Obesity

Clinical Outcome Measures for Monitoring Physical Function in Pediatric Obesity: An Integrative Review

Ryan Mahaffey[1], Stewart C. Morrison[2], David Stephensen[3], and Wendy I. Drechsler[1]

Objective: Measuring physical function in children with obesity is important to provide targets for clinical intervention to reduce impairments and increase participation in activities. The objective of this integrative review was to evaluate measurement properties of performance-based measures of physical function in children with overweight and obesity.

Design and Methods: An integrative review of literature published in Cochrane Reviews, SPORTDiscus, CINAHL, PLoS, Medline, and Scopus was conducted.

Results: Twenty-eight studies were eligible and represented 66 performance-based measures of physical function. Assessments of repeatability and feasibility were not conducted in the majority of performance measures reported; only 6-min-timed walk (6MTW) was examined for test-retest repeatability. Measures of flexibility, strength, aerobic performance, anaerobic performance, coordination, and balance demonstrated construct validity and responsiveness; however, findings were inconsistent across all performance-based measures. Multi-item tests of physical function demonstrated acceptable construct validity and responsiveness; however, internal consistency was not determined.

Conclusions: There is moderate evidence that 6MTW is suitable for the measurement of physical function in children with obesity. However, evidence is low for the use of aerobic and anaerobic performance, muscle strength, Movement Assessment Battery for Children, and Bruininks-Oseretsky Test of Motor Proficiency multi-item performance instruments and very low for flexibility, coordination, and balance tests. Based on this review, measurement of physical function using 6MTW is recommended.

Obesity (2016) **00**, 00–00. doi:10.1002/oby.21468

Introduction

Increasing recognition of the health burden and long-term economic challenges of childhood obesity underpins the importance of optimum and effective healthcare interventions for children with obesity. Evidence shows that childhood obesity is associated with compromised musculoskeletal function and musculoskeletal pathologies (1-4) and warrants access to rehabilitation services. Childhood obesity is associated with poorer health-related quality of life (5-7) and linked with a range of functional problems, including an increased risk of pain, discomfort and joint stiffness (particularly in the feet and lower extremities; refs. 8 and 9), postural deformities (10), and orthopedic complications (11,12). Findings from a recent review looking at the impact of childhood obesity on physical function and disability further highlighted problems in children with obesity with deficits in function that included impaired cardiorespiratory fitness, performance in motor tasks, and decrements in muscle strength, gait, and balance (13). Physical activity is key to the prevention and management of obesity (14), maintaining good musculoskeletal health (1,2), and mitigating orthopedic comorbidity and functional impairment (1). Despite this, the complexities and poor robustness of measuring physical function and the ability to measure effective rehabilitation remain a challenge. There is a lack of consensus on what tools are appropriate to characterize and monitor physical function in children with obesity (13). Furthermore, current understanding on the measurement properties of clinical tools used to evaluate the features of physical function is lacking.

Physical function is a complex phenomenon and challenging to measure because of the multidimensional constructs (15). Within the International Classification of Functioning, Disability and Health (ICF) Framework (16), the term function encompasses complex interactions between any given health condition, body structures and functions, activity and participation, and contextual factors (Figure 1). Physical function fits into the ''activity'' domain of the ICF framework, which relates to the execution of a task or action by an individual and has been described as the ability to ''move around'' and

[1] School of Health, Sport and Bioscience, University of East London, London, UK. Correspondence: Ryan Mahaffey (r.mahaffey@uel.ac.uk) [2] School of Health Sciences, University of Brighton, Brighton, UK [3] East Kent University Hospitals NHS Trust, Canterbury, UK.

Funding agencies: Ryan Mahaffey is funded on a 3-year postdoctoral research grant from the William M. Scholl Endowment Fund.
Disclosure: The authors declared no conflict of interest.
Additional Supporting Information may be found in the online version of this article.
Received: 1 June 2015; **Accepted:** 1 December 2015; **Published online 00 Month 2016. doi:10.1002/oby.21468**

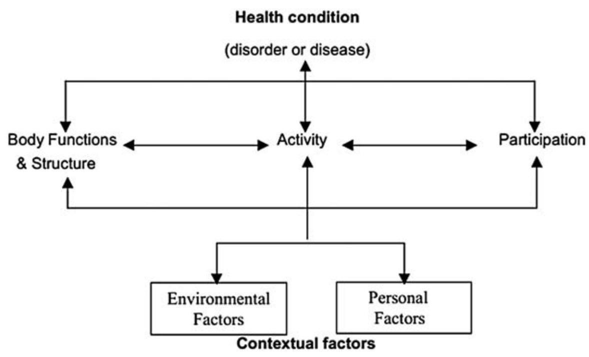

Figure 1 International Classification of Functioning, Disability and Health Framework (16). Reprinted by permission from Towards a Common Language for Functioning, Disability and Health, World Health Organization, pg. 9, copyright 2002.

perform ''daily activities'' (16). The choice of outcome measure for evaluating physical function is important to objectively determine baseline function prior to treatment and to evaluate the efficacy of intervention. Physical ability and function vary widely during childhood growth, and outcome measures appropriate for both the young and adolescent child are necessary. Outcome measures for assessing physical function can be self-reported or standardized performance measures; however, given the challenges with defining function, it is no surprise that multiple outcome measures have emerged in the scientific literature.

The aim of this study was to undertake an integrative review of the literature on the measurement properties (i.e., validity, internal consistency, repeatability, responsiveness, and feasibility) of performance-based methods that have been used to assess physical function in children and adolescents with obesity and overweight. The findings from this integrative review will inform discussion around the optimum measures to use in this field through a robust synthesis and critical evaluation of current knowledge. This in turn will provide clinicians and investigators a basis to choose performance-based methods for use in clinical practice or for future research.

Methods

Using the search strategy described in Supporting Information Table S1, a review of the following data sources from January 1990 to February 2014 was undertaken: Cochrane Library, EBSCO (SPORTDiscus and CINAHL), PLoS, PubMed (Medline), and Elsevier (Scopus). All titles and abstracts generated by the search were independently screened for inclusion by two authors (R.M. and S.C.M.). Disagreement between authors was discussed and consensus reached. The search was restricted to English language and met the following criteria: 1) cross-sectional study design and 2) described the use of a clinician-assessed method, clinical evaluation, or a measurement tool (multi-item/task scored assessment) to record an aspect of physical function in children and adolescents age $\geq$5 to $\leq$19 years with overweight/obesity (OW/OB). Methods recording 1) self-perceived patient reported physical function; 2) longitudinal or intervention studies; 3) measures requiring expensive sophisticated equipment such as three-dimensional gait analysis, isokinetic dynamometers, or accelerometers; 4) no specific aim to study physical function in children with OW/OB; and 5) endogenous OB (e.g., Down's syndrome) were excluded. In addition, conference proceedings, unpublished reports and case series reports, and studies where OW/OB were not standardized to reference population-derived data sets and cutoffs (e.g., BMI *z*-scores) or adiposity estimates (e.g., skin folds and DEXA) were excluded.

Evaluation of measurement properties

The consensus-based standards for the selection of health status measurement instruments (COSMIN) study determined validity, reliability, and responsiveness and were key to evaluate the methodological quality of studies on measurement properties. Standardized definitions and criteria to assess construct validity, internal consistency, repeatability, and responsiveness together with feasibility were applied using previously established guidelines (17,18). Measurement properties were rated as positive (+), negative (−), or indeterminate (?) depending on the methods and results of the study. If no information was available, a zero was recorded.

Construct validity. The extent to which the result of a particular measure is related to other measures, disease severity or clinical evidence. Results were determined using correlation or regression analysis (where there was a correlation with other measures of physical performance or disease activity, ''a+'' was scored; if alternate statistical method undertaken or statistical significance not reported, ''a?'' was scored; where there was no correlation with other measures of physical performance or disease activity, ''a−'' was scored; and if there was no information about construct validity, a ''0'' was scored).

Internal consistency. The extent to which items in a subscale were interrelated and measured the same dimension; determined using Cronbach's alpha (where Cronbach's alpha $\geq$ 0.70, ''a+'' was scored; if alternate statistical method undertaken, ''a?'' was scored; where Cronbach's alpha was <0.70, ''a−'' was scored; and if there was no information about internal consistency, a ''0'' was scored).

Repeatability. The extent to which repeated measurements were reported to be the same. Intraclass correlation coefficient (ICC) or kappa for ordinal/dichotomous data were considered an appropriate measure, and the use of Pearson's/Spearman's correlation coefficient and coefficient of variation (CoV) were considered inadequate due to neglect of systematic error (where ICC or k was $\geq$ 0.70, ''a+'' was scored; if an alternate statistical method undertaken, ''a?'' was scored; where ICC or k was < 0.70, ''a−'' was scored; and if there was no information about test-retest reliability, a ''0'' was scored).

Responsiveness. The ability of the measure to detect difference in the concept being measured or to identify impairment (if significant P, 0.05 group difference in mean $\pm$ SD or median $\pm$ IQR of group comparison, ''a+'' was scored; if no significant group difference, ''a−'' was scored; and where there was no statistical analysis or no information on responsiveness, a ''0'' was scored).

Feasibility. The equipment, space, time, cost, and training required to administer, record, score, and interpret the outcome measure was recorded. The feasibility of the measures was not rated; however, the appropriate considerations are presented. Rating was not undertaken to determine feasibility because the relevance is dependent on the application and intended use of the outcome measure.

Data extraction

A description of the outcome measure was extracted from the included studies along with a description of the patient characteristics: age, gender, number of children assessed, including number with OW/OB, and study design. Data on measurement properties were extracted and recorded in accordance with the criteria previously stated. ''Measures'' was defined as the physical function concept being measured (e.g., strength and aerobic performance), and ''test'' refers to the method, protocol, or technique used (e.g., 6-min-timed walk test [6MTW] and sit-up) to quantify the physical function measure.

Data synthesis

Identified outcome measures are described together with details of their use. Evidence of construct validity, internal consistency, test-retest repeatability, responsiveness, and feasibility is reported and pooled data summarized. Risk of bias in relation to study design and the review process is discussed. Methodological quality of individual studies was assessed using the COSMIN guidelines and summarized using the Cochrane Reviews GRADE criteria. Four levels of quality are specified:

- High: randomized studies with low risk of bias OR methodologically well-designed observational studies with large, consistent, and precise estimates of the magnitude of an effect, and there is a clear consistent dose-outcome response gradient.
- Moderate: randomized studies with unclear bias OR well-designed observational studies with large, consistent, and precise estimates of the magnitude of an intervention effect.
- Low: randomized studies with high risk of bias OR sound observational studies with consistent estimates of the magnitude and direction of an effect.
- Very low: randomized or observational studies with serious methodological limitations and high risk of bias.

Summary of Evidence

Search strategy and selection of articles

The initial search strategy resulted in 4,511 publications. Following screening of titles and abstracts, 103 studies met the inclusion criteria and were accessed for review of the full text, of which 28 eligible studies were included in the review (see Table 1 and Figure 2). Full-text studies were excluded for a number of reasons: 1) the study lacked an aim relating to performance-based or measurement properties of physical function (34 studies); 2) the study adopted sophisticated equipment which would not typically be available in the clinical environment (17 studies); and 3) the study lacked definition of OW/OB according to referenced population data sets and cutoffs (14 studies).

Participant characteristics

Outcome measures were administered to 36,279 children (18,262 girls) aged between 5 and 18 years (Table 1). The number of children with OW/OB in all studies was 9,806 based on BMI *z*-scores and percentage body fat (% BF). The proportion of children with OW/OB in each study ranged from 12 to 100% of the total participants. Only two of the 28 studies evaluated children with OW/OB exclusively (19,20). Of the remaining studies, 25.7% of the participants had OW/OB, and the remainder were children with healthy weight (HW).

Outcome measures

Table 2 provides a description of the 66 performance tests used to evaluate physical function in children with OW/OB children. The measurement domains can be summarized as those evaluating

- flexibility (eight studies, 26,487 children): sit-and-reach, v-sit, and shoulder stretch (21-28);
- strength (16 studies, 32,739 children): sit-ups, partial sit-ups, bent-leg sit-up, horizontal jump, vertical jump, push-ups, bent-arm hang, trunk lifts, sit-to-stand, muscle force (Bruininks-Oseretsky Test of Motor Proficiency [BOT2]), and throwing (21-36);
- aerobic performance (17 studies, 34,185 children): time- and distance-limited run/walks (20-28,30,37-42);
- anaerobic performance (nine studies, 6,545 children): shuttle and sprint runs, timed up-and-down stairs, and timed up-and-go (19,22,26,27,29,31-34);
- coordination and balance (nine studies, 2,914 children): plate tapping, agility tasks, one-leg balance, timed up-and-go, and step tests (22,23,30,34,41,43-46); and
- multi-item instruments recording performance scores (eight studies, 2,334 children; refs. 27,28,30,32,38,43,45,46).

The multi-item instruments can be divided into two general areas of physical function: first, physical fitness (general physical fitness [GPF], The President's Physical Fitness [PPF] Awards, Fitnessgram), and second, motor skills (Körperkoordinationstest für Kinder [KTK], Movement Assessment Battery for Children [MABC], BOT2, and Community Balance and Mobility Scale [CB&M]). Four studies normalized the performance measure to scores based on variation within the cohort ((29,31,36,41); timed up-and-down stairs only in ref. 41), and nine studies normalized the performance measure to referenced standardized scores (27,28,30,34,38,42,43,45,46).

Measurement properties

The measurement properties for the 66 performance tests are detailed in Table 1 and summarized in Table 2. When comparing between studies, the term ''performance'' was used rather than the specific units used to measure physical performance to avoid confusion between different measurement concepts (e.g., aerobic performance can be greater if more distance is covered in a set time or less time is taken to cover a set distance).

Construct validity. Construct validity was assessed in all six of the domains of physical function.

Flexibility. Reduced sit-and-reach distance was associated with OW/OB in boys and girls determined by % BF (21,24); however, when determined by BMI *z*-score, no relationship was found with OW/OB (21). This highlights the challenge of synthesizing data where OW/OB is determined by different methods. No correlation was found when the v-sit version of the sit-and-reach test was used (27). The sit-and-reach and v-sit protocol both reflect flexibility of the back, legs, and shoulders; however, the v-sit version of the test was performed without the need for a box to standardize and stabilize the position of the foot, which may have contributed to the lack of association.

Strength. Reduced sit-up performance was associated with OW/OB determined by BMI *z*-score (21,30), but not when OW/OB was determined by % BF (21), further highlighting discrepancies when

TABLE 1 Measurement properties and outcome measures

Study	*N* (*n* girls) (HW/OW/OB) % Groups	Age range or mean ± SD (years)	Measure of OB	Study design	Construct validity (r, r^2, or OR and *P* values where reported)	Internal consistency and repeatability (ICC or *k* where reported)	Responsiveness (*P* values where reported)
Andreasi et al. (21)	988 (466 girls) (63.8/15.9/16.9)	7-15	BMI reference data (CDC), cutoffs for OW = 85th percentile, OB = 95th percentile (WHO); %BF (SKF); two sites, Lohman equation, girls ≥ 25% and boys ≥ 20% = moderately elevated BF (no cutoff for excessive)	Relationships between categorical measures of physical fitness and groups (UW/HW/OW/OB and %BF normal/moderately excessive/excessive)	Sit-and-reach negatively associated with %BF groups ($P = 0.02$); no association with BMI groups ($P = 0.99$). 1-min abdominal test negatively associated with BMI groups ($P < 0.0001$) and positively associated with %BF groups ($P < 0.0001$). 9-min run/walk negatively associated with BMI groups ($P < 0.0001$) and %BF groups ($P < 0.0001$).		
Ara et al. (31)	715 (373 girls) (33.3 OW and OB)	7-12	BMI reference data and cutoffs (IOTF)	Relationships between categorical measures of physical fitness and groups (HW/OW and OB)	10 × 5 shuttle run time negatively associated with BMI groups in boys ($P = 0.0031$) and girls ($P = 0.002$). Horizontal jump positively associated with BMI groups in boys ($P = 0.008$) and girls ($P < 0.001$).		
Bovet et al. (29)	4,599 (2,262 girls) (8.1 OW/3.1 OB boys) (13.1 OW/4.4 OB girls)	12-15	BMI reference data and cutoffs (IOTF); lean = <10th percentile	Relationships and comparisons between categorical measures of physical fitness and groups (lean/HW/OW/OB)			Significantly ($P < 0.05$) more boys and girls in "good" percentiles for fitness tests were HW compared with OB and OW in sit-ups, horizontal and vertical jump, push-ups, shuttle run, agility run, sprint but not small ball throw and basketball throw. Significantly more HW boys in "good" percentiles than OB boys for small ball throw.

TABLE 1. (continued).

Study	*N* (*n* girls) (HW/OW/OB) % Groups	Age range or mean ± SD (years)	Measure of OB	Study design	Construct validity (*r*, r^2, or OR and *P* values where reported)	Internal consistency and repeatability (ICC or *k* where reported)	Responsiveness (*P* values where reported)
							Significantly more OB and OW boys and girls in "good" percentiles compared with HW for basketball throw.
Burke et al. (37)	602 (321 girls) (18.3 OW/5.0 OB boys) (16.3 OW/4.4 OB girls)	11-14	BMI reference data and cutoffs (IOTF)	Relationships between fitness and groups (HW/OW and OB)	Greater fitness (number of laps in Leger test) associated with decreased risk of OW and OB group in boys (OR = 0.74; 95% CI = 0.55-0.99) and girls (OR = 0.93; 95% CI = 0.91-0.99).		
Čapkauskienė et al. (33)	216 (104 girls) (56.2/15.2/3.7 boys) (59.6/9.6/1.0 girls)	5-7	%BF (SKF); two sites, Lohman equation, girls ≥ 25% and boys ≥ 20% = high BF	Relationships between categorical measures of physical fitness and %BF (no groups)	Girls correlations: 20-m run with %BF ($r = 0.262$; $P < 0.001$). Force with %BF ($r = -0.359$; $P < 0.001$). General fitness with %BF ($r = -0.322$; $P < 0.001$). Boys correlations: 20-m run with %BF ($r = 0.204$; $P < 0.05$). Nonsignificant results not reported.		
Casajus et al. (22)	1,068 (518 girls) (69.1/24.2/6.6)	7-12	BMI reference data and cutoffs (IOTF)	Comparison between physical performance with groups (HW/OW and OB)			HW had higher performances in horizontal jump, sit-ups, bent-arm hang, speed shuttle run, endurance shuttle run (all $P < 0.001$) but not sit-and-reach and plate tapping compared with OW and OB.
D'Hondt et al. (43)	954 (500 girls) (50.0/37.7/12.3)	5-12	BMI reference data and cutoffs (IOTF)	Relationships and comparisons between categorical measures of gross motor coordination with groups (HW/OW/OB)	Moderate-to-severe motor impairment (<15th percentile) 18.9% in HW, 43.3% in OW, 70.8% in OB (Pearson's χ^2 test = 120.9, $P < 0.001$).		In all tests (beam, hop, board, jump, total performance), HW significantly ($P < 0.001$) outperformed OW who Significantly ($P < 0.001$) outperformed OB.

TABLE 1. (continued).

Study	*N* (*n* girls) (HW/OW/OB) % Groups	Age range or mean ± SD (years)	Measure of OB	Study design	Construct validity (r, r^2, or OR and P values where reported)	Internal consistency and repeatability (ICC or k where reported)	Responsiveness (P values where reported)
D'Hondt et al. (45)	117 (60 girls) (52.1/18.8/29.1)	5-10	BMI reference data (Flemish reference data), cutoffs (IOTF)	Relationships with and comparisons between categorical measures of gross and fine motor skill with groups (HW/OW/OB)	Correlation between BMI z-scores with ball skills ($r^2 = 0.04$; $P = 0.03$), static and dynamic balance ($r^2 = 0.21$; $P < 0.01$), total score ($r^2 = 0.12$; $P < 0.01$). Frequency of total score to BMI groups, greater numbers of OB in lower score percentiles for MABC (Pearson's χ^2 test = 17.63; $P < 0.01$).		Significantly better scores for ball skills ($P < 0.05$), static and dynamic balance ($P < 0.01$), total score ($P < 0.01$) in HW and OW compared with OB.
Gonzalez-Suarez et al. (33)	380 (213 girls) (68.2/21.0/10.8 boys) (76.1/16.9/7.0 girls)	11-12	BMI reference data and cutoffs (IOTF)	Relationships with and comparisons between categorical measures of physical fitness with groups (HW/OW/OB)	OR that low-standing broad jump performance associated with OW compared with HW was 2.2 (1.3-3.7) and OB compared with HW was 4.9 (2.1-11.8). OR that low 50-m sprint performance associated with OW compared with HW was 0.8 (0.5-1.3) and OB compared with HW was 2.0 (0.9-4.3).	Intrarater $P = 0.95$, Interrater $P = 0.90$	Significantly greater standing broad jump in HW compared with OW ($P = 0.001$) but not OW compared with OB. Significantly faster 50-m sprint in HW compared with OW ($P = 0.02$) and OW compared with OB ($P = 0.03$).
Graf et al. (23)	344 (167 girls) (77.2/5.5/6.5)	5-8	BMI reference data and cutoffs (German reference values); BMI values ≥10 to < 90 = HW, ≥90 to<97 = OW, ≥97 = OB	Comparisons between measures of motor abilities with groups (UW/HW/OW/OB)			Significant differences found between UW/HW/OW/OB groups for 6-min run, sit-and-reach, horizontal jump, push-ups, sit-ups, one-leg balance ($P < 0.05$) but not side-to-side jumps. Post hoc analysis showed HW ran significantly further in 6 min than OB/OW; HW

TABLE 1. (continued).

Study	*N* (*n* girls) (HW/OW/OB) % Groups	Age range or mean ± SD (years)	Measure of OB	Study design	Construct validity (r, r^2, or OR and *P* values where reported)	Internal consistency and repeatability (ICC or *k* where reported)	Responsiveness (*P* values where reported)
							jumped significantly further standing long jump than OB; UW balanced better on one leg than OB (all $P < 0.05$). Nonsignificant post hoc results not reported.
Graf et al. (38)	668 (327 girls) (78.1/8.1/5.7)	5-8	BMI reference data and cutoffs (German reference values); BMI values ≥10 to < 90 =HW, ≥90 to<97 = OW, ≥97 = OB	Comparisons between categorical measures of motor abilities with groups (UW/HW/OW/OB)			Significant difference across groups for KTK score and 6-min run (both $P < 0.001$). HW and UW performed better than OW and OB in KTK and 6-min run.
Gray and Smith (27)	82 (gender not differentiated) (37.5/21.9/40.6	5-18	BMI reference data and cutoffs (CDC); HW (5th to 85th percentile), at risk for OW (85th to 95th percentile), or OW (>95th percentile)	Relationships with categorical measures of physical fitness (no groups)	Significant correlations with BMI percentiles: physical fitness $r = -25$ ($P < 0.05$), distance $r = -0.24$ ($P < 0.05$). Nonsignificant correlations ($P > 0.05$) between v-sit $r = -0.16$, partial sit-ups $r = -0.27$, right-angle push-ups $r = -0.15$, shuttle run $r = -0.11$, one-leg balance $r = 0.02$.		
Guinhouya (39)	51 (27 girls) (34 HW/17 OW)	10.1 ± 0.8	BMI reference data and cutoffs (IOTF)	Comparisons between physical capacity with groups (HW/OW)			OW produced more work during 6MWT ($P < 0.0001$). No difference between HW and OW for 6MWT distance ($P = 0.88$).
Hamlin et al. (36)	54 (28 girls) (81 HW/19 OW boys) (86 HW/14 OW girls)	8-13	%BF from BIA (no reference data sets for cutoffs)	Relationships with and comparisons between measures of fitness (no groups)	Girls: correlation between CMJ and %BF ($r = -0.57$), $P < 0.05$. Boys: no correlation between CMJ and BF ($r = 0.19$).		
Joshi et al. (28)	7,230 (3,541 girls) (60/18/22)	5-17	BMI reference data and cutoffs (CDC)	Relationships and comparisons between categorical measures of fitness with groups (HW/OW/OB)	Correlation between BMI categories and overall fitness scores ($P < 0.05$). HW achieved higher score. No		Significant differences in achieving healthy fitness zones for PACER less in OB compared with HW ($P < 0.0001$) but not compared with OW; curl-up

TABLE 1. (continued).

Study	*N* (*n* girls) (HW/OW/OB) % Groups	Age range or mean ± SD (years)	Measure of OB	Study design	Construct validity (r, r^2, or OR and *P* values where reported)	Internal consistency and repeatability (ICC or k where reported)	Responsiveness (*P* values where reported)
					post hoc analysis between groups.		less in OB compared with HW ($P < 0.0001$) but not compared with OW; push-up less in OB compared with HW ($P < 0.0001$) and OW ($P = 0.044$); shoulder stretch less in OB compared with HW ($P < 0.0001$) and OW ($P = 0.021$). No significant differences in trunk lift between groups ($P > 0.05$).
Liao et al. (24)	13,500 (6,750 girls) (13.8 OW/19.6 OB boys) (12.7 OW/11.4 OB girls	10-18	BMI reference data (Taiwanese growth charts), cutoffs (HW = ≤84th percentile, OW = ≥85th ≥ 94th percentile, OB = ≥ 95th percentile)	Relationships with and comparisons between categorical measures of physical fitness with groups (HW/OW/OB)	OR of OW being in least fit quintile for flexibility 0.56 boys and 0.55 girls, lower body strength 3.36 boys and 1.95 girls, and CR endurance 2.68 boys and 2.02 girls (all $P < 0.001$). For ab curls 0.89 boys and 0.91 girls, not significant ($P > 0.05$) OR of OB being least fit quintile for flexibility 0.51 ($P < 0.001$) boys and 0.68 ($P = 0.006$) girls, ab strength 1.48 ($P = 0.007$) boys and 1.62 ($P = 0.002$) girls, lower body strength 6.99 ($P < 0.001$) boys and 3.66 ($P < 0.001$) girls, CR endurance 10.44 ($P < 0.001$) boys and 5.40 ($P < 0.001$) girls.		Significant difference of prevalence of HW, OW, OB between most fit and least fit quintile in: flexibility ($P < 0.05$ boys and $< .001$ girls), ab strength ($P < 0.001$ boys and girls), lower body explosive strength ($P < 0.001$ boys and girls, CR endurance ($P < 0.001$ boys and girls). In all cases, prevalence of OW and OB higher in least fit quintile, except flexibility which was variable.
Mak et al. (25)	3,204 (1,578 girls) (65.4/15.4/2.6 boys) (69.0/7.8/0.9 girls)	12-18	BMI reference data and cutoffs (IOTF)	Comparisons between measures of physical fitness with groups (UW/HW/OW and OB)			Significant difference from HW to OW and OB in push-ups and 9-min run ($P < 0.001$) and sit-up ($P < 0.05$) but not sit-and-reach ($P = 0.79$) for

TABLE 1. (continued).

Study	*N* (*n* girls) (HW/OW/OB) % Groups	Age range or mean ± SD (years)	Measure of OB	Study design	Construct validity (r, r^2, or OR and *P* values where reported)	Internal consistency and repeatability (ICC or *k* where reported)	Responsiveness (*P* values where reported)
							boys and sit-up and 9-min run ($P < 0.001$) and push-up ($P < 0.05$) but not sit-and-reach ($P = 0.53$) for girls. In all cases, OW and OB performed poorer.
Malina et al. (26)	688 (171 girls) (88 HW/12 OW and OB)	6-13	BMI reference data and cutoffs (IOTF)	Comparisons between measures of physical fitness with groups (HW/OW and OB)			HW vs. OW and OB (significance at $P < 0.05$). Boys grade 1-3: greater distance run in HW but no difference in sit-and-reach and sit-ups. Boys grade 4-6: greater sit-and-reach and greater distance run in HW but no difference in sit-ups. Girls grade 1-3: no significant differences between groups. Girls grade 4-6: greater distance run in HW but not sit-and-reach and sit-ups. No significant differences between groups for horizontal jump or sprint.
Morinder et al. (40)	49 (19 girls) all OB (test-retest) 250 (124 girls) OB 97 (49 girls) HW (comparison)	8-16	BMI reference data and cutoffs (IOTF)	Repeatability, relationships with, and comparisons between measures of physical capacity with groups (HW/OB)	Correlations between 6MWT and BMI *z*-score $r = -0.42$ ($P < 0.001$).	ICC 0.84, CV 4.3%, R 67.6 m, LoA (2SD) ~68 m ($n = 49$)	Significant difference in 6MTW between OB and HW in boys and girls ($P < 0.001$).
Mota et al. (42)	255 (128 girls) (30.5 OW/13.2 OB boys) (29.1 OW/12.6 OB girls)	8-10	BMI reference data and cutoffs (IOTF)	Relationships with and comparisons between categorical measures of physical fitness with groups (HW/OW/OB)	Girls with OW (OR 0.05, $P < 0.001$) and OB (OR 0.09, $P = 0.001$) more likely to be unfit. No significant likelihood of boys with OW (OR 0.30, $P = 0.06$) and OB (OR 0.47, $P < 0.25$) being unfit.		Statistical differences in meeting healthy fitness zones between weight groups in girls ($P < 0.05$) but not boys ($P = 0.15$).

TABLE 1. (continued).

Study	N (n girls) (HW/OW/OB) % Groups	Age range or mean ± SD (years)	Measure of OB	Study design	Construct validity (r, r^2, or OR and P values where reported)	Internal consistency and repeatability (ICC or k where reported)	Responsiveness (P values where reported)
Nunez-Gaunaurd et al. (30)	86 (39 girls) (55/23/22)	10-15	BMI reference data and cutoffs (CDC)	Relationships with and comparisons between categorical measures of motor proficiency and measures of physical fitness (HW/OW/OB)	Negative correlations with BMI percentiles in OW and OW group (girls only $n = 39$); sit-to-stand $r = -0.36$, $P = 0.024$, BOT-2 SF standard score $r = -0.47$, $P = 0.002$, BOT-2 SF score% $r = -0.47$, $P = 0.003$, ab curls $r = -0.33$, $P = 0.041$. No significant correlations between 6MWT $r = -0.18$, $P = 0.279$, ttimed up-and-down stairs $r = -0.27$, $P = 0.095$, push-ups $r = -0.15$, $P = 0.373$ with BMI percentiles in OW and OB groups.		Sit-to-stand reps differences between all groups ($P = 0.045$), BOT-2 SF standard score differences between HW and OB ($P = 0.003$), BOT-2 SF scores% difference between HW and OB ($P = 0.005$), 6MWT difference between HW and OB ($P = 0.034$), ab curls difference between HW and OB ($P = 0.001$). No significant difference in push-ups ($P = 0.492$) or timed up-and-down stairs ($P = 0.211$) between groups. HW performed better than OW who performed better than OB in all cases.
Pathare et al. (44)	70 (41 girls) (58.6 HW/41.4 OW and OB)	5-9	BMI z-scores (CDC) cutoffs (> 5th <85th = HW, ≥85th percentile = OW and OB)	Comparisons between measures of balance with groups (HW/OW and OB)			Significantly poorer performance in balance of OW and OB compared with HW $P < 0.05$).
Poulsen et al. (34)	116 (56 girls) (23.3 HW/76.7 OW and OB	6-11	BMI reference data (Australian growth charts), cutoffs (IOTF)	Relationships with and comparisons between categorical measures of fundamental movement skills with groups (HW/OW and OB)	Significant negative linear relationships between BMI z-score and balance ($P < 0.01$), strength ($P < 0.001$), running speed and agility ($P < .001$), bilateral coordination ($P < 0.01$) but not upper limb coordination ($P > 0.05$).		Significantly lower scores for OW compared with HW in bilateral coordination, strength, balance, running speed and agility ($P < 0.0001$), upper limb coordination ($P = 0.012$).
Riddiford-Harland et al. (35)	86 (36 girls) (43 HW/43 OB)	8.49 ± 0.5	BMI reference data and cutoffs (IOTF)	Comparisons between measures of functional strength and power with groups (HW/OB)			OB: Significantly higher basketball throw distance ($P = 0.015$) but lower vertical jump ($P = 0.005$) and lower standing long jump ($P = 0.016$) performance compared with HW.

TABLE 1. (continued).

Study	*N* (*n* girls) (HW/OW/OB) % Groups	Age range or mean ± SD (years)	Measure of OB	Study design	Construct validity (r, r^2, or OR and P values where reported)	Internal consistency and repeatability (ICC or k where reported)	Responsiveness (P values where reported)
Sartorio et al. (19)	306 (165 girls) All OB	10-17	BMI reference data (Italian growth charts), cutoffs, low BMI (<2.5 *z*-score), medium BMI (2.5 to 3.0 *z*-score), higher BMI (>3.0 *z*-score) %BF BIA ($n = 77$; ref. (47))	Relationships with and comparisons between measures of leg power with groups (low BMI/medium BMI/high BMI)	Linear regression ($n = 77$) revealed significant association ($r^2 = 0.66$) between FFM (kg) and W for boys and girls ($P < 0.001$).	($n = 13$) same day test-retest (2 h between tests) CV 2.5%	Significantly higher W in higher BMI *z*-score groups ($P < 0.001$).
Tsiros et al. (41)	239 (107 girls) (55.2 HW/44.8 OB)	10-13	BMI reference data (CDC), cutoffs (IOTF), %BF (DEXA)	Relationships with and comparisons between categorical/absolute measures of physical functioning with groups (HW/OB)	Significant relationships between unadjusted timed up-and-down stairs *z*-score and %BF ($r = 0.59$; $P < 0.01$), timed-up-and-go and %BF ($r = 0.26$; $P < 0.01$), 6MWT and %BF ($r = -0.51$; $P < 0.01$).		Significantly lower performance in timed up-and-go, 6MWT, timed up-and-down stairs (all $P < 0.01$) in OB compared with HW.
Vanhelt et al. (20)	97 (52 girls) All OB	7-18	BMI reference data (no data set), cutoffs (97th percentile)	Repeatability of 6MWT in one OB group		ICC 0.99, LoA ± 21.8 m	
Wright et al. (46)	84 (40 girls) (18.1 OW/14.5 OB)	8-11	BMI reference data and cutoffs (CDC) ≥5th to <85th percentile = HW, ≥85th to <95th percentile = OW, ≥95th percentile = OB	Comparisons between categorical measures of balance and mobility (HW and UW/OW/OB)			Significantly poorer performance in OW and OB compared with HW in tandem walking ($P = 0.019$), hopping forward: left ($P = 0.001$) and right ($P < 0.001$), running with controlled stop ($P = 0.021$), forward to backward walk ($P = 0.006$). No significant differences in unilateral stance, 180° tandem pivot, lateral foot scooting, crouch and walk, lateral dodging, walking and looking, walk, look, and carry, descending stairs, step-ups (all $P > 0.05$).

TABLE 2 Summary of measurement properties of outcome measures

Measurement domain	Performance test	Study	Construct validity	Internal consistency	Repeatability	Responsiveness
Flexibility	Sit-and-reach	Andreasi et al. (21)	+ (BF) − BMI	NA	0	0
		Casajus et al. (22)	0	NA	0	−
		Graf et al. (23)	0	NA	0	−
		Liao et al. (24)	+	NA	0	−
		Mak et al. (25)	0	NA	0	−
		Malina et al. (26)	0	NA	0	+ (b 9-12) − (b 6-9, g 6-12)
	v-sit	Gray and Smith (27)	−	NA	0	0
	Shoulder stretches	Joshi et al. (28)	?	NA	0	+
Strength/ resistance	Sit-ups	Andreasi et al. (21)	+ (BF and BMI)	NA	0	0
		Bovet et al. (29)	0	NA	0	+
		Casajus et al. (22)	0	NA	0	+
		Graf et al. (23)	0	NA	0	−
		Joshi et al. (28)	?	NA	0	+
		Mak et al. (25)	0	NA	0	+
		Malina et al. (26)	0	NA	0	−
		Nunez-Gaunaurd et al. (30)	+	NA	0	+
	Partial sit-ups	Gray and Smith (27)	−	NA	0	0
	Bent-leg sit-up	Liao et al. (24)	+	NA	0	+
	Horizontal jump	Ara et al. (31)	+	NA	0	0
		Bovet et al. (29)	0	NA	0	+
		Čapkauskienė et al. (32)	+ (g) − (b)	NA	0	0
		Casajus et al. (22)	0	NA	0	+
		Gonzalez-Suarez et al. (33)	?	NA	?	+
		Graf et al. (23)	0	NA	0	+
		Liao et al. (24)	+	NA	0	+
		Malina et al. (26)	0	NA	0	−
		Riddiford-Harland et al. (35)	0	NA	0	+
	Vertical jump	Bovet et al. (29)	0	NA	0	+
		Hamlin et al. (36)	+ (g) − (b)	NA	0	−
		Riddiford-Harland et al. (35)	0	NA	0	+
	Push-ups	Bovet et al. (29)	0	NA	0	+
		Graf et al. (23)	0	NA	0	−
		Joshi et al. (28)	?	NA	0	+
		Mak et al. (25)	0	NA	0	+
		Nunez-Gaunaurd et al. (30)	−	NA	0	−
	Right-angle push-ups	Gray and Smith (27)	−	NA	0	0
	Bent-arm hang	Casajus et al. (22)	0	NA	0	+
	Trunk lifts	Joshi et al. (28)	?	NA	0	−
	Sit-to-stand	Nunez-Gaunaurd et al. (30)	+	NA	0	+
	Strength (BOT2)	Poulsen et al. (34)	+	NA	0	+
	Small ball throw	Bovet et al. (29)	0	NA	0	+
	Basketball throw	Bovet et al. (29)	0	NA	0	+
		Riddiford-Harland et al. (35)	0	NA	0	+
	1 kg ball throw	Čapkauskienė et al. (32)	+ (g) + (b)	NA	0	0

TABLE 2. (continued).

Measurement domain	Performance test	Study	Construct validity	Internal consistency	Repeatability	Responsiveness
Aerobic performance	9-min run/walk	Andreasi et al. (21)	+ (BF and BMI)	NA	0	0
		Mak et al. (25)	0	NA	0	+
	Multistage/Leger/shuttle run	Bovet et al. (29)	0	NA	0	+
		Burke et al. (37)	?	NA	0	0
		Casajus et al. (22)	0	NA	0	+
	PACER (Fitnessgram)	Joshi et al. (28)	?	NA	0	+
	6-min run	Graf et al. (23)	0	NA	0	+
		Graf et al. (38)	0	NA	0	+
	1/4-, 1/2-, or 1-mile run	Gray and Smith (27)	+	NA	0	0
	6-min timed walk	Guinhouya (39)	0	NA	0	+ (work) − (distance)
		Morinder et al. (40)	+	NA	+	+
		Nunez-Gaunaurd et al. (30)	−	NA	0	+
		Tsiros et al. (41)	+	NA	0	+
		Vanhelt et al. (20)	0	NA	+	0
	800/1,600 m run/walk	Liao et al. (24)	+	NA	0	+
	8/12 run	Malina et al. (26)	0	NA	0	+ (b 6-12, g 9-12) − (g 6-9)
	1-mile run	Mota et al. (42)	?	NA	0	+ (g) − (b)
Anaerobic performance	10 × 5 m shuttle run	Ara et al. (31)	+	NA	0	0
		Bovet et al. (29)	0	NA	0	+
		Casajus et al. (22)	0	NA	0	+
	4 m × 5 m shuttle run	Čapkauskienė et al. (32)	− (g) − (b)	NA	0	0
	No details	Gray and Smith (27)	−	NA	0	0
	Running speed/agility (BOT2)	Poulsen et al. (34)	+	NA	0	+
	Sprint: 40 m	Bovet et al. (29)	0	NA	0	+
	Sprint: 20 m	Čapkauskienė et al. (32)	+ (g) + (b)	NA	0	0
	Sprint: 50 m	Gonzalez-Suarez et al. (33)	?	NA	0	+
	Sprint: ~32 m	Malina et al. (26)	0	NA	0	−
	Margaria step climbing test	Sartorio et al. (19)	+	NA	?	+
Coordination and balance	Plate tapping	Casajus et al. (22)	0	NA	0	−
	Jumping side-to-side	Graf et al. (23)	0	NA	0	−
		D'Hondt et al. (43)	0	NA	0	+
	One-leg balance	Graf et al. (23)	0	NA	0	−
		Gray and Smith (27)	−	NA	0	0
		Pathare et al. (44)	0	NA	0	+
	Walking backward	D'Hondt et al. (43)	0	NA	0	+
	One-leg hop		0	NA	0	+
	Moving sideways		0	NA	0	+
	One-leg/jumping/walking	D'Hondt et al. (45)	+	NA	0	−
	Catching/rolling/throwing		+	NA	0	+
	Posting/threading/writing		−	NA	0	−
	Timed up-and-down stairs	Nunez-Gaunaurd et al. (30)	−	NA	0	−
		Tsiros et al. (41)	+	NA	0	+

TABLE 2. (continued).

Measurement domain	Performance test	Study	Construct validity	Internal consistency	Repeatability	Responsiveness
	Balance	Poulsen et al. (34)	+	NA	0	+
	Bilateral coordination		+	NA	0	+
	Upper limb coordination		−	NA	0	+
	Timed up-and-go test	Tsiros et al. (41)	+	NA	0	+
	Unilateral stance (L and R)	Wright et al. (46)	0	NA	0	−
	Tandem walking		0	NA	0	+
	180° tandem pivot		0	NA	0	−
	Lateral foot scooting (L and R)		0	NA	0	−
	Hopping forward (L and R)		0	NA	0	+
	Crouch and walk		0	NA	0	−
	Lateral dodging		0	NA	0	−
	Walking and looking (L and R)		0	NA	0	−
	Running with controlled stop		0	NA	0	+
	Forward to backward walking		0	NA	0	+
	Walk, look, and carry (L and R)		0	NA	0	−
	Descending stairs		0	NA	0	−
	Step-ups × 1 step (L and R)		0	NA	0	−
Combined performance score	General physical fitness	Čapkauskienė et al. (32)	+ (g) − (b)	0	0	0
	KTK	Graf et al. (38)	0	0	0	+
		D'Hondt et al. (43)	+	0	0	+
	MABC	D'Hondt et al. (45)	+	0	0	+
	The PPF Awards	Gray and Smith (27)	+	0	0	0
	The Fitnessgram: Healthy Fitness Zones	Joshi et al. (28)	+	0	0	0
	BOT2 Short Form	Nunez-Gaunaurd et al. (30)	+	0	0	+
	CB&M	Wright et al. (46)	?	0	0	0

different definitions of OW/OB are used. No correlation was found when partial sit-ups (27) and push-ups (27,30) were used as measures of strength. OW/OB was associated with reduced performance of a horizontal jump in boys and girls (24,31), vertical jump (36) in girls but not boys (32), and sit-to-stand where only girls were assessed (30). An explanation for the contrast in findings is not clear; however, studies showing no association recruited lower participant numbers (i.e., $n < 100$) when compared with the studies which did find associations (n range, 116-13,500; refs. 21,24,31,32,34). Reduced strength, measured by the BOT2 including push-ups, sit-ups, and horizontal jumps, was associated with OW/OB (34). One study found reduced throwing performance was associated with OW/OB in girls measured by % BF (32); however, one study found no association in girls and boys measured by BMI z-score (35). The reason for gender differences is not clear; however, the comparison of strength values and performance in boys is challenging because of pubertal influences; the gender differences found indicates the importance of controlling for pubertal status when studying boys.

Aerobic performance. Reduced performance of 9-min run/walk (21), 1/4/1/2/1-mile run (27), 800/1,600-m walk/run (24), and Leger shuttle run (37) was associated with children with OW/OB. Distance walked in 6 min correlated with OB/OB in two studies (40,41), but not a third study which recruited a smaller cohort ($n = 86$ compared with $n = 347$ and 239 in studies 40 and 41, respectively; ref. 30).

Anaerobic performance. Mixed findings on relationships between anaerobic performance and OW/OB were found. Reduced shuttle run performance was associated with OW/OB in two studies (31,34) but not associated in two other studies (27,32). The two studies to find associations both used 10 m × 5 m shuttles, whereas the two studies that did not find associations used only 4 m × 5 m shuttles (32), or no information was provided (27). It is possible that fewer shuttle runs were insufficient to draw out performance-related relationships with OW/OB. Reduced sprint performance was associated with OW/OB measured by % BF (32) and power output during step climbing measured by fat-free mass in OW/OB (19).

Coordination and balance. Reduced static balance (34) and time taken to ascend and descend a set of stairs (41) was associated with OW/OB and was not associated in other studies (27,30). Differences in timed up-and-down stairs may be related to sample size. Tsiros et al. (41) reported associations with BMI SD on 239

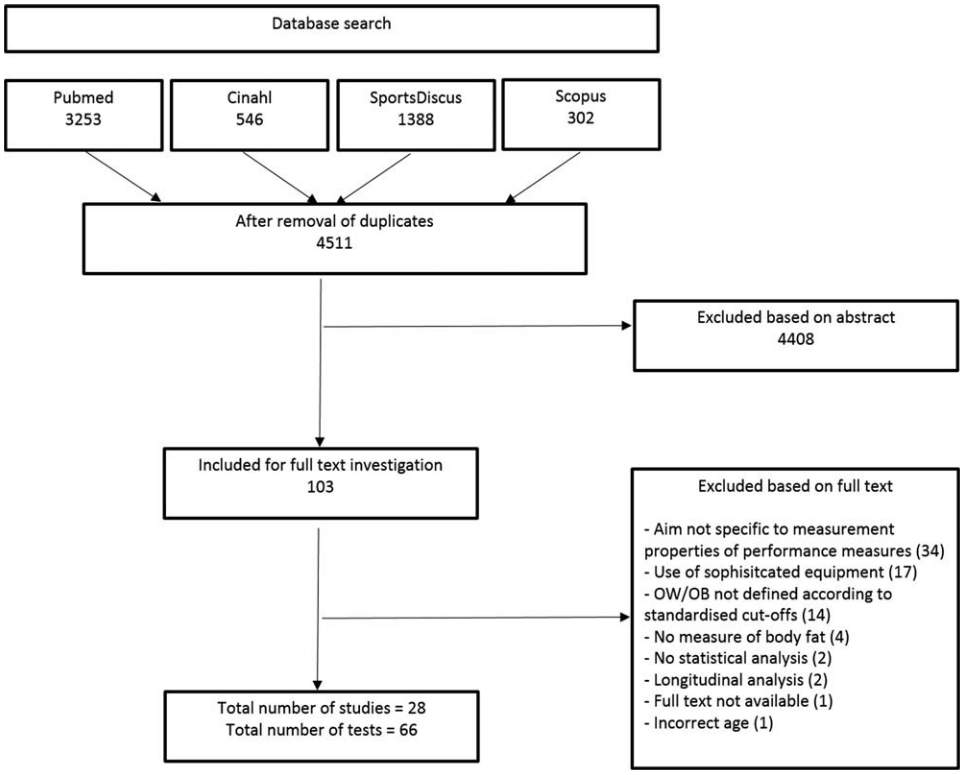

Figure 2 Study selection diagram.

children, whereas Nunez-Gaunaurd et al. (30) reported for HW ($n = 48$) and OW/OB ($n = 39$) separately. Reduced performance of dynamic coordination tasks in the MABC (one-leg balance/jumping/walking; ref. 45) and the BOT2 (tapping foot and finger, jumping jacks, walking forward on a line, standing on one leg on a balance beam, throwing a ball at a target, and catching a tossed ball; ref. 34) along with ball skills (catching/rolling/throwing) were associated with OW/OB. However, manual dexterity (posting/threading/writing) from the MABC was not associated with OW/OB (45).

Multi-item performance instruments. All instruments were associated with poorer performance in children with OW/OB when compared with HW (27,28,30,32,38,43,45,46), although Čapkauskienė et al. (32) found associations in girls but not in boys.

Internal consistency. No study investigated internal consistency of the seven multi-item outcome measures (GPF, KTK, MABC, PPF Awards, Fitnessgram, BOT2, and CB&M) in children with OW/OB.

Test-retest repeatability. Of the 66 measures, repeatability in children with OW/OB was only evaluated for the 6MTW (20,40). Indeterminate ratings were given for many studies because ICC or Kappa statistics were not reported. The use of CoV is considered as an inadequate measure of test-retest repeatability because of its dependency on the scale of measurement (47).

Responsiveness. Responsiveness was assessed in all six of the domains of physical function.

Flexibility. Findings were inconsistent for flexibility performance with two studies reporting poorer performance in OW/OB when compared with HW (26,28), and five studies found no differences between OW/OB and HW (22-26). The study of Joshi et al. (28) was the only study to measure flexibility of the upper body (shoulder stretches); the difference in this flexibility measure may have led to the contrasting finding between other flexibility studies which used sit-and-reach technique. The one study to find significant differences in sit-and-reach performance between HW and OW/OB stratified the sample by gender and age group (6-9 years and 10-13 years; ref. 26). Differences between groups in other studies may have been hidden in variation between age groups and genders.

Strength. Most studies ($n = 21$) reported that strength performance was significantly lower in OW/OB (22-25,28-30,33-35); however, two studies found significantly higher performance in OW/OB (29,35), and seven studies reported no differences between OW/OB and HW groups (23,28,30,36,42). Responsiveness of strength tests for sit-ups, horizontal jump, vertical jump, push-ups, bent-arm hangs, and sit-to-stand demonstrated lower performances in children with OW/OB when compared with HW (22-25,28,30,33-35). Only strength measures where body mass was not lifted against gravity (i.e., throwing), OW/OB performed better than HW (29,35).

TABLE 3 Characteristics of outcome measures

Study	Performance test (measured function)	Scoring method	Summary of procedure	Feasibility
Andreasi et al. (21)	Sit-and-reach (flexibility)	Distance (cm)	Procedures in accordance with Manual de Aplicação de Medidas e Testes, Normas e Critérios de Avaliação	Data collected by physical education professionals
	Sit-ups (abdominal strength/resistance)	Repetitions (1 min)		
	9-min run/walk (aerobic resistance)	Distance (m)		
Ara et al. (31)	Horizontal jump (leg and lower body explosive strength/power)	Distance (cm) converted to cohort quartiles	-	-
	10 m × 5 m shuttle run and turn (running speed and agility)	Time (s) converted to cohort quartiles	Time to complete five cycles (back and forth) between parallel lines separated by 5 m, all children motivated to run as fast as possible	
Bovet et al. (29)	Sit-ups (abdominal strength and endurance)	Repetitions in 30 s converted to cohort percentiles	-	Performed during regular school hours by 25 specially trained instructors over 6-month period
	Horizontal jump (leg and lower body explosive strength/power)	Distance (m) converted to cohort percentiles	Better of two attempts taken	
	Vertical jump (leg and lower body explosive strength/power)	Distance (m) converted to cohort percentiles	Better of two attempts taken	
	Push-ups (upper body muscular strength and endurance)	Repetitions in 60 s converted to cohort percentiles	-	
	Small ball throw (upper body strength/power and coordination)	Distance (m) converted to cohort percentiles	Better of two attempts taken	
	Basketball throw (upper body strength/power and coordination)	Distance (m) converted to cohort percentiles	Better of two attempts taken	
	Multistage run (cardiovascular fitness/endurance)	Laps/distance (m) converted to cohort percentiles	Run between two lines 20 m apart at running pace increased by 0.5 km/h every minute using audio tape	
Burke et al. (37)	Leger shuttle run (cardiovascular fitness/endurance)	Distance (laps/m)	-	-
Čapkauskienė et al. (33)	General physical fitness	Each test result equated to points; point total evaluated against referenced charts and children categorized "low," "right enough," "good," and "high"	-	All tests performed indoors (except 20-m) in morning
	Force test (horizontal jump)	Distance (cm) converted to standardized points	Better of three attempts taken	
	Speed (4 m × 5 m shuttle run)	Time (s) converted to standardized points	Better of two attempts taken	
	Quickness test (20-m sprint)	Time (s) converted to standardized points	Better of two attempts taken	
	Power test (1 kg ball throw)	Distance (cm) converted to standardized points	Better of three attempts taken	

TABLE 3. (continued).

Study	Performance test (measured function)	Scoring method	Summary of procedure	Feasibility
Casajus et al. (22)	Sit-and-reach (flexibility)	Distance (cm)	Reach forward as far as possible from seated position: better of two attempts taken	Physical fitness assessed using European physical test battery (Eurofit)
	Sit-ups (trunk strength and endurance)	Correct repetitions in 30 s	Number of performance sit-ups achieved in 30 s	
	Standing broad jump (explosive strength)	Distance (cm)	Jump for distance from standing start: better of two attempts taken	
	Bent-arm hang (upper body strength and endurance)	Time (0.1 s)	Maintain bent-arm position while hanging from bar	
	Multistage shuttle run (cardiorespiratory endurance)	Steps	Progressively faster walk/run between two 20-m reversing and changing direction	
	10 m × 5 m run and turn (running speed and agility)	Time to complete five cycles (0.1 s)	Run and turn (shuttle) test at maximum speed, two parallel lines drawn on floor 5 m apart	
	Plate tapping (coordination and speed of limb movements)	Time to touch each disc 25 times	Rapid tapping of two plates alternately with preferred hand	
D'Hondt et al. (43)	Two-legged jumping from side to side during 15 s	Raw performance score of each test item converted into standardized Motor Quotient adjusted for age (all items) and gender (hop and jump), converted to percentiles based on German norms; below 15th percentile equates to gross motor coordination problems	Dynamic body coordination, gross motor, dynamic balance skills (in accordance with KTK)	Group of trained examiners, over 6-month period; children wore light sportswear and were barefoot
	Walking backward along balance beams of decreasing width: 6.0, 4.5, 3.0 cm			
	One-legged hopping over foam obstacle with increasing height in consecutive steps of 5 cm			
	Moving sideways on wooden boards during 20 s			
D'Hondt et al. (45)	Manual dexterity, ball skills, static and dynamic balance (in accordance with MABC)	Raw performance score converted into scaled score ranging between 0 and 5, cluster scores and total MABC score calculated from summed percentile scores based on Dutch standardization sample of norms	Eight age-specific motor skill test items, brought together in three clusters	Tests conducted in quiet room at children's schools
Gonzalez-Suarez et al. (33)	Horizontal jump (leg and lower body explosive strength/power)	Distance (m)	Start with bent knees, feet close together, hand by sides, jump as far as possible, measured from heel to start line: average of three attempts taken	Faculty members and interns of College Sports Science, training session prior to data collection, all data collected on same day
	50-m sprint (running speed and anaerobic power)	Time (s)	Two laps of 25 m, run as fast as possible	

TABLE 3. (continued).

Study	Performance test (measured function)	Scoring method	Summary of procedure	Feasibility
Graf et al. (23)	Sit-and-reach (flexibility)	Distance (cm)	Reach forward as far as possible with straight legs	-
	Sit-ups (strength)	Repetitions in 40 s	Feet held down, raise chest until elbow touches knees, lowers without shoulder blades touching floor	
	Standing long jump	Distance (line to rear foot on landing)	Take off and land with both feet, swing arms: better of two attempts taken	
	Push-ups (endurance strength)	Correct repetitions in 40 s	Hands begin on buttocks, move to under shoulders for push-up, hands reduced to buttocks after decent	
	6-min run (endurance)	Distance (m)	54-m "round" ran in small groups (up to eight children)	
	Lateral jumping: side-to-side (coordination, speed, strength, and endurance)	Correct repetitions in 15 s	Jump back and forth over line on floor: better of two attempts taken	
	One-leg balance (coordination and balance)	Number of ground contacts (discontinued if >15 in 30 s)	Stand on small rope barefoot on either leg for 1 min	
Graf et al. (38)	Balancing backward One-legged obstacle jumping Jumping from side to side Sideway movements	For each task, points given that made up overall Motor Quotient under consideration of gender and age factor, categorized "not possible," "severe motor disorder," "moderate motor disorder," "normal," "good," "high"	KTK (dynamic body coordination, gross motor, dynamic balance skills)	Data collected over 5-month period
Gray and Smith (27)	V-sit (flexibility) Partial curl-ups (abdominal strength) Right-angle push-ups (upper body strength) 1/4-, 1/2-, or 1-mile run (aerobic endurance) Shuttle run (agility and strength)	Scored 1 (below standard) to 3 (PPF standard) percentile levels of (age- and sex-specific) achievement	In accordance with PPF Awards	Certified researcher
	One-leg balance		Stand on one foot with arms at right angles to their side without jumping or placing second foot on floor for up to 60 s	
Guinhouya (39)	6-min walk (functional performance measure of walking endurance)	Distance (laps/m), work (distance × body mass in kg)	30-m track in corridor, walk as far as possible in 6 min, children walked alone, standardized encouragement	One examiner, testing took place between 8:30 and 11:30

TABLE 3. (continued).

Study	Performance test (measured function)	Scoring method	Summary of procedure	Feasibility
Hamlin et al. (36)	Vertical jump (leg and lower body explosive strength/power)	Distance (m) converted to cohort percentiles	Maximal effort, practice attempt, better of three attempts taken, 2-min recovery between attempts	Performed during regular school hours by 25 specially trained instructors over 6-month period on concrete nonslip surface; children wore shorts, light T-shirt; for running shoes
Joshi et al. (28)	Flexibility (shoulder stretches) Sit-ups (muscular strength and endurance) Push-ups (as above) Trunk lifts (as above) PACER multistage run (aerobic capacity)	Overall fitness score: adding results (pass or fail) of five subtests assessed according to healthy fitness zone	According to Fitnessgram, PACER adapted from 20-m shuttle run test, progressive in intensity (easy at beginning and harder at end), set to music	Data collected by trained researchers and local school personnel trained on Fitnessgram administration
Liao et al. (24)	Sit-and-reach (lower back and upper thigh flexibility)	Distance (cm) converted to standardized quintiles	Reach forward from seated position, holding stretch for 2 s: better of two attempts taken	Data collected by schoolteachers trained with standardized instruction manual; board and 25-cm ruler used in sit-and-reach tests
	Sit-ups (abdominal strength and endurance)	Correct repetitions in 60 s converted to standardized quintiles	Knee at right angles, arm across chest, ankles held down by assistant, touch knee with elbows	
	Horizontal jump (leg and lower body explosive strength/power)	Distance (m) converted to standardized quintiles	Two-foot take-off and landing, arm assistance: better of two attempts taken (tape measure)	
	800/1,600-m run/walk (cardiovascular fitness/endurance)	Time (s) converted to standardized quintiles	800-m run/walk for all girls and boys 10-12 years, 1,600-m run/walk for boys 13-18 years, complete distance as fast as possible (stopwatch, running track)	
Mak et al. (25)	Sit-and-reach (lower back and upper thigh flexibility)	Distance (cm)	Reach smoothly forward from seated position, holding stretch for 2 s	Data collection carried out during physical education lessons in schools; for sit-and-reach, standard reach box with 23 cm marked at level of feet
	Sit ups (abdominal strength and endurance)	Repetitions in 1 min	Knees bent at 90° and feet flat on floor	
	Push-ups (upper body muscular strength and endurance)	Repetitions in 1 min	Extended legs for boys, bent knees for girls	
	9-min run/walk (cardiovascular fitness/endurance)	Distance (m)	15 m × 25 m basketball court	

TABLE 3. (continued).

Study	Performance test (measured function)	Scoring method	Summary of procedure	Feasibility
Malina et al. (26)	Sit-and-reach (lower back and upper thigh flexibility)	Distance (cm)	-	-
	Sit-ups (abdominal strength and endurance)	Repetitions in 30 s	-	
	Horizontal jump (leg and lower body explosive strength/power)	Distance (m)	Bend knees, arms in front, swing arms, push off, land with both feet: better of three attempts taken	
	8/12-min run (cardiovascular fitness/ endurance)	Distance (m)	Distance run in 8 min (grades 1-3), 12 min (grades 4-6)	
	~32-m sprint (running speed and anaerobic power)	Time (s)	35-yard dash (~32 m)	
Morinder et al. (40)	6-min walk (functional performance measure of walking endurance)	Distance (m)	70-m indoor corridor, marks every 2 m, wear comfortable shoes, walk as far as possible, standardized encouragement	Mean test-retest interval was 4 days; same physiotherapist conducted test on both occasions.
Mota et al. (42)	1-mile run (cardiovascular fitness/ endurance)	Time (min) converted to age- and gender-specific standardized categories (healthy fitness zone: fit or unfit)	Fitnessgram procedures: run/walk as fast as possible for 1 mile, 250-m running track, motivated children	Completed on school playground field; data collected by researchers or physical education staff
Nunez-Gaunaurd et al. (30)	Sit-ups (abdominal strength and endurance)	Repetitions in 30 s converted to standardized percentiles	Complete as many abdominal curls as possible in 30 s	Adjustable chair for timed sit-to-stand
	Push-ups (upper body muscular strength and endurance)	Repetitions in 60 s converted to standardized percentiles	Complete as many 90° push-ups with or without knees supported (child's choice) in 30 s, start and end in plank with elbows extended, bend elbows to 90°	
	Sit-to-stand (functional performance measure of lower extremity strength and endurance)	Repetitions in 60 s converted to standardized percentiles	Starting position with hip flexion at 90°, knee flexion at 90°, feet parallel and flat on floor, trunk erect, hands on waist or crossing chest, stand to full trunk and lower extremities extension	
	6-min walk (functional performance measure of walking endurance)	Distance (m) converted to standardized percentiles	Walk as far as possible in 6 min	
	Timed up-and-down stairs (functional mobility of lower extremities and trunk strength, coordination, postural control)	Steps per second converted to standardized percentiles	How quickly child ascends and descends flight of stairs (12-14 steps: different for each school)	

TABLE 3. (continued).

Study	Performance test (measured function)	Scoring method	Summary of procedure	Feasibility
	BOT2 SF (fine and gross motor skills and motor proficiency)	Raw scores for individual measures converted to point scores and added for total BOT2 SF point score, percentile rank obtained from age- and sex-specific tables	Fine motor precision, fine motor integration, manual dexterity, bilateral coordination, balance, running speed and agility, upper limb coordination and strength	
Pathare et al. (44)	One-leg standing balance	Time (s)	Eyes open, hands placed on hips, criteria to stop timing included movement of supporting foot, hooking free leg behind supporting leg, dropping free leg below 45° of knee flexion, moving arms from start position, tilting trunk, or looking away from visual target	Standardized instructions and adequate training for researchers
Poulsen et al. (34)	BOT2 (no fine motor and dexterity measures; fundamental movement skills proficiency)	Raw scores for individual measures converted to point scores and added for total BOT2 SF point score; percentile rank obtained from age- and sex-specific tables	Characterize children's fundamental movement skills proficiency in four motor-area composites: bilateral coordination, upper limb coordination, balance, running speed and agility, strength	Data collected over 13-month period by occupational therapist during one visit to outpatient hospital clinic
Riddiford-Harland et al. (35)	Horizontal jump (leg and lower body explosive strength/power)	Distance from heel to line (cm)	From behind line, feet parallel, arms against torso, maximal jump, land with feet together, minimal arm motion, one practice jump: better of two attempts taken	-
	Vertical jump (leg and lower body explosive strength/power)	Distance between standing and jumping reach (nearest 0.5 cm)	From static position, left arm by side, right arm pointing upward, feet flat, chalk-dusted fingers, maximal jump, one practice jump: better of two attempts taken	
	Ball throwing (upper body strength/power and coordination)	Distance (m)	Seated on floor, head, shoulders, back, buttocks against wall, legs extended at knees, arms abducted, forearms parallel to floor, using only upper limb movement, basic chest pass using size 7 basketball	
Sartorio et al. (19)	Margaria step climbing test (anaerobic performance	Average mechanical power (W) during stair climbing, calculations referenced	30 min prior to testing one to two practice trials taken	-

TABLE 3. (continued).

Study	Performance test (measured function)	Scoring method	Summary of procedure	Feasibility
Tsiros et al. (41)	6-min walk (functional performance measure of walking endurance)	Distance (nearest 0.1 m)	Demonstration, walk between two marks on 10-m track, standardized encouragement	-
	Timed up-and-down stairs (functional mobility of lower extremities and trunk strength, coordination, postural control)	Time (nearest 0.1 s) converted to cohort *z*-score (to account for variation in step height between sites)	Demonstration, ascend, and descend 15 steps "quickly, but safely": average of two trials taken	
	Timed up-and-go test (functional mobility)	Time (nearest 0.1 s)	Demonstration, seated with knee angle of 90° in height-adjustable armless chair, move from sitting to standing, walk 3 m, touch wall, return to sitting: average of three taken	
Vanhelt et al. (20)	6-min walk (functional performance measure of walking endurance)	Distance (m)	Walk as far as possible in 6 min, around border of handball field (40 m × 20 m), comfortable clothing, verbal encouragement, standardized outdoor environment, repeated tests 1 week apart	Data collected over 12-month period; participants recruited by four pediatricians
Wright et al. (46)	13 tasks with six items measured bilaterally: unilateral stance, tandem walking, 180° tandem pivot, lateral foot scooting, hopping forward, crouch and walk, lateral dodging, walking and looking, running with controlled stop, forward to backward walking, walk, look, and carry, descending stairs, step-ups × 1 step	Each task rated on 6-point scale (0-5), criteria for scores 1-5 progressively demanding with regards to time, distance, quality of performance; score of "0" denotes inability to perform task	In accordance with CB&M	Testing conducted in church/school hallways and gymnasiums; data collected by two pediatric physiotherapists, >5 years of experience, and four physiotherapists who received standardized training; testing time was 15-20 min per child

Aerobic performance. Most studies ($n = 14$) reported that aerobic performance was significantly lower in OW/OB (22,24-26,28-30,38,40-42); however, three studies found no differences between OW/OB and HW groups (26,39,42). Three studies found significant differences in 6MTW performance between OW, OB, and HW groups (30,40,41). In contrast, Guinhouya (39) found no significant difference in 6MTW when comparing a group of OW girls with a healthy group of girls; however, the lack of an OB group may have influenced these findings.

Anaerobic performance. Performance of anaerobic tests was significantly lower in OW/OB in five tests (22,29,33,34), and no differences were found in one test (26). The possible reason for finding no significant difference in anaerobic performance in the study by Malina et al. (26) was the low sample size of participants with OW/OB; of girls and boys age 6-7 to 8-9 years, 15 participants with OW/OB were reported in each gender, and of girls and boys age 9-10 to 11-12 years, 27 and 22 participants with OW/OB, respectively, were reported.

Coordination and balance. There was significantly poorer performance in coordination and balance tests in children with OW/OB for 15 tests (34,41,43-46), and no differences were found in 15 tests (22,23,30,45,46). The range and variation of tests for coordination and balance may underlie the reason for contrasting findings between studies. The tests ranged from predominantly coordination-based measures such as plate tapping and posting/treading/writing to whole body dynamic balance measures such as jumping side-to-side and hopping. Although coordination and balance are commonly examined together during the same test, there is a need to appropriately define the measure for between-study comparisons.

Multi-item performance instruments. All instruments that measured responsiveness (four tests) found significantly poorer performance of children with OW/OB when compared with HW children (30,38,43,45).

Feasibility. Table 3 presents the data provided on the conditions under which measurements were recorded: the time taken to carry out the procedures, the expertise of the person(s) administering the tests, and the equipment requirements. In most of the studies, measures of time (stopwatch), distance (tape measure, sit-and-reach box), and repetitions were included; however, information on the specific measurements was often lacking in many of the studies reviewed.

Study design

Insufficient reporting of study population in terms of BMI *z*-scores reference data and % BF measures, together with differences in the reference values and cutoffs used for OW/OB categories limit interpretability of this review. Of the five studies to define OW/OB by % BF measures, 71% of the ratings were positive when compared with only 27% rating when OW/OB was defined by BMI *z*-score. Sophisticated body composition measures such as DEXA (41) may provide better sensitivity for defining OW, OB, and HW groups when compared with skin fold, bioelectrical impedance analysis (BIA), and BMI *z*-score, but places participants at risk of radiation exposure thereby reducing its clinical application. Eleven studies (44%) used the International Obesity Task Force (IOTF) reference data set to calculate BMI *z*-scores, seven studies (28%) used the Centre for Disease Control and Prevention (CDC) reference data set, six (24%) referred BMI to national data sets, and one study did not reported the reference data set. Cutoffs were also defined differently between studies, 15 studies (60%) used the IOTF cutoffs for HW, OW, and OB (based on adult cutoffs of 25 and 30 kg/m^2 for women and men at 18 years, respectively); five (20%) used CDC cutoffs ($\geq$95th percentile = OB, 85th to $<$95th percentile = OW); two (8%) referenced national cutoffs; and three (12%) did not reference their cutoffs. Many regression-type analysis of physical function were excluded from this review because BMI was not presented as a BMI *z*-score despite the studies using BMI *z*-score to define the number of participants with OW/OB.

All studies included boys and girls in their samples; however, only nine studies (14%) separated results by gender. Many studies included a comparison of physical function between boys and girls but did not separate the cohort into gender categories for comparison between groups of HW and OW/OB. Performance differences in all physical function domains (except coordination and balance) were found to differ in the studies that stratified their sample into boys and girls, suggesting gender differences in physical function exist in children with OW/OB.

The majority of the studies analyzed performance using the raw outcome data, 10 studies converted the raw data to scores/centiles based on standardized reference values, and four studies referenced the raw data to centiles within the cohort of children in the study sample. Of the four studies to report outcome data to within cohort variation, three (29,31,41) found significant construct validity and/or responsiveness results for all measures of physical function. Hamlin et al. (36) found no construct validity for boys and no responsiveness for boys or girls; however, the sample size was low ($n = 54$), which means that cohort variation may have been small. Of the nine studies to report outcome data relative to standardized reference values, seven studies (27,28,30,38,43,45,46) produced multi-item scores all of which demonstrated significant construct validity and/or responsiveness. The two studies that did not produce multi-item scores (34,42) were based on reported multi-item tests (BOT-2 and Fitnessgram), and only the individual tests were reported. The benefit of referring outcome measures to standardized values is that cutoffs for the level of performance (e.g., poor, normal, and high) can be used to define the amount of physical function. However, standardization relies on the reference sample being of large enough numbers and spread of values that cutoffs (which maybe arbitrarily chosen) can be derived. Consideration of the type of analysis should be undertaken based on the design and aim of the study.

Limitations of the summary of evidence

It is acknowledged that there are no standardized criteria to assess the measurement properties of performance-based physical function and alternative criteria applied to the data may produce different interpretations. Non-English publications, conference proceedings, unpublished reports, and case series reports were not included. The search was limited to cross-sectional studies to establish the baseline measurement properties. It is acknowledged that the omission of longitudinal and intervention studies may lead to bias and an underestimation of the responsiveness of measures. However, understanding the effects of interventions on measurement properties would require further review and analysis. A further review of longitudinal and intervention studies should be conducted to fully analyze outcome measures of physical function in children with OW/OB. Only Malina et al. (26) stratified their sample population by age and found differences between age groups 6-7 to 8-9 years and 9-10 to 11-12 years in boys and girls (Mexican school grades 1-3 and 4-6), respectively. Differences in age groups may underlie some of the contrasting findings in this review. Fourteen studies (50%) sampled only children aged 5-12 years, 12 studies (43%) sampled children and adolescents (age range crossed 12-13 years), and two studies (7%) sampled only adolescents aged 13-18 years. Future studies should consider recruiting or stratifying samples by child/adolescent age ranges. Of particular interest to this review was the exclusion of sophisticated equipment such as three-dimensional (3D) motion capture. Although the inclusion of measurement properties of sophisticated equipment would be useful, it was beyond the scope of this review which focused on tools commonly used in clinical practice. Most studies in this review were observational cohort studies and did not set out any hypotheses to evaluate measurement properties, which may explain why many received indeterminate or negative ratings. Hypothesis testing is needed to establish which outcomes should be used for answering specific clinical and research questions.

Key Summary of Evidence

- Of the 66 measures identified, only the 6MTW test was examined for all measurement properties (internal consistency not appropriate).
- There is moderate evidence that the 6MTW test is suitable for monitoring physical function.
- There is low evidence that time and distance measures of aerobic performance, anaerobic performance measured with 10 m × 5 m sprints (but not when few repetitions are utilized), and multi-item performance instruments (in particular the MABC and BOT2) may be suitable for monitoring physical function in children with OW/OB (but test-retest repeatability is lacking). Internal consistency of the multi-item tests has been determined in other pediatric populations (48,49); however, it has not been evaluated specifically in children with OW/OB.
- There is low evidence in girls with OW/OB and insufficient evidence in boys that strength measured with a full sit-up and horizontal jump is suitable for monitoring physical function in children with OW/OB (but test-retest repeatability is lacking).
- There is very low evidence that measurement of flexibility, coordination, and balance tests are suitable methods of monitoring physical function in children with OW/OB.
- More positive ratings were given to studies that measured OW/OB by % BF when compared with studies that used BMI *z*-score; although to date, no established cutoffs for defining OW/OB by % BF have been determined.
- Few of the studies included in this review set out with the aim to investigate the measurement properties of outcome measures for physical function. Future work is needed, in particular, to understand the repeatability of most of the tests included in this review (with the exception of 6MWT). Similarly, no study on multi-item test scores for physical function examined internal consistency specifically in pediatric OW/OB. Rigorous evaluation of the measurement properties of outcome measures used to monitor physical function in children with OW/OB in larger collaborative studies is required.

Conclusions and Recommendations

The synthesis of current evidence on the measurement properties of physical function will assist measurement selection by clinicians and researchers with an interest in evaluating physical function in children with OW/OB. Few of the studies included in this review set out with the aim to investigate the measurement properties of outcome measure for physical function. Rigorous evaluation of the measurement properties of outcome measures used to monitor physical function in children with OW/OB in larger collaborative studies is required with a particular focus on those not assessed in the current review: measurement error, validity (content, structural, cross-cultural, and criterion), and interpretability. There is moderate evidence that the 6MTW is suitable for measurement of physical function in children with OB. However, evidence is low for the use of time and distance measures of aerobic and anaerobic performance, muscle strength, and the MABC and BOT2 multi-item performance instruments and very low for tests of flexibility, coordination, and balance. Based on this review, measurement of physical function using the 6MTW is recommended.

All tests, except the 6MTW, require further testing for repeatability and multi-item instruments require assessment of internal consistency in children with OW/OB. Larger gender- and age-specific studies using both BMI *z*-score and % BF methods together with standardized reference cutoffs to categorize OW/OB are needed to determine physical function outcome measures in children with OW/OB. The selection of age- and gender-matched reference data and cutoffs for OW/OB requires careful consideration as this appears to influence comparisons of physical function tests. O

References

1. Shultz SP, Anner J, Hills AP. Paediatric obesity, physical activity and the musculoskeletal system. *Obes Rev* 2009;10:576-582.
2. Ebbeling CB, Pawla DB, Ludwig DS. Childhood obesity: public-health crisis, common sense cure. *Lancet* 2002;360:473-482.
3. Paulis WD, Silva S, Koes BW, Middelkoop M. Overweight and obesity are associated with musculoskeletal complaints as early as childhood: a systematic review. *Obesity Rev* 2014;15:52-67.
4. Krul M, van der Wouden JC, Schellevis FG, van Suijlekom-Smit LW, Koes BW. Musculoskeletal problems in overweight and obese children. *Ann Fam Med* 2009;7:352-356.
5. Pinhas-Hamiel O, Singer S, Pilpel N, Fradkin A, Modan D, Reichman B. Health-related quality of life among children and adolescents: associations with obesity. *Int J Pediatric Obes* 2006;30:267-272.
6. Schwimmer JB, Burwinkle TM, Varni JW. Health-related quality of life of severely obese children and adolescents. *JAMA* 2003;289:1813-1819.
7. Bout-Tabaku S, Briggs MS, Schmitt LC. Lower extremity pain is associated with reduced function and psychosocial health in obese children. *Clin Orthop Relat Res* 2013;471:1236-1244.
8. De Sá Pinto AL, De Barros Holanda PM, Radu AS, Villares SM, Lima FR. Musculoskeletal findings in obese children. *Paediatr Child Health* 2006;42:341-344.
9. Stovitz SD, Pardee PE, Vazquez G, Duval S, Schwimmer JB. Musculoskeletal pain in obese children and adolescents. *Acta Paediatr* 2008;97:489-493.
10. Wearing SC, Hennig EM, Byrne NM, Steele JR, Hills AP. The impact of childhood obesity on musculoskeletal form. *Obesity Rev* 2006;7:209-218.
11. Wills M. Orthopedic complications of childhood obesity. *Pediatr Phys Ther* 2004;16:230-235.

12. Taylor ED, Theim KR, Mirch MC, et al. Orthopedic complications of overweight in children and adolescents. *Pediatrics* 2006;117:2167-2174.
13. Tsiros MD, Coates AM, Howe PR, Grimshaw PN, Buckley JD. Obesity: the new childhood disability? *Obes Rev* 2011;12:26-36.
14. Trost SG, Kerr LM, Ward DS, Pate RR. Physical activity and determinants of physical activity in obese and non-obese children. *Int J Obes Relat Metab Disord* 2001;25:822-829.
15. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage* 2013;21:1042-1052.
16. World Health Organization. Towards a Common Language for Functioning, Disability and Health: International Classification for Functioning, Disability and Health. World Health Organization: Geneva; 2002.
17. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:1605.
18. Terwee CB, Mokkink LB, Steultjens MP, Dekker J. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Rheumatology* 2006;45:890-902.
19. Sartorio A, Agosti F, De Col A, Lafortuna CL. Age- and gender-related variations of leg power output and body composition in severely obese children and adolescents. *J Endocrinol Invest* 2006;29:48-54.
20. Vanhelst J, Fardy PS, Salleron J, Beghin L. The six-minute walk test in obese youth: reproducibility, validity, and prediction equation to assess aerobic power. *Disabil Rehabil* 2013;35:479-482.
21. Andreasi V, Michelin E, Rinaldi AE, Burini RC. Physical fitness and associations with anthropometric measurements in 7 to 15-year-old school children. *J Pediatr* 2010;86:497-502.
22. Casajus JA, Leiva MT, Villarroya A, Legaz A, Moreno LA. Physical performance and school physical education in overweight Spanish children. *Ann Nutr Metab* 2007;51:288-296.
23. Graf C, Koch B, Dordel S, et al. Physical activity, leisure habits and obesity in first-grade children. *Eur J Cardiovasc Prev Rehabil* 2004;11:284-290.
24. Liao Y, Chang SH, Miyashita M, et al. Associations between health-related physical fitness and obesity in Taiwanese youth. *J Sports Sci* 2013;31:1797-1804.
25. Mak KK, Ho SY, Lo WS, et al. Health-related physical fitness and weight status in Hong Kong adolescents. *BMC Public Health* 2010;10:88.
26. Malina RM, Pena Reyes ME, Tan SK, Little BB. Physical fitness of normal, stunted and overweight children 6-13 years in Oaxaca, Mexico. *Eur J Clin Nutr* 2011;65:826-834.
27. Gray A, Smith C. Fitness, dietary intake, and body mass index in urban Native American youth. *J Am Diet Assoc* 2003;103:1187-1191.
28. Joshi P, Bryan C, Howat H. Relationship of body mass index and fitness levels among schoolchildren. *J Strength Cond Res*– 2012;26:1006-1014.
29. Bovet P, Auguste R, Burdette H. Strong inverse association between physical fitness and overweight in adolescents: a large school-based survey. *Int J Behav Nutr Phys Act* 2007;4:24.
30. Nunez-Gaunaurd A, Moore JG, Roach KE, Miller TL, Kirk-Sanchez NJ. Motor proficiency, strength, endurance, and physical activity among middle school children who are healthy, overweight, and obese. *Pediatr Phys Ther* 2013;25:130-138.
31. Ara I, Sanchez-Villegas A, Vicente-Rodriguez G, et al. Physical fitness and obesity are associated in a dose-dependent manner in children. *Ann Nutr Metab* 2010;57:251-259.
32. Čapkauskienė S, Visagurskienė K, Bakienė R, Vitkienė I, Vizbaraitė D. Peculiarities of physical fitness and body composition of 5-7 year-old children of several Kaunas preschools and interaction between those indexes. *Educ Physical Train Sport* 2009;73:14-20.
33. Gonzalez-Suarez CB, Grimmer-Somers K. The association of physical activity and physical fitness with pre-adolescent obesity: an observational study in Metromanila, Philippines. *J Phys Act Health* 2011;8:804-810.
34. Poulsen AA, Desha L, Ziviani J, et al. Fundamental movement skills and self-concept of children who are overweight. *Int J Pediatr Obes* 2011;6:e464-e471.
35. Riddiford-Harland DL, Steele JR, Baur LA. Upper and lower limb functionality: are these compromised in obese children? *Int J Pediatr Obes* 2006;1:42-49.
36. Hamlin MJ, Fraser M, Lizamore CA, Draper N, Blackwell G, Shearman J. Effects of bioelectrical impedance-derived fat and lean mass on fitness levels in 8- to 13-year-old children. *Asian J Exercise Sports Sci* 2014;11:36-45.
37. Burke V, Beilin LJ, Durkin K, Stritzke WG, Houghton S, Cameron CA. Television, computer use, physical activity, diet and fatness in Australian adolescents. *Int J Pediatr Obes* 2006;1:248-255.
38. Graf C, Koch B, Kretschmann-Kandel E, et al. Correlation between BMI, leisure habits and motor abilities in childhood (CHILT-project). *Int J Obes Relat Metab Disord* 2004;28:22-26.
39. Guinhouya B. Outcomes and cardiac response of overweight prepubescent to the 6 minutes walk test. *Minerva Pediatr* 2011;63:375-384.
40. Morinder G, Mattsson E, Sollander C, Marcus C, Larsson UE. Six-minute walk test in obese children and adolescents: reproducibility and validity. *Physiother Res Int* 2009;14:91-104.
41. Tsiros MD, Buckley JD, Howe PR, et al. Day-to-day physical functioning and disability in obese 10- to 13-year-olds. *Pediatr Obes* 2013;8:31-41.
42. Mota J, Flores L, Flores L, Ribeiro JC, Santos MP. Relationship of single measures of cardiorespiratory fitness and obesity in young schoolchildren. *Am J Hum Biol* 2006;18:335-341.
43. D'Hondt E, Deforche B, Vaeyens R, et al. Gross motor coordination in relation to weight status and age in 5- to 12-year-old boys and girls: a cross-sectional study. *Int J Pediatr Obes* 2011;6:e556-e564.
44. Pathare N, Haskvitz EM, Selleck M. Comparison of measures of physical performance among young children who are healthy weight, overweight, or obese. *Pediatr Phys Ther* 2013;25:291-296.
45. D'Hondt E, Deforche B, De Bourdeaudhuij I, Lenoir M. Relationship between motor skill and body mass index in 5- to 10-year-old children. *Adapt Phys Activ Q* 2009;26:21-37.
46. Wright MJ, Bos C. Performance of children on the Community Balance and Mobility Scale. *Phys Occup Ther Pediatr* 2012;32:416-429.
47. Stephensen D, Drechsler WI, Scott OM. Outcome measures monitoring physical function in children with haemophilia: a systematic review. *Haemophilia* 2014;20:306-321.
48. Wuang YP, Su CY. Reliability and responsiveness of the Bruininks-Oseretsky Test of Motor Proficiency-Second Edition in children with intellectual disability. *Res Dev Disabil* 2009;30:847-855.
49. Wuang YP, Su JH, Su CY. Reliability and responsiveness of the Movement Assessment Battery for Children-Second Edition Test in children with developmental coordination disorder. *Dev Med Child Neurol* 2012;54:160-165.