagues (18) and this may explain the slower HRR observed in that study.

Our lower HRR results compared with Mimura and colleagues (18) can be interpreted in at least two ways. First, our participants may be less fit than that reported in children of similar age, and second, different exercise intensities may
result in different HRR, which ultimately make it difficult to compare studies. We cannot confidently determine which is the case in our study. The limitation to using HRR is the lack of a standard protocol to evaluate HRR. A reliable and standardized method to access HRR in preschoolers could be useful in predicting future health outcomes.

Recognizing that the early years are an important time of motor skill development, we thought it was important to include a test of gross motor skill and assess its reliability. Running tests are commonly used in tests of motor skill development and motor proficiency in preschoolers. We found that the time to complete a 25-m dash was highly reliable. This demonstrates that sprints can provide consistency as a gross motor skill and, to some extent, probably reflect anaerobic fitness. There are several studies that have included running tests in preschoolers (18,23,11). To make comparisons between our study and those in the literature, we calculated speed (distance/time). On average, our participants completed the 25-m with a speed of $11.4 \pm 1.7$ kph. The only study to look at the same age group as the current study (3–5 year olds) had an average speed of 11.6 kph ($n = 3358$ boys and girls) during a 20-m dash (11), a finding that is remarkably similar to our results. However, our results are slower than the 13 kph previously reported for 4- to 5-year-olds (18) and 4- to 6-year-olds (23). This discrepancy is most likely due to our inclusion of younger children (3 year olds). Regardless of the possible impact of age on running speed, our results indicate that this type of gross motor skill is highly reliable in preschool-age children.

Testing preschoolers was an enjoyable yet demanding task. Remaining motionless for the measurement of HRR was difficult for many of the children. To overcome this we had them pretend to be a statue, or count to see how long they could hold still. We also found their attention was short lived. Between tests and during set-up we engaged the children by conversing on topics they found enjoyable. During the adapted Wingate test, the children pretended they were being chased and had to cycle as fast as possible. Walking and running on the treadmill test was the most difficult for the children. We found that parent’s presence was an integral part of our success in testing this young population. The task enjoyed the most by all children was the 25-m dash.

In summary, this is the first study to assess the reliability of health-related fitness tests in preschool-age children. Measures of short-term muscle power output from a modified Wingate test, and the time to complete a 25-m dash were found to be highly reliable. Conversely, the reliability of the HR response to treadmill exercise and heart rate recovery was not as strong. By establishing the reliability of these tests, future goals are to determine the association between these measures of fitness and measures of health and physical activity. By establishing these relationships, regular assessment of fitness among preschoolers may become a more accepted practice.

Acknowledgments

We wish to thank our participants and their families for their enthusiasm and dedication. Funding for this study was made possible with a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to B.W. Timmons.
References


