

Feasibility and reliability of tests measuring health-related physical fitness in children with moderate to severe intellectual disabilities

Marieke Wouters, Anna M. van der Zanden, Heleen M. Evenhuis,
and Thessa I. M. Hilgenkamp

American Journal on Intellectual and Developmental Disabilities 122(5): 422-438.

ABSTRACT

Physical fitness is an important marker for health. In this study we investigated the feasibility and reliability of health-related physical fitness tests in children with moderate to severe intellectual disability. Thirty-nine children (2-18 yrs) performed tests for muscular strength and endurance, the modified 6-minute walk test (6MWT) for cardiorespiratory fitness, and body composition tests, and 30-97% of the tests were successfully completed. Short-term test-retest reliability of all tests was good (Intraclass Correlation Coefficient [ICC] >.8), long-term test-retest reliability was good for most tests (ICC>.7), but low ICCs were found for most strength tests. Measuring body composition and cardiorespiratory fitness is feasible and reliable. Measuring muscle endurance is fairly feasible and reliable.

INTRODUCTION

High levels of physical fitness are associated with lower risks for a range of negative health outcomes in children and adolescents [1-4]. Children with intellectual disability (ID) may have more chronic health conditions than children who develop typically [5], and their physical fitness seems to be lower than that of typically developing children as well [6-9]. Measuring physical fitness is important to identify children in need of interventions and to evaluate these interventions. Moreover, objective information on the level of fitness of these vulnerable children can be useful for care organizations to set priorities for policy and practice.

However, measuring physical fitness in this population is challenging. Limitations in intellectual functioning and communication, sensory impairments, limb malfunctions, delayed growth and motor development, autism spectrum disorder and challenging behavior, and lack of understanding or motivation to perform the test with the required complete exertion or maximum performance [10-16] can all influence test performance. Field-based tests are often the tests of choice for measuring physical fitness in this population [8, 17, 18]. These tests are cost and time efficient and can be conducted in a familiar environment for the participants, in contrast to a test conducted in a laboratory. Lab-based tests provide more objective and accurate measurements, therefore, it is important to establish the psychometric properties of field-based tests. In our literature review we studied the properties of tests in children with ID, and we concluded that feasibility, reliability and validity are not well established in children with ID [19]. Furthermore, younger children and children with more severe levels of ID were underrepresented in published studies, which limits the generalizability of the results and the evidenced-based application of physical fitness tests in prevention and treatment of chronic disability or disease in all children with ID.

Therefore, potentially suitable tests should be studied to determine their feasibility and psychometric properties in young children and children with more severe ID. Based on the literature review conducted by the authors [19] and focus groups, we selected body mass index (BMI) and waist circumference for body composition, and four tests from the Functional Strength Measurement (FSM) for muscular strength and endurance. The FSM is a test battery developed and tested in children aged 4-10 years [20]. The six minute walk test (6MWT) was modified to measure cardiorespiratory fitness. For these selected tests, positive results on feasibility, reliability and/or validity had been found in adolescents with mild to moderate ID [8, 21, 22] or school-aged children with mild ID [23], but no information is yet available on the feasibility and psychometric properties in children with moderate to severe levels of ID.

For that reason, this study focusses on the feasibility and short and long-term reliability of these field-based physical fitness tests in children with moderate to severe levels of ID, in order to determine the suitability of these tests for application in clinical practice and in research.

METHODS

Participants

This study was performed in collaboration with a service provider for people with disabilities in the Netherlands. Three children's daycare facilities were selected, in which 130 children received daycare support.

We invited all children aged 2-18 years with a moderate or severe level of ID and sufficient motor capacities to walk independently to participate. The first inclusion criteria (level of ID) was evaluated by the behavioral therapist or psychologist of the child by reviewing existing testing results from the Bayley Scale of Infant and Toddler Development, Third Edition (BSID-III) [24], the Snijders-Oomen Nonverbal Intelligence Test [25] or comparable tests. Children were invited if their testing results were corresponding to an IQ of 20-55. The second inclusion criteria (motor capacity) was evaluated by the physical therapist of the child by answering the question, "Can this child walk independently?" After this first screening, parents or legal representatives of the children who met the inclusion criteria received an invitation letter. Children were not invited if the behavioral therapist, psychologist, or the medical staff advised against participation in the study because of a negative influence on the behavior or physical health of the child (exclusion criteria).

Seventy-one of 130 children (55%) were invited to participate. Reasons for not inviting were a mild or profound level of ID ($n = 20$), insufficient motor capacities ($n = 9$), or both ($n = 16$). One child could not be invited because of contra-indications due to a severe autism spectrum disorder (ASD), and 13 children were not present during the testing period. For 41 children (58%), we received informed consent and they were enrolled in the study (Figure 1). The parents or legal representatives of these children gave informed consent and completed the Physical Activity Readiness Questionnaire (parQ) for further medical clearance. The included participants were not different than the invited nonparticipants on demographics (age and sex, data not shown).

Ethical approval was obtained (MEC-2013-491) from the ethics committee of the Erasmus Medical Center. The study adheres to the Declaration of Helsinki for research involving human subjects [26].

Measurements

Information on the level of ID, autism spectrum disorder, and behavioral problems was provided by the behavioral therapist or psychologist of the participants. Information on age, Down syndrome, and physical problems was extracted from the medical file. The gross motor scale of the BSID-III [24] was completed by physical therapists and was used to give insight into their motor development. Because this test was developed for children up to 42 months, raw scores were not converted into standardized norm scores. We used the raw scores to check if completion rates were significantly different for relative low and high motor development.

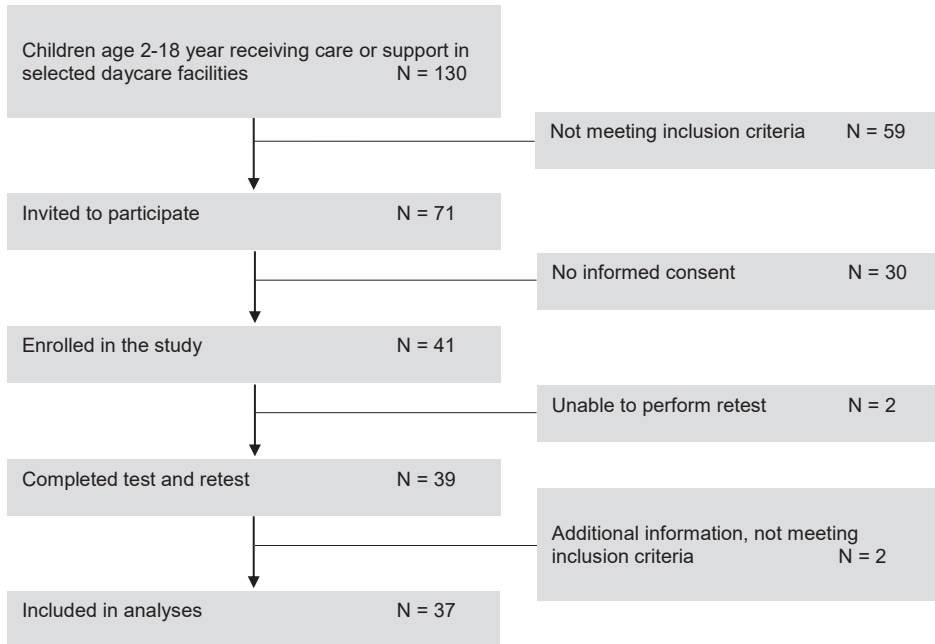


Figure 1. Flow diagram

Physical fitness tests

Health-related physical fitness can be divided in five measurable components: body composition, muscular strength, muscular endurance, cardiorespiratory fitness and flexibility [27]. Flexibility has not been included in this study, because the relationship between flexibility and health has not been confirmed in children [28, 29]. Potentially suitable tests were selected based on our literature review [19], with the addition of extra muscular strength and endurance tests tested in other pediatric populations. All tests were discussed with a focus group of 10 physical therapists working with individuals with ID for more than 3 years. In the final selection, some tests with positive psychometric properties found in the review were eliminated, based on the experience and estimation of the focus group.

Body composition

Body mass index (BMI) was calculated as body weight in kilogram divided by height in meters squared. The participant was on bare feet and wore light clothes. Height was measured with a portable stadiometer accurate at 0.1 cm level. The participant was asked to stand with his or her back against the stadiometer. The participant had to stand straight, with the feet flat on the ground, and the face in the "Frankfurt plane," with the lower margin of the left orbit and the upper margin of each ear canal at the same plane. Body weight was measured using an electronic calibrated scale accurate at 0.1 kg level. The participant had to stand independently and motionless for at least 2 seconds to properly measure the weight.

The BMI scores were compared to the World Health Organization (WHO) growth references [30, 31]. The WHO classification for underweight, overweight, and obesity is based on BMI-for-age z-scores. BMI $-2SD$ is classified as underweight. Children at age 0-5 years with BMI $+2SD$ are described as overweight, and $+3SD$ as obese. For older children, $+1SD$ is classified as overweight and $+2SD$ as obese [32].

Waist circumference provides a measure of abdominal adiposity [33] and was measured using a flexible measuring tape accurate at the 0.1 cm level. The participant was asked to stand still, with the arms down. The circumference was measured half way between the iliac crest and tenth rib, directly on the skin. This measurement was found reliable (ICC = .98) in 17 adolescents with mild to moderate levels of ID [22].

Muscular strength and endurance

The muscular strength and endurance tests were selected from the Functional Strength Measurements (FSM) [34]. This test battery consists of eight items assessing the functional strength of the upper and lower extremities, and was developed for children 4 to 10 years old. It was studied in children with no disability [34] and with mild ID [23]. Good test retest reliability (ICC = .7 - .9) [23, 34], internal consistency ($\alpha = .7$) [34] and fair validity ($r = .5-.7$) [34] were found. We selected three throwing tasks and the standing long jump for muscular strength and stair climbing for muscular endurance, because these tests were the most functional and, thereby, understandable tests according to the focus group. For all of the muscular tasks, the participant practiced at least once prior to the test starting, with a maximum of five practice attempts prior to the test. Improvement was usually seen between the first and third practice of the task, but, when necessary, participants were allowed to practice up to five times to ensure correct execution or to verify that the participant was not able to execute the task correctly. After the practice session, the participant performed the test three times, with at least 30 seconds rest in between. The maximum score of these three attempts was the final score. For the throwing tasks, different bags were used (1, 2, and 3 kg), and bag selection was based on the age of the participant. Due to the fact that the FSM is developed for children ≤ 10 years, we added a bag of 4 kg for children of ages 11 and above.

During the overarm throwing task, the participant stood behind a line marked on the ground with the feet slightly apart, raised the bag behind his or her head, and threw the bag as far as possible. The distance from the line to the proximal side of the bag was measured. In the underarm throwing task, the participant stood behind a marked line with legs spread. The participant bent forward, while holding the bag, and swung the bag back, between the legs, and then threw it forward. The distance between the marked line and the proximal side of the bag was measured in centimeters.

For the chest pass test, the participant sat on the ground with his or her back against the wall and legs spread on the ground. The participant held the bag in both hands, with the back of his or her hands against the chest and the elbows pointed sideways. The participant

passed the bag as far as possible, without bending forwards. The distance between the proximal side of the bag and the wall was measured in centimeters.

For the standing long jump, the participant stood behind a line marked on the ground, with his or her feet slightly apart. The participant jumped as far as possible with both feet together by bending his or her knees and swinging the arms. The distance between the line and the proximal side of the heels was measured in centimeters.

For muscular endurance, we selected the stair climbing test. Six-step stairs (three up and three down) of the BSID-III were used for the stair climbing test. The participant walked the stairs as fast as possible up and down and repeated this as many times as possible in 30 seconds. The amount of stair steps walked in 30 seconds was the score.

Cardiorespiratory fitness

The six-minute walk test (6MWT) is a test of functional capacity, and is recommended by the American College of Sports Medicine (ACSM) to use as a field-based test for cardiorespiratory fitness in several populations with low fitness levels [27]. During the 6MWT, the participants walk as many meters as possible in 6 minutes. A review of Bartels, de Groot [35] indicated good test-retest reliability for children with chronic health conditions, and in some of these populations criterion validity was sufficient as well. In adolescents who were overweight and had mild to moderate levels of ID, Elmahgoub, Van de Velde [36] found good reliability ($ICC=.82$) and fair validity ($r=.69$). Better test-retest reliability was found in adults with ID ($ICC=.98$) [37]. They also found good concurrent validity with the maximum oxygen uptake ($r=.84$). Sufficient concurrent validity was also found in adults with Down syndrome (DS; $r=.78$) [38].

In the original 6MWT, the participant independently walks 30 meters back and forth in a corridor to reach as many meters as possible in 6 minutes, with standardized encouragement [39]. We modified the test to make it more feasible for children with ID, because more than half of the participants were unable to perform the test without physical guidance (i.e., for direction; unpublished data). Therefore, the instructor gave the participants a loose hand, without pulling along. Furthermore, the participants walked a square of 20 meters, and were encouraged individually to keep walking as fast as they could. The total distance covered in 6 minutes was recorded in meters. Before and after the test, the participant sat on a bench. The heart rate was continuously monitored, before (1 minute), during, and after the test (3 minutes). Peak and average heart rate were calculated afterwards.

Procedure

The health-related physical fitness of the participants was tested on two occasions, with 2 to 4 weeks between sessions 1 and 2, and the third session within an hour after session 2 (Figure 2), to study both the long-term and short-term reliability. The three test sessions were conducted at the same location at the same time of the day and tested by the same instruc-

tor. Body composition was only measured in session 1 and 2. The instructors were physical therapists or human movement scientists, all experienced in working with children with ID. In a 1-2 hour workshop, all instructors were taught how to conduct the tests according to protocol. An extra instructor was always present to note the scores and check for protocol deviations.

The tests were performed in a gymnasium at the three different daycare facilities during daytime. The tests were performed in a fixed order within each test session: body composition, modified 6-minute walk test, overarm throwing, stair climbing, underarm throwing, standing long jump, and chest pass. In this sequence, we took into account the load on respectively the upper and lower extremities and familiarization with the tester and the test environment by starting with relatively easy tests (i.e., body composition and the modified 6MWT). Resting between tests was allowed. The instructors explained the tests and demonstrated how to do them. For the muscular strength tests, the participants could practice up to five times, as previously discussed. When they performed the task correctly during this practice phase, the actual test started. During all tests, the instructors encouraged the participants to perform maximally. No standardized encouragement was used, because every child has their own preferences and communication style. If a test was not performed according to protocol, or not performed at all, the presumed reason for dropout was noted (e.g., cognitive issues, motor development issues, challenging behavior, concentration issues, or issues with logistics). The instructors also indicated the level of effort during the tests. If a participant was unable to perform the test with maximum effort, the test was scored as invalid and the presumed reason was noted. The dropout statistics can be found in Table 2.

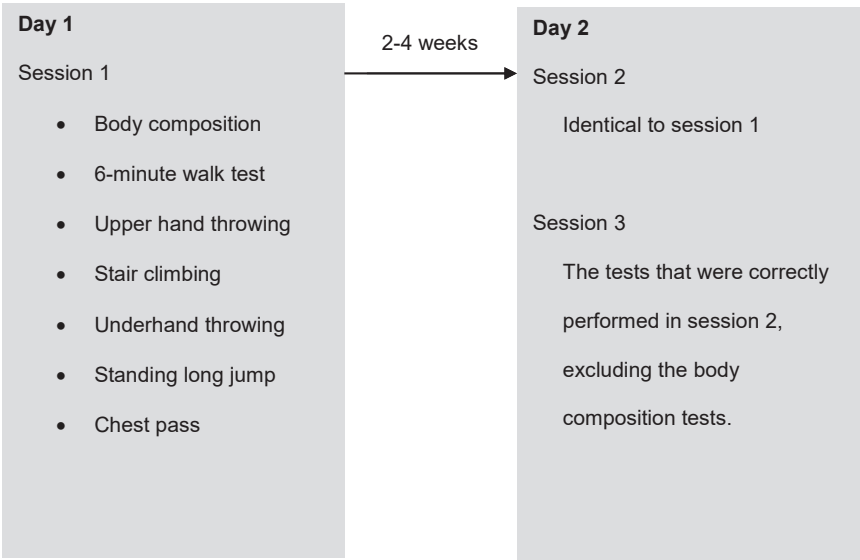


Figure 2. Test procedure

Analysis

Descriptive statistics were used to describe the participants and the feasibility outcomes. Feasibility of the test was based on the completion rate. When completion rates were $\geq 75\%$ the test was feasible; a test was considered fairly feasible when completion rates were 50-75%, and not feasible when completion rates were $< 50\%$. Chi-square tests were used to check if completion rates were significantly different for sex, age groups (2-4; 5-8; 9-12; 13-17 y), level of ID, overweight ($zBMI \leq 1$ vs. > 1), diagnosed with ASD, challenging behavior, and motor development groups (raw BSID-III score < 62 vs. ≥ 62). A raw BSID-III score of 62 is comparable to a motor development of a typically developing child of 33 months.

Normality of the fitness parameters was checked by use of a Kolmogorov-Smirnov test. If the parameters were normally distributed, systematic difference in mean scores over time was checked by a paired T-test. The Wilcoxon Signed Rank test was used as the nonparametric alternative.

The test-retest reliability was studied by use of the ICC (two-way mixed model, absolute agreement, single measures). The ICC provides insight into the relative reliability by comparing the within-subject variability to the between-subject variability [40]. An ICC $> .7$ was indicated as sufficient test-retest reliability [41].

Measurement error was calculated to give information on the absolute reliability: the degree to which repeated measurements vary for individuals [40]. The Standard Error of Measurement (SEM) estimates the standard error in a set of repeated scores. It is calculated by $SEM = SD \times \sqrt{1-ICC}$, where SD is the pooled standard deviation of the measurements. In a reliable measurement, the SEM would be small. The Minimal Detectable Change (MDC) at 95% CI is defined as the minimum change that falls outside the measurement error in the score of an instrument and facilitates the interpretation of changes in test scores. It is calculated as $MDC = 1.96 * SEM * \sqrt{2}$. The SEM and MDC are depicted in the measurement unit of the test [41].

A Bland-Altman plot is a graphical method to compare two test results [42] and gives insight in the relationship between averages and differences, systematic biases and possible outliers. The difference between the two tests is plotted against the average of the two tests. Horizontal lines are plotted at the mean difference and at the limits of agreement (LOA), which are defined as mean difference plus or minus $1.96 * \text{standard deviation of the difference}$. The plots are depicted in appendix 1.

The analyses were performed for session 1 compared to session 2 to establish the long-term reliability, and for session 2 compared to session 3 to establish the short-term reliability. The data was analyzed by use of SPSS 22 (IBM statistics) and alpha was set at 5%.

RESULTS

Participants

Initially, 41 children were included, but two participants were not able to participate in the second test day. One participant moved to a different region of the country and changed daycare facilities, and one participant was recovering from a planned operation during the retest. Thirty-nine participants completed the test and retest, but two participants were excluded from analysis: as part of usual care, their level of ID was retested after the data collection and they were evaluated to have a mild level of ID (Figure 1). The characteristics of the remaining 37 participants can be found in Table 1.

Table 1. Characteristics of the Participants (N=37)

	N	%
Sex		
Male	25	68
Female	12	32
Age		
2-4 years	5	14
5-8 years	13	35
9-12 years	8	22
13-17 years	11	30
Level of intellectual disability		
Moderate	14	38
Severe	23	62
Weight class*		
Underweight	0	0
Normal weight	23	62
Overweight	12	33
Unknown	2	5
Diagnoses		
Autism spectrum disorder	14	38
Down syndrome	4	11
Challenging behavior	27	73
Motor Development		
Low (raw BSID-III score < 62)	21	57
High (raw BSID-III score ≥ 62)	16	43

Notes. BSID-III = Bayley Scale of Infant and Toddler Development, Third Edition. * Based on BMI-for-age z-scores compared with WHO growth references [30, 31].

Feasibility

The average time to complete the test battery on the first day was 42 minutes ($SD=13$), with a range of 12 to 67 minutes. Table 2 depicts the number of successful completions and reasons for dropout in the first session. The body composition tests and the walk test appeared to be feasible with completion rates of 76-97%, and the stair climbing appeared fairly feasible (59%). The muscular strength tests had the lowest completion rates (30-41%) and were considered not feasible for this population.

The most common reason for dropping out was the cognitive abilities of the participants—they did not understand the instructions or aim of the test and were therefore not able to perform the test according to the protocol. In the standing long jump and the stair climbing test, motor development was the limiting factor. The dropout rate was not significantly different for sex, age, level of ID, being overweight, or having DS, ASD, or challenging behavior. However, more participants with low motor development could not perform the muscular strength tests: overarm throwing ($\chi^2=8.428$; $df=1$; $p=.004$), underarm throwing ($\chi^2=4.985$; $df=1$; $p=.026$), chest pass ($\chi^2=5.072$; $df=1$; $p=.029$), and standing long jump ($\chi^2=8.397$; $df=1$; $p=.004$).

Table 2. Completion Rates and Reasons for Dropout of the First Session (N=37)

	Successful completion		Cognitive development	Challenging behavior	Motor development	Concentration	Logistics
Height	35	95%	2	0	0	0	0
Weight	36	97%	1	0	0	0	0
Waist circumference	35	95%	2	0	0	0	0
Overarm throwing	11	30%	16	4	4	1	1
Underarm throwing	11	30%	18	3	3	1	1
Chest pass	11	30%	18	3	3	1	1
Standing long jump	15	41%	8	3	10	1	0
Stair climbing	22	59%	5	3	6	1	0
Modified 6MWT	30	81%	2	5	0	0	0
Heart rate monitor	28	76%	3	5	0	0	1

Notes. 6MWT = six-minute walk test

Test-retest reliability

The number of days between the first and second test sessions was, on average, 18 days ($SD=4$). Table 3 shows the results of the three sessions and the long- and short-term reliability coefficients (ICC, SEM, MDC). All short-term ICCs (between session 2 and 3) were sufficient ($ICC>.7$) and the Bland-Altman plots revealed no systematic bias, but some outliers were seen (see Appendix 1).

The long-term test-retest reliability (between session 1 and 2) for the body composition measurements were very high ($ICC=.99$). Of note, the average body weight increased by 0.6

Table 3. Mean Values and Reliability Results

	Session 1		Session 2		Session 3		Long-term reliability (1 vs 2)			Short-term reliability (2 vs 3)		
	n	M (sd)	n	M (sd)	n	M (sd)	ICC ₁₂	(95%CI)	SEM	MDC	ICC ₂₃	(95%CI)
Body composition												
Body mass index (kg/m ²)	35	19.5 (5.2)	35	19.6 (5.2)	-	-	.99	(.99-1.00)	0.5	1.4	-	-
Height (cm)	35	134 (25)	35	134 (25)	-	-	.99	(.99-1.00)	2.5	7.0	-	-
Weight (kg)	36	37.9 (23.2)	35	38.5 (23.5)*	-	-	.99	(.99-1.00)	2.3	6.5	-	-
Waist circumference (cm)	35	65.4 (18.2)	35	66 (18)	-	-	.99	(.99-1.00)	0.5	1.4	-	-
Muscular strength												
Overarm throwing (cm)	11	151 (46)	13	151 (46)	13	162 (49)	.76	(.32-.93)	23	64	.79	(.60-.94)
Underarm throwing (cm)	11	171 (88)	13	187 (84)	14	182 (84)	.64	(.10-.89)	51	142	.85	(.57-.95)
Chest pass (cm)	11	161 (41)	9	152 (39)	8	142 (41)	.55	(-.12-.88)	27	75	.91	(.63-.98)
Standing long jump (cm)	15	26 (19)	14	37 (23)*	12	37 (18)	.63	(.17-.87)	12	34	.87	(.62-.96)
Muscular endurance												
Stair climbing (n)	22	24 (10)	19	26 (12)*	15	26 (8)	.92	(.76-.97)	3	8	.85	(.62-.95)
Cardiorespiratory fitness												
Modified 6MWT (m)	30	331 (63)	31	330 (79)	26	342 (72)	.78	(.60-.89)	33	92	.95	(.88-.98)
Peak heart rate (bpm)	28	138 (17)	30	133 (16)	24	132 (12)	.60	(.29-.80)	10	26	.87	(.73-.94)
Average heart rate (bpm)	28	124 (13)	30	122 (13)	24	123 (11)	.74	(.50-.87)	7	18	.88	(.73-.94)

Notes: 6MWT = six-minute walk test; ICC = Intraclass Correlation Coefficient; SEM = Standard Error of Measurement; MDC = Minimal Detectable Change; * Significant difference between session 1 and 2: weight: W=348.5; p=.048; standing long jump t₁₃=2.347; p=.035; stair climbing t₂₀=2.22; p=.038.

kg over the two sessions ($W=348.5$; $p=.048$), but this did not significantly increase BMI. The overarm throwing was sufficiently reliable for the long-term period ($ICC=.76$), with $SEM=21$ cm and $MDC=57$ cm. The other muscular strength tests had low ICCs (.55-.64). Moreover, the average standing long jump score increased significantly from the first to the second session, by 11 cm ($p=.035$). The walking stair test, for measuring muscular endurance, had good test-retest reliability ($ICC=.92$), but relatively high measurement error ($SEM=3$; $MDC=8$) and showed significant increase over time between session 1 and 2 ($t_{20}=-2.22$; $p=.038$). No difference between session 2 and 3 was found. The modified 6MWT was found reliable ($ICC=.78$), with a SEM of 29m and an MDC of 80m. Measuring the average heart rate during this test was reliable ($ICC=.74$).

DISCUSSION

In this study, we assessed the feasibility and test-retest reliability of nine field-based physical fitness tests in 37 children with moderate or severe levels of ID, with the aim to find appropriate tests to use in this specific population. Body composition (BMI, waist circumference) and cardiorespiratory fitness (6MWT with modifications) can be measured feasibly and reliably, whereas muscular strength and strength endurance tests are hardly feasible and hardly reliable in this group. Only overarm throwing (strength) and stair climbing (endurance) had sufficient long-term reliability.

The reliability coefficients found for body composition measurements are comparable with those in typically developing children [43, 44], adolescents with mild to moderate levels of ID [8, 22] and adults with severe levels of ID and sensory disabilities [45]. We found a small, but significant, increase of weight between sessions (0.6kg), not effecting the reliability of BMI. The weight increase is likely to be caused by the day-to-day variation within an individual [46], or difference in the clothes they were wearing. The maximum difference was 1.5 kg, and increases as well as decreases in body weight were seen (see Bland-Altman plot 3.A, appendix 1).

For muscular strength, low feasibility and reliability were found. The throwing tests were hard to understand for the participants (cognitive issues were the main reason for dropping out), and participants with a higher motor development were more likely to perform the test according to protocol. As these tests ask a lot of motor coordination and task understanding, it is not surprising that we found lower reliability than the values found in typically developing children or children with mild levels of ID. In these populations, ICCs of .7-.9 were found [23, 34]. Only the overarm throwing task had sufficient reliability, but the SEM was large (23 cm) and the completion rate low (30%). Possibly, throwing a bag is not an activity that these children are used to. We believe that practicing these tasks will eventually increase the feasibility. The motor skills will develop, and task understanding could possibly be improved

in part of the group of children. However, if we want feasible tests to be applicable in a wide population, tests need to be easy to understand and perform, and not require much practice and familiarization. Because there is no alternative test that has been shown feasible and reliable in children with moderate to severe levels of ID, overarm throwing can be used for children that can perform this task, until a more suitable test is found.

The reliability of the walking stair test was sufficient and comparable with studies in children with no ID [34] and children with mild levels of ID [23]. Stair climbing is a functional task that children often perform during their activities of daily living, and, therefore, the task understanding was supposed to be greater. In this sample, the score on the walking stair test increased significantly between session 1 and 2, and not between session 2 and 3. Therefore, we recommend practicing the whole test with children at least one time before scoring. A limitation of this test is that children with low motor development cannot perform this test, as can be seen in the results.

The 6MWT was modified to make it more feasible for children with moderate to severe levels of ID. The results we found are comparable to the results found in adolescents who were overweight and had mild to moderate levels of ID ($ICC=.82$, $SEM=30m$, $MDC=83m$) [36]. In typically developing children [47], in children with a chronic health condition [35], and in adults with ID [37] ICCs of .84 - .98 were found.

We measured heart rate during the 6MWT with two purposes. For safety purposes, we monitored heart rate during the test to check for unusual high or low heart rates. Secondly, the achieved heart rate gives a rough estimation of the amount of effort during the test. The 6MWT is developed as a submaximal test [39], but people with low physical fitness or disease can reach near-maximum levels [48]. The peak heart rates we found were submaximal and comparable to those found in typically developing children, where average peak heart rates of 136 – 153 bpm were found during the 6MWT [49-51] indicating that the amount of effort is comparable to these studies. One study demonstrated lower peak heart rates in typically developing children, with an average of 112 bpm during the 6MWT [52].

The modifications we made in the 6MWT were successful in decreasing the dropout rate for this specific pediatric group, but they are likely to influence the walking distance. Walking on a continuous track [53, 54] and different task instructions [55] are known to increase the test score. Holding the hand of the pacer can increase stability and confidence, and it is possible that participants adapted their walking speed to that of the tester, even though the testers were instructed not to pull them along. Therefore, the results of the modified 6MWT are more likely to overestimate than underestimate the cardiorespiratory fitness of the child. To date, no studies have studied the effect of walking with a pacer, although this has been done more often in individuals with ID [37, 56, 57]. Ideally no modifications are made to the original test protocol, in order to enable comparison between studies. Yet, when modifications are inevitable, clear description of the adaptations and consistent use of the modification over time is important.

Study limitations

The study sample was a heterogeneous group with an age range of 2 to 17 years. The heterogeneity was also seen in the wide range of scores on the tests, which inflates the ICC results, because the ICC compares the variance between subjects with the variance within subjects. On the other hand, because the target population is a heterogeneous group, this study sample gives a good reflection of the actual situation and improves generalizability of the study results. The ICCs might, therefore, be less suitable for this study than the SEM and MCD.

Due to the high dropout rates in the muscular strength tests, the reliability results are based on small sample sizes. Because of this, the power of the results to detect differences was decreased and estimation of the ICCs less precise (e.g., large confidence interval).

All tests were performed in a fixed sequence. In our sequence, we took into account the load on upper and lower extremities and familiarization with the tester and the test environment by starting with the easy tests of body composition and the modified 6MWT. The sequence of tests can be of influence on the results (i.e., at the end of the test session, energy levels, motivation, or concentration can be reduced). In our results, the reliability of the muscular strength tests at the end of the session was lower, but the dropout rate was the same.

Directions for future research

We indicated an ICC score of .7 to be sufficient for group analysis. However, for individual analysis an ICC of .9 is required [58]. Therefore, and because of the high MDCs we found, we recommend using the selected tests only for group analysis, and not for individual evaluation at this point. Further research is required in subgroups to find out if the tests can be used in specific groups based on age or level of ID.

The validity of the selected tests has not been assessed in children with moderate to severe levels of ID yet. The assessment of criterion validity will be difficult, because gold standard tests are often complicated lab-based tests. Possibly, construct validity is a feasible alternative.

Conclusion

Measuring physical fitness of children with moderate to severe levels of ID with field-based tests is feasible, and can be reliably measured for the purpose of group analysis by using BMI, waist circumference, overarm throwing, stair climbing, and the modified 6-minute walk test.

REFERENCES

1. Smith, J.J., et al., *The health benefits of muscular fitness for children and adolescents: a systematic review and meta-analysis*. Sports Med, 2014. **44**(9): p. 1209-23.
2. Ortega, F.B., et al., *Physical fitness in childhood and adolescence: a powerful marker of health*. Int J Obes (Lond), 2008. **32**(1): p. 1-11.
3. Ortega, F.B., et al., *Cardiorespiratory fitness and sedentary activities are associated with adiposity in adolescents*. Obesity (Silver Spring), 2007. **15**(6): p. 1589-99.
4. Hurtig-Wennlof, A., et al., *Cardiorespiratory fitness relates more strongly than physical activity to cardiovascular disease risk factors in healthy children and adolescents: the European Youth Heart Study*. Eur J Cardiovasc Prev Rehabil, 2007. **14**(4): p. 575-81.
5. Oeseburg, B., et al., *Prevalence of chronic health conditions in children with intellectual disability: a systematic literature review*. Intellect Dev Disabil, 2011. **49**(2): p. 59-85.
6. Gillespie, M., *Cardiovascular fitness of young Canadian children with and without mental retardation*. Education and Training in Developmental Disabilities, 2003. **38**(3): p. 296-301.
7. Golubovic, S., et al., *Effects of exercise on physical fitness in children with intellectual disability*. Res Dev Disabil, 2012. **33**(2): p. 608-14.
8. MacDonncha, C., et al., *Reliability of Eurofit physical fitness items for adolescent males with and without mental retardation*. Adapted Physical Activity Quarterly, 1999. **16**(1): p. 86-95.
9. Mercer, V.S. and C.L. Lewis, *Hip Abductor and Knee Extensor Muscle Strength of Children with and without Down Syndrome*. Pediatr Phys Ther, 2001. **13**(1): p. 18-26.
10. Halle, J.W., D. Gabler-Halle, and Y.B. Chung, *Effects of a peer-mediated aerobic conditioning program on fitness levels of youth with mental retardation: two systematic replications*. Ment Retard, 1999. **37**(6): p. 435-48.
11. Frey, G.C. and B. Chow, *Relationship between BMI, physical fitness, and motor skills in youth with mild intellectual disabilities*. International Journal of Obesity, 2006. **30**(5): p. 861-867.
12. Hartman, E., et al., *On the relationship between motor performance and executive functioning in children with intellectual disabilities*. J Intellect Disabil Res, 2010. **54**(5): p. 468-77.
13. Shea, S.E., *Mental retardation in children ages 6 to 16*. Semin Pediatr Neurol, 2006. **13**(4): p. 262-70.
14. Simons, J., et al., *Validity and reliability of the TGMD-2 in 7-10-year-old Flemish children with intellectual disability*. Adapt Phys Activ Q, 2008. **25**(1): p. 71-82.
15. Westendorp, M., et al., *The relationship between gross motor skills and academic achievement in children with learning disabilities*. Res Dev Disabil, 2011. **32**(6): p. 2773-9.
16. Westendorp, M., et al., *Are gross motor skills and sports participation related in children with intellectual disabilities?* Res Dev Disabil, 2011. **32**(3): p. 1147-53.
17. Fernhall, B., et al., *Validation of cardiovascular fitness field tests in children with mental retardation*. Am J Ment Retard, 1998. **102**(6): p. 602-12.
18. Pizarro, D.C., *Reliability of the health related fitness test for mainstreamed educable and trainable mentally handicapped adolescents*. ADAPT PHYS ACT Q, 1990. **7**(3): p. 240-248.
19. Wouters, M., H.M. Evenhuis, and T.I.M. Hilgenkamp, *Systematic review of field-based physical fitness tests for children and adolescents with intellectual disabilities*. Res Dev Disabil, 2017. **61**: p. 77-94.
20. Smits-Engelsman, B.C.M. and W.F.M. Verhoef-Aertssen, *Functional Strength Measurement (FSM): Manual*. . 2012, Meteren, the Netherlands: FSM Production.

21. Bandini, L.G., et al., *Is body mass index a useful measure of excess body fatness in adolescents and young adults with Down syndrome?* J Intellect Disabil Res, 2012.
22. Tejero-Gonzalez, C.M., et al., *Reliability of the Alpha Health-Related Fitness Test Battery in Adolescents with down Syndrome.* Journal of Strength and Conditioning Research, 2013. **27**(11): p. 3221-3224.
23. Bullens, H., *[Functional strength in children with mild intellectual disability measured with the FSM, a test-retest reliability study]*, in *Master Pediatric Physical Therapy (MPPT)*. 2012, Avans +: Breda, the Netherlands.
24. Bayley, N., *Bayley scales of infant and toddler development, third edition* 2006, San Antonio, TX: Pearson Education, Inc.
25. Tellegen, P.J., et al., *Snijders-Oomen Nonverbal Intelligence Test, SON -R 2½-7, Manual & Research Report.* . 1998, Lisse, the Netherlands: Swets & Zeitlinger.
26. World Medical Association. *Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects.* 2013 [cited 2018 December 12]; Available from: <https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/doh-oct2013-jama-2/>.
27. American College of Sports Medicine, *ACSM's Guidelines for Exercise Testing and Prescription.* 10th ed. 2017, Philadelphia, PA: Lippincott Williams & Wilkins.
28. Ganley, K.J., et al., *Health-related fitness in children and adolescents.* Pediatr Phys Ther, 2011. **23**(3): p. 208-20.
29. Ruiz, J.R., et al., *Predictive validity of health-related fitness in youth: a systematic review.* Br J Sports Med, 2009. **43**: p. 909-923.
30. de Onis, M., et al., *Development of a WHO growth reference for school-aged children and adolescents.* Bull World Health Organ, 2007(85): p. 661-8.
31. WHO Multicentre Growth Reference Study Group, *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development.* . 2006, World Health Organization: Geneva. p. 312.
32. de Onis, M. and T. Lobstein, *Defining obesity risk status in the general childhood population: which cut-offs should we use?* Int J Pediatr Obes, 2010. **5**(6): p. 458-60.
33. Taylor, R.W., et al., *Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy X-ray absorptiometry, in children aged 3-19 y.* Am J Clin Nutr, 2000. **72**(2): p. 490-5.
34. Aertssen, W.F., G.D. Ferguson, and B.C. Smits-Engelsman, *Reliability and Structural and Construct Validity of the Functional Strength Measurement in Children Aged 4 to 10 Years.* Phys Ther, 2016. **96**(6): p. 888-97.
35. Bartels, B., J.F. de Groot, and C.B. Terwee, *The six-minute walk test in chronic pediatric conditions: a systematic review of measurement properties.* Phys Ther, 2013. **93**(4): p. 529-41.
36. Elmahgoub, S.S., et al., *Reproducibility, validity and predictors of six-minute walk test in overweight and obese adolescents with intellectual disability.* Disabil Rehabil, 2012. **34**(10): p. 846-51.
37. Nasuti, G., L. Stuart-Hill, and V.A. Temple, *The Six-Minute Walk Test for adults with intellectual disability: a study of validity and reliability.* J Intellect Dev Disabil, 2013. **38**(1): p. 31-8.
38. Boer, P.H. and S.J. Moss, *Validity of the 16-metre PACER and six-minute walk test in adults with Down syndrome.* Disabil Rehabil, 2016: p. 1-9.
39. American Thorax Society, *Guidelines for the sixminute walk test.* . Am J Respir Crit Care Med, 2002. **166**: p. 111-117.

40. Atkinson, G. and A.M. Nevill, *Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine*. Sports Med, 1998. **26**(4): p. 217-38.
41. De Vet, H.C.W., et al., *Mesurement in Medicine*. 2011, Cambridge: Cambridge University Press.
42. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement*. Lancet, 1986. **1**(8476): p. 307-10.
43. Berkson, S.S., et al., *Reliability of height and weight measurements collected by physical education teachers for a school-based body mass index surveillance and screening system*. J Sch Health, 2013. **83**(1): p. 21-7.
44. Lohman, T., A. Roche, and R. Martorell, *Anthropometric standardization reference manual*. . 1988, Champagne, Illinois: Human Kinetic Books.
45. Waninge, A., et al., *Feasibility and reliability of body composition measurements in adults with severe intellectual and sensory disabilities*. J Intellect Disabil Res, 2009. **53**(4): p. 377-88.
46. Khosla, T. and W.Z. Billewicz, *Measurement of Change in Body-Weight*. Br J Nutr, 1964. **18**: p. 227-39.
47. Li, A.M., et al., *The six-minute walk test in healthy children: reliability and validity*. Eur Respir J, 2005. **25**(6): p. 1057-60.
48. Jehn, M., et al., *The 6-min walk test in heart failure: is it a max or sub-maximum exercise test?* Eur J Appl Physiol, 2009. **107**(3): p. 317-23.
49. Lammers, A.E., et al., *The 6-minute walk test: normal values for children of 4-11 years of age*. Arch Dis Child, 2008. **93**(6): p. 464-8.
50. Nixon, P.A., M.L. Joswiak, and F.J. Fricker, *A six-minute walk test for assessing exercise tolerance in severely ill children*. J Pediatr, 1996. **129**(3): p. 362-6.
51. Li, A.M., et al., *Standard reference for the six-minute-walk test in healthy children aged 7 to 16 years*. Am J Respir Crit Care Med, 2007. **176**(2): p. 174-80.
52. Priesnitz, C.V., et al., *Reference values for the 6-min walk test in healthy children aged 6-12 years*. Pediatr Pulmonol, 2009. **44**(12): p. 1174-9.
53. Sciurba, F., et al., *Six-minute walk distance in chronic obstructive pulmonary disease: reproducibility and effect of walking course layout and length*. Am J Respir Crit Care Med, 2003. **167**(11): p. 1522-7.
54. Bansal, V., et al., *Modifying track layout from straight to circular has a modest effect on the 6-min walk distance*. Chest, 2008. **133**(5): p. 1155-60.
55. Weir, N.A., et al., *The influence of alternative instruction on 6-min walk test distance*. Chest, 2013. **144**(6): p. 1900-5.
56. Lavay, B., G. Reid, and M. Cressler-Chaviz, *Measuring the cardiovascular endurance of persons with mental retardation: a critical review*. Exerc Sport Sci Rev, 1990. **18**: p. 263-90.
57. Oppewal, A., et al., *Cardiorespiratory fitness in individuals with intellectual disabilities-A review*. Res Dev Disabil, 2013. **34**(10): p. 3301-3316.
58. Nunnally, J.C. and I.H. Bernstein, *Psychometric Theory*. 3rd ed. 1994, New York: McGraw Hill.

APPENDIX

Bland-Altman Plots of Long-Term (A) and Short-Term (B) Results

