Systematic review of field-based physical fitness tests for children and adolescents with intellectual disabilities

Marieke Wouters, Heleen M. Evenhuis, Thessa I. M. Hilgenkamp.

Research in Developmental Disabilities (2017) 61: 77-94.

ABSTRACT

Background: Testing physical fitness in children and adolescents with intellectual disabilities (ID) can be challenging. This review provides an overview on psychometric properties of field-based physical fitness tests studied in children and adolescents with ID.

Methods: A literature search was performed in March 2014. Studies were included if they evaluated feasibility, reliability and/or validity of a field-based physical fitness test in children and adolescents with ID.

Results: Twenty-six papers met the inclusion criteria and described 18 tests on body composition (4), muscular strength (4), muscular endurance (6), and cardiorespiratory fitness (4). Best results on feasibility, reliability and/or validity were found for bioelectric impedance analysis, body mass index, grip strength, arm hang and distance run/walk tests. These results were mainly found in adolescents with mild to moderate ID.

Conclusion: Some tests were found feasible, reliable and/or valid in subgroups of children and adolescents with ID, but not in children and adolescents with all ages and levels of ID. Further assessment is needed before wider application in all children and adolescents with ID.

INTRODUCTION

In children, having an intellectual disability is associated with many comorbidities like epilepsy, cerebral palsy (CP) and anxiety disorders and can influence health [8]. Based on the available literature, their physical fitness also seems to be lower than that of their typically developing peers [27, 31, 34, 62-66], making them particularly vulnerable to health problems in adulthood. In the general pediatric population, high levels of physical fitness are associated with lower risks of a range of health outcomes, such as cardiovascular diseases, diabetes mellitus and mental health [18-20, 67].

Measuring physical fitness in children with ID has its challenges. First of all, any limitations in intellectual functioning (e.g. learning, problem solving and adaptive behavior) [3] will influence children's understanding of standardized task instructions and their ability to perform the test according to these instructions. They often need adapted communication, such as sign language and icons. Also, these children often lack the understanding or motivation to perform the test with the required complete exertion or maximum performance [56]. Next to these cognitive problems, children with ID often have additional sensory impairments, limb malfunctions, growth or motor development delays that can interfere with test outcomes [47, 53, 57-60, 68]. Test results can also be influenced by autism spectrum disorder or challenging behavior, which both are more common in children with ID [8]. Besides the above-mentioned difficulties, visiting unfamiliar settings or being subject to new equipment for test assessments can be stressful and thus hamper test performance.

In previous research, both field-based and lab-based physical fitness tests were used in children and adolescents with ID. Many researchers preferred to use field-based instead of lab-based tests [31, 34, 62-64], most likely because they are related to functional tasks, relatively easy to understand, and can be conducted on-site with simple and inexpensive equipment. However, the field-based tests that were applied, have been developed for the general population. An overview of the psychometric properties in children with ID is required. Existing information on psychometric properties in adults with ID [69] is not applicable, because children are still in the process of developing their motor skills, which results in age-related differences in the skills they are expected to be able to do, as well as in differences in interpretation of results.

With this paper we aim to provide an overview of published psychometric properties of field-based physical fitness tests in children with ID, to facilitate both clinical therapists and researchers in choosing appropriate tests to measure physical fitness in children with ID.

METHODS

Definition physical fitness

In this review we focused on health-related physical fitness, which contains "those specific components of physical fitness that have a relationship with good health" [15]. According to the American College of Sports Medicine [70], health-related physical fitness consists of five measurable components: body composition, muscular strength, muscular endurance, cardiorespiratory fitness, and flexibility. Flexibility has not been included in this review, since the relationship between flexibility and health has not been confirmed in children [16, 17].

Search strategy

A literature search was performed using the following databases in March 2014 without any time constriction: Pubmed, Embase, PsychINFO, Web of Science, Cochrane, Medline and Google Scholar. The search strategy and main key words were "intellectual disability" AND "child" AND ("physical fitness" OR "body composition" OR "muscular strength" OR "endurance") AND "psychometric properties". Adjunct synonyms were used (Table 1). The whole search strategy is available on demand from the corresponding author.

Table 1. Key words and important synonyms used in the search strategy

KEY WORD	SYNONYMS
Intellectual disability	Mental retardation; learning disorder; developmental disability
Child	Infant; adolescent; youth; teenager; pediatric
Physical fitness	Athletic performance; physiology; performance; condition; examination
Body composition	Anthropometry; weight; length; circumference; skinfold; body fat
Muscular strength	Power; force; hand grip; dynamometer
Endurance	Heart function test; cardiorespiratory; ergometry; aerobic capacity
Psychometric property	Reproducibility; reliability; feasibility; validity; accuracy

Inclusion criteria

Papers were included in this review when the following criteria were met:

- The study population consisted of at least 50% children and adolescents (<18 years) with an ID;
- The paper described a study in which the feasibility, reliability, and/or validity of a physical fitness test on body composition, muscular strength, muscular endurance and/or cardio-respiratory fitness was evaluated;
- The studied test was field-based: the test materials can be easily transported, are inexpensive, and the test can be conducted in a gymnasium;
- The full text paper was traceable and written in English.

Reviews on physical fitness tests were not included, but the reference lists of reviews were studied and applicable papers were included in this review.

Data extraction

The papers were screened in a three-step process: firstly on title, secondly on abstract, and finally on the full text. Reference lists of the included papers were checked for additional eligible studies. The selection process was carried out by the first author (MW). Uncertainties about inclusion were discussed with the third author (TH) until consensus was reached. The flow diagram of the search process is depicted in Figure 1.

The first author systematically and critically read the selected papers and extracted the following data: study population (sample size, level of ID, age, gender, specific syndromes or physical disabilities); physical fitness test studied; statistical analysis used; results on feasibility, reliability, and/or validity.

Results are presented in tables per physical fitness test, divided into four components of physical fitness: body composition, muscle strength, muscle endurance and cardiorespiratory fitness. These results are depicted in Table 3 to 6.

When different names were used for the same test, the most frequently used name was chosen. Furthermore, the following categories were used in this review: borderline ID (IQ 70-85), mild ID (IQ 55-70), moderate ID (IQ 35-55), severe ID (IQ 20-35) and profound ID (IQ <20) [71].

After the data was extracted, the results were labeled "sufficient" or "insufficient" (see chapter 2.5), the methodological quality of the papers was assessed (see chapter 2.6), the level of evidence was determined (see chapter 2.7), and this resulted in the summary of the results per test (Table 2). This extraction process can be found in Figure 1.

Labelling results

The results of the individual studies were scored "sufficient" or "insufficient". This was based on criteria of Terwee, Bot [72]. For reliability studies, results were considered sufficient if ICC was ≥0.70, weighted kappa ≥0.70, or Pearson's r ≥0.80. Criterion validity was considered sufficient if the correlation with the gold standard was ≥0.70. Feasibility results were evaluated based on success rates (≥75% sufficient). If a study only reported the number of participants, and gave no information on success rates or drop-outs per test, feasibility was labeled as 'not available'.

Methodological quality assessment

The methodological quality of the papers was assessed using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist, developed for evaluating studies on measurement properties [73]. The items in box B (reliability), box C (measurement error), and/or box H (criterion validity) were rated as "poor", "fair", "good"

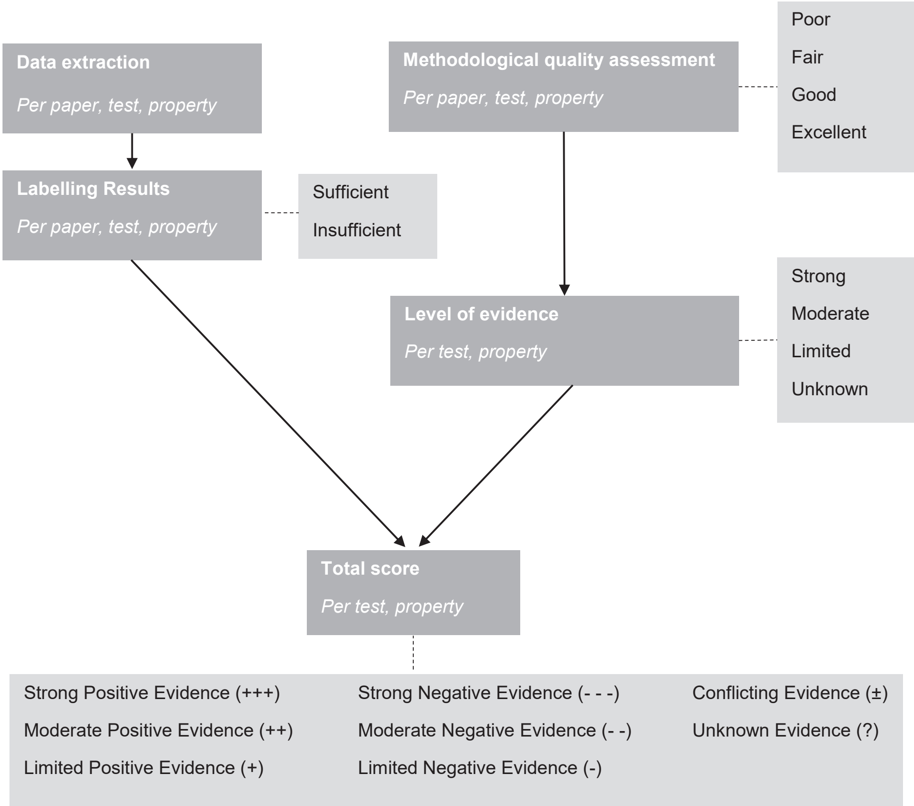

Figure 1. The result extraction process.
The dark grey frameworks display the steps in the extraction process, the light gray frameworks display the outcome possibilities in these steps.

or "excellent", if applicable. An overall score was given per box by taking the lowest score of any of the items within that box. Measurement error gives information on the absolute reliability: the degree to which repeated measurements vary for individuals. This makes comparison of different measurement tools more simple than using only the relative reliability, which indicates the degree to which individuals maintain their position in a sample with repeated measures, measured by e.g. Intraclass Correlation Coefficient (ICC) [74].

Papers only investigating feasibility could not be assessed with the COSMIN checklist. Papers with "poor" quality were not used for the results and conclusions of this review.

Level of evidence

Only the papers with "fair" to "excellent" methodological quality were used to determine the level of evidence per test for each studied psychometric property. The overall level of evidence per test was scored as "strong" when consistent findings in multiple studies of "good" methodological quality were found, or in one study of "excellent" methodological quality; "moderate" when consistent findings in multiple studies of "fair" methodological quality were found, or in one study of "good" methodological quality; "limited" if only one study of "fair" methodological quality was found; and "unknown" if only studies of "poor" methodological quality were found. If no studies were found, the level of evidence was labeled as "not available". This approach was based on the strategy of the Cochrane Back Review Group [75].

In Table 2, the level of evidence is combined with the results on feasibility, reliability and validity per test, and could be "positive", "negative" or "conflicting". The levels of evidence are labeled "positive", when only sufficient results were found; "negative" when only insufficient results were found; and "conflicting" if both sufficient and insufficient results were found.

RESULTS

Twenty-six papers, published between 1973 and 2013, met the inclusion criteria (Figure 2). They describe 18 tests in total, concerning body composition (BIA, BMI, skinfold measurements, waist circumference), muscular strength (grip strength, hand held dynamometry, softball throw, standing long jump), muscular endurance (arm hang, bench press, dumbbell press, isometric push-up, pull-up, sit-up) and cardiorespiratory fitness (fixed distance run/walk, fixed time run/walk, SRT, step test). Table 2 summarizes the results per test and Table 3 to 6 present the characteristics of the papers on respectively body composition tests, muscular strength tests, muscular endurance tests, and cardiorespiratory fitness tests.

Study characteristics

The studies were conducted in North America (n=15), Europe (n=7), Asia (n=2), and New Zealand (n=1). One study included participants from both North America and Europe [76]. Sample sizes ranged from 6 [77] to 81 [78] and age ranged from 2 to 22 years. In a majority of the papers (20/26) the study population had an average age of 12 years or older. In seven papers, children with borderline or mild ID were included, in 11 papers children with mild to moderate ID were included, and in three papers the sample included children with moderate to profound ID. In five papers the level of ID was not reported. Children with Down syndrome (DS) were the exclusive study population in seven papers, and children with ID and CP in four papers.

All tests were originally developed for the general population. The original test protocols were mostly used, but in some papers small tests adjustments were made. The hand grip test was performed sitting instead of standing, and more attempts were allowed during the standing long jump test [79]. Winnick and Short [80] modified the sit-up to a curl-up test, in which the children did not had to come to full sit-up position, and they experimented with different arm positions (study #5). Pizarro [78] extended the length of the sit-up test from 30 to 60 s. For the cardiorespiratory fitness tests a pacer was used in some of the fixed distance tests [81-83]; children were continuously encouraged during the fixed time test [84]; group size was minimalized to one or two children in the Shuttle Run Test (SRT) [79, 85, 86]; Fernhall, Pitetti [81] shortened the track of the SRT to 16 m; physical assistance in the SRT was provided [80]; and the frequencies in the step test were lowered [87].

One paper focused exclusively on feasibility, eight papers described the test-retest reliability, and seven described the criterion validity of one or more tests. Ten papers described more than one psychometric property of one or more tests. For skinfold measurements and cardiorespiratory fitness tests, results on feasibility, reliability and validity were found. For the other tests, results on one or two psychometric properties were missing, especially validity was hardly studied.

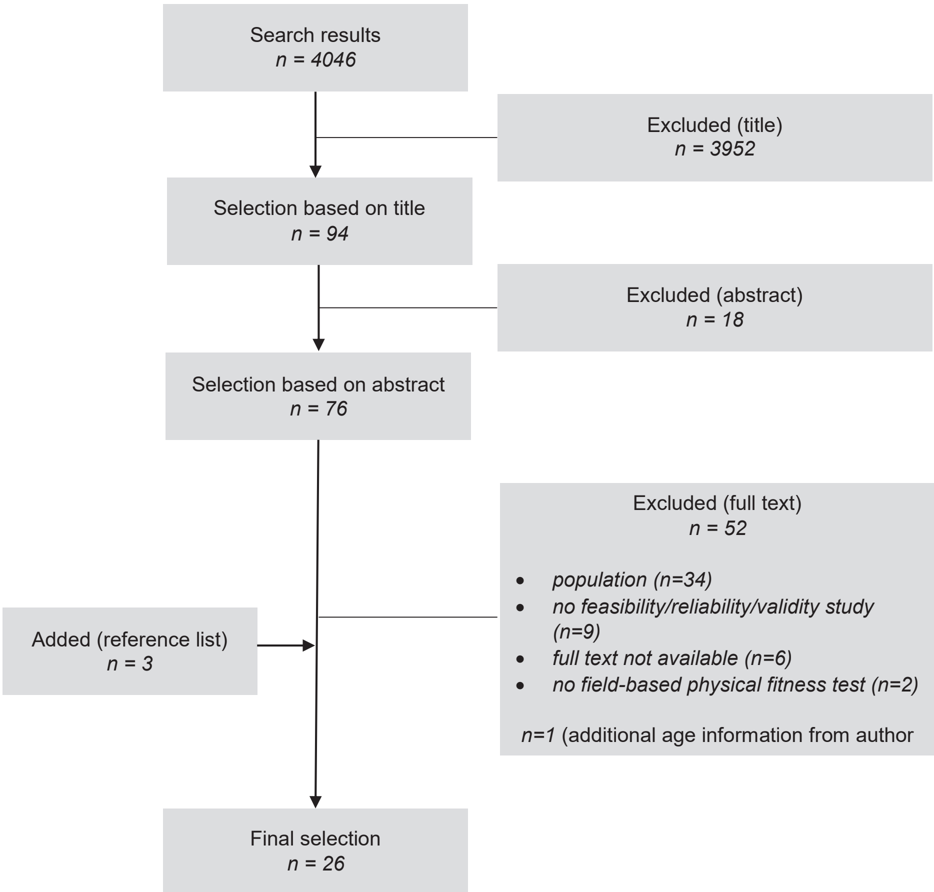

Figure 2. Flow diagram of the paper selection

Table 2. Summary and levels of evidence per test

Test	Feasibility*	Reliability	Measurement error	Criterion validity
Body composition				
Bioelectrical impedance analysis	+	N/A	N/A	+++
Body mass index	N/A	++	N/A	++
Skinfold measurements	±	+++	N/A	−−
Waist circumference	N/A	?	N/A	N/A
Muscular strength				
Grip strength	+	++	N/A	N/A
Hand-held dynamometry	N/A	++	?	N/A
Softball throw	N/A	+	N/A	N/A
Standing long jump	N/A	++	N/A	N/A
Muscular endurance				
Arm hang	+	+	N/A	N/A
Bench press	+	?	N/A	N/A
Dumbbell press	-	?	N/A	N/A
Isometric push-up	-	?	N/A	N/A
Pull-up	-	?	N/A	N/A
Sit-up	±	±	N/A	N/A
Cardiorespiratory fitness				
Fixed distance run/walk	+	++	N/A	+
Fixed time run/walk	±	+	?	−−
Shuttle run	+	±	N/A	±
Step test	±	?	N/A	?

Notes. +++ or −− − strong positive or negative evidence; ++ or −− moderate positive or negative evidence; + or – limited positive or negative evidence; ± conflicting evidence; ? unknown; N/A not available ; * based on results only, no COSMIN-score was available for the feasibility studies.

Quality assessment

According to the COSMIN quality criteria, 13 papers scored "poor", three "fair" and seven "good" on methodological quality. Two papers had different scores on their reliability and validity study. No paper scored "excellent". COSMIN-labels can be found in the final columns in Table 3 to 6. In 25 of the 36 assessments, sample size had the lowest score and was therefore determining for the total score. Within the boxes, high scores ("excellent" or "good") were given to items concerning study design and statistical methods used. We saw methodological flaws in the description of the response rate and study population: in only two papers information on the number of invited children was available [88, 89]. Also, the selection methods were rarely described, as well as the study settings and handling of missing items.

Synthesis of the results

Table 2 summarizes the level of evidence for each of the 18 physical fitness tests. A positive rating, meaning sufficient feasibility, reliability or validity, was given to 25% (18/72) of all possible ratings.

Body composition

Body composition, BIA and BMI were the only tests with positive results with moderate to strong level of evidence (Table 2, Table 3). Feasibility of BIA in children with moderate to profound ID and CP was sufficient [90]. Sufficient validity was found children with DS [88], and children with moderate to severe ID and CP [91]. Sufficient reliability and validity of BMI was found in adolescents with mild ID [64] and DS [92].

Muscular strength

Grip strength was the only muscular strength test with positive feasibility and reliability results (Table 2, Table 4). Sufficient reliability was found in adolescent boys with mild ID [64] and no feasibility problems were found [80].

Muscular endurance

For muscular endurance tests, best results were found for the arm hang test, but only with limited evidence (Table 2, Table 5). Reliability was sufficient in 16 year olds with borderline to moderate ID [93] and Winnick and Short [80] described no feasibility problems.

Cardiorespiratory fitness

Only for the fixed distance run/walk test, solely positive results were found, with limited to moderate levels of evidence (Table 2, Table 6). Feasibility was sufficient, although pacing problems were described [78]. Sufficient reliability was found for children (10-17 y) with borderline to moderate ID [78, 81, 83, 93]. Fernhall, Pitetti [81] also demonstrated sufficient validity.

DISCUSSION

This systematic review shows that a range of field-based physical fitness tests focusing on body composition, muscular strength and endurance, and cardiorespiratory fitness has been evaluated in children with ID. The tests with the most positive scores on feasibility, reliability and/or validity are BMI, BIA, grip strength, arm hang, and fixed distance walk/run test. These tests were mainly studied in adolescents with mild to moderate ID. Younger children and children with more severe ID were underrepresented. Only one study with acceptable ("fair" or "good") quality included children with moderate to severe ID [49]. The sufficient validity

Table 3. Characteristics of included studies on body composition

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score reliability	Validity (criterion)	COSMIN-score validity
<u>Bioelectrical impedance analysis</u>							
Loveday, Thompson [88]	70(30)	-; M: 12.1±3.2/ F: 12.6±3.5y; DS	N/A	N/A	N/A	Highest correlation %BF and **DXA** by Schaefer's equation [94] M: r=.89; MD=-0.6% ; LOA=±11.3%; F: r=.91; MD=-5.9%; LOA= ±13.2%. BIA in girls underestimated %BF compared with DXA	Good
Rieken, Van Goudoever [91]	61(32)	MoID-SID; 10.1±4.3y; CP	N/A	N/A	N/A	Equation to predict TBW based on tibia length. Weight and resistance by BIA ICC=.96; R^2=92%; SEE=1.7kg, compared to **DLW** technique. LOA=±4.4 kg	Good
Veugelers, Penning [90]	35(19)	MoID-PID; 8.7±4.0y; CP	97% completion; 71% felt comfortable. FFM and TBW was calculated for 83%.	N/A	N/A	N/A	N/A
<u>Body mass index</u>							
Bandini, Fleming [92]	32(20)	-; M: 17.4±2.6/ F: 17.1±2.5y; DS	N/A	N/A	N/A	95^{th} percentile cut off vs 85^{th} cut off in relation to %BF by **DXA**: sensitivity 71% vs 100%, specificity 96% vs 60%, positive predictive value 83% vs 41%; negative predictive value 92% vs 100%, efficiency 91% vs 69%	Good

Table 3. Characteristics of included studies on body composition (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score reliability	Validity (criterion)	COSMIN-score validity
MacDonncha, Watson [64]	63(63)	MilD; 15.5±1.2y; -	N/A	1 week; height: ICC=.99; 95%CI=8.4mm; %EoM=0.5mm; weight ICC=.99; 95%CI=1.6; %EoM=2.8kg	Good	N/A	N/A
Tejero-Gonzalez, Martinez-Gomez [79]	17(12)	MilD-MoID; 15.4±2.0y; DS	N/A	1 month; Non-significant difference between test and retest; ICC=.95	Poor	N/A	N/A
Skinfold							
Gonzalez-Aguero, Vicente-Rodriguez [95]	28(16)	-; 16.3±2.6y; DS	N/A	N/A	N/A	Prediction %BF most accurate using Slaugther's equation [96] based on 4 skinfolds compared to **ADP** (r=.105, ns; LOA=±25.8%).	Poor
MacDonncha, Watson [64]	20(20)	MilD; 15.5±1.2y; -	N/A	1 week; sum of 4 skinfolds ICC=.99; 95%CI=2.4mm; %EoM=6.7mm	Good	N/A	N/A
Pizarro [78]	20(10)	MilD-MoID; 12-15y; -	Problems with subscapular and abdominal skinfolds.	N/A	N/A	N/A	N/A
	81(45)	MilD-MoID; 12-15y; -	N/A	1 week; triceps skinfold ICC_{milD}=.98; ICC_{moID}=.99	Good	N/A	N/A
Rieken, Van Goudoever [91]	61(32)	MoID-SID; 10.1±4.3y; CP	80% completion; 20% problems with subscapular, triceps or all skinfolds.	N/A	N/A	Gurka's equation [97] to predict %BF based on sum 4 skinfolds ICC=.51, R^2=27%; Developed equation to predict %BF based on sum 4 skinfolds ICC=.59; R^2=44%; both compared to **DLW** technique.	Good

Table 3. Characteristics of included studies on body composition (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score reliability	Validity (criterion)	COSMIN-score validity
Stallings, Cronk [89]	28(11)	50% SID-PID 50% noID; 7.5±2.3y; CP, 90% wheelchair users	N/A	N/A	N/A	Significant correlations of FM (**D_2O** method) with log-values of single skinfold thickness r=.57-.70, sum of 4 skinfolds r=.67, Slaughter's equation [96] r=.69. In the group without gastrostomy (n=19) only the log-value of the supra-iliac skinfold remained significant (r=.57).	Poor
Tejero-Gonzalez, Martinez-Gomez [79]	16(11)	MilD-MoID; 15.4±2.0y; DS	N/A	1 month; no difference between test and retest; $ICC_{subscapular}$=.64; $ICC_{tricpes}$=.85	Poor	N/A	N/A
van den Berg-Emons, van Baak [98]	22(11)	50% MilD; 10.0±1.5y; CP	N/A	N/A	N/A	%BF by the Durnin & Rahaman's [99] and Slaugther's [96] equations based on 4 skinfolds were significantly lower than the %BF by the **D_2O** method.	Poor
Waist circumference							
Tejero-Gonzalez, Martinez-Gomez [79]	17(12)	MilD-MoID; 15.4±2.0y; DS	N/A	1 month; Non-significant difference between test and retest; ICC=.98	Poor	N/A	N/A

Notes. ID=intellectual disability; MilD=mild ID; MoID=moderate ID; SID=severe ID; PID=profound ID; noID=no ID; DS=Down syndrome; M=male; F=female; y=year; CP=cerebral palsy; BIA=bioelectric impedance analysis; BMI=body mass index; %BF=percentage body fat; DXA=Dual-energy X-ray absorptiometry; FM=fat mass; TBW=total body water; FFM=fat free mass; ADP=air displacement plethysmography; D_2O=deuterium oxide; DLW=double labeled water; %EoM=percentage error of the mean; CI=confidence interval; ICC=intraclass correlation coefficient; LOA=limits of agreement; r=correlation coefficient; R^2=explained variance; SEE=standard error of the estimates; N/A=not available.

Table 4. Characteristics of included studies on muscular strength.

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score reliability	Validity	COSMIN-score validity
Grip strength							
George [100]	48(31)	BoID-MiID; 9-20y; -	N/A	Within 1 session; ICC=.51-.71. Decrease in grip strength over trials(significant F > 5.14).	Poor	N/A	N/A
MacDonncha, Watson [64]	63(63)	MiID; 15.5±1.2y; -	N/A	1 week; ICC=.95; 95%CI=6.9kg; %EoM=22.7kg	Good	N/A	N/A
Tejero-Gonzalez, Martinez-Gomez [79]	16(12)	MiID-MoID; 15.4±2.0y; DS	N/A	1 month; no difference between test and retest; ICC=.86	Poor	N/A	N/A
Winnick and Short [80] Study #1	44(18)	-; 10-17y; 24 DS, -	100% completion	N/A	N/A	N/A	N/A
Study #5	36(14)	-; 11-17y; -	N/A	2 weeks; α=.96; P=.92	Poor	N/A	N/A
Study Houston	8(3)	-; M: 14.8y/ F: 15.3y; -	100% completion	N/A	N/A	N/A	N/A
Hand-held dynamometry							
Mercer and Lewis [33]	17(6)	-; 11.2±2.4y; DS	N/A	6-11 days; hip abductor: ICC=.94, SEM=17.4N, 95%CI=12.9-26.7N; Knee Extensor: ICC=.89, SEM=29.4N, 95%CI=22.2-44.9N	Poor	N/A	N/A
Wuang, Chang [101]	61(30)	-; 14.1±3.3; -	N/A	20 days; ICC=.81 - .96; SEM=0.40-0.57 N; CV=14-38%, except for the ankle plantar flexors (ICC=.69; SEM=0.72N; CV=55%)	Good	N/A	N/A
Softball throw							
Aufsesser [93]	36(26)	BoID-MoID; 16y; -	N/A	5 weeks, once a week; no difference between 5 trial means (F=1.62); r=.83-.94	Fair	N/A	N/A

Table 4. Characteristics of included studies on muscular strength. (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score reliability	Validity	COSMIN-score validity
Standing long jump							
Aufsesser [93]	36(26)	BoID-MoID; 16y; -	N/A	5 weeks, once a week; no difference between 5 trial means(F=0.14); r=.75-.84	Fair	N/A	N/A
MacDonncha, Watson [64]	63(63)	MiID; 15.5±1.2y; -	N/A	1 week; ICC=.98; 95%CI=19.6cm; %EoM=13.5cm	Good	N/A	N/A
Tejero-Gonzalez, Martinez-Gomez [79]	16(12)	MiID-MoID; 15.4±2.0y; DS	N/A	1 month; no difference between test and retest; ICC=.84	Poor	N/A	N/A

Notes. ID=intellectual disability; BoID= borderline ID; MiID=mild ID; MoID=moderate ID; DS=Down syndrome; M=male; F=female; y=years;; %EoM=percentage error of the mean; α=Chronbach's alpha; CI=confidence interval; CV=coefficient of variation; ICC=intraclass correlation coefficient; F=ANOVA test statistic; P=proportion; r=correlation coefficient; SEM=standard error of measurement; N/A=not available.

Table 5. Characteristics of included studies on muscular endurance

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score Reliability	Validity	COSMIN-score Validity
Arm hang							
Aufsesser [93]	36(26)	BoID-MoID; 16y; -	N/A	5 weeks, once a week; Flexed arm hang. no difference between 5 trial means(F=0.73); r=.86-.94	Fair	N/A	N/A
Winnick and Short [80] Study #1	44(18)	-; 10-17y; 24 DS, -	95% completion(at least 1 second).	N/A	N/A	N/A	N/A
Study #5	11(4)	-; 10-12y; -	N/A	2 weeks; Extended arm hang α=.85; P=.72	Poor	N/A	N/A
Study #5	17(8)	-; 13-17y; -	N/A	2 weeks; Flexed arm hang. α=.93; P=.82	Poor	N/A	N/A
Bench press							
Winnick and Short [80] Study #1	25(≤18)	-; 10-17y; 24 DS, -	84% completion; repeated instruction on different days necessary for many participants.	N/A	N/A	N/A	N/A
Study #5	23(9)	-; 10-17y; -	N/A	2 weeks; α=.91; P=.82	Poor	N/A	N/A
Dumbbell press							
Winnick and Short [80] Study #4	12(5)	-; 13-17y; -	58% completion(at least one repetition).	1-7 days; α=.98; P=.91; k=0.61	Poor	N/A	N/A
Isometric push-up							
Winnick and Short [80] Study #1	18(-)	-; 10-17y; 24 DS, -	83% completion(at least 1 second).	N/A	N/A	N/A	N/A
Study #4	38(18)	-; 13-17y; -	29% completion(at least 1 second).	1-7 days; α=.83; P=.94; k=0.77	Poor	N/A	N/A
Study #5	8(1)	-; 10-12y; -	N/A	2 weeks; α=.83	Poor	N/A	N/A

Table 5. Characteristics of included studies on muscular endurance (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score Reliability	Validity	COSMIN-score Validity
Pull up							
Winnick and Short [80] Study #1	44(18)	-; 10-17y; 24 DS, -	45% completion(at least one pull up).	N/A	N/A	N/A	N/A
Sit-up							
Aufsesser [93]	36(26)	BoID-MoID; 16y; -	N/A	5 weeks, once a week; no difference between 5 trial means (F=0.84); r=.60-.86	Fair	N/A	N/A
MacDonncha, Watson [64]	63(63)	MiID; 15.5±1.2y; -	N/A	1 week; 30s. ICC=.95; 95%CI=4.7 repetitions; %EoM=26.1 repetitions	Good	N/A	N/A
Pizarro [78]	81(45)	MiID-MoID; 12-15y; -	82% completion; position infractions.	1 week; Modified sit-up (60s). ICC_{miID}=.83; $ICC_{moID=}$.94	Good	N/A	N/A
Winnick and Short [80] Study #2	40(16)	-; 10-17y; -	25% correct performance; problems in cadence and arm position.	N/A	N/A	N/A	N/A
Study #3	33(14)	-; 10-17y; -	61% correct performance; 51% correct the first time.	N/A	N/A	N/A	N/A
Study #5	25(8)	-; 10-17y; -	95% correct performance; 73% correct the first time.	2 weeks; α=.82; P=.72	Poor	N/A	N/A

Notes. ID=intellectual disability; BoID= boarderline ID; MiID=mild ID; MoID=moderate ID; DS=Down syndrome; y=years; α=chronbach's alpha; %EoM=percentage error of the mean; CI=confidence interval; ICC=intraclass correlation coefficient; F=ANOVA test statistic; k=kappa; P=proportion; r=correlation coefficient; N/A=not available.

Table 6. Characteristics of included studies on cardiorespiratory fitness

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score Reliability	Validity (criterion)	COSMIN-score Validity
FIXED DISTANCE							
300-yard run							
Aufsesser [93]	36(26)	BoID-MoID; 16y; -	N/A	5 weeks, once a week; Run/walk. no difference between 5 trial means(F=0.35); r=.79-.93	Fair	N/A	N/A
Baumgartner and Horvat [77] Group 1	23(15)	MiID-MoID; 12-14y; -	N/A	1-7 days; ICC=.63; no difference among the 3 trial means.	Poor	N/A	N/A
Group 5	6(6)	MiID; 12-14y; -	N/A	1-7 days; ICC=.47; no difference among the 3 trial means.	Poor	N/A	N/A
Group 6	7(7)	MoID; 12-14y; -	N/A	1-7 days; ICC=.84; no difference among the 3 trial means.	Poor	N/A	N/A
600-yard run							
Baumgartner and Horvat [77] Group 2	14(9)	MiID-MoID; 9-12y; -	N/A	1-7 days; ICC=.69; no difference among the 3 trial means.	Poor	N/A	N/A
Group 7	6(6)	MoID; 12-14y; -	N/A	1-7 days; ICC=.25; no difference among the 3 trial means.	Poor	N/A	N/A
Fernhall, Pitetti [81]	34(22)	MiID-MoID; 14.3±2.3y; 8 DS	N/A	2-5 days; Run/walk. ICC=.98	Fair	Correlation time with TM $\mathbf{VO_{2peak}}$ r=-.80; equation R^2=74%; SEE=4.8ml/kg/min	Fair
880-yard run							
Baumgartner and Horvat [35] Group 3	26(16)	MoID; 10-14y; -	N/A	1-7 days; ICC=.75; no difference among the 3 trial means.	Poor	N/A	N/A

Table 6. Characteristics of included studies on cardiorespiratory fitness (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score Reliability	Validity (criterion)	COSMIN-score Validity
Group 4	12(6)	MoID; 15-18y; -	N/A	1-7 days; ICC=.49; decline among the 2 trial means ($p<.05$).	Poor	N/A	N/A
Group 8	15(15)	MoID; 12-14y; -	N/A	1-7 days; ICC=.94; no difference among the 3 trial means.	Poor	N/A	N/A
Fernhall, Pitetti [40]	23(13)	MiID-MoID; 10-17; 8 DS	N/A	2-3 days; Run/walk. r=.96	Poor	Correlation time with TM **VO_{2max}** r=-.60; The equation with time and BMI had multiple R=.67; SEE=7.3ml/min/kg	Poor
Pizarro [36]	81(45)	MiID-MoID; 12-15y; -	89% completion; 75% without stopping; 104 deviations in ID groups(poor pacing, sprinting & walking)	1 week; ICC_{miID}=.80; ICC_{moID}=.90	Good	N/A	N/A
1-mile walk							
Teo-Koh and McCubbin [41]	40(40)	MiID-MoID; 14.1±1.3y; 4 DS	N/A	1-2 weeks; ICC_{time}=.97; ICC_{HRend}=.88; ICC_{HRpeak}=.92	Fair	N/A	N/A
	22(22)	MiID-MoID; 14.2±1.2y; 2 DS	N/A	N/A	N/A	Correlation time with TM relative **VO_{2peak}** r=-.76; The equation with time and weight R^2=64%; SEE=3.84ml/kg/min; Over estimation of 8.0ml/kg/min.	Poor

Table 6. Characteristics of included studies on cardiorespiratory fitness (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score Reliability	Validity (criterion)	COSMIN-score Validity
SHUTTLE RUN							
16-meter SRT							
Fernhall, Pitetti [39]	34(22)	MilD-MoID; 14.3±2.3y; 8 DS	N/A	2-5 days; ICC=.96	Fair	Correlation level with TM **VO_{2peak}** r=.77; equation R^2=79%, SEE=4.4ml/kg/min	Fair
Winnick and Short [38] Study #3	20(-)	-; 10-17y; -	A longer distance was run when the length of the laps was reduced from 20m to 16m.	4 trials with 2 weeks; α=.98; P=1.0; k=1.015	Poor	N/A	N/A
Study #5	35(13)	-; 11-17y; -	N/A	2 weeks; α=.98; P=.93	Poor	N/A	N/A
Study Houston	8(3)	-; M: 14.8/ F: 15.3y; -	100% completion with physical assistance.	N/A	N/A	N/A	N/A
20-meter SRT							
Agiovlasitis, Pitetti [66]	53(28)	MilD-MoID; 14.5±2.9y; DS	N/A	N/A	N/A	Correlation level with TM **VO_{2peak}** r=.45; significant prediction of VO_{2peak}, R^2=23%; SEE=4.8ml/kg/min. LOA=±9.3ml/kg/min	Good
Aufsesser [51]	36(26)	BolD-MoID; 16y; -	N/A	5 weeks, once a week; no difference between 5 trial means(F=0.43); r=.64-.86	Fair	N/A	N/A

Table 6. Characteristics of included studies on cardiorespiratory fitness (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score Reliability	Validity (criterion)	COSMIN-score Validity
Beets, Pitetti [44]	42(31)	MiID; 15.0±3.4y; 3 DS	N/A	≥ 48 h; ICC_{laps}=.90-.91; ICC_{HRpeak}=.62-.78;	Poor	Correlation HR_{peak} with TM **HR_{peak}** M: r=.76; F: r=.81	Good
Fernhall, Pitetti [39]	34(22)	MiID-MoID; 14.3±2.3y; 8 DS	N/A	2-5 days; ICC=.97	Fair	Correlation level with TM **VO_{2peak}** r=.74; equation R^2=77%; SEE=4.5ml/kg/min	Fair
Fernhall, Millar [67]	17(9)	MiID-MoID; 13.7±2.9y; 6 DS	N/A	N/A	N/A	Correlation predicted VO_{2peak} with TM **VO_{2peak}** r=.86; SEE=6.2ml/kg/min, TE=5.6ml/kg/min. LOA=-13.5-10.8ml/kg/min	Poor
Gillespie [43]	30(15)	MiID; 8.0±0.6y; -	N/A	72h; ICC=.53	Fair	N/A	
Guerra, Pitetti [34]	26(15)	MiID; 15.3±2.7y; DS	N/A	N/A	N/A	Correlation predicted VO_{2peak} with TM **VO_{2peak}** r=.54; SE=4.43ml/kg/min, TE=40.43ml/kg/min. Significant difference measured and predicted VO_{2peak} 31%(t(25)=9.8). LOA=1.6-17.5ml/kg/min	Poor
MacDonncha, Watson [5]	63(63)	MiID; 15.5±1.2y; -	N/A	1 week; ICC=.94; 95%CI=21.6 laps; %EoM=36.5 laps	Good	N/A	

Table 6. Characteristics of included studies on cardiorespiratory fitness (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score Reliability	Validity (criterion)	COSMIN-score Validity
Tejero-Gonzalez, Martinez-Gomez [37]	17(12)	MilD-MoID; 15.4±2.0y; DS	N/A	1 month; Non-significant difference between test and retest; ICC=.86	Poor	N/A	N/A
Winnick and Short [38] Study #1	44(18)	-; 10-17y; 24 DS, -	100% completion with physical assistance(at least one lap); 64% completed 6 or more laps.	N/A	N/A	N/A	N/A
STEP TEST							
Pitetti, Fernhall [45]	22(13)	MilD-MoID; 14.7±2.7y; 6 DS	91% completion; physical reasons in DS participants.	3 times with 3-5 day interval; 13 ascents/min. r=.48(ns)	Poor	Correlation HR_{end} with TM **VO_{2peak}**: r=-.31(ns); SEE=10.2ml/kg/min; TE=20.8ml/kg/min	Poor
	22(13)	MilD-MoID; 14.7±2.7y; 6 DS	91% completion; physical reasons in DS participants.	3 times with 3-5 day interval; 15 ascents/min. r=.79	Poor	Correlation predicted VO_{2peak} with TM **VO_{2peak}**: r=-.48; SEE=8.7ml/kg/min; TE=18.2ml/kg/min	Poor
	19(13)	MilD-MoID; 14.7±2.7y; 4 DS	74% completion; 21% physical reasons, 5% motivation.	3 times with 3-5 day interval; 17 ascents/min; r=.83	Poor	Correlation predicted VO_{2peak} with TM **VO_{2peak}**: r=-.46; SEE=8.8ml/kg/min; TE=16.1ml/kg/min	Poor

Table 6. Characteristics of included studies on cardiorespiratory fitness (continued)

Study	N (male)	Population (Level of ID; age; other)	Feasibility	Reliability (test-retest) and measurement error	COSMIN-score Reliability	Validity (criterion)	COSMIN-score Validity
FIXED TIME							
6-minute walk							
Elmahgoub, Van de Velde [42]	61(22)	MilD-MoID; 16.9 ±2.5y; 16 DS, overweight	N/A	N/A	N/A	Correlation distance and bicycle ergometer relative $\mathbf{VO_{2peak}}$ r=.69.	Good
	39(15)	MilD-MoID; 16.8 ±2.6y; ≤16 DS, overweight	N/A	1 week; no difference between test and retest; ICC=.82; SEM=29.8m; SRD=82.6m; LOA=-88.2 and 77.9m.	Fair	N/A	N/A
Winnick and Short [38] Study #1	44(18)	-; 10-17y; 24 DS, -	Run/walk. 100% completion; concerns on motivation and pacing.	N/A	N/A	N/A	N/A
9-minute run							
Pizarro [36]	20(10)	MilD-MoID; 12-15y; -	50% completion; 70% quit running/ walking or completed less than 1000 yards.	N/A	N/A	N/A	N/A

Notes. ID=intellectual disability; BoID= boarderline ID; MilD=mild ID; MoID=moderate ID; DS=Down syndrome; M=male; F=female; y=years; HR=heart rate; SRT=shuttle run test; TM=treadmill; TE=total error; VO_{2peak}=peak oxygen uptake; %EoM=percentage error of the mean; α=Chronbach's alpha; CI=confidence interval; F=ANOVA test statistic; ICC=intraclass correlation coefficient; k=kappa; LOA=limits of agreement; P=proportion; r=correlation coefficient; R^2=explained variance; SEE=standard error of the estimate; SEM=standard error of measurement; SRD=smallest real difference; N/A=not available results of BIA in that study were limited to the selected population, since all the children had severe CP. Insufficient feasibility results were found for several muscular endurance tests (dumbbell press, isometric push-up, and pull-up tests), indicating that these tests are not suitable for children and adolescents with ID. Insufficient validity results were found for skinfold measurements and fixed time run/walk test, questioning the accuracy of these tests for measuring the concepts. One should be cautious using these tests. If desired, adjustments should be made and the psychometric properties should be restudied.

As can been seen in Table 2, no test was thoroughly studied in children with ID. We found only few studies per test, ranging from one (softball throw) to 14 (fixed distance and SRT). The amount of "fair" or "good" quality studies per individual test ranged from zero to five. No "excellent" studies were found, and description of methods was often missing. Studies with good methodological quality are required to enhance the knowledge in this population. We acknowledge doing research in people with ID comes with methodological challenges, but clear description of the study population, setting, selection methods, response rate and statistical methods is achievable and necessary for judging the usefulness and quality of a study.

Only test-retest and not inter-observer reliability was assessed, and measurement error was only studied in hand-held dynamometry and fixed time run/walk test. For the majority of the tests validity had not been studied and the available validity studies for body composition used various gold standards to analyze the criterion validity. A lack of validity studies was to be expected, since the cognitive, motivational, and physical problems can make it difficult to tests these children, and gold standard tests are often complicated lab-based tests. Other validity concepts, like construct validity, should be explored in individuals with ID.

Previous research in adults with ID [52] and younger children without disabilities [53] confirms the results found in this review regarding the feasibility problems of skinfold measurements, and showed comparable results regarding feasibility and reliability of strength tests in older adults with ID [54]. Regarding cardiorespiratory fitness in adults with ID, the review of Pitetti, Rimmer [55], updated by Oppewal, Hilgenkamp [56] concluded that SRT, fixed distance and fixed time tests are reliable, but that it is problematic to estimate VO_{2max} from these field-based tests. We found conflicting evidence for the reliability of SRT, and sufficient criterion validity of fixed distance run. These differences can be specific for the population (children vs. adults) or can be caused by the selection we made based on methodological quality. Pitetti, Rimmer [55] and Oppewal, Hilgenkamp [56] included papers with all methodological quality.

A limitation to this study was that the papers were read and scored by one author. Only in case of doubt did a second author read the paper. This could have had influence on the study selection. Relevant reports could have been missed. However, we estimated that the selection process was framed clearly a priori and did not require difficult judgments.

Conclusion

This review provides insight in the diversity and the level of evidence of field-based physical fitness tests studied children with ID. The results indicated that some tests, originally developed for TD children, were feasible, reliable and/or valid in subgroups of children and adolescents with ID, but these tests were not studied in children and adolescents with all ages and levels of ID. Future research should assess the psychometric properties of the tests more broadly. Researchers and professionals in clinical practice should be cautious by using field-based tests for diagnostics and evaluation.

REFERENCES

1. 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.
2. Casey, A.F., X. Wang, and K. Osterling, *Test-retest reliability of the 6-minute walk test in individuals with down syndrome.* Arch Phys Med Rehabil, 2012. **93**(11): p. 2068-2074.
3. 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.
4. Golubovic, S., et al., *Effects of exercise on physical fitness in children with intellectual disability.* Res Dev Disabil, 2012. **33**(2): p. 608-14.
5. 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.
6. Calders, P., S. Elmahgoub, and D. Cambier, *Physical and Metabolic Fitness of Children and Adolescents with Intellectual Disability - How to Rehabilitate?*, in *Latest Findings in Intellectual and Developmental Disabilities Research*, P.U. Tan, Editor. 2012, InTech.
7. Hartman, E., et al., *Development of physical fitness in children with intellectual disabilities.* J Intellect Disabil Res, 2015. **59**(5): p. 439-49.
8. Salaun, L. and S.E. Berthouze-Aranda, *Physical fitness and fatness in adolescents with intellectual disabilities.* J Appl Res Intellect Disabil, 2012. **25**(3): p. 231-9.
9. Pitetti, K.H., D.A. Yarmer, and B. Fernhall, *Cardiovascular fitness and body composition of youth with and without mental retardation.* Adapted Physical Activity Quarterly, 2001. **18**(2): p. 127-141.
10. Fernhall, B. and K.H. Pitetti, *Limitations to physical work capacity in individuals with mental retardation.* Clinical Exercise Physiology, 2001. **3**: p. 176-185.
11. 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.
12. 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.
13. 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.
14. 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.
15. Schalock, R.L., et al., *Intellectual Disability: Definition, Classification, and Systems of Supports*. 11th ed. Vol. 2016. 2010, Washington, D.C: American Association on Intellectual and Developmental Disabilities
16. 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.
17. Frey, G.C. and B. Chow, *Relationship between BMI, physical fitness, and motor skills in youth with mild intellectual disabilities.* Int J Obes (Lond), 2006. **30**(5): p. 861-7.
18. 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.
19. 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.

20. 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.
21. 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.
22. 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.
23. Shea, S.E., *Mental retardation in children ages 6 to 16.* Semin Pediatr Neurol, 2006. **13**(4): p. 262-70.
24. Hilgenkamp, T.I.M., R. van Wijck, and H.M. Evenhuis, *Physical fitness in older people with ID-Concept and measuring instruments: a review.* Res Dev Disabil, 2010. **31**(5): p. 1027-38.
25. Caspersen, C.J., K.E. Powell, and G.M. Christenson, *Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research.* Public Health Rep, 1985. **100**(2): p. 126-31.
26. American College of Sports Medicine, *ACSM's Guidelines for Exercise Testing and Prescription*. 10th ed. 2017, Philadelphia, PA: Lippincott Williams & Wilkins.
27. 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.
28. Ganley, K.J., et al., *Health-related fitness in children and adolescents.* Pediatr Phys Ther, 2011. **23**(3): p. 208-20.
29. American Psychiatric Association, *Diagnostic and statistical manual of mental disorders: DSM-IV* 1994, American Psychiatric Association,: 4th ed. Washington (DC).
30. Terwee, C.B., et al., *Quality criteria were proposed for measurement properties of health status questionnaires.* J Clin Epidemiol, 2007. **60**(1): p. 34-42.
31. Terwee, C.B., et al., *Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist.* Qual Life Res, 2012. **21**(4): p. 651-7.
32. 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.
33. van Tulder, M., et al., *Updated method guidelines for systematic reviews in the cochrane collaboration back review group*, in *Spine (Phila Pa 1976)*, E.B.o.t.C.C.B.R. Group, Editor. 2003. p. 1290-9.
34. Guerra, M., K.H. Pitetti, and B. Fernhall, *Cross validation of the 20-meter shuttle run test for adolescents with down syndrome.* ADAPT PHYS ACT Q, 2003. **20**(1): p. 70-79.
35. Baumgartner, T. and M. Horvat, *Reliability of field based cardiovascular fitness running tests for individuals with mental retardation.* ADAPT PHYS ACT Q, 1991. **8**(2): p. 107-114.
36. 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.
37. 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.
38. Winnick, J.P. and F.X. Short, *Project Target: Criterion-referenced physical fitness. Standards for adolescents with disabilities. Final Report (Project No. H023C30091-95).* 1998.
39. 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.
40. Fernhall, B., et al., *Validity and Reliability of the 1/2-Mile Run-Walk as an Indicator of Aerobic Fitness in Children With Mental Retardation* Pediatr Exerc Sci, 1996. **8**(2): p. 130-142.

41. Teo-Koh, S.M. and J.A. McCubbin, *Relationship between peak VO2 and 1-mile walk test performance of adolescent males with mental retardation.* Pediatr Exerc Sci, 1999. **11**(2): p. 144-157.
42. 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.
43. Gillespie, M., *Reliability of the 20-Metre Shuttle Run for Children With Intellectual Disabilities.* European Journal of Adapted Physical Activity, 2009. **2**(2): p. 7-13.
44. Beets, M.W., K.H. Pitetti, and B. Fernhall, *Peak heart rates in youth with mental retardation: Pacer vs. treadmill.* Pediatr Exerc Sci, 2005. **17**(1): p. 51-61.
45. Pitetti, K.H., et al., *A step test for evaluating the aerobic fitness of children and adolescents with mental retardation.* Pediatric Exercise Science, 1997. **9**(2): p. 127-135.
46. Loveday, S.J., J.M.D. Thompson, and E.A. Mitchell, *Bioelectrical impedance for measuring percentage body fat in young persons with Down syndrome: Validation with dual-energy absorptiometry.* Acta Paediatr Int J Paediatr, 2012. **101**(11): p. e491-e495.
47. Stallings, V.A., et al., *Body composition in children with spastic quadriplegic cerebral palsy.* J Pediatr, 1995. **126**(5 Pt 1): p. 833-9.
48. Veugelers, R., et al., *Feasibility of bioelectrical impedance analysis in children with a severe generalized cerebral palsy.* Nutrition, 2006. **22**(1): p. 16-22.
49. Rieken, R., et al., *Measuring body composition and energy expenditure in children with severe neurologic impairment and intellectual disability.* Am J Clin Nutr, 2011. **94**(3): p. 759-766.
50. 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.
51. Aufsesser, P.M., *Effects of repeated testing on the reliability of fitness scores of institutionalized mentally retarded individuals.* AM J MENT DEFIC, 1979. **84**(3): p. 313-315.
52. Casey, A.F., *Measuring body composition in individuals with intellectual disability: a scoping review.* J Obes, 2013. **2013**: p. 628428.
53. Castro-Pinero, J., et al., *Criterion-related validity of field-based fitness tests in youth: a systematic review.* Br J Sports Med, 2010. **44**(13): p. 934-43.
54. Hilgenkamp, T.I.M., R. van Wijck, and H.M. Evenhuis, *Feasibility and reliability of physical fitness tests in older adults with intellectual disability: a pilot study.* J Intellect Dev Disabil, 2012. **37**(2): p. 158-62.
55. Pitetti, K.H., J.H. Rimmer, and B. Fernhall, *Physical fitness and adults with mental retardation. An overview of current research and future directions.* Sports Med, 1993. **16**(1): p. 23-56.
56. Oppewal, A., et al., *Cardiorespiratory fitness in individuals with intellectual disabilities-A review.* Res Dev Disabil, 2013. **34**(10): p. 3301-3316.
57. Schaefer, F., et al., *Usefulness of bioelectric impedance and skinfold measurements in predicting fat-free mass derived from total body potassium in children.* Pediatr Res, 1994. **35**(5): p. 617-24.
58. Gonzalez-Aguero, A., et al., *Accuracy of prediction equations to assess percentage of body fat in children and adolescents with Down syndrome compared to air displacement plethysmography.* Res Dev Disabil, 2011. **32**(5): p. 1764-1769.
59. Slaughter, M.H., et al., *Skinfold equations for estimation of body fatness in children and youth.* Hum Biol, 1988. **60**(5): p. 709-23.
60. Gurka, M.J., et al., *Assessment and correction of skinfold thickness equations in estimating body fat in children with cerebral palsy.* DEV MED CHILD NEUROL, 2010. **52**(2): p. e35-41.

61. van den Berg-Emons, R.J., M.A. van Baak, and K.R. Westerterp, *Are skinfold measurements suitable to compare body fat between children with spastic cerebral palsy and healthy controls?* DEV MED CHILD NEUROL, 1998. **40**(5): p. 335-9.
62. Durnin, J.V. and M.M. Rahaman, *The assessment of the amount of fat in the human body from measurements of skinfold thickness.* Br J Nutr, 1967. **21**(3): p. 681-9.
63. George, C., *Reliability estimates of grip strength assessments in the institutionalized mentally retarded.* AM CORRECT THER J, 1973. **27**(6): p. 173-175.
64. 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.
65. Wuang, Y.P., et al., *Test-retest reliabilities of hand-held dynamometer for lower-limb muscle strength in intellectual disabilities.* Res Dev Disabil, 2013. **34**(8): p. 2281-2290.
66. Agiovlasitis, S., et al., *Prediction of VO2peak from the 20-m shuttle-run test in youth with Down syndrome.* Adapt Phys Activ Q, 2011. **28**(2): p. 146-156.
67. Fernhall, B., et al., *Cross validation of the 20-m shuttle run test for children and adolescents with mental retardation.* ADAPT PHYS ACT Q, 2000. **17**(4): p. 402-412.