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

Peer-reviewed author version

JOHNSON, Charlotte; Hallemans, Ann; GOETSCHALCKX, Mieke; MEYNS, Pieter; RAMECKERS, Eugene; KLINGELS, Katrijn & VERBECQUE, Evi (2023) Psychometric properties of functional postural control tests in children: A systematic review. In: Annals of Physical and Rehabilitation Medicine, 66 (4) (Art N° 101729).

DOI: 10.1016/j.rehab.2022.101729
Handle: http://hdl.handle.net/1942/39581

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

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

*corresponding author: e-mail: charlotte.johnson@uantwerpen.be

Abstract

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

Objective: The aim of this systematic review was therefore to facilitate the selection of an appropriate pediatric functional postural control test in research and clinical practice.

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

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

Conclusion: Currently, two functional tests comprehend the entire construct of postural control. Although reliability is overall good, validity results depend upon task, age and pathology. Future research should focus on test batteries, especially further exploring structural validity and responsiveness in various populations with methodologically strong study designs.

Keywords: "postural balance"[MeSH Terms], "psychometrics"[MeSH Terms], assessment, "child"[MeSH Terms], "child, preschool"[MeSH Terms])

Abbreviations

Abbreviation used in text	*Explanation*	*Language*
APA	Anticipatory postural adjustments	Eng.
BBS	Berg Balance Scale	Eng.
BBW	Balance Beam Walking	Eng.
BESS	Balance Error Scoring System	Eng.
BESTest	Balance Evaluation Systems Test	Eng.
BOT-2	Bruininks-Oseretsky Test for Motor proficiency, 2nd edition	Eng.
CB&M	Community Balance & Mobility Scale	Eng.
CGT	Complex Gait Test	Eng.
COSMIN	COnsensus-based Standard for the selection of health Measurement Instruments	Eng.
CP	Cerebral palsy	Eng.
CT-SIB	(Modified) Clinical Test of Sensory Interaction in Balance	Eng.
DGI	Dynamic Gait Index	Eng.
ECAB	Early Clinical Assessment of Balance	Eng.
FAB	Fullerton Advanced Balance Scale	Eng.
FRT	Forward Reach Test	Eng.
FSST	Four Square Step Test	Eng.
GAS	Goal Attainment Scale	Eng.
GDBT	Ghent Developmental Balance Test	Eng.
GMFM-66/88	gross motor function measure with 66/88 items;	Eng.

GRADE	Grading of Recommendations Assessment, Development and Evaluation	
ICC	intraclass correlation coefficient	Eng.
ICF	International Classification of Functioning, Disability and Health	Eng.
Kids-BESTest	Kids-Balance Evaluation Systems Test	Eng.
K_w	weighted kappa	Eng.
LoE	Level of Evidence	Eng.
LRT	Lateral Reach Test	Eng.
MABC(-2)	Movement Assessment Battery for children (2nd edition)	Eng.
MABC(-2) B	balance domain	Eng.
MCID	minimal clinically important difference	
MRT	Multidirectional Reach Test	Eng.
OLS	One-Leg-Stance	Eng.
PBS	Pediatric Balance Scale	Eng.
PDMS-2	Peabody Developmental Motor Scales, 2nd edition;	Eng.
PDMS-2 L/S	locomotion domain; stationary domain	Eng.
PEDI	Pediatric Evaluation Disability Inventory;	Eng.
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analyses	Eng.
PRT	Pediatric Reach Test	Eng.
RT	Reach Tests	Eng.
SBST	Stork Balance Stand Test	Eng.
SEBT	Star Excursion Balance Test	Eng.
SEM	standard error of measurement	Eng.

SRM	standardized response mean	
SOT	Sensory Organization Test	Eng.
SWOC	Standardized Walking Obstacle Course	Eng.
TDC	Typically developing children	Eng.
TS	Tandem-Stance	Eng.
TUDS	Timed Up and Down Stairs test	Eng.
TUG	Timed Up and Go test	Eng.
YBT	Y-Balance Test	Eng.

Introduction

Postural control deficits are one of the most common impairments in physiotherapeutic practice in a variety of pediatric populations, such as in cerebral palsy (CP), traumatic brain injury or developmental coordination disorder [1-3]. Due to its impact on the children's motor development and daily activities, the identification of postural control deficits is critical in order to plan 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 show consensus 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]. However, the theoretical construct of postural control still lacks consensus, despite the limited 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 lacking consensus resulted in a large diversity of functional pediatric postural control tests and a lacking gold standard [1, 4]. Thus, a guide for selecting the most appropriate functional postural control test is valuable, which is determined by 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]. Due to the task-specificity [8] and knowledge that different tasks tap into different systems [9, 10], the identification of these postural control systems based on task-type may aid us in understanding the underlying construct of the test [4, 5]. Tests comprising multiple tasks (test batteries) tend to evaluate multiple postural control systems. When more systems are covered, the closer a test comes to evaluating the entire construct of postural control, thereby potentially increasing the ability of the test to identify deficient underlying systems as opposed to tests assessing only one system [4, 5].

The quality of the test is determined by its psychometric properties, which refer to 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 the 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 on psychometric properties of the available functional pediatric postural control tests, revealing 25 studies, covering 14 different functional tests. Overall, structural validity and responsiveness of these tests were underexposed. Since then, studies on this research topic 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 is to facilitate the selection of an appropriate pediatric functional postural control test by mapping its psychometric properties and **feasibility**. This aim is obtained by answering following research questions: 1) What are the existing psychometric properties of each postural control test? 2) What are the underlying systems evaluated in the test regarding the multisystemic framework [5] and 3) what are the feasibility features of each test.

Methodology

This systematic review is 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 showed either typical development or had 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 of 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 two independent reviewers (CJ, EV) in the same sequence: population, intervention, outcome, study design, language. Studies were selected in two 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). The original search strategy was used as [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 on eligibility.

Data collection process and data items

Data from each individual study, initially [2] and the newly included references, were extracted by two independent reviewers (CJ, EV) . Each reviewer extracted data from half of the included papers 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, sex distribution.
2. ***Assessment characteristics***: name of the functional test, the test items included and whether it intends 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 one task, the dominant system was identified and classified as such.
3. ***Psychometric properties*** were extracted 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 is summarized in tables.

For each test an overall judgement is indicated in tables with color coding in line with the COSMIN criteria for good psychometric properties [6, 7, 13]. These judgements 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 given if correlations were ≥0.70 and “-” of <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 do not carry a specific symbol or color [6, 7, 13].

Study quality was assessed by determining risk of bias using the COSMIN checklist. Afterwards the quality of results per 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 per functional postural control tests 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 the 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, MG) assessed risk of bias: MG scored all studies and CJ and EV each scored 50%. Two independent reviewers (CJ, EV) performed grading. Consensus was reached in a consensus meeting. Appendix B provides details on how COSMIN and GRADE 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 seven test batteries and 19 tests covering one dominant system. Twelve tests were new since 2014. Twelve tests were new since 2014. Appendix D provides a detailed description of each test. Table 2 proves an overview of the reliability results and Table 3 the validity results. Figure 2 depicts the level of evidence and an overall quality of results per functional postural control test. **Feasibility** features are presented in Table 4. Figure 1 summarizes the postural control systems assessed in each test.

Population characteristics

Populations investigated vary 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 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 samples were reported less frequently, such as Down syndrome [14-17] or hearing impairment [18, 19] or 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 of the Pediatric Reach Test (PRT) [21-23], the Timed Up Down Stairs test (TUDS) [24], the Timed Up and Go (TUG) [17, 24-30] and the Balance Beam Walking (BBW). Additionally, nine 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 records 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 dominant system

The following tests mainly assess APA (Figure 1) by recording the time children maintain tandem-stance (TS) [18, 19, 49], one-leg-stance (OLS) [18, 19, 49, 50] or variant like 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 due to a narrowed base of support and SEBT/YBT also evaluate orientation in space (Figure 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 for 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 records 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 for TDC between 7-11 years [38] and 10-17 years [37] (Table 4).

Tests assessing orientation in space as dominant system

The functional ***Reach Tests*** *(RT)* and its variants primarily evaluate orientation in space (Figure 1) by estimating the child's maximum stability limits. Additionally, APA and control of dynamics are required (Figure 1). Fourteen studies comprised the different variants of RT [22, 23, 39, 51-60], four 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 dominant system

Since 2014, two tests found their way into 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 low for children showing 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 specificity of 43% in children with CP [40] (Table 3). The ***BESS*** showed good test-retest, intra- and interrater reliability [41, 42] (Table 2). Reference values are 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 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 records 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 DS [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 low, indicating the FSST measures a different construct [30] (Table 3).

One new record published on the **TUDS** [24] showing 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 poor 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] CP children. Concurrent validity between TUG versus FSST and TUDS are discussed elsewhere. 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 during which the child walks at maximum speed [45]. The CGT has good test-retest reliability (Table 2), but poor 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 poor 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 three new records on the previously reported Pediatric Balance Scale (PBS) [15, 65, 66] and three 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 papers 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], six new papers 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 lesser extent sensory orientation (Figure 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 is present [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 poor 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 GMFCS levels I-III from each other using the PBS [82, 84] and from TDC [84] (Table 3). Next to the Brazilian version [81], a Korean [70, 82], a Persian [85] and a Turkish version is available now as well [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 expect for the orientation in space (Figure 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 distinguish children with CP from those with typical development [74] (Table 3). Children with CP with a Gross Motor Function Classification System (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 (Figure 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 two 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 (Figure 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 four 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 so far 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 are overall characterized by very low (red color) or low (pink color) level of evidence (Figure 2), which means that the true measurement property is likely to 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 in-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 severity of the risk of bias in validity studies increased due to a small study sample size (4 studies), insufficient description of the comparator (6 studies) or the lacking justification of choice of statistical analysis (6). All included functional postural control tests, except the Flamingo test, were investigated at least once for *reliability* (Figure 2). Most ICCs or K_w reached the 0.70 criterion or more except for TS and BBW due 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 very low (7/16 tests), low (6/16 tests) or moderate (3/16 tests) body of evidence (Figure 2). The PBS, ECAB and TUG were most extensively investigated, nevertheless, the overall evidence for their psychometric properties is rather low. Only responsiveness of the TUG shows high evidence. The ECAB shows high evidence on its internal consistency, but results should be interpreted cautiously as no studies on its structural validity are reported yet. Exclusively for the PBS qualitative research is available on its structural validity and known-group validity (Figure 2).

Discussion

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.**

Twenty-six functional postural control tests were identified, meaning that 12 tests were newly developed since 2014 [2]. Studies overall showed a large variety considering the types of functional postural control tests (one vs. multiple systems), and the investigated psychometric properties, population, and age-ranges. Likewise, assessment time, test protocols and required equipment varied frequently. Currently there are reference data 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 the measurement error and concurrent validity were reported frequently but not for all (Figure 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. Exclusively for structural validity and known-group validity of the PBS conclusive results with high evidence were found. 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 should be covered by its assessment tools. It was assumed that test batteries approximate this theoretically (Figure 1), but only two 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 understand the underlying constructs of the test which determines the content validity.

In these 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 (Figure 1), while the PBS does not assess reactive postural responses. Nevertheless, not all supposed systems (Figure 1) were translated into actual dimensions statistically [70]: all tasks of the PBS belong to the same dimensions [65], and the FAB appeared consists of two dimensions [70]. thus, these dimensions do not reflect the multiple systems needed for postural control, indicating other factors may be in play like the included populations.

Previous explorative research showed consistent findings regarding unidimensionality and **task-specificity** in studies including **healthy individuals** [8, 65, 86] or in heterogeneous pathological populations [65, 87], proven by small-sized 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 small-sized 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 upon the child's developmental stage [86], which is invigorated by the availability of various age-norms (Table 4). For example, for BBW significant differences were found between the ages 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.**

Contrarily, the dimensionality investigated with structural validity analysis changes in homogeneous pathological populations [70, 88-91]. In children with **CP**, the FAB consists of two 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 two 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: one dimension for the FAB with changed item-hierarchy in individuals with stroke [89], four dimensions for the BESTest in individuals with stroke [90] and three 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 upon the severity of the pathology. Healthy children indeed 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 also among children with different functional levels, such as GMFCS levels I-III (CP) with PBS endorsing these tests' *known-group validity* [82, 84] (Table 3). Furthermore, children with CP [15, 52, 82] show higher correlations between functional postural control tests and GMFM total scores compared to 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 inferior to 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) poor. Overall, (very) serious risk of bias was the main cause of its low evidence, but inconsistency, shown by either conflicting results (Appendix E) or very wide confidence intervals, and imprecision caused by small sample sizes played a role as well. 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 changed score can only be attributed relatively to the amount of error, therefore lacking the clinical meaningfulness of the change (responsiveness). Hence, COSMIN guidelines prescribe 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 only available records on the ECAB [75], PBS [78] and TUG [27, 29]. Determining a test's 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, while with the distribution-method (calculation of SRM) only statistically significant results are obtained, without the possibility to clinically interpret 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 upon 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], which is related to the comprehensiveness of the test. Therefore, more time should not be considered as a limitation. All tests were explained and demonstrated in advance to the children, and a practice trial is often allowed to familiarize 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 their balance performance as naturally as possible [94].

Study strengths and limitations

We search three databases systematically 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 indeed limit a good 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 make 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 indeed relates to the underlying deficient neurological pathways. Hence, thorough structural validity assessment may aid in disentangling whether indeed 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

Summarized 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, its body of evidence and its **feasibility**.

Due to 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. Both tests were psychometrically investigated in different populations and combined comprise most systems of the multisystemic balance framework. Still, the 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 evaluate children's postural control comprehensively, but with the present evidence we cannot recommend one over the other. For children younger than 4 years, the ECAB is comprehensive and shows good psychometric properties already from the age of 0.5 years with an overall moderate body of evidence.

Conclusion

Validity results of 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 emphasizes that postural control should be evaluated comprehensively covering 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 one system should be avoided, and clinicians should choose tests considering the entire postural control construct, age-differences, pathology-specific deficits and good psychometric properties. However, the available functional postural control tests show moderate to good psychometric properties, but level of evidence is poor and thus should be interpreted with caution. Moreover, structural validity research is presently lacking to make firm conclusions on the most favorable functional postural control test. Presently children with CP and TD are most extensively investigated. However, ideally, a functional postural control test is able to distinguish between different degrees of postural control performances (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 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 the 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.

Acknowledgements: none

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

Declaration of interest: none

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 AS-C, Majorie H. Motor control2016.

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. Physical Therapy. 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. Developmental Medicine and Child Neurology. 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 G, 3rd, 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 (Lond). 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. TÜZÜN EH, Levent E, UZUNER S, SEZEREL B, TOMAÇ H, MIHÇIOĞLU 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, and Education. 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. Pediatric Physical 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. Neurorehabilitation and 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 Injury. 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. International Journal of Pediatric Otorhinolaryngology. 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. Physical & Occupational Therapy in Pediatrics. 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. Developmental Medicine and Child Neurology. 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;Advance online publication:1-8.

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 Msc PT, Kim Gm PhD PT, Yi Ch PhD PT. 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. Journal of Developmental and Physical Disabilities. 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. Physical Therapy. 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 Injury. 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. Research in Developmental Disabilities. 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. Journal of Physical Therapy Science. 2012;24(4):301-5.

81. Ries LG, Michaelsen SM, Soares PS, Monteiro VC, Allegretti KM. Cross-cultural adaptation and reliability analysis of the Brazilian version of Pediatric Balance Scale (PBS). Rev Bras Fisioter. 2012;16(3):205-15.

82. Yi SH, Hwang JH, Kim SJ, Kwon JY. Validity of Pediatric Balance Scales in Children with Spastic Cerebral Palsy. Neuropediatrics. 2012;43(6):307-13.

83. Franjoine MR, Darr N, Held SL, Kott K, Young BL. The performance of children developing typically on the pediatric balance scale. Pediatr Phys Ther. 2010;22(4):350-9.

84. Franjoine MR, Darr N, Young B, McCoy SW, Fiss AL. Examination of the effects of age, sex, and motor ability level on balance capabilities in children with cerebral palsy GMFCS levels I, II, III and typical development using the Pediatric Balance Scale. Dev Neurorehabil. 2022;25(2):115-24.

85. Alimi E, Kalantari M, Nazeri AR, Akbarzade Baghban A. Test-retest & Inter-rater Reliability of Persian Version of Pediatric Balance Scale in Children with Spastic Cerebral Palsy. Iran J Child Neurol. 2019;13(4):163-71.

86. Verbecque E, Klingels K, Rameckers E, Ferguson G, Smits-Engelsman B. The construct of balance control in primary school-aged children: Unidimensional and task-specific. Hum Mov Sci. 2021;79:102847.

87. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med. 2010;42(4):323-31.

88. Klein PJ, Fiedler RC, Rose DJ. Rasch Analysis of the Fullerton Advanced Balance (FAB) Scale. Physiother Can. 2011;63(1):115-25.

89. Jeon Y-j, Kim G-m. Rasch analysis of the Korean version of the Fullerton Advanced Balance scale. Physical Therapy Korea. 2017;24(4):20-8.

90. Miyata K, Hasegawa S, Iwamoto H, Otani T, Kaizu Y, Shinohara T, et al. Structural validity of Balance Evaluation Systems Test assessed using factor and Rasch analyses in patients with stroke. J Phys Ther Sci. 2018;30(12):1446-54.

91. Benka Wallén M, Sorjonen K, Löfgren N, Franzén E. Structural Validity of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in People With Mild to Moderate Parkinson Disease. Phys Ther. 2016;96(11):1799-806.

92. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102-9.

93. Sedaghat AR. Understanding the Minimal Clinically Important Difference (MCID) of Patient-Reported Outcome Measures. Otolaryngol Head Neck Surg. 2019;161(4):551-60.

94. Reutimann S, Hill-Strathy M, Krewer C, Bergmann J, Müller F, Jahn K, et al. Influence of footwear on postural sway: A systematic review and meta-analysis on barefoot and shod bipedal static posturography in patients and healthy subjects. Gait & Posture. 2022;92:302-14.

76. 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.

77. 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.

78. 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.

79. 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.

80. 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 Injury. 2010;24(13-14):1585-94.

Tables

Table 1: definitions of psychometric properties according to COSMIN taxonomy

Table 2: summary of relative reliability and measurement error per functional postural control test

Table 3: summary of structural, cross-cultural, concurrent and construct validity per functional postural control test

Table 4: feasibility features of the functional postural control tests

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

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

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)[28], CP (14, 7-12)[49]			**[28, 49]**
	(m)CTSIB$_{OLS}$	DS (9, 8-17)[16]			*SEM: 0.21-0.59*
	BESS	TDC (381, 5-14)[50, 51]	MDC: 4.58[51]	MDC: 9.57[51] **[50, 51]**	MDC: 7.33 [51]
	$TS_{on\ beam}$	TDC (237, 10)[58]			**[58]**
	SBST	TDC (90, 3-6)[43]			0.99-1.00 [43]
	OLS	TDC (294, 4-12)[20, 58, 59], TDC&CP (25, 8-14)[60], HI (23, 6-12)[20]	**[60]**	SEM: 2.63 [60] **[59]**	SEM TDC: 10.16-13.37[20], SEM HI: 8.71-8.83[20] **[58, 59]**
	YBT	TDC (188, 7-11)[47]		ME: 2.68-3.13 [47]	ME: 16.41 [47]
	SEBT	CP (8, 6-12)[45]	**[45]**	SEM: 2.63 [45]	
	PRT	CP (38, 2-12)[34, 65]	**[65]**	**[65]**	SEM: 16.8 [34]
	FRT	TDC (93, 7-16)[24, 64], TDC&CP (25, 8-14)[60], TBI (24, 7-14)[24], HI (65, 6-11)[63], CP (22, 5-12)[61]	SEM TDC : 1.41 [24], SEM TBI : 0.97 [24], SEM HI: 0.29-0.51 [63] **[60]**	**[60, 61, 63]**	**[61, 64]**
	LRT	TDC (24, 7-14)[24], TBI (24, 7-14)[24] HI (65, 6-11)[63]	SEM TDC: 0.80-0.97 [24], SEM TBI: 0.72-0.90 [24], SEM HI: 0.28-0.32[63]	**[63]**	
	FSST	TDC (179, 5-12)[25, 53], CP (16, 5-12)[19], DS (27, 5-17)[18, 19]	**[53]**	ME TDC:-1.11-0.87 [53] **[18, 19]**	SEM TDC: 0.96-0.98[25], SEM CP: 1.34[19], SEM DS: 2.32[19] **[18]**
	TUDS	TDC&CP (25, 8-14)[60], DS (8, 3-17)[35]	**[60]**	**[60]**	MDC DS: 12.52 [35] **[60]**
	TUG	TDC (226, 3-14)[24, 29, 36], TDC&CP (25, 8-14)[60], ABI (54, 7-16)[24, 26], CP (95, 3-14)[38, 61, 68], CP&BD (41, 3-19)[29], DS (12, 3-17)[35]	SEM TDC: 0.60 [24], SEM ABI: 0.23 [24] **[29, 36, 60]**	SEM TDC: 0.14-0.15 [36] **[60]**	SEM TDC: 0.67-0.83 [36], SEM CP: 0.46; 0.42[26], SEM CP: 0.51-3.15 [38, 61, 68], MDC DS: 1.26 [35] **[24, 29]**
	SWOC	TDC (50, 4-11)[30], DD (23, 6-21)[30]	**[30]**	**[30]**	
	CGT	TDC (90, 3-6)[54]			**[54]**
	DGI	TDC&FASD (11, 8-15)[27]			**[27]**
	BBW	TDC (601, 3-6; 237, 10)[41, 54, 58]			SEM: 0.35-4.01 [41] **[54, 58]**
MULTIPLE SYS	**CB&M**	ABI (32, 7-18)[80]	SEM: 3.7§[80]	SEM: 4.8; 3.9[80]	Sem: 5.8, 5.6[80]
	GDBT	TDC (144, 1.5-6)[32], MR (22, 1.5-6)[32]		SEM MR: 0.78	SEM TDC: 0.21
	PBS	TDC (40, 5-7)[31], CP (146, 5-13)[82, 83, 87], BD (34, 4-18)[70], MI (20, 5-15)[31]	SEM CP: 0.37-0.43[82] **[70, 83]**	SEM CP: 0.65[82]; 1.78-1.80[87] **[70, 83]**	SEM CP: 0.61[82]; 1.79[87] **[31]**
	BBS	CP (50, 5-14) [61, 68]	SEM: 0[68]	**[61, 68]**	SEM: 0.18-0.22[68] **[61]**
	ECAB	CP (575, 1-12)[34, 75, 76, 78]	**[76]**	**[34, 76]**	SEM: 0-3.6 [34, 75]
	FAB	CP (40, 5-16)[74]			**[74]**

T E M S	**Kids-BESTest**	TDC (34, 7-17)[3], CP (18, 8-17)[71]	SEM TDC - Full/Mini: 0.81/0.54 [3] SEM CP – Full/Mini: 1.98/0.88 [71]	SEM TDC – Full/Mini: 1.45/0.96 [3]; SEM CP – Full/Mini: 3.08/1.20 [71]	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 [71]

Legend: Underlined = video recordings; 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

Interpretation: bold references: reliability was calculated, but no measurement error was analyzed. Light grey cells: the majority of the findings showed an ICC/κw≥0.70; dark grey cells: the majority of the findings showed showed ICC/κw<0.70; medium grey cells: there was no majority making conclusions indeterminate; empty cells: property was not investigated

Table 3: summary of structural, cross-cultural, concurrent and construct validity per 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)[49]			SOT (-)		
	$TS_{on\ beam}$	TDC (237, 10)[58]			**TS_{EO}/TS_{EC}** vs OLS_{EO}/OLS_{EC} (-)		
	OLS	TDC (80, 6-14) [21, 60], HI (23, 6-12) [21], CP (20, 8-14) [60]			**$OLS_{EO/EC}$** vs $COP_{(F)EO/(F)EC}$, COP_{TSvel}, COP_{SLSvel} (±), TUDS (±)		HI<TDC***
	PRT	CP (38, 2-12)[34, 65]			COP_{AP-ML} (±)[65], ECAB (+) [34]	GMFM-66 B&C (+) [34]	
	FRT	TDC (27, 8-14) [60], CP (50, 5-14)[60, 61], TBI (24, 7-14)[23], DS (13, 8-17) [18]			TUG (+)[61], TUDS (-)[60], FSST (-)[18], BBS (+)[61]	GMFM-88 (D-E) (-)[61], STS (-)[61], Walking speed (-)[61], Step length (±)[23]	GMFCS III<II<I **[61], TBI<TDC***[23]
	LRT	TBI (24, 7-14)[23]				Step length (-)	TBI<TDC***
	FSST	TDC (30, 6-12)[25], CP (36, 6-12)[19, 25], DS (27, 5-17)[18, 19]			TUG (±)[19, 25], FRT (-)[18], BOT-2 (-)[25]		CP=TBI<TDC*[25]
	TUDS	TDC (27, 8-14)[60], CP (20, 8-14)[60],			TUG (±), FRT (±), OLS_{EO} (±)[60]		GMFCS II/III>I>TDC[60]
	TUG	TDC (112, 3-12)[25, 30, 36], TBI (24, 7-14)[23], CP (66, 5-14)[19, 39, 60, 61], CP&BD (41, 3-19)[29], DS (14, 6-12)[19], DD (23, 6-21)[30]			MABC-2 B (-)[36], TUDS (±) [60], FSST (±)[19, 25], SWOC (+) [30], BBS (+) [61], FRT (+)[61]	Step length (+) [23], GMFM-88 (D-E) (-) [61], STS (-)[61], Walking speed (+) [61]	TBI<TDC***[23] III>II>I**[61]; III>II/I**[39]; NS[38];
	SWOC	TDC (50, 4-11)[30], DD (23, 6-21)[30]			TUG (+)		
	CGT	TDC (80, 3-6)[54]			BBW (-)		
	DGI	TDC&FASD (20, 8-15)[27]					FASD<TDC**
	BBW	TDC (593, 3-6)[41], TDC (237, 10)[58]			SBST (-)[41], $OLS_{EO/EC}$ (-), $TS_{beamEO/EC}$ (-)[58]		
MULTIP	**GDBT**	TDC (28, 1.5-6)[32], MR (20, 1.5-6)[32]			PDMS-2 S-L (-), MABC B (-)	PDMS-2 (-), MABC (-), BOT-2 (-)	MR<TDC***
	PBS	TDC (258, 2-4) [86] CP (342, 1-16) [74, 81, 84, 86], BD & TDC (138 & 685, 2-13)[69], BD (34, 4-18)[70], DS (44, 2-10)[17]	Uni-dimensional[69]	[70]	FRT (+) [70], SOT (-)[84], FAB (-)[74]	GMFM-66 (+)[81, 84], WeeFIM (+)[81, 84], GMFM-88(D-E) (+)[17, 84], PEDI-mobility (+)[84]	GMFCS I>II>III**[84] TD>GMFCS I>II>III* [86]

LE SYSTEMS	**BBS**	CP (30, 5-12)[61]			FRT (+), TUG (+)	GMFM-88(D-E) (-), STS (-), Walking speed (-)	GMFCS I/II>III>VI*
	ECAB	CP (575, 1-12) [34, 75-78] CP (37, 0.5-3)-TDC (13, 0.5-3) [79]		[77]	PRT (+)[34]	GMFM-66 (+) [79], GMFM-66-B&C (+)[77], GMFM-88(A-B-C-D-E) (±)[75, 79]	CP<TDC [78]; GMFCS I>III/IV/V***; II>V**[75]; I>II>III>IV>V***[77]; II>IV*; III>IV**[78]
	FAB	CP (40, 5-16)[74]	Two-dimensional [74]	[74]	PBS (-)[74]		
	Kids-BESTest	TDC (41, 7-18) [72, 73], CP (17, 7-18) [72, 73]			**FRT/LRT vs COP measures,** Δ trunk, knee, ankle angle (-)[73]; **mCTSIB: FEC vs COP measures** (-)[72]		

Legend: 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, Δ: change.

Interpretation: * $p<0.05$; ** $p<0.01$; *** $p<0.001$; [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 (±).

Table 4: feasibility features of the functional postural control tests									
	Test	Population	Age (years)	Reference values	Cross-cultural adaptation	Administration time (minutes)	Equipment	Outcome measure (s)	Price *
	(m)CT-SIB [16, 28, 49]	TDC [28]	4-12			20	medium density foam pad; stopwatch [16, 28, 49]; <u>optional</u>: *visual conflict dome* [28]	time	free
		GDD [28]	4-12	X [28]					
		CP [49]	7-12						
		DS [16]	8-17						
	BESS [50-52]	TDC [50-52]	5-14	X [52]		15	medium density foam pad; stopwatch; score card [50-52]	score	free
	TS [44, 58]	TDC [44, 58]	3-10	X [44]		<5	tape; stopwatch [44, 58]; <u>optional</u>: *visual target on wall; balance beam (2x0.05m) [58];*	time	free
	Flamingo test [40]	TDC [40]	6-10	X [40]		<5	stopwatch [40]	# attempts [40]	free
	SBST [41-43]	TDC [41-43]	3-6	X [41]		<5	stopwatch [41-43]	time	free
	OLS$_{EO/EC}$ [20, 21, 44, 58-60]	TDC [20, 21, 33, 44, 58-60]	3-12			<5	stopwatch; [20, 21, 44, 58-60] <u>optional</u>: *tape; visual target on wall [44, 59, 60]*	time	free
		HI [20, 21]	6-12						
		CP [60]	8-14						
	SEBT [44, 45] / **YBT** [46, 47]	TDC [44, 46, 47]	3-19	X [44]		10	PVC pipe; platform [46, 47]; <u>or</u> tape [44, 45]; length measure [44-47]; footwear [47] or barefoot [44-46]; <u>optional</u>: *plastic alligator toys [47]*	distance [45, 47]; % of leg length [44, 46]	free
		CP [45]	6-12						
ONE SYSTEM	**PRT** [22, 33, 34, 65, 66] / **FRT** [18, 23, 24, 60-64] / **LRT** [23, 24, 63] / **MRT** [48]	TDC [23, 24, 33, 48, 60, 62, 64-66]	2-18	X (3-12y) [33]		5-10	length measure, yardstick or ruler [18, 22-24, 33, 48, 60-66]; <u>optional</u>: *stool; score sheet [63, 65]*	distance	free
		CP [61]	2-18						
		knee hypermobility [22]	6-12	X [22]					
		HI [63]	6-11						
		TBI [23, 24]	7-14						
		DS [18]	8-17						
	FSST [18, 19, 25, 53]	TDC [25, 53]	5-12			<5	canes/rods (4x); stopwatch [18, 19, 25, 53]; footwear [19, 53] or barefoot [18]; <u>optional</u>: *non-skid rubber mat; visual target on wall; red sequence-numbers in squares [19]*	time	free
		CP [19, 25]	5-12						
		ABI [25]	6-12						
		DS [18, 19]	6-17						
	TUDS [35, 60]	TDC [60]	8-15			<5	14-steps flight stairs with handrails; regular footwear; stopwatch [35, 60]	time	free
		CP [60]	8-15						
	TUG [19, 23-26, 29, 30, 35-39, 55-57, 60, 61, 68]	TDC [23-25, 29, 36, 37, 55-57, 60, 68]	3-19	X [37, 55-57]		5	(adjustable) chair; length measure; tape; stopwatch [23-26, 29, 30, 35-39, 55-57, 60, 61, 68]; regular footwear [23, 24, 35-37, 39, 55] or barefoot [60]; <u>optional</u>: *tapeline [23, 24, 26, 30, 39, 55, 56, 60, 68]* [30]*, cone [56], target on the wall [29, 35, 37, 39, 56, 57] or Duplo brick for transportation [36]*	time	free
		CP [19, 38, 39, 60, 61, 68]	3-19						
		ABI [26]	8-16						
		TBI [23, 24]	7-14						
		CP&BD [29]	3-19						
		DD [30]	6-21						
		DS [19]	6-12						
	SWOC [30]	TDC [30]	4-11			15	free walkway; axillary crutch; visually stimulating mat; trash can; shag rug; chair with(out) armrest (2x); tray; glasses; stopwatch [30]	time; steps	free
		DD [30]	6-21						
	CGT [54]	TDC [54]	3-6	X [54]		<5	free walkway; tape; length measure; cones (≥11); stopwatch [54]	time	free
	DGI [27]	TDC [27]	8-15			15	free walkway; shoe box; cones (2x); stairs [27]	Criterion scores	free
		FASD [27]	8-15						

	BBW [41, 54, 58]	TDC [41, 54, 58]	3-6 [41]; 10 [58]	X		<5	balance beam (2.5x0.04x0.12 m) [41, 54, 58]; stopwatch [41, 58]; length measure [41, 54]	time [41, 58]; distance [41, 54]; steps [41]	free
MULTIPLE SYSTEMS	**CB&M** [80]	ABI [80]	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 footwear; stopwatch [80]	Criterion scores	free
	PBS [17, 31, 69, 70, 81-86]	TDC [31, 85, 86]	2-13	X [85, 86]	Brazilian [83], Turkish [70] Korean [74, 84] Persian [87]	10-20	Pediatric version of the BBS. adjustable bench; chair with back support and arm rests; step; chalkboard eraser; yardstick; small level [17, 31, 69, 70, 81-85]; <u>*optional*</u>*: flash cards; blindfold, footprints; visual colored target; Velcro [70]*	Criterion scores	free
		MI [31]	5-15						
		DS [17]	2-10						
		BD [69, 70]	2-18						
		CP [74, 81-84, 86, 87]	2-16	X [86]					
	BBS [61, 68]	CP [61, 68]	6-14			10-20	2 standard chairs, with back support, one with arm rests, one without; step; a ruler; stopwatch [61, 68]	Criterion scores	free
	GDBT [32]	TDC [32]	1.5-6	X [32]		20	GDBT manual; thin mat; tape; ball; medium density foam; stopwatch; scoring sheet [32]	Criterion scores	€ 24.99
		MR [32]	4-5						
	ECAB [34, 75-79]	TDC [78]	0.5-3		Turkish [75]	30	bench; step; mat; stopwatch; test form [34, 75-79]	Criterion scores	free
		CP [34, 75-79]	0.5-12						
	FAB [74]	CP [74]	5-16		Korean	10-12	bench; medium density foam (2x); length measure; pencil; stopwatch [74]	Criterion scores	free
	Kids-BESTest [3, 71-73]	TDC [3, 72, 73]	7-18			30	free walkway; tape; blindfold; length measure; medium density foam; incline ramp; bench; shoe box; weight (1kg); adjustable chair; stopwatch [3, 71]	Criterion scores	free
		CP [71-73]	7-18						

<u>**Legend**</u>: * prices refer to manual. 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
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.

Figures

Figure 1: postural control systems per test

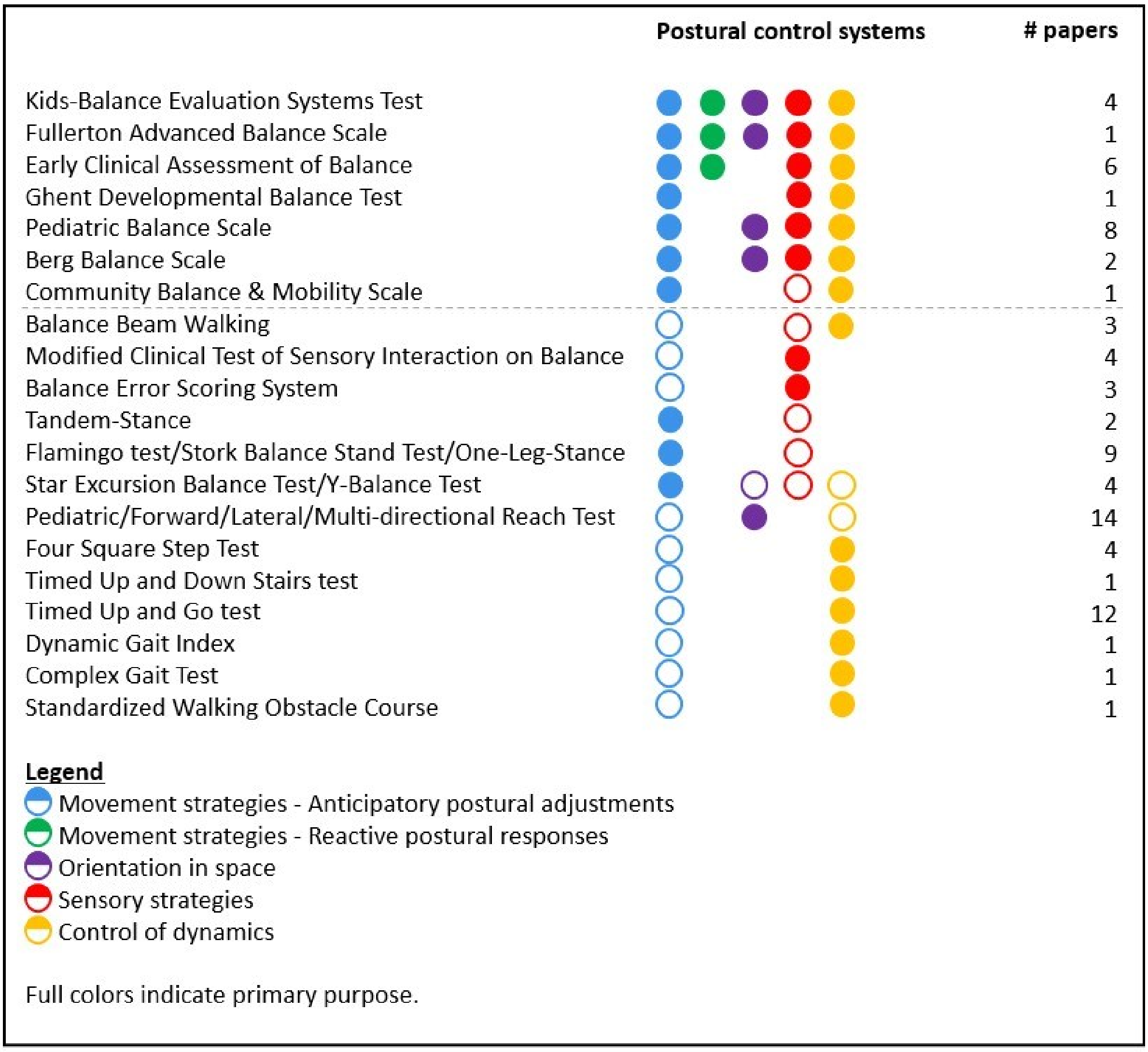

Figure 2: level of evidence (GRADE) and quality of measurement properties rated against COSMIN criteria

Level of evidence (GRADE) and quality of measurement properties rated against COSMIN criteria										
	Test	Reliability		Re-sponsiveness	Validity					
		Reliability	Measurement error		Structural	Internal consistency	Concurrent (comparator)	Hypothesis testing (comparator)	Known group (comparison)	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	+	?				-		+	

Appendices

Appendix A: search strategy for Pubmed, Web of Science and Scopus

Appendix B: Methodology of quality assessment of studies and results

Appendix C: flowchart of the selection process

Appendix D: description of the functional postural control tests

Appendix E: GRADE rating

Appendix F: risk of bias assessment using the COSMIN checklist

Appendix G: reference list remaining reference