Psychometric properties of functional postural control tests in children: A systematic review

Annals of Physical and Rehabilitation Medicine 66 (2023) 101729

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Review

Psychometric properties of functional postural control tests in children: A systematic review

Charlotte Johnson[a,b,*], Ann Hallemans[a], Mieke Goetschalckx[b], Pieter Meyns[b], Eugene Rameckers[b,c,d], Katrijn Klingels[b], Evi Verbecque[b]

[a] Research Group MOVANT, Department of Rehabilitation Sciences and Physiotherapy (REVAKI), University of Antwerp, 2610 Wilrijk, Belgium
[b] Rehabilitation Research Centre (REVAL), Faculty of Rehabilitation Sciences and Physiotherapy, Hasselt University, 3590 Diepenbeek, Belgium
[c] Department of Rehabilitation Medicine, Research School CAPHRI, Maastricht University, 6229 Maastricht, the Netherlands
[d] Center of expertise Adelante rehabilitation, 6301 KA Valkenburg, the Netherlands

A R T I C L E I N F O

Article History:
Received 29 April 2022
Accepted 4 November 2022

A B S T R A C T

Background: Postural control deficits are one of the most common impairments treated in pediatric physiotherapeutic practice. Adequate evaluation of these deficits is imperative to identify postural control deficits, plan treatment and assess efficacy. Currently, there is no gold standard evaluation for postural control deficits. However, the number of studies investigating the psychometric properties of functional pediatric postural control tests has increased significantly.

Objective: To facilitate the selection of an appropriate pediatric functional postural control test in research and clinical practice.

Methods: Systematic review following the PRISMA guidelines. PubMed, Web of Science and Scopus were systematically searched (last update: June 2022; PROSPERO: CRD42021246995). Studies were selected using the PICOs-method (pediatric populations (P), functional assessment tools for postural control (I) and psychometric properties (O). The risk of bias was rated with the COSMIN checklist and the level of evidence was determined with GRADE. For each test, the postural control systems were mapped, and the psychometric properties were extracted.

Results: Seventy studies investigating 26 different postural control tests were included. Most children were healthy or had cerebral palsy. Overall, the evidence for all measurement properties was low to very low. Most tests (95%) showed good reliability (ICC>0.70), but inconsistent validity results. Structural validity, internal consistency and responsiveness were only available for 3 tests. Only the Kids-BESTest and FAB covered all postural control systems.

Conclusion: Currently, 2 functional tests encompass the entire construct of postural control. Although reliability is overall good, validity results depend on task, age and pathology. Future research should focus on test batteries and should particularly explore structural validity and responsiveness in different populations with methodologically strong study designs.

Abbreviations: APA, Anticipatory postural adjustments; BBS, Berg Balance Scale; BBW, Balance Beam Walking; BESS, Balance Error Scoring System; BESTest, Balance Evaluation Systems Test; BOT-2, Bruininks-Oseretsky Test for Motor pProficiency, 2nd edition; CB&M, Community Balance & Mobility Scale; CGT, Complex Gait Test; COSMIN, COnsensus-based Standard for the selection of health Measurement Instruments; CP, Cerebral palsy; (m)CT-SIB, (Modified) Clinical Test of Sensory Interaction in Balance; DGI, Dynamic Gait Index; ECAB, Early Clinical Assessment of Balance; FAB, Fullerton Advanced Balance Scale; FRT, Forward Reach Test; FSST, Four Square Step Test; GAS, Goal Attainment Scale; GDBT, Ghent Developmental Balance Test; GMFCS, Gross Motor Function Classification System; GMFM-66/88, Gross Motor Function Measure with 66/88 items; GRADE, Grading of Recommendations Assessment, Development and Evaluation; ICC, Intraclass correlation coefficient; ICF, International Classification of Functioning, Disability and Health; Kids-BESTest, Kids-Balance Evaluation Systems Test; Kw, Wweighted kappa; LoE, Level of Evidence; LRT, Lateral Reach Test; MABC(-2), Movement Assessment Battery for Cchildren (2nd edition); MABC(-2) B, Balance domain; MCID, Minimal clinically important difference; MRT, Multidirectional Reach Test; OLS, One-Leg-Stance; PBS, Pediatric Balance Scale; PDMS-2, Peabody Developmental Motor Scales, 2nd edition; PDMS-2 L/S, Llocomotion domain stationary domain; PEDI, Pediatric Evaluation Disability Inventory; PRISMA, Preferred Reporting Items for Systematic Review and Meta-Analyses; PRT, Pediatric Reach Test; RT, Reach Tests; SBST, Stork Balance Stand Test; SEBT, Star Excursion Balance Test; SEM, Standard error of measurement; SRM, Standardized response mean; SOT, Sensory Organization Test; SWOC, Standardized Walking Obstacle Course; TDC, Typically developing children; TS, Tandem-Stance; TUDS, Timed Up and Down Stairs test; TUG, Timed Up and Go test; YBT, Y-Balance Test

* Corresponding author at: Universiteitsplein 1, 2610 Wilrijk, Antwerp, Belgium.
E-mail address: charlotte.johnson@uantwerpen.be (C. Johnson).

https://doi.org/10.1016/j.rehab.2022.101729
1877-0657/

Introduction

Postural control deficits are one of the most common impairments treated in physiotherapeutic practice in a variety of pediatric populations, such as cerebral palsy (CP), traumatic brain injury or developmental coordination disorder [1–3]. Because of the impact of postural control deficits on motor development and daily activities, their identification is critical to planning treatment.

Clinicians most often use functional assessment tools to evaluate postural control because they are intended to represent the functional deficits children encounter in daily life, do not require expensive equipment and are easy to apply [4]. Currently, clinicians and researchers agree on the definition of postural control, i.e., the control of the body's position in space for postural orientation and within the base of support for postural stability [1, 5]. Although the theoretical construct of postural control still lacks consensus, there is some support for the implementation of the International Classification of Functioning, Disability and Health (ICF) framework, task-oriented approach and the systems approach defined by Shumway-Cook and Woollacott [1]. This lack of consensus has led to a large diversity of functional pediatric postural control tests with no gold standard [1, 4]. Thus, a guide to aid selection of the most appropriate functional postural control test would be valuable, including the underlying construct, the quality of the test and its feasibility [6, 7]. To correctly identify postural control deficits, the applied test should reflect the underlying construct adequately [7]. As several systems are involved in postural control, they should all be addressed during assessment [5]. Owing to task-specificity [8] and the fact that different tasks involve different systems [9, 10], identification of postural control systems based on task-type could improve understanding of the underlying construct of the test [4, 5]. Tests comprising multiple tasks (test batteries) often evaluate multiple postural control systems. Tests that cover more systems more closely evaluate the entire construct of postural control, thereby potentially increasing the ability of the test to identify deficient underlying systems in contrast with tests that assess only one system [4, 5].

The quality of the test is determined by its psychometric properties, which include reliability, responsiveness, and validity [6, 7]. Since there are no formal measurement properties related to feasibility, we refer to it as the ease with which the test is applied in its intended context, given specific constraints, such as population type, cost price, time or equipment needed to perform the test [6, 7].

In 2014, our research group [2] published a narrative literature review of the psychometric properties of the available functional pediatric postural control tests, revealing 25 studies covering 14 different functional tests. Overall, the structural validity and responsiveness of these tests were under reported. Since then, the number of studies on this research topic has increased considerably. Therefore, the narrative review was updated and transformed into a systematic review following this Population Intervention Comparison Outcome Study design or PICOs-question: What are the functional postural control tests (I) available for children (P) that have been investigated regarding their psychometric properties such as reliability, validity and responsiveness (O)? The aim of this systematic review was to facilitate the selection of an appropriate pediatric functional postural control test by mapping its psychometric properties and feasibility. This aim was reached by answering the following research questions: 1) What are the existing psychometric properties of each postural control test? 2) Which underlying systems are evaluated in the test with regard to the multisystemic framework [5]. 3) What are the feasibility features of each test.

Methodology

This systematic review was conducted and written following the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) [11] and was registered in the international database of prospectively registered systematic reviews or PROSPERO (registration number: CRD42021246995). This systematic review is an update and expansion of the narrative review of Verbecque et al. (2015) [2]. Details of the protocol can be retrieved on the PROSPERO database (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021246995).

Eligibility criteria and selection process

To identify relevant studies, predefined eligibility criteria were applied according to the PICOs method in line with the narrative literature review [2]. A detailed description of the eligibility criteria is available in the PROSPERO protocol (CRD42021246995). Studies were included if:

1. **Population:** children included had either typical development or postural control deficits of any origin. Children were aged between 18 months and 12 years. If the sample comprised children between 0-18 years, most children (>2/3) had to be 18 months-12 years (i.e. <1/3 was 0-18 months old and/or 12-18 years), for the entire sample to be included.
2. **Intervention:** postural control was assessed with functional postural control tests.
3. **Outcome:** the article provided an assessment of at least one psychometric property such as reliability, validity, responsiveness or reference values for the functional postural control test for which numerical data had to be available.
4. **Study design:** studies covered original peer-reviewed research with the purpose of investigating psychometric properties e.g., validity, reliability, responsiveness or reference values.
5. **Language:** written in English, Dutch, French or German.

Eligibility was assessed by 2 independent reviewers (CJ and EV) in the same sequence: population, intervention, outcome, study design, and language. Studies were selected in 2 phases: phase 1, on title and abstract and phase 2, on full text. After phase 2, the references of all included studies were hand-searched to ensure all relevant studies were included. Consensus was reached in a meeting after each phase.

Information sources and search strategy

A systematic literature search was conducted in PubMed, Scopus and Web of Science (WoS) (last update on June 30th, 2022). We followed the original search strategy used in Verbecque et al. [2], comprising terms related to "child", "postural balance" and "psychometric properties" and adapted to the database requirements (Appendix A). Since this is an update, only studies published after the 31st of December 2013 were searched in PubMed and WoS. Scopus was searched without date restrictions because this database was not searched in the original literature review [2]. All citations were exported to EndNote to remove duplicates manually and subsequently all unique hits were screened for eligibility.

Data collection process and data items

Data from each article, including the initial review [2] and the newly included references, were extracted by 2 independent reviewers (CJ and EV). Each reviewer extracted data from half of the included articles and checked the other half. Discrepancies were discussed in a consensus meeting.

The extracted data concerned:

1. **General population characteristics**: pathology or typical development, number of participants per group, age range and sex distribution.

2. **Assessment characteristics**: name of the functional test, the test items included, and whether it intended to cover one or multiple postural control systems. This information was used to map the ***underlying postural control systems*** assessed: movement strategies - anticipatory postural adjustments (APA)/reactive postural responses, orientation in space, sensory strategies and control of dynamics [5]. If the test consisted of 1 task, the dominant system was identified and classified as such.
3. **Psychometric properties** extracted were expressed as numeric values of the functional postural control test, such as intra-class correlation coefficients (ICCs), standard error of measurement (SEMs), correlation coefficients and p-values. To minimize publication bias, data were only extracted if all data were numerically provided in the study, if values were only given visually or partially, results were not extracted. Significance levels were set at 0.001, 0.01 and 0.05 to ensure uniform reporting. The COnsensus-based Standard for the Selection of Health Measurement Instruments (COSMIN) definitions were applied to identify correct psychometric properties [7] and are listed in Table 1.
4. **Feasibility parameters**: presence of pediatric reference values, cross-cultural adaptation, time to administer, equipment and cost of the test [7].

Results could not be pooled due to diversity, i.e., different populations, different ages, different postural control tests or different measurement properties investigated. Therefore, a meta-analysis was not performed [12]. Reliability, measurement errors, validity and internal consistency data per postural control test and population are summarized in the tables.

For each test, an overall judgment is indicated in the tables with color coding in line with the COSMIN criteria for good psychometric properties [6, 7, 13]. These judgments are required to establish the level of evidence. For ***reliability***, green indicates ICC/weighted kappa (κ_w) of ≥0.70 and orange ICCs/κ_w of <0.70 in the majority of cases. Validity was considered per individual property. For ***concurrent validity***, "+" was attributed for correlations ≥0.70 and "-" for <0.70. For ***construct validity*** – hypothesis testing, "+" confirming the same construct with correlations of ≥0.50 and "-" confirming a different construct with correlations of <0.30 and ± if correlations were between 0.30 and 0.50 indicating related constructs [6, 7, 13]. Other types of validity and responsiveness were not attributed a specific symbol or color [6, 7, 13].

Study quality was assessed by determining risk of bias using the COSMIN checklist. Then, the quality of results for each psychometric property was rated across studies [6, 7, 13]. Last, the COSMIN modifications of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) principles were applied to estimate the level of evidence (LoE) of the psychometric properties for each functional postural control test across studies. Four elements are considered in GRADE: 1) risk of bias, 2) inconsistency, 3) imprecision and 4) indirectness. Each element should meet specific criteria to have the highest level of evidence (=) and elements were downgraded (↓) if element-specific criteria were not met. The quality of studies providing only reference values was not assessed because this is not possible with the COSMIN scoring system [6, 7, 13]. Three independent reviewers (CJ, EV and MG) assessed risk of bias: MG rated all studies and CJ and EV each rated 50%. Two independent reviewers (CJ and EV) performed the grading. Consensus was reached in a consensus meeting. Appendix B provides details on how COSMIN and GRADE principles were applied.

Results

Study selection

The selection process of relevant studies is presented in Appendix C. A total of 72 studies were included in this systematic review, 25 studies from the original narrative review [2] and 47 newly published.

Results of studies

We identified 26 different functional postural control tests including 7 test batteries and 19 tests covering 1 dominant system. Twelve tests were new since 2014. Appendix D provides a detailed description of each test. Fig. 1 summarizes the postural control systems assessed in each test. Table 2 provides an overview of the reliability results and Table 3 the validity results. Fig. 2 depicts the level of evidence and the overall quality of results for each functional postural control test. Feasibility features are presented in Table 4.

Population characteristics

Populations investigated varied from typically developing children (TDC) to children with mild motor impairments, such as hearing-impaired children or children with global developmental delays to more severely affected children, such as those with CP or traumatic brain injury. Children with CP (22/72 studies; 0.5-18 years) and TDC (40/72 studies; 0.5-19 years) were reported most frequently. Other populations were reported less frequently, such as Down syndrome

Table 1
Definitions of psychometric properties according to COSMIN taxonomy [7].

Property	*Definition*
Structural validity	The degree to which the scores of a test are an adequate reflection of the dimensionality of the construct to be measured; typically assessed with a classical test theory such as confirmatory factor analysis or using the item response theory or Rasch analysis [7].
Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted test are an adequate reflection of the performance of the items of the original version of the test.
Construct validity – hypothesis testing/ known-group validity	The degree to which the scores of the postural control test are consistent with hypotheses (for instance regarding internal relationships, relationships to scores of other instruments (hypothesis testing), or differences between relevant groups (known-group validity) based on the assumption that the test validly measures the construct to be measured.
Criterion/concurrent validity	Originally COSMIN suggests criterion validity which refers to the degree to which the scores of a test are an adequate reflection of a 'gold standard'. However, since no gold standard exists in the field of functional postural control tests, the term concurrent validity is used instead of criterion validity. Concurrent validity thus refers to the degree to which a functional postural control tests reflects the other.
Internal consistency	The degree of the interrelatedness among test items.
Reliability	the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same PROM (internal consistency); over time (test–retest); by different persons on the same occasion (inter– rater); or by the same persons (i.e. raters or responders) on different occasions (intra–rater).
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured.
Responsiveness	The ability of a test to detect change over time in the construct to be measured

Abbreviation: COSMIN: COnsensus-based Standard for the selection of health Measurement Instruments.

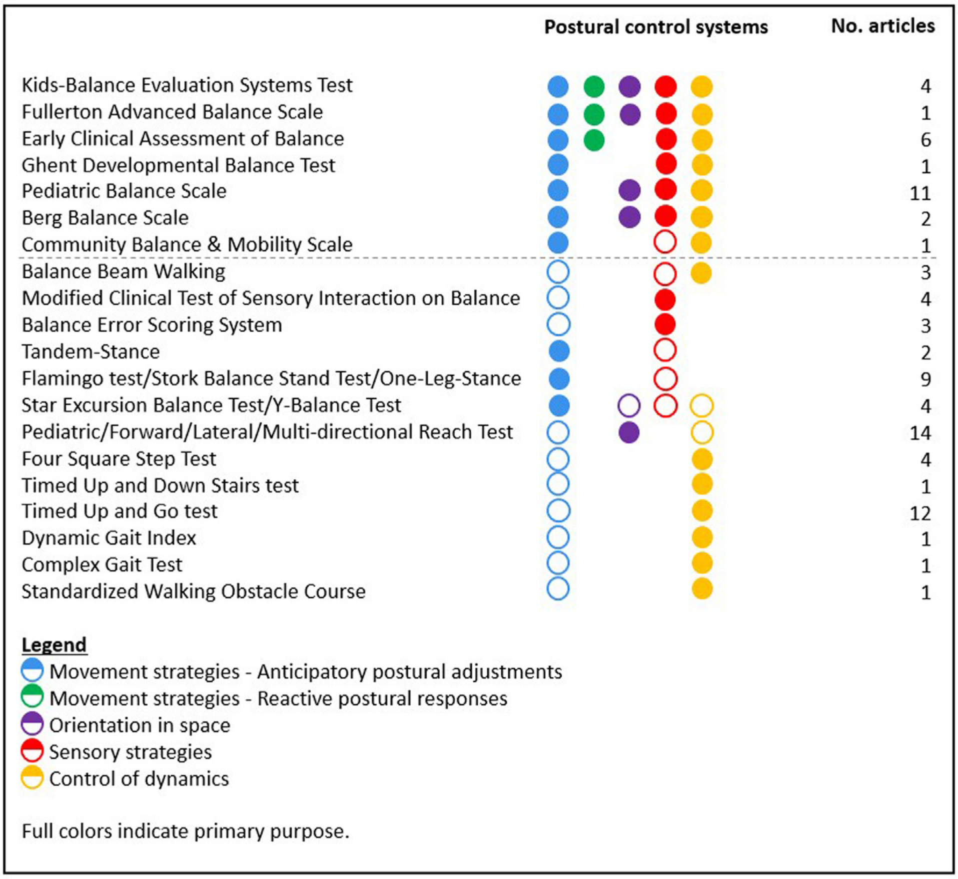

Fig. 1. Postural control systems evaluated by test.

[14–17], hearing impairment [18, 19] and heterogeneous samples such as global developmental delays [20]. Overall age varied from 0.5 to 21 years.

Postural control tests assessing one dominant system

This literature update revealed new research on the Pediatric Reach Test (PRT) [21–23], the Timed Up Down Stairs test (TUDS) [24], the Timed Up and Go test (TUG) [17, 24-30] and the Balance Beam Walking test (BBW) [32, 45]. Additionally, 9 new postural control tests were found: Flamingo test [31], Stork Balance Stand test (SBST) [32–34], Star Excursion Balance Test (SEBT) [35, 36], Y-balance test (YBT) [37, 38], Multidirectional Reach Test (MRT) [39], (modified) Clinical Test of Sensory Interaction on Balance ((m)CT-SIB) [14, 20, 40], Balance Error Scoring System (BESS) test [41–43], Four Square Stepping Test (FSST) [16, 17, 30, 44] and Complex Gait Test (CGT) [45]. Since 2014, 14 new articles were published providing reference data [23, 26, 31-33, 35, 37-39, 43, 45-48]. All tests are freely available and take less than 5 to 20 minutes to administer (Table 4).

Tests assessing anticipatory postural adjustments as the dominant system. The following tests mainly assess APA (Fig. 1) by recording the time children maintain tandem-stance (TS) [18, 19, 49], one-leg-stance (OLS) [18, 19, 49, 50] or variants like the Flamingo test [31], SBST [32–34] or by estimating the reach-distance of the free foot during OLS: SEBT [35, 36] and YBT [37, 38]. To a lesser extent, the tests evaluate sensory strategies with a narrowed base of support and SEBT and YBT also evaluate orientation in space (Fig. 1).

No new records on reliability or validity were identified for the **TS** [18, 19, 49] and the traditional timed **OLS** [18, 19, 49, 50], but were found for the **SBST** [32–34]. The SBST has good test-retest reliability [34] (Table 2), but poor concurrent validity with backward BBW [32] (Table 3) in TDC (age 3-6). Reference values are developed for both TS and OLS for 3- to 19-year-old TDC [35], for the **Flamingo test** (TDC; ages 6-10) [31] and the SBST (TDC; ages 3-6) [33] (Table 4).

The **SEBT** has good intrarater and interrater reliability in children with CP [36] and the **YBT** has good interrater and test-retest reliability in TDC [38] (Table 2), but no articles on validity or responsiveness are available yet. For the SEBT, reference values have been established for 3- to 19-year-old TDC [35] and for the YBT, reference values are available for TDC aged between 7 and 11 years [38] and 10 and 17 years [37] (Table 4).

Tests assessing orientation in space as the dominant system. The functional **Reach Test (RT)** and its variants primarily evaluate orientation in space (Fig. 1) by estimating the child's maximum stability limits. Additionally, APA and control of dynamics are required (Fig. 1). Fourteen studies investigated the different variants of RT [22, 23, 39, 51-60], 4 are new since 2014 [22, 23, 39, 60], adding good test-retest reliability for the **PRT** in children with CP (age 2-7) [22] (Table 2) and reference values for the **PRT** in Turkish TDC with [21] and without knee hypermobility aged 6-12 years [60]. Reference values for the **MRT** exist for 5- to 12-year-old TDC [39] (Table 4).

Tests assessing sensory strategies as the dominant system. Since 2014, 2 tests have been studied for pediatric rehabilitation: the (m)CT-SIB [14, 20, 40] and the BESS test [41–43]. Different sensory conditions require the use of different sensory strategies. Test-retest reliability of the **(m)CT-SIB** for CP and Down syndrome is good [14, 40], but poor for children with global developmental delay [20] (Table 2). Concurrent validity of the (m)CT-SIB with the Sensory Organization test (SOT) was poor, but significant [40]. The mCT-SIB showed a sensitivity of 95% and a specificity of 43% in children with CP [40]

Table 2
Summary of relative reliability and measurement error.

Functional postural control test		Group (N, age range in years)	Reliability and measurement error (numeric values: SEM/MDC/ME)		
			Intrarater/Within session	Interrater	Test-retest
ONE SYSTEM	(m)CTSIB	GDD (20, 4-12) [20], CP (14, 7-12) [40]			(+) [20, 40]
	(m)CTSIB$_{OLS}$	DS (9, 8-17) [14]			(+) [14] *SEM: 0.21-0.59* [14]
	BESS	TDC (381, 5-14) [41, 42]	(+) [42] *MDC: 4.58* [42]	(+) [41, 41] *MDC: 9.57* [42]	(+) [42] *MDC: 7.33* [42]
	TS$_{on\ beam}$	TDC (237, 10) [49]			(-) [49]
	SBST	TDC (90, 3-6) [34]			(+) [34]
	OLS	TDC (294, 4-12) [18, 49, 50], TDC&CP (25, 8-14) [51], HI (23, 6-12) [18]	(+) [51]	(?) [50] *SEM: 2.63* [51]	(+) [49, 50] *SEM TDC: 10.16-13.37* [18], *SEM HI: 8.71-8.83* [18]
	YBT	TDC (188, 7-11) [38]		(+) [38] *ME: 2.68-3.13* [38]	(+) [38] *ME: 16.41* [38]
	SEBT	CP (8, 6-12) [36]	(+) [36]	(+) [36] *SEM: 2.63* [36]	
	PRT	CP (38, 2-12) [22, 58]	(-) [58]	(-) [58]	(+) [22] *SEM: 16.8* [22]
	FRT	TDC (93, 7-16) [54, 57], TDC&CP (25, 8-14) [51], TBI (24, 7-14) [54], HI (65, 6-11) [56], CP (22, 5-12) [52]	(+) [51] *SEM TDC : 1.41* [54], *SEM TBI : 0.97* [54], *SEM HI: 0.29-0.51* [56]	(+) [51, 52, 56]	(+) [52, 57]
	LRT	TDC (24, 7-14) [54], TBI (24, 7-14) [54] HI (65, 6-11) [56]	(+) [54] [56] *SEM TDC: 0.80-0.97* [54], *SEM TBI: 0.72-0.90* [54], *SEM HI: 0.28-0.32* [56]	(+) [56]	
	FSST	TDC (179, 5-12) [30, 44], CP (16, 5-12) [17], DS (27, 5-17) [16, 17]	(+) [44]	(+) [16, 17] *ME TDC:-1.11-0.87* [44]	(+) [16] *SEM TDC: 0.96-0.98 [30], SEM CP: 1.34* [17], *SEM DS: 2.32* [17]
	TUDS	TDC&CP (25, 8-14) [51], DS (8, 3-17) [24]	(+) [51]	(+) [51]	(+) [51] *MDC DS: 12.52* [24]
	TUG	TDC (226, 3-14) [25, 54, 61], TDC&CP (25, 8-14) [51], ABI (54, 7-16) [28, 54], CP (95, 3-14) [27, 52, 62], CP&BD (41, 3-19) [61], DS (12, 3-17) [24]	(+) [25, 51, 61] *SEM TDC: 0.60* [54], *SEM ABI: 0.23* [54]	(+) [51] *SEM TDC: 0.14-0.15* [25]	(+) [54, 61] *SEM TDC: 0.67-0.83 [25], SEM CP: 0.46; 0.42$^{\mu}$ [28], SEM CP: 0.51-3.15* [27, 52, 62], *MDC DS: 1.26* [24]
	SWOC	TDC (50, 4-11) [63], DD (23, 6-21) [63]	(+) [63]	(+) [63]	
	CGT	TDC (90, 3-6) [45]			(+) [45]
	DGI	TDC&FASD (11, 8-15) [64]			(+) [64]
	BBW	TDC (601, 3-6; 237, 10) [32, 45, 49]			(?) [45, 49] SEM: 0.35-4.01 [32]
MULTI PLE SYSTEMS	CB&M	ABI (32, 7-18) [77]	(+) [77] *SEM: 3.7§* [77]	(+) [77] *SEM: 4.8; 3.9§* [77]	(+) [77] *SEM: 5.8; 5.6$^{\mu}$* [77]
	GDBT	TDC (144, 1.5-6) [76], MR (22, 1.5-6) [76]		(+) [76] *SEM MR: 0.78*	(+) [76] *SEM TDC: 0.21*
	PBS	TDC (40, 5-7) [79], CP (146, 5-13) [80, 81, 85], BD (34, 4-18) [95], MI (20, 5-15) [79]	(+) [81, 95] *SEM CP: 0.37-0.43* [80]	(+) [81, 95] *SEM CP: 0.65§* [80]; 1.78-1.80 [85]	(+) [79] SEM CP: 0.61 [80]; 1.79 [85]
	BBS	CP (50, 5-14) [52, 62]	(+) [62] *SEM: 0* [62]	(+) [52, 62]	(+) [52] *SEM: 0.18-0.22§* [62]
	ECAB	CP (575, 1-12) [22, 71, 72, 74]	(+) [72]	(+) [22, 72]	[22, 71] (+) *SEM: 0-3.6]* [22, 71
	FAB	CP (40, 5-16) [70]			(+) [70]
	Kids-BESTest	TDC (34, 7-17) [3], CP (18, 8-17) [67]	(+) [3, 67] *SEM TDC - Full/Mini: 0.81/0.54* [3] *SEM CP – Full/Mini: 1.98/0.88* [67]	+) [3, 67] *SEM TDC – Full/Mini: 1.45/0.96* [3]; *SEM CP – Full/Mini: 3.08/1.20* [67]	(+) [3, 67] *SEM TDC – Full/Mini: 2.38;1.77§/0.47;0.86§* [3]; *SEM CP – Full/Mini: 2.03;2.19§/1.67;1.43§* [67]

Legend: §: video recordings; an overall judgement of reliability is given with a (+), (-) and (?) in bold considering all references; (+): the majority of the findings showed an ICC/ w≥0.70; (-): the majority of the findings showed showed ICC/ w<0.70; (?): there was no majority making conclusions indeterminate; empty cells: property was not investigated; measurement errors (i.e. ME, SEM, MDC) are given in numbers in italic.

Abbreviations: ABI: acquired brain injury, BBS: Berg Balance Scale, BBW: Balance Beam Walking, BD: balance disabilities, BESS: Balance Error Scoring System, CB&M: Community Balance & Mobility Scale, CGT: Complex Gait Test, CP: cerebral palsy, (m)CT-SIB: (Modified) Clinical Test of Sensory Interaction in Balance, DD: developmental disabilities, DGI: Dynamic Gait Index, DS: Down syndrome, ECAB: Early Clinical Assessment of Balance, FAB: Fullerton Advanced Balance Scale, FASD: fetal alcohol spectrum disorder, FRT: Forward Reach Test, FSST: Four Square Step Test, GDBT: Ghent Developmental Balance Test, GDD: global developmental delay, HI: hearing impairment; Kids-BESTest: Kids-Balance Evaluation Systems Test, LRT: Lateral Reach Test, MI: motor impairment, MR: motor retardation, MRT: Multidirectional Reach Test, OLS$_{EO/EC}$: One-Leg-Stance with eyes or eyes closed, PBS: Pediatric Balance Scale, PRT: Pediatric Reach Test, SBST: Stork Balance Stand Test, SEBT: Star Excursion Balance Test, SWOC: Standardized Walking Obstacle Course, TBI: traumatic brain injury, TDC: typically developing children, TS: Tandem-Stance, TUDS: Timed Up and Down Stairs test, TUG: Timed Up and Go test, YBT: Y-Balance Test.

Table 3
Summary of structural, cross-cultural, concurrent and construct validity for each functional postural control test.

Functional postural control test		Group (n, age range in years)	Structural validity	Internal consistency (Cronbach α: 95% CI)	Concurrent validity (comparator, +/ -/ ±[A])	Construct validity Hypothesis testing (comparator, +/ -/ ±[B])	Known group validity (comparison)
ONE SYSTEM	**(m)CTSIB**	CP (32, 7-12) [40]			SOT (-)		
	$TS_{on\ beam}$	TDC (237, 10) [49]			**TS_{EO}/TS_{EC}** vs OLS_{EO}/OLS_{EC} (-)		
	OLS	TDC (80, 6-14) [19, 51], HI (23, 6-12) [19], CP (20, 8-14) [51]			**$OLS_{EO/EC}$** vs $COP_{(F)EO/(F)EC}$, COP_{TSvel}, COP_{SLSvel} (±), TUDS (±)		HI<TDC***
	PRT	CP (38, 2-12) [22, 58]			COP_{AP-ML} (±) [58], ECAB (+) [22]	GMFM-66 B&C (+) [22]	
	FRT	TDC (27, 8-14) [51], CP (50, 5-14) [51, 52], TBI (24, 7-14) [53], DS (13, 8-17) [16]			TUG (+) [52], TUDS (-) [51], FSST (-) [16], BBS (+) [52]	GMFM-88 (D-E) (-) [52], STS (-) [52], Walking speed (-) [52], Step length (±) [53]	GMFCS III<II<I ** [52], TBI<TDC*** [53]
	LRT	TBI (24, 7-14) [53]				Step length (-)	TBI<TDC***
	FSST	TDC (30, 6-12) [30], CP (36, 6-12) [17, 30], DS (27, 5-17) [16, 17]			TUG (±) [17, 30], FRT (-) [16], BOT-2 (-) [30]		CP=TBI<TDC* [30]
	TUDS	TDC (27, 8-14) [51], CP (20, 8-14) [51],			TUG (±), FRT (±), OLS_{EO} (±) [51]		GMFCS II/III>I>TDC [51]
	TUG	TDC (112, 3-12) [25, 30, 63], TBI (24, 7-14) [53], CP (66, 5-14) [17, 29, 51, 52], CP&BD (41, 3-19) [61], DS (14, 6-12) [17], DD (23, 6-21) [63]			MABC-2 B (-) [25], TUDS (±) [51], FSST (±) [17, 30], SWOC (+) [63], BBS (+) [52], FRT (+) [52]	Step length (+) [52], GMFM-88 (D-E) (-) [52], STS (-) [52], Walking speed (+) [52]	TBI<TDC*** [53] III>II>I** [52]; III>II/I** [29]; NS [27];
	SWOC	TDC (50, 4-11) [63], DD (23, 6-21) [63]			TUG (+)		
	CGT	TDC (80, 3-6) [45]			BBW (-)		
	DGI	TDC&FASD (20, 8-15) [64]					FASD<TDC**
	BBW	TDC (593, 3-6) [32], TDC (237, 10) [49]			SBST (-) [32], $OLS_{EO/EC}$ (-), $TS_{beamEO/EC}$ (-) [49]		
MULTIPLE SYSTEMS	**GDBT**	TDC (28, 1.5-6) [76], MR (20, 1.5-6) [76]			PDMS-2 S-L (-), MABC B (-)	PDMS-2 (-), MABC (-), BOT-2 (-)	MR<TDC***
	PBS	TDC (258, 2-4) [84] CP (342, 1-16) [70, 78, 82, 84], BD & TDC (138 & 685, 2-13) [65], BD (34, 4-18) [66], DS (44, 2-10) [15]	Uni-dimensional [65]	[66]	FRT (+) [66], SOT (-) [82], FAB (-) [70]	GMFM-66 (+) [78, 82], WeeFIM (+) [78, 82], GMFM-88(D-E) (+) [15, 82], PEDI-mobility (+) [82]	GMFCS I>II>III** [82] TD>GMFCS I>II>III* [84]
	BBS	CP (30, 5-12) [52]			FRT (+), TUG (+)	GMFM-88(D-E) (-), STS (-), Walking speed (-)	GMFCS I/II>III>VI*
	ECAB	CP (575, 1-12) [22, 71-74] CP (37, 0.5-3)-TDC (13, 0.5-3) [75]		[73]	PRT (+) [22]	GMFM-66 (+) [75], GMFM-66-B&C (+) [73], GMFM-88(A-B-C-D-E) (±) [71, 75]	CP<TDC [74]; GMFCS I>III/IV/V***; II>V** [71]; I>II>III>IV>V*** [73]; II>IV*; III>IV** [74]
	FAB	CP (40, 5-16) [70]	Two- dimensional [70]	[70]	PBS (-) [70]		
	Kids-BESTest	TDC (41, 7-18) [68, 69], CP (17, 7-18) [68, 69]			**FRT/LRT vs COP measures,** Δ trunk, knee, ankle angle (-) [69]; **mCTSIB: FEC vs COP measures** (-) [68]		

* $p<0.05$; ** $p<0.01$; *** $p<0.001$.
[A] indicates strength of correlation coefficients: (+) if ≥0.7, (-) if <0.7;
[B] indicates confirmation (+) or rejection (-) of the hypothesis (correlation coefficients indicating same construct >0.5, related construct: 0.3-0.5 or different: <0.3) or inconsistent results (±), Δ: change.
Abbreviations: AP: anteroposterior direction, BBS: Berg Balance Scale, BBW: Balance Beam Walking, BD: balance disabilities, BESS: Balance Error Scoring System, BOT-2: Bruininks-Oseretsky Test for Motor proficiency, 2nd edition, CB&M: Community Balance & Mobility Scale, CGT: Complex Gait Test, CP: cerebral palsy, COP: center of pressure, (m)CT-SIB: (Modified) Clinical Test of Sensory Interaction in Balance, DD: developmental disabilities, DGI: Dynamic Gait Index, DS: Down syndrome, ECAB: Early Clinical Assessment of Balance, EO: eyes open, EC: eyes closed, FAB: Fullerton Advanced Balance Scale, FASD: fetal alcohol spectrum disorder, FEO: bilateral stance on foam with eyes open, FEC: bilateral stance on foam with eyes closed, FRT: Forward Reach Test, FSST: Four Square Step Test, GDBT: Ghent Developmental Balance Test, GMFM-66/88: gross motor function measure with 66/88 items, HI: hearing impairment; Kids-BESTest: Kids-Balance Evaluation Systems Test, LRT: Lateral Reach Test, MABC(-2): Movement Assessment Battery for children (2nd edition); MABC B: balance domain, ML: mediolateral direction, MR: motor retardation, MRT: Multidirectional Reach Test, OLSEO/EC: One-Leg-Stance with eyes or eyes closed, PBS: Pediatric Balance Scale, PDMS-2: Peabody Developmental Motor Scales, 2nd editions, PDMS-2 L: locomotion domain, PDMS-2 S: stationary domain, PEDI: Pediatric Evaluation Disability Inventory, PRT: Pediatric Reach Test, SBST: Stork Balance Stand Test, SEBT: Star Excursion Balance Test, SOT: Sensory Organization Test, SWOC: Standardized Walking Obstacle Course, STS: Sit to stand, TBI: traumatic brain injury, TDC: typically developing children, TS: Tandem-Stance, TUDS: Timed Up and Down Stairs test, TUG: Timed Up and Go test, vel: velocity, YBT: Y-Balance Test.

	Test	Reliability: Reliability	Reliability: Measurement error	Re-sponsiveness	Validity: Structural	Validity: Internal consistency	Validity: Concurrent (comparator)	Validity: Hypothesis testing (comparator)	Validity: Known group (comparison)	Validity: Cross-cultural
ONE SYSTEM	mCT-SIB	+	?		N/A		-		?	
	BESS	+	?		N/A					
	TS	-			N/A		-			
	Flamingo test									
	SBST	+	?		N/A		-			
	OLS	+	?		N/A		-		-	
	SEBT/YBT	+	?		N/A					
	PRT/FRT/ LRT/MRT	+	?		N/A		?	?	+	
	FSST	+	?		N/A		-		+	
	TUDS	+			N/A		-		+	
	TUG	+	+	?	N/A		+	?	+	
	SWOC	+			N/A		+			
	CGT	+	?		N/A		-			
	DGI	+			N/A				+	
	BBW	?	?		N/A		-			
MULTIPLE	CB&M	+	?							
	PBS	+	?	+	+	+	-	-	+	
	BBS	+	?				+	-	+	
	GDBT	+	?				+	+		
	ECAB	+	?	+		?	+	+	+	
	FAB	+			-	?	-			
	Kids-BESTest	+	?				-		+	

Fig. 2. level of evidence (GRADE) and quality of measurement properties rated with the COSMIN criteria.

Legend: N/A: not applicable.

Interpretation colors: colors are the consensus of GRADE scoring per functional postural control test.

red: very low level of evidence; pink: low level of evidence; orange: moderate level of evidence; green: high level of evidence.

Interpretation ±/-/?: "+": sufficient, "-": insufficient, "?": indeterminate determined by COSMIN criteria.

Reliability**:** "+": ICC or weighted Kappa ≥ 0.70, "-": ICC or weighted Kappa <0.70; "?": ICC or weighted Kappa not reported.

Measurement error: "+": SDC or LoA<MIC, "-": SDC or LoA>MIC, "?": MIC not defined.

Responsiveness: "+": The result is in accordance with the hypothesis OR AUC ≥ 0.70, "-": The result is not in accordance with the hypothesis OR AUC <0.70, "?": No hypothesis defined (by the review team).

Structural validity: "+": CTT: confirmatory factor analysis: comparative fit index (CFI) or Tucker-Lewis index (TLI) or comparable measure >0.95 or Root Mean Square Error of Approximation (RMSEA) <0.06 OR Standardized Root Mean Residuals (SRMR) <0.082; IRT/Rasch: No violation of *unidimensionality*: CFI or TLI or comparable measure >0.95 or RMSEA <0.06 or SRMR <0.08 and no violation of *local independence*: residual correlations among the items after controlling for the dominant factor < 0.20 or Q3′s <0.37 and no violation of *monotonicity*: adequate looking graphs OR item scalability >0.30 AND adequate *model fit*: IRT: $\chi 2$ >0.01 Rasch: infit and outfit mean squares ≥0.5 and ≤1.5 or Z− standardized values >-2 and <2; "-": Criteria for + not met; "?": CTT: not all information for "+" reported, IRT/Rasch: Model fit not reported.

Internal consistency: "+": At least low evidence for sufficient structural validity and Cronbach's alpha(s) ≥0.70 for each unidimensional scale or subscale, "-": At least low evidence for sufficient structural validity and Cronbach's alpha(s) <0.70 for each unidimensional scale or subscale, "?": Criteria for "At least low evidence for sufficient structural validity" not met.

Concurrent validity: "+": Correlation with gold standard ≥0.70 OR AUC ≥0.70, "-": Correlation with gold standard <0.70 OR AUC <0.70, "?": Not all information for '+' reported

Hypothesis testing/known group validity: "+": The result is in accordance with the hypothesis, "-": The result is not in accordance with the hypothesis, "?": No hypothesis defined (by the research team).

Abbreviations: BBS: Berg Balance Scale, BBW: Balance Beam Walking, BESS: Balance Error Scoring System, CB&M: Community Balance & Mobility Scale, CGT: Complex Gait Test, (m)CT-SIB: (Modified) Clinical Test of Sensory Interaction in Balance, DGI: Dynamic Gait Index, ECAB: Early Clinical Assessment of Balance, FAB: Fullerton Advanced Balance Scale, FRT: Forward Reach Test, FSST: Four Square Step Test, GDBT: Ghent Developmental Balance Test, GRADE: Grading of Recommendations Assessment, Development and Evaluation, Kids-BESTest: Kids-Balance Evaluation Systems Test, LRT: Lateral Reach Test, MRT: Multidirectional Reach Test, OLSEO/EC: One-Leg-Stance with eyes or eyes closed, PBS: Pediatric Balance Scale, PRT: Pediatric Reach Test, SBST: Stork Balance Stand Test, SEBT: Star Excursion Balance Test, SWOC: Standardized Walking Obstacle Course, TS: Tandem-Stance, TUDS: Timed Up and Down Stairs test, TUG: Timed Up and Go test, YBT: Y-Balance Test.

(Table 3). The **BESS** showed good test-retest, intra- and interrater reliability [41, 42] (Table 2). Reference values have been determined for the (m)CT-SIB for children with global developmental delay (4-12 years) [20] and for the BESS for TDC aged 5-14 [43] (Table 4).

Tests assessing control of dynamics as the dominant system. The FSST [16, 17, 30, 44], TUDS [24, 51], TUG [17, 24-30, 46-48, 51-54, 61-63], Standardized Walking Obstacle Course (SWOC) [63], CGT [45], Dynamic Gait Index (DGI) [64] and BBW [45] all require dynamic control because of a changing base of support. All movements are self-induced, requiring APA.

No new articles were published on the ***SWOC*** [63] or the ***DGI*** [64].

The ***FSST*** [16, 17, 30, 44] has good intrarater [30], interrater [17, 44] and test-retest reliability [17, 30] in TDC and children with CP, but poor test-retest reliability in children with Down syndrome [16, 17] (Table 2). Concurrent validity was good with the TUG in children with CP [17], but poor with the TUG in TDC [30, 44] and children with Down syndrome [17] and with the forward RT (FRT) in children with Down syndrome [16] (Table 3). Correlations with the Bruininks-Oseretsky Test of Motor Proficiency, 2nd edition (BOT-2) were weak, indicating the FSST measures a different construct [30] (Table 3).

Table 4
Feasibility features of the functional postural control tests.

Test		Population	Age (years)	Ref-erence values	Cross-cultural adaptation	Admini-stration time (minutes)	Equipment	Outcome measure(s)	Price *
ONE SYSTEM	**(m)CT-SIB** [14, 20, 40]	TDC [20] GDD [20] CP [40] DS [14]	4-12 4-12 7-12 8-17	 X [20]		20	medium density foam pad; stopwatch [14, 20, 40]; *optional: visual conflict dome* [20]	time	free
	BESS [41–43]	TDC [41–43]	5-14	X [43]		15	medium density foam pad; stopwatch; score card [41–43]	score	free
	TS [35, 49]	TDC [35, 49]	3-10	X [35]		<5	tape; stopwatch [35, 49]; *optional: visual target on wall; balance beam (2 × 0.05m)* [49]*;*	time	free
	Flamingo test [31]	TDC [31]	6-10	X [31]		<5	stopwatch [31]	# attempts [31]	free
	SBST [32–34]	TDC [32–34]	3-6	X [32]		<5	stopwatch [32–34]	time	free
	OLS$_{EO/EC}$ [18, 19, 35, 49-51]	TDC [18, 19, 21, 35, 49-51] HI [18, 19] CP [51]	3-12 6-12 8-14			<5	stopwatch; [18, 19, 35, 49-51] *optional: tape; visual target on wall* [35, 50, 51]	time	free
	SEBT [35, 36] / **YBT** [37, 38]	TDC [35, 37, 38] CP [36]	3-19 6-12	X [35]		10	PVC pipe; platform [37, 38]; or tape [35, 36]; length measure [35–38]; footwear [38] or barefoot [35–37]; *optional: plastic alligator toys* [38]	distance [36, 38]; % of leg length [35, 37]	free
	PRT [21-23, 58, 59] / **FRT** [16, 51-57] / **LRT** [53, 54, 56] / **MRT** [39]	TDC [21, 39, 51, 53-55, 57-59] CP [52] knee hypermobility [23] HI [56] TBI [53, 54] DS [16]	2-18 2-18 6-12 6-11 7-14 8-17	X (3-12y) [21] X [23]		5-10	length measure, yardstick or ruler [16, 21, 23, 39, 51-59]; *optional: stool; score sheet* [56, 58]	distance	free
	FSST [16, 17, 30, 44]	TDC [30, 44] CP [17, 30] ABI [30] DS [16, 17]	5-12 5-12 6-12 6-17			<5	canes/rods (4x); stopwatch [16, 17, 30, 44]; footwear [17, 44] or barefoot [16]; *optional: non-skid rubber mat; visual target on wall; red sequence-numbers in squares* [17]	time	free
	TUDS [24, 51]	TDC [51] CP [51]	8-15 8-15			<5	14-steps flight stairs with handrails; regular footwear; stopwatch [24, 51]	time	free
	TUG [17, 24-30, 46-48, 51-54, 61-63]	TDC [25, 26, 30, 46-48, 51, 53, 54, 61, 62] CP [17, 27, 29, 51, 52, 62] ABI [28] TBI [53, 54] CP&BD [61] DD [63] DS [17]	3-19 3-19 8-16 7-14 3-19 6-21 6-12	X [26, 46-48]		5	(adjustable) chair; length measure; tape; stopwatch [24-30, 46-48, 51-54, 61-63]; regular footwear [24-26, 29, 46, 53, 54] or barefoot [51]; *optional: tape-line* [28, 29, 46, 47, 51, 53, 54, 62, 63] [63]*, cone* [47]*, target on the wall* [24, 26, 29, 47, 48, 61] *or Duplo brick for transportation* [25]	time	free
	SWOC [63]	TDC [63] DD [63]	4-11 6-21			15	free walkway; axillary crutch; visually stimulating mat; trash can; shag rug; chair with(out) armrest (2x); tray; glasses; stopwatch [63]	time; steps	free
	CGT [45]	TDC [45]	3-6	X [45]		<5	free walkway; tape; length measure; cones (≥11); stopwatch [45]	time	free
	DGI [64]	TDC [64] FASD [64]	8-15 8-15			15	free walkway; shoe box; cones (2x); stairs [64]	Criterion scores	free
	BBW [32, 45, 49]	TDC [32, 45, 49]	3-6 [32]; 10 [49]	X		<5	balance beam (2.5 × 0.04 × 0.12 m) [32, 45, 49]; stopwatch [32, 49]; length measure [32, 45]	time [32, 49]; distance [32, 45]; steps [32]	free

(*continued on next page*)

Table 4 (*Continued*)

Test		Population	Age (years)	Ref-erence values	Cross-cultural adaptation	Admini-stration time (minutes)	Equipment	Outcome measure(s)	Price *
MULTIPLE SYSTEMS	**CB&M** [77]	ABI [77]	8-18			30	free walkway; laundry basket; weights (2 lbs/1 kg; 7.5 lbs/3.5 kg; visual target on ground; bean bag; regular foot-wear; stopwatch [77]	Criterion scores	free
	PBS [15, 65, 66, 78-84]	TDC [79, 83, 84] MI [79] DS [15] BD [65, 66] CP [70, 78, 80-82, 84, 85]	2-13 5-15 2-10 2-18 2-16	X [83, 84] X [84]	Brazilian [81], Turkish [66] Korean [70, 82] Persian [85]	10-20	Pediatric version of the BBS. adjustable bench; chair with back support and arm rests; step; chalkboard eraser; yardstick; small level [15, 65, 66, 78-83]; *optional: flash cards; blindfold, footprints; visual colored target; Velcro [66]*	Criterion scores	free
	BBS [52, 62]	CP [52, 62]	6-14			10-20	2 standard chairs, with back support, one with arm rests, one without; step; a ruler; stopwatch [52, 62]	Criterion scores	free
	GDBT [76]	TDC [76] MR [76]	1.5-6 4-5	X [76]		20	GDBT manual; thin mat; tape; ball; medium density foam; stopwatch; scoring sheet [76]	Criterion scores	€ 24.99
	ECAB [22, 71-75]	TDC [74] CP [22, 71-75]	0.5-3 0.5-12		Turkish [71]	30	bench; step; mat; stopwatch; test form [22, 71-75]	Criterion scores	free
	FAB [70]	CP [70]	5-16		Korean	10-12	bench; medium density foam (2x); length measure; pencil; stopwatch [70]	Criterion scores	free
	Kids-BESTest [3, 67-69]	TDC [3, 68, 69] CP [67–69]	7-18 7-18			30	free walkway; tape; blindfold; length measure; medium density foam; incline ramp; bench; shoe box; weight (1kg); adjustable chair; stopwatch [3, 67]	Criterion scores	free

Legend: * prices refer to manual. Criterion scores: for each item, the performance is scored against a predetermined criterion using an ordinal scale varying from 3 to 5-point rating scales. These scores are summed to determine the final test score. Usually, higher scores represent better postural control performance.

Abbreviations: ABI: acquired brain injury, BBS: Berg Balance Scale, BBW: Balance Beam Walking, BD: balance disabilities, BESS: Balance Error Scoring System, CB&M: Community Balance & Mobility Scale, CGT: Complex Gait Test, CP: cerebral palsy, (m)CT-SIB: (Modified) Clinical Test of Sensory Interaction in Balance, DD: developmental disabilities, DGI: Dynamic Gait Index, DS: Down syndrome, ECAB: Early Clinical Assessment of Balance, FAB: Fullerton Advanced Balance Scale, FASD: fetal alcohol spectrum disorder, FRT: Forward Reach Test, FSST: Four Square Step Test, GDBT: Ghent Developmental Balance Test, GDD: global developmental delay, HI: hearing impairment; Kids-BESTest: Kids-Balance Evaluation Systems Test, LRT: Lateral Reach Test, MI: motor impairment, MR: motor retardation, MRT: Multidirectional Reach Test, $OLS_{EO/EC:}$ One-Leg-Stance with eyes or eyes closed, PBS: Pediatric Balance Scale, PRT: Pediatric Reach Test, SBST: Stork Balance Stand Test, SEBT: Star Excursion Balance Test, SWOC: Standardized Walking Obstacle Course, TBI: traumatic brain injury, TDC: typically developing children, TS: Tandem-Stance, TUDS: Timed Up and Down Stairs test, TUG: Timed Up and Go test, YBT: Y-Balance Test.

One new article published on the **TUDS** [24] showed good test-retest reliability in children with Down syndrome [24] (Table 2).

Eleven new studies investigated the ***TUG*** [17, 24-30, 46-48]. The TUG has good intrarater, interrater and test-retest reliability for TDC [25, 54, 61], children with CP [27, 52, 62], children with acquired brain injuries [28, 54], children with Down syndrome and heterogeneous groups consisting of children with CP and balance disabilities [61] or CP and TDC [51] (Table 2). Concurrent validity with the Movement Assessment Battery for Children, 2nd edition (MABC-2) balance subscale was low in TDC (age 3-6) [25] but correlated well with the SWOC [63] in TDC and children with developmental disabilities and with BBS [52] and FRT [52] in children with CP. Concurrent validity between TUG versus FSST is discussed in the FSST section above. However, contradictory results were found between TUG versus FSST [17] and TUDS [51] (Table 3). In children with CP, TUG performance correlated strongly with Gross Motor Function Measure (GMFM) scores indicating a similar construct in children with CP [52]. The responsiveness of the TUG was confirmed for children with a mild to moderate form of CP using an anchor-based method. The researchers used a 1-point change on the Goal Attainment Scale (GAS) to evaluate the responsiveness of the TUG (12).

Two new studies were available on **BBW** [32, 45] showing good test-retest reliability [32, 45] (Table 2), but poor concurrent validity with SBST in TDC (age 3-6) [32] (Table 3).

One study proposed the **CGT**, a rectangular-shaped walking course around which the child walks at maximum speed [45]. The CGT has good test-retest reliability (Table 2), but low concurrent validity with the BBW in TDC (age 3-6) [45] (Table 3).

Two new studies were available on the **BBW** [32, 45], showing good test-retest reliability [32, 45] (Table 2), but low concurrent validity with the SBST in TDC (age 3-6 years) [32] (Table 3).

Reference values for the TUG, CGT and BBW exist. The update revealed new reference values: 1) TUG for TDC with ages 3-18 [26], 4-11 [48], 6-12 [47] and 5-13 years [46] and 2) BBW [32] and CGT [45] for 3- to 6-year-old TDC (Table 4).

Postural control tests assessing multiple systems

The update revealed 3 new records on the previously reported Pediatric Balance Scale (PBS) [15, 65, 66] and 3 new test batteries: the Balance Evaluation Systems Test for children (Kids-BESTest) [3, 67-69], the Fullerton Advanced Balance scale (FAB) [70] and the Early Clinical Assessment of Balance (ECAB) [22, 71-75]. All manuals of these tests, except for the Ghent Developmental Balance Test (GDBT) [76], are freely available and test administration time varies between 10 to 30 minutes (Table 4).

No new articles were published on the **Community Balance and Mobility Scale (CB&M)** [77], the **GDBT** [76] or the **Berg Balance Scale (BBS)** [52, 62]. In addition to the previously reported studies [78-83], 6 new articles were published on the **PBS** [15, 65, 66, 70, 84, 85]. The PBS items primarily assess APA, orientation in space and control of dynamics, and to a lesser extent sensory orientation (Fig. 1). In children with known balance disabilities, good internal consistency, intrarater and interrater reliability were reported [66] and in children with CP, excellent test-retest and interrater reliability were reported [85] (Table 2). The scale is unidimensional, and the difficulty level of the items was established with Rasch analysis (structural validity) in a large sample of children with known balance deficits and TDC [65] (Table 3). Concurrent validity with the FAB is low in CP children (5-6 years) [70]. Scores of the PBS correlate well with GMFM scores, indicating a similar underlying construct in children with Down syndrome [15] (Table 3). The PBS distinguishes children with Gross Motor Function Classification System (GMFCS) levels I-III from each other [82, 84] and TDC [84] (Table 3). In addition to the Brazilian version [81], Korean [70, 82], Persian [85] and Turkish versions are now available [66] (Table 4). Reference values exist for 2- to 5-year-old TDC and CP children [84].

The **ECAB** [22, 71-75] assesses all domains of balance control except for orientation in space (Fig. 1). The test has good internal consistency [73] (Table 3), intrarater [72], interrater [22, 72] and test-retest reliability [22, 71] in children with CP (Table 2). Concurrent validity with the PRT is good in children with CP [22] (Table 3). The ECAB scores correlate strongly with GMFM scores in children with CP [75], suggesting a similar construct, and they distinguish children with CP from those with typical development [74] (Table 3). Children with CP with a GMFCS level I can be distinguished from the other levels [71, 73], but differences between other GMFCS levels are inconsistent [71, 73, 74] (Table 2). The standardized response mean in children with CP after 3 and 6 months of intervention was medium and large respectively [75].

The **FAB** [70] covers the entire construct of postural control (Fig. 1). The FAB has good internal consistency and test-retest reliability, but poor concurrent validity with the PBS in children with CP (Tables 2 and 3). The FAB for children consists of 2 dimensions, labelled by the authors as “static and quasi-dynamic” and “stability of gait” and is available in Korean (Table 3).

The **Kids-BESTest** [3, 67-69] also evaluates the entire construct of postural control (Fig. 1). The test has good intrarater, interrater and test-retest reliability in TDC and children with CP (age 7-18) [3, 67]. The Mini-Kids-BESTest, a shortened version of the Kids-BESTest comprising 14 items across 4 domains (APA/transitions, reactive postural responses, sensory orientation, and stability in gait), has poorer interrater reliability compared to the full version (Table 2) [3, 67]. Concurrent validity has been investigated for specific Kids-BESTest items, showing poor concurrent validity for the FRT, lateral RT (LRT) and mCT-SIB, with center of pressure measures (Table 3) [68, 69].

Risk of bias and level of evidence

The investigated psychometric properties of the tests were overall characterized by a very low or low level of evidence (Fig. 2), which means that the true measurement property is likely to be, or may be, substantially different from the estimate of the measurement property. This was mainly caused by downgrading for risk of bias (70% of scores) and imprecision (57% of scores) (Appendix E). The severity of the risk of bias (Appendix F) for reliability and measurement errors mainly increased because of inappropriately short times between test-sessions to determine test-retest reliability and its measurement error (<14 days) (33 studies) and/or because administration and test conditions were not (thoroughly) explained in the study (22 studies). The risk of bias in validity studies increased for a small study sample size (4 studies), insufficient description of the comparator (6 studies) or a lack of justification of choice of statistical analysis (6 studies). All included functional postural control tests, except the Flamingo test, were investigated at least once for reliability (Fig. 2). Most ICCs or K_w reached the 0.70 criterion or higher, except for TS and BBW owing to respectively low or contradictory results, but the majority had a very low level of evidence (12/20 tests). Measurement errors were predominantly rated indeterminate (13/15 tests) since they can only be correctly interpreted if the minimal important change is properly calculated. The body of evidence varied from very low (9/15 tests) to moderate (4/15 tests). Concurrent validity showed overall correlations below 0.70 (-) with a very low (7/16 tests), low (6/16 tests) or moderate (3/16 tests) body of evidence (Fig. 2). The PBS, ECAB and TUG were most extensively investigated, nevertheless, the overall evidence for their psychometric properties is rather low. Strong evidence of responsiveness is only available for the TUG. Strong evidence was available for internal consistency for the ECAB, but results should be interpreted cautiously as no studies on its structural validity have yet been reported. Only qualitative evidence is available for the structural validity and known-group validity of the PBS (Fig. 2).

Discussion

The aim of this systematic review was to facilitate the selection of an appropriate pediatric functional postural control test by mapping the psychometric properties and feasibility of available tests.

Twenty-six functional postural control tests were identified, 12 of which were newly developed since our narrative review in 2014 [2]. Overall, the articles were heterogeneous in terms of the types of functional postural control tests (one vs. multiple systems), and the psychometric properties, population, and age ranges investigated. Likewise, assessment time, test protocols and required equipment were highly varied. Reference data are available for most postural control tests, except for the OLS, FSST, SWOC, DGI, FAB and Kids-BESTest (Table 4).

For each test, except for the Flamingo test, reliability was investigated at least once, whereas measurement error and concurrent validity were reported frequently but not for all (Fig. 2). Structural validity, internal consistency, hypothesis testing, cross-cultural validity and responsiveness remain understudied. The PBS, ECAB and TUG were most extensively investigated, nevertheless, the evidence for their psychometric properties remains low. Conclusive results with high evidence were only found for structural validity and known-group validity of the PBS. For all other psychometric properties, new methodologically sound research would likely change their estimates.

Validity

The construct of postural control is hypothesized to be multisystemic [5], and this should be covered by assessment tools. It was assumed that test batteries approximate this theoretically (Fig. 1), but only 2 tests, the Kids-BESTest and FAB, cover the entire multisystemic framework of postural control [5]. The identification of the postural control systems based on task-type was crucial to understanding the underlying constructs of the test, which determines the content validity.

In pediatric populations, structural validity was only investigated for the FAB and PBS [65, 70], not for the Kids-BESTest [3, 67-69]. The FAB covers the entire multisystemic framework (Fig. 1), whereas the PBS does not assess reactive postural responses. Nevertheless, not all supposed systems (Fig. 1) were translated into actual dimensions statistically [70]: all PBS tasks belong to the same dimensions [65], and the FAB appeared to consist of 2 dimensions [70]. Thus, these dimensions do not reflect the multiple systems needed for postural control, indicating that other factors may be involved, such as the included populations.

Previous exploratory research showed consistent findings regarding one-dimensionality and task-specificity in studies of healthy individuals [8, 65, 86] or heterogeneous pathological populations [65, 87], demonstrated by weak correlations across different tasks (anticipatory, reactive, steady-state and dynamic balance) [8, 86] and correlations ≥0.70 for similar tasks (control of dynamics: TUG vs. SWOC [63], vs. FSST [17] and vs. FRT [52]). The weak, significant correlations representing concurrent validity (Table 3) imply that the different tasks are significantly interrelated but depict another dominant system depending upon the task. For instance, correlations in TDC between FRT (orientation in space) and TUDS (control of dynamics) (r=-0.32) or between BBW (control of dynamics) and SBST (APA) (r=-0.26) [32, 51] underpin the multisystemic nature of postural control [5]. Recent evidence stressed that postural control performance also depends on the child's developmental stage [86], which is highlighted by the availability of various age norms (Table 4). For example, for BBW significant differences were found between the ages of 3, 4, 5 and 6 years [32], and in the PBS even significant 6-month differences were found between 2.5- and 5-year-olds [83]. Hence, in healthy children, postural control is both task- and age-specific.

In contrast, the dimensionality investigated with structural validity analysis changes in homogeneous pathological populations [70, 88-91]. In children with CP, the FAB consists of 2 dimensions described by the authors as "static and quasi-dynamic balance function" versus "stability in gait", although the last dimension includes OLS, which seems to be more related to the first than to the second dimension [70]. Seemingly, the 2 dimensions are not determined by task-type, but by tasks that are perceived as easier or more difficult for these children. This shift in dimensionality has been shown in other exploratory research: 1 dimension for the FAB with a different item hierarchy in individuals with stroke [89], 4 dimensions for the BESTest in individuals with stroke [90] and 3 dimensions for the mini-BESTest in individuals with Parkinson's disease [91]. Thus, each specific pathology determines both dimensionality and item hierarchy.

Our findings indeed confirm that postural control performance depends on the severity of the pathology. Healthy children can be distinguished from children with mild, e.g. hearing impairment with OLS [19] and severe motor deficits, e.g. traumatic brain injury with FSST [30] but tests can also differentiate between children with different functional levels, such as GMFCS levels I-III (CP) with the PBS endorsing the known-group validity of these tests [82, 84] (Table 3). Furthermore, higher correlations between functional postural control tests and GMFM total scores were found in children with CP [15, 52, 82] as compared with more heterogeneous groups, like children with balance disabilities [61] or mildly affected groups, like children with Down syndrome [71]. This indicates that all motor constructs are more strongly related to postural control when movement disorders become more severe. As such, the task-specificity found in healthy individuals becomes less important than the severity of the underlying pathology. Both should be considered.

Reliability, measurement errors and responsiveness

Except for the PBS, BESS, FSST and ECAB, the evidence for good reliability of all other functional postural control tests is (very) low. Overall, (very) serious risk of bias was the main cause of low evidence, but inconsistency, shown by either conflicting results (Appendix E) or very wide confidence intervals, and imprecision caused by small sample sizes also played a role. Especially in younger children, more inconsistent results were present, which may be caused by the typical day-to-day variability in their performance resulting from their developmental stage [25], whereas more consistent results were found for children with severe movement disorders like CP or traumatic brain injuries. Not all types of reliability were investigated for each test, which is important if measures need to be repeated, used interchangeably between healthcare professionals, or to determine the effect of therapy [6, 7, 13]. Moreover, reproducibility errors were only investigated in 21/26 functional postural control tests, with overall (very) serious risk of bias. Although measurement errors based on SEMs (calculated from the test-retest reliability) can aid in interpreting physiotherapeutic treatment outcomes, a change score can only be attributed relative to the amount of error, therefore lacking the clinical meaningfulness of the change (responsiveness). Hence, COSMIN guidelines state that the level of evidence for measurement errors decreases if the responsiveness is not determined for the test at hand. However, responsiveness is still insufficiently investigated, with data only available for the ECAB [75], PBS [78] and TUG [27, 29]. Test responsiveness was either investigated by calculating the standardized response mean (SRM), following the distribution method [29, 75, 78], or by calculating the minimal clinically important difference (MCID), following the anchor-based method [27, 75, 78]. By applying the anchor-based method, a clinician or researcher can immediately interpret if the change is clinically meaningful or not, whereas with the distribution method (calculation of SRM) only statistically significant results are obtained, without the possibility of clinically interpreting the change on the test scores [92, 93]. Furthermore, the distribution-method is based on the SDs of that population, making results insufficiently generalizable [93]. There is contradictory evidence on MCIDs [27, 29], because they may depend on the population of interest, the reference test (GAS and WeeFIM, both reliable, valid and responsive tests) or a combination of both. To summarize, reproducibility, SEMs and MCIDS should all be considered.

Feasibility

Administration times varied from less than 5 minutes for some of the single tests to 30 minutes for the ECAB [22, 71-75] and Kids-BESTest [3, 67]. This difference is related to the comprehensiveness of the test, therefore, more time should not be considered a limitation. All tests were explained and demonstrated in advance to the children, and a practice trial was often allowed to familiarize them with the test. This way the motor function was assessed rather than cognitive abilities [17]. Functional postural control tests are performed barefoot unless stated otherwise to represent balance performance as naturally as possible [94].

Study strengths and limitations

We systematically searched 3 databases using a comprehensive search query. Two independent reviewers assessed risk of bias and extracted data. The risk of missing potentially relevant articles was minimized by adding hand searching. The COSMIN checklist, recommended for evaluating methodology in psychometric studies, was applied to establish the level of evidence for each test [6, 7, 13]. Therefore, the poor-quality scores limit the interpretation of the results of the included studies [6, 7, 13]. The large diversity of available functional postural control tests, which leads to a variety of investigations of different measurement properties of the different tests, makes interpretation of test results challenging.

Recommendations for future research

It is crucial that future research focuses on exploring the structural validity of the most comprehensive test batteries in methodologically strong study designs to draw firm conclusions concerning the degree to which the scores of a test are an adequate reflection of the dimensionality of the construct to be measured. Currently, it is unknown if the multi-systemic framework relates to the underlying deficient neurological pathways. Hence, thorough structural validity assessment may help to disentangle whether all systems are being addressed as theoretically hypothesized [6, 7, 13], preferably combining functional assessment with brain imaging techniques. This research must include both TDC and mildly to severely affected pathological groups that could all benefit from physiotherapeutic treatment planning related to postural control. Age-differences should be considered. Second, responsiveness based on the anchor-based method of those tests that are structurally valid deserves attention.

Recommendations for clinical practice

The most appropriate functional postural control test should be selected considering: the entire framework of postural control, task-specificity, age-appropriate items, pathology-specific characteristics, psychometric properties, the level of evidence and feasibility.

Owing to a lack of methodologically strong research on the psychometric properties of functional postural control tests in children, the most appropriate functional postural control test can only be suggested cautiously. In line with Verbecque et al.[2], the PBS combined with the TUG can be used for children from the age of 4. The psychometric properties of both tests have been investigated in different populations and, when combined, they cover most systems of the multisystemic balance framework. However, reactive movement strategies are not evaluated with these tests. From the age of 8 years, both the FAB and Kids-BESTest are promising tools to comprehensively evaluate postural control in children, but with the currently available evidence, we cannot recommend one over the other. For children younger than 4 years, the ECAB is comprehensive and shows good psychometric properties from the age of 0.5 years, with an overall moderate level of evidence.

Conclusion

Validity results for the functional postural control tests emphasize that postural control is task-specific in healthy children but strongly relates to the severity of the underlying pathology. This underlines that postural control should be evaluated comprehensively with tests that cover its entire construct: movement strategies: APA and reactive postural responses, sensory strategies, orientation in space and control of dynamics. Hence, the use of tests assessing a single system should be avoided, and clinicians should choose tests that cover the entire postural control construct, differentiate between age-group and pathology-specific deficits, and have good psychometric properties. However, the available functional postural control tests show moderate to good psychometric properties, but the level of evidence is low and thus should be interpreted with caution. Moreover, structural validity research is presently lacking to draw firm conclusions on the most favorable functional postural control test. Currently, children with CP and TD are most extensively investigated. However, ideally, a functional postural control test should distinguish between different degrees of postural control performance (e.g., normal-mild-moderate-severe) allowing targeted identification of postural control deficits in a large variety of children. Based on current evidence, it seems that for children aged 4-8 years the PBS may be combined with the TUG, for children aged 8 years or older the Kids-BESTest or FAB are both promising, and for children below 4 years, the ECAB could be useful. Future research should focus on exploring structural validity and responsiveness with an anchor-based method in methodologically sound study designs including a variety of population types, from mild to moderate to severely affected children considering age-appropriate test items, covering the entire multisystemic framework.

Declaration of Competing Interest

None.

Acknowledgments

None.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rehab.2022.101729.

References

[1] Dewar R, Claus AP, Tucker K, Johnston LM. Perspectives on postural control dysfunction to inform future research: a Delphi study for children with cerebral palsy. Arch Phys Med Rehabil 2017;98(3):463–79.

[2] Verbecque E, Lobo Da Costa PH, Vereeck L, Hallemans A. Psychometric properties of functional balance tests in children: a literature review. Dev Med Child Neurol 2015;57(6):521–9.

[3] Dewar R, Claus AP, Tucker K, Ware R, Johnston LM. Reproducibility of the Balance Evaluation Systems Test (BESTest) and the mini-BESTest in school-aged children. Gait Posture 2017;55:68–74.

[4] Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med 2010;46(2):239–48.

[5] Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? Age Ageing 2006;35(Suppl 2):ii7–ii11.

[6] Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol 2020;20(1):293.

[7] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27(5):1147–57.

[8] Kiss R, Schedler S, Muehlbauer T. Associations between types of balance performance in healthy individuals across the lifespan: a systematic review and meta-analysis. Front Physiol 2018;9:1366.

[9] Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther 2009;89(5):484–98.

[10] Woollacott ASC, Majorie H. Motor control 2016.

[11] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

[12] Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. John Wiley & Sons; 2019.

[13] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27(5):1171–9.

[14] Aranha VP, Samuel AJ, Saxena S. Reliability and sensitivity to change of the timed standing balance test in children with Down syndrome. J Neurosci Rural Pract 2016;7(1):77–82.

[15] Malak R, Kostiukow A, Krawczyk-Wasielewska A, Mojs E, Samborski W. Delays in motor development in children with Down syndrome. Med Sci Monit 2015;21:1904–10.

[16] Verma A, Samuel AJ, Aranha VP. The four square step test in children with Down syndrome: reliability and concurrent validity. J Pediatr Neurosci 2014;9(3):221–6.

[17] Bandong AN, Madriaga GO, Gorgon EJ. Reliability and validity of the Four Square Step Test in children with cerebral palsy and Down syndrome. Res Dev Disabil 2015;47:39–47.

[18] De Kegel A, Dhooge I, Cambier D, Baetens T, Palmans T, Van Waelvelde H. Test-retest reliability of the assessment of postural stability in typically developing children and in hearing impaired children. Gait Posture 2011;33(4):679–85.

[19] De Kegel A, Dhooge I, Peersman W, Rijckaert J, Baetens T, Cambier D, et al. Construct validity of the assessment of balance in children who are developing typically and in children with hearing impairments. Phys Ther 2010;90(12):1783–94.

[20] Dannenbaum E, Horne V, Malik F, Villeneuve M, Salvo L, Chilingaryan G, et al. Vestibular assessments in children with global developmental delay: an exploratory study. Pediatr Phys Ther 2016;28(2):171–8.

[21] Yuksel E, Ozcan Kahraman B, Nalbant A, Kocak UZ, Unver B. Functional reach and lateral reach tests in Turkish children. Phys Occup Ther Pediatr 2017;37(4):389–98.

[22] Randall KE, Bartlett DJ, McCoy SW. Measuring postural stability in young children with cerebral palsy: a comparison of 2 instruments. Pediatr Phys Ther 2014;26(3):332–7.

[23] Deshmukh AA. Normal values of functional reach and lateral reach tests in children with knee hypermobility. Pediatr Phys Ther 2014;26(2):230–6.

[24] Martin K, Natarus M, Martin J, Henderson S. Minimal detectable change for TUG and TUDS tests for children with Down syndrome. Pediatr Phys Ther 2017;29(1):77–82.

[25] Hallemans A, Klingels K, Van Criekinge T, Vereeck L, Verbecque E. Reliability and concurrent validity of a modified timed up and go test for healthy preschoolers. Eur J Pediatr 2020;179(10):1579–86.

[26] Nicolini-Panisson RD, Donadio MVF. Normative values for the Timed 'Up and Go' test in children and adolescents and validation for individuals with Down syndrome. Dev Med Child Neurol 2014;56(5):490–7.

[27] Carey H, Martin K, Combs-Miller S, Heathcock JC. Reliability and responsiveness of the timed up and go test in children with cerebral palsy. Pediatr Phys Ther 2016;28(4):401–8.

[28] Baque E, Barber L, Sakzewski L, Boyd RN. Test-re-test reproducibility of activity capacity measures for children with an acquired brain injury. Brain Inj 2016;30(9):1143–9.

[29] Hassani S, Krzak JJ, Johnson B, Flanagan A, Gorton 3rd G, Bagley A, et al. One-Minute Walk and modified Timed Up and Go tests in children with cerebral palsy: performance and minimum clinically important differences. Dev Med Child Neurol 2014;56(5):482–9.

[30] Leizerowitz G, Katz-Leurer M. Feasibility, stability and validity of the four square step test in typically developed children and children with brain damage. Brain Inj 2017;31(10):1356–61.

[31] De Miguel-Etayo P, Gracia-Marco L, Ortega FB, Intemann T, Foraita R, Lissner L, et al. Physical fitness reference standards in European children: the IDEFICS study. Int J Obes 2014;38(Suppl 2):S57–66.

[32] Latorre-Román PA, Martínez-Redondo M, Párraga-Montilla JA, Lucena-Zurita M, Manjón-Pozas D, González PJC, et al. Analysis of dynamic balance in preschool children through the balance beam test: a cross-sectional study providing reference values. Gait Posture 2021;83:294–9.

[33] Latorre Román P, Mora López D, Robles Fuentes A, García Pinillos F. Reference values of static balance in Spanish preschool children. Percept Mot Skills 2017;124(4):740–53.

[34] Latorre Román P, Mora López D, Fernández Sánchez M, Salas Sánchez J, Moriana Coronas F, García-Pinillos F. Test-retest reliability of a field-based physical fitness assessment for children aged 3-6 years. Nutr Hosp 2015;32(4):1683–8.

[35] McKay MJ, Baldwin JN, Ferreira P, Simic M, Vanicek N, Burns J, et al. Reference values for developing responsive functional outcome measures across the lifespan. Neurology 2017;88(16):1512–9.

[36] Kim DH, An DH, Yoo WG. Reliability, standard error of measurement, and minimal detectable change of the star excursion balance test in children with cerebral palsy. J Back Musculoskelet Rehabil 2020;33(6):909–12.

[37] Schwiertz G, Brueckner D, Beurskens R, Muehlbauer T. Lower quarter Y balance test performance: reference values for healthy youth aged 10 to 17 years. Gait Posture 2020;80:148–54.

[38] Faigenbaum AD, Myer GD, Fernandez IP, Carrasco EG, Bates N, Farrell A, et al. Feasibility and reliability of dynamic postural control measures in children in first through fifth grades. Int J Sports Phys Ther 2014;9(2):140–8.

[39] Sharma K, Samuel AJ, Midha D, Aranha VP, Narkeesh K, Arumugam N. Multi-directional reach test in South Asian children: normative reference scores from 5 year to 12 years old. Homo 2018;69(1-2):62–9.

[40] Almutairi AB, Christy JB, Vogtle L. Psychometric properties of clinical tests of balance and vestibular-related function in children with cerebral palsy. Pediatr Phys Ther 2020;32(2):144–50.

[41] Kuo KT, Hunter BC, Obayashi M, Lider J, Teramoto M, Cortez M, et al. Novice vs expert inter-rater reliability of the balance error scoring system in children between the ages of 5 and 14. Gait Posture 2021;86:13–6.

[42] Hansen C, Cushman D, Chen W, Bounsanga J, Hung M. Reliability testing of the balance error scoring system in children between the ages of 5 and 14. Clin J Sport Med 2017;27(1):64–8.

[43] Hansen C, Cushman D, Anderson N, Chen W, Cheng C, Hon SD, et al. A normative dataset of the balance error scoring system in children aged between 5 and 14. Clin J Sport Med 2016;26(6):497–501.

[44] Tüzun EH, Levent E, Uzuner S, Sezerel B, Tomaç H, Mihçioglu S, et al. Validity and reliability of the four square step test in typically developed children. Türk Fizyoterapi ve Rehabilitasyon Dergisi.31(3):240–6.

[45] Latorre-Román PÁ, Consuegra González PJ, Martínez-Redondo M, Cardona Linares AJ, Salas-Sánchez J, Lucena Zurita M, et al. Complex Gait in preschool children in a dual-task paradigm is related to sex and cognitive functioning: a cross-sectional study providing an innovative test and reference values. Mind, Brain, Educ 2020;14(4):351–60.

[46] Itzkowitz A, Kaplan S, Doyle M, Weingarten G, Lieberstein M, Covino F, et al. Timed up and go: reference data for children who are school age. Pediatr Phys Therapy 2016;28(2):239–46.

[47] Bustam IG, Suriyaamarit D, Boonyong S. Timed Up and Go test in typically developing children: protocol choice influences the outcome. Gait Posture 2019;73:258–61.

[48] Panchal A, Tedla JS, Ghatamaneni D, Reddy RS, Sangadala DR, Alshahrani MS. Normative reference values for the timed up-and-go test in Indian children aged four to 11 years old and their correlation with demographic characteristics: a cross-sectional study. Niger J Clin Pract 2021;24(4):569–75.

[49] Humphriss R, Hall A, May M, Macleod J. Balance ability of 7 and 10 year old children in the population: results from a large UK birth cohort study. Int J Pediatr Otorhinolaryngol 2011;75(1):106–13.

[50] Atwater SW, Crowe TK, Deitz JC, Richardson PK. Interrater and test-retest reliability of two pediatric balance tests. Phys Ther 1990;70(2):79–87.

[51] Zaino CA, Marchese VG, Westcott SL. Timed up and down stairs test: preliminary reliability and validity of a new measure of functional mobility. Pediatr Phys Ther 2004;16(2):90–8.

[52] Gan SM, Tung LC, Tang YH, Wang CH. Psychometric properties of functional balance assessment in children with cerebral palsy. Neurorehabil Neural Repair 2008;22(6):745–53.

[53] Katz-Leurer M, Rotem H, Lewitus H, Keren O, Meyer S. Relationship between balance abilities and gait characteristics in children with post-traumatic brain injury. Brain Inj 2008;22(2):153–9.

[54] Katz-Leurer M, Rotem H, Lewitus H, Keren O, Meyer S. Functional balance tests for children with traumatic brain injury: within-session reliability. Pediatr Phys Ther 2008;20(3):254–8.

[55] Norris RA, Wilder E, Norton J. The functional reach test in 3- to 5-year-old children without disabilities. Pediatr Phys Ther 2008;20(1):47–52.

[56] Rajendran V, Roy FG, Jeevanantham D. Reliability of pediatric reach test in children with hearing impairment. Int J Pediatr Otorhinolaryngol 2012;76(6):901–5.

[57] Volkman KG, Stergiou N, Stuberg W, Blanke D, Stoner J. Methods to improve the reliability of the functional reach test in children and adolescents with typical development. Pediatr Phys Ther 2007;19(1):20–7.

[58] Bartlett D, Birmingham T. Validity and reliability of a pediatric reach test. Pediatr Phys Ther 2003;15(2):84–92.

[59] Deshmukh AA, Ganesan S, Tedla JS. Normal values of functional reach and lateral reach tests in Indian school children. Pediatr Phys Ther 2011;23(1):23–30.

[60] Yuksel E, Kahraman BO, Nalbant A, Kocak UZ, Unver B. Functional reach and lateral reach tests in Turkish children. Phys Occup Ther Pediatr 2017;37(4):389–98.

[61] Williams EN, Carroll SG, Reddihough DS, Phillips BA, Galea MP. Investigation of the timed 'up & go' test in children. Dev Med Child Neurol 2005;47(8):518–24.
[62] Iatridou G, Dionyssiotis Y. Reliability of balance evaluation in children with cerebral palsy. Hippokratia 2013;17(4):303–6.
[63] Held SL, Kott KM, Young BL. Standardized Walking Obstacle Course (SWOC): reliability and validity of a new functional measurement tool for children. Pediatr Phys Ther 2006;18(1):23–30.
[64] Lubetzky-Vilnai A, Jirikowic TL, McCoy SW. Investigation of the dynamic Gait index in children: a pilot study. Pediatr Phys Ther 2011;23(3):268–73.
[65] Darr N, Franjoine MR, Campbell SK, Smith E. Psychometric properties of the pediatric balance scale using Rasch analysis. Pediatr Phys Ther 2015;27 (4):337–48.
[66] Erden A, Acar Arslan E, Dündar B, Topbaş M, Cavlak U. Reliability and validity of Turkish version of pediatric balance scale. Acta Neurol Belg 2020;121(3):669–73.
[67] Dewar R, Claus AP, Tucker K, Ware RS, Johnston LM. Reproducibility of the kids-BESTest and the kids-MIni-BESTest for children with cerebral palsy. Arch Phys Med Rehabil 2019;100(4):695–702.
[68] Dewar RM, Tucker K, Claus AP, van den Hoorn W, Ware RS, Johnston LM. Evaluating validity of the Kids-Balance Evaluation Systems Test (Kids-BESTest) Clinical Test of Sensory Integration of Balance (CTSIB) criteria to categorise stance postural control of ambulant children with CP. Disabil Rehabil 2021:1–8 Advance online publication.
[69] Dewar RM, Tucker K, Claus AP, Ware RS, Johnston LM. Postural control performance on the functional reach test: validity of the Kids-Balance Evaluation Systems Test (Kids-BESTest) criteria. Arch Phys Med Rehabil 2021;102 (6):1170–9.
[70] Sim YJ, Kim GM, Yi CH. The reliability and validity of the Korean version of the Fullerton Advanced Balance scale in children with cerebral palsy. Physiother Theory Pract 2019;35(11):1087–93.
[71] Aras B, Seyyar GK, Kayan D, Aras O. Reliability and validity of the Turkish version of the Early Clinical Assessment of Balance (ECAB) for young children with cerebral palsy. J Dev Phys Disabil 2019;31(3):347–57.
[72] Gontijo APB, Starling JMP, Oliveira GD, Meier D, Mancini MC. Cultural adaptation and reliability analysis of the early clinical assessment of balance. Rev Paul Pediatr 2019;37(3):325–31.
[73] McCoy SW, Bartlett DJ, Yocum A, Jeffries L, Fiss AL, Chiarello L, et al. Development and validity of the early clinical assessment of balance for young children with cerebral palsy. Dev Neurorehabil 2014;17(6):375–83.
[74] Pierce SR, Kornafel T, Skorup J, Paremski AC, Prosser LA. Construct validity of the early clinical assessment of balance in toddlers with cerebral palsy: brief report. Dev Neurorehabil 2020;23(2):137–9.
[75] Pierce SR, Skorup J, Miller A, Paremski AC, Prosser LA. The responsiveness and validity of the Early Clinical Assessment of Balance in toddlers with cerebral palsy: brief report. Dev Neurorehabil 2019;22(7):496–8.
[76] De Kegel A, Baetens T, Peersman W, Maes L, Dhooge I, Van Waelvelde H. Ghent developmental balance test: a new tool to evaluate balance performance in toddlers and preschool children. Phys Ther 2012;92(6):841–52.
[77] Wright FV, Ryan J, Brewer K. Reliability of the Community Balance and Mobility Scale (CB&M) in high-functioning school-aged children and adolescents who have an acquired brain injury. Brain Inj 2010;24(13-14):1585–94.
[78] Chen CL, Shen IH, Chen CY, Wu CY, Liu WY, Chung CY. Validity, responsiveness, minimal detectable change, and minimal clinically important change of Pediatric Balance Scale in children with cerebral palsy. Res Dev Disabil 2013;34(3):916–22.
[79] Franjoine MR, Gunther JS, Taylor MJ. Pediatric balance scale: a modified version of the berg balance scale for the school-age child with mild to moderate motor impairment. Pediatr Phys Ther 2003;15(2):114–28.
[80] Her JG, Woo JH, Ko J. Reliability of the pediatric balance scale in the assessment of the children with cerebral palsy. J Phys Therapy Sci 2012;24(4):301–5.