Proxy-reported sensory measures for children and adolescents with neurodevelopmental disorders: A systematic review

Nicolette Soler[1,2] | Reinie Cordier[3,4] | Iain E. Perkes[1,5,6] | Russell C. Dale[2,7,8] | Paula Bray[9]

[1]School of Health Sciences, University of Newcastle, Newcastle, NSW, Australia

[2]Children's Hospital Westmead Clinical School, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

[3]Curtin School of Allied Health, Faculty of Health Sciences, Curtin University, Perth, WA, Australia

[4]Department of Social Work, Education and Community Wellbeing, Northumbria University, Newcastle, UK

[5]Discipline of Psychiatry & Mental Health and Discipline of Paediatrics & Children's Health, School of Clinical Medicine, University of New South Wales Medicine & Health, University of New South Wales, Sydney, NSW, Australia

[6]School of Women's and Children's Health, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia

[7]Department of Paediatric Neurology, The Children's Hospital at Westmead, Sydney, NSW, Australia

[8]Kids Neuroscience Centre, and Brain and Mind Centre, University of Sydney, Sydney, NSW, Australia

[9]Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

Correspondence
Paula Bray, The University of Sydney and Sydney Children's Hospitals Network, Locked Bag 4001, Westmead, 2145, Sydney, NSW, Australia.
Email: paula.bray@health.nsw.gov.au

Funding information
National Health and Medical Research Council; Petre Foundation; Petre Foundation; Indiana University; University of Sydney; Sydney Medical School

Abstract

AIM: To determine the quality and utility of proxy-reported sensory measures for children and adolescents with neurodevelopmental disorders (such as autism spectrum disorder, attention-deficit/hyperactivity disorder, movement disorders, and intellectual disability).

METHOD: We systematically searched 11 databases. We applied the updated Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist and criteria for good measurement properties to evaluate instrument development and psychometric properties. Findings were summarized using a COSMIN adaptation of Grading of Recommendations, Assessment, Development and Evaluations.

RESULTS: From 11 databases, 6748 articles were screened. Ninety-one full-length articles were reviewed after removing excluded studies and manual searches conducted by two reviewers. Data were extracted for 12 measures from 20 articles. Of the 12 measures, only three provided sufficient data to evaluate content validity and psychometric measurement properties. The Participation and Sensory Environment Questionnaire-Home (PSEQ-H) was the only measure that satisfied moderate content validity and moderate-to-high quality for measurement properties. These properties included: structural validity, hypothesis testing for construct validity, internal consistency, reliability, and measurement error.

INTERPRETATION: One measure, the PSEQ-H, met eight criteria for good measurement properties. To facilitate evidence-informed clinical decision-making, all psychometric properties of all 12 sensory-based, proxy-reported measures were presented. The importance of consumer engagement in measure development and the need for ongoing evaluation of measures against contemporaneous standards is recommended.

Abbreviations: COSMIN, Consensus-based Standards for the selection of health Measurement INstruments; CSP2, Child Sensory Profile 2; EPYFEI, Assessment of Sensory Processing and Executive Functions in Childhood; GRADE, Grading of Recommendations, Assessment, Development and Evaluations; PSEQ, Participation and Sensory Environment Questionnaire; PSEQ-H, Participation and Sensory Environment Questionnaire-Home; SEQ-3.0, Sensory Experiences Questionnaire-Version 3; SP2, Sensory Profile 2; SPM, Sensory Processing Measure; SPM-H, Sensory Processing Measure-Home; SPM-P, Sensory Processing Measure-Preschool; SPSRC, Sensory Processing and Self-Regulation Checklist; SSP2, Short Sensory Profile 2.

- -

The ability to adaptively organize and regulate responses to sensory stimuli (including hearing, vision, touch, smell, taste, movement and balance [vestibular], body awareness [proprioception], and interoception) in one's environment is critical to participation in everyday activities.[1] Atypical responses to sensory stimuli are observed in behaviours incongruent to the sensation experienced.[2] Terminology used to describe these observed behaviours to sensory stimuli in children with neurodevelopmental disorders includes sensory dysregulation, sensory processing, and atypical sensory reactivity.[1,3–5] In this review, we use the term 'sensory dysregulation'.[4,6,7]

Sensory dysregulation is common in people with neurodevelopmental disorders[8–13] and is associated with impaired participation in activities of daily living.[14–18] Sensory dysregulation is a recognized diagnostic feature of autism spectrum disorder (ASD).[19] However, children with other neurodevelopmental disorders also experience sensory dysregulation. For instance, approximately 90% of children with tic disorders and other comorbid neurodevelopmental disorders experience sensory dysregulation.[20–22] Increased sensory dysregulation has also been reported in individuals with obsessive–compulsive disorder[23] and attention-deficit/hyperactivity disorder (ADHD).[24]

Sensory dysregulation is associated with decreased school participation, reduced enjoyment and engagement in daily tasks, and increased parental stress.[10,14,25,26] Accordingly, assessment and management of sensory dysregulation is an accepted part of comprehensive care for children with neurodevelopmental disorders.[27] Therapeutic approaches are commonly used to address sensory dysregulation in children with neurodevelopmental disorders, with most of these strategies having been developed for children with ASD.[5,28–31] Validated, sensitive, reliable, and responsive clinician-, teacher-, patient-, and proxy-reported outcome measures to assess treatment efficiency are necessary for clinical use in sensory dysregulation.[32]

There have been three previous systematic reviews of sensory measures.[33–35] However, two of these reviews[33,34] omitted the analysis of measure design.[36–39] Moreover, these reviews were undertaken between 2013 and 2017. Measurement evaluation methods have since progressed to incorporate criteria of measure relevance, comprehensiveness, comprehensibility, sensitivity, and fitness for purpose.[36–40] These criteria warrant consideration for existing sensory measures to improve the selection of instruments for research and clinical practice.

There is discordance in the literature about the most cited sensory outcome measures,[41] with measures often not covering the depth and breadth of patient symptoms.[20] The comprehensiveness of the Sensory Profile 2 (SP2) and Sensory Processing Measure (SPM) in children with tic disorders and comorbid neurodevelopmental conditions were brought into question because study participants reported sensory dysregulation symptoms that were not rated on either measure.[20] This brings into question the measurement design, construct, fitness for purpose, and validity of the psychometric

properties of the available proxy-reported, sensory-based measures available to clinicians and researchers. Therefore, in the absence of such a review, there is a need to synthesize the available evidence to guide clinicians and researchers in selecting measures to evaluate sensory dysregulation.

This systematic review evaluates proxy-reported, sensory-based measures for children and adolescents with neurodevelopmental disorders using the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN).[36–39] The complexity and volume of measures precluded appraisal of clinician- and teacher-rated sensory-based measures for children and adolescents; therefore, they are outside the scope of this systematic review.

This study was conducted between March 2020 and September 2021 and aimed to (1) identify all current proxy-reported measures relating to sensory dysregulation in children and adolescents with a neurodevelopmental disorder and (2) comprehensively evaluate the development and psychometric properties of these measures.

METHOD

The systematic review protocol was developed and registered with Prospero (CRD42020158005). COSMIN[36–39] was used to appraise the measurement properties of the proxy-reported sensory measures used with children with neurodevelopmental disorders. PRISMA 2020 standards were used to report guidelines (Appendix S1 and Table S1).[42,43]

Literature search

A search using subject heading and free text search terms relating to the population, sensory dysregulation, measures, and measurement properties was conducted on 3rd March 2020 across 11 databases. All retrieved articles were stored

---

What this paper adds

- Three measures provided studies on content validity and psychometric measurement properties.
- The Participation and Sensory Environment Questionnaire-Home had moderate quality for content validity studies and high-to-moderate quality evidence for psychometric properties.
- The Participation and Sensory Environment Questionnaire was the only measure that included consumer involvement through qualitative interviews and pilot testing.
- Consumer involvement in measure development is important for content validity.
- Ongoing evaluation of measures against contemporaneous standards is recommended.

in EndNote X9 (Clarivate, London, UK).[44] These databases included: Allied and Complementary Medicine Database, CINAHL, Cochrane, Complementary and Alternative Medicine, Embase, InformIT, MEDLINE, Pre-MEDLINE, PsycINFO, Scopus, and Web of Science (Tables S2 and S3). A manual search was also conducted with Google Scholar using keyword search terms, the name and abbreviations of measures, and by following the publication history of the authors of the identified measures. A manual search of the databases and websites of relevant publication companies (i.e. Acer, Pearson Clinical, Pro-Ed, Psychological Assessment Australia, Wiley, and WPS) was undertaken to ensure no measure or measurement manual was omitted (Figure S1).

Eligibility criteria

Articles were included if the study reported the development of (1) a child-, proxy/parent-, or caregiver-rated (2) multi-sensory measure (3) for children and young people aged 3 to 18 years (4) diagnosed with a neurodevelopmental disorder. The lower age of 3 years was selected because a systematic review of sensory-based measures used in infants had already been conducted.[45] Second, many measures are designed for children aged 3 years and older.[46–49] Therefore, different questionnaires would be used for children younger than 3 years.[45] The upper age of 18 years was used because the reviewers wanted to capture all assessments developed for children or adolescents.[50]

Studies reporting on participants of an age or diagnostic range broader than our inclusion criteria were included if a subgroup analysis was published or available on request. There was no limit regarding the year of publication and no restriction on publication language.

Independent reviewers (PB and NS) determined article eligibility using a two-step process (Figure S2). First, the title, keywords, and abstracts were reviewed to designate articles as duplicate, excluded, or included. Manuscripts of articles that passed this screening were then reviewed for final allocation as included or excluded. Discrepancies were resolved through discussion and consensus.

Evaluation of the quality of measurement properties

COSMIN, the accepted approach to appraise measures, was used to evaluate both the quality of studies and the quality of psychometric measurement properties of sensory-based measures through a multi-step process.[36–39] The study reviewers (PB and NS) evaluated content validity and then the psychometric measurement properties of the measures using the 10 COSMIN sequential steps.

The three sequential evaluation COSMIN processes were completed using the COSMIN methodology: (1) content validity, (2) internal structure, and (3) remaining measurement properties. Content validity is the degree to which the instrument's content represents the construct reported to be measured.[36–39] Through a measure having adequate content validity, the clinician or researcher is assured that the items on the questionnaire are relevant, comprehensive, and comprehensible regarding the construct being tested and the target population.[37] Therefore, content validity is the most important measurement property.[37] The COSMIN manual suggests that measures with high-quality evidence of inadequate content validity can be excluded from any further assessment in the systematic review.[37]

Internal structure refers to how the individual items in the measure relate to one another.[36–39] The evaluation of the remaining measurement properties mainly assesses the quality of the scale, or subscale, as a whole as opposed to each individual item on the scale.[36–39]

First, two independent reviewers (ND and PB) independently evaluated (step 1) content validity, which assesses the quality of (1) measure development and (2) content validity. The reviewers then (step 2) evaluated the internal structure of these measures, which included: (1) structural validity, (2) internal consistency, and (3) cross-cultural validity. Finally, (step 3) the following remaining measurement properties were evaluated: (1) reliability, (2) measurement error, (3) criterion validity, and (4) hypothesis testing for construct validity, which consists of convergent and discriminant validity and responsiveness (Figure S1).[36–39]

We evaluated all 12 measures in relation to all the psychometric properties (steps 2 and 3) as per our study protocol, which aimed to compare all available sensory-based measures. The COSMIN methodology suggests that only measures that score 'adequate' on content validity (step 1) should be evaluated further.[36–39,51] Many commonly used sensory-based measures would be excluded from further review.[47,51,52] Through a comprehensive evaluative approach of all measures, evidence is provided to compare clinical utility and guide the selection of measures across all psychometric properties. However, measures without evidence of content validity cannot be recommended for clinical use.

All three evaluation steps (i.e. content validity, internal structure, and measurement properties) include (1) evaluation of the methodological quality of the studies using the COSMIN Risk of Bias checklist, (2) application of criteria for good measurement properties using the COSMIN criteria, and (3) summarization[36–39] and grading the quality of evidence using the COSMIN adaptation of the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach (Figure S1).[52,53]

The COSMIN Risk of Bias checklist was used to assess the methodological quality and screen for risk of bias in each included study to determine the trustworthiness of the reported study results.[36–39] The studies were rated on a 4-point score: very good (V), adequate (A), doubtful (D), and inadequate (I) for each standard. An overall score was determined by taking the lowest score across all items scored in each domain.[37]

To evaluate the quality of measurement properties, psychometric results published for each study were graded as

sufficient (+), insufficient (−), inconsistent (±), or indeterminate (?) using domain-specific COSMIN 'good measurement properties' criteria (Table S4).[36–39]

Then reviewers graded the pooled and summarized quality of evidence for each measurement property for each measure; an overall rating was determined using a COSMIN adaptation of GRADE[52,53] as specified by COSMIN.[36–38] The GRADE approach was developed for clinical trials but the COSMIN adaptation of GRADE outlined by COSMIN was developed for systematic reviews of patient-reported outcome measures. The quality of evidence refers to the confidence in the trustworthiness of the pooled or summarized result. The COSMIN adaptation of GRADE was applied to each property of each measure.

The quality of evidence was rated across five factors. These were: the risk of bias (i.e. the methodological quality of the studies); inconsistencies (i.e. unexplained inconsistency of results across the studies); imprecision (i.e. the total sample size of the available studies); indirectness (i.e. evidence from different populations other than the population of interest in the review); and publication bias (i.e. negative results are published less often).[36–38] For content validity, three factors were considered: risk of bias, inconsistency, and indirectness. For the internal structure and other measurement properties, all five factors were considered.

The quality of evidence was graded as high, moderate, low, or very low evidence according to the COSMIN adaptation of GRADE (Table S5).[36–38] It was always assumed that the quality of evidence was high. The COSMIN adaptation of GRADE has been implemented to downgrade the evidence by one or two levels per factor (i.e. moderate, low, or very low evidence) where concerns relating to the aforementioned factors exist in relation to the quality of evidence. When only a single study of inadequate quality of evidence existed, the evidence was downgraded by three levels (i.e. very low quality of evidence) (Table S5).[36–38]

After these steps, reviewers evaluated the feasibility of using these measures, formulated recommendations, and reported on the systematic review (Figure S1).

RESULTS

The results for the literature search and content validity (step 1) are discussed first. Thereafter, the results for each of the 10 measurement properties is addressed for internal structure (step 2) and other measurement properties (step 3).

Literature search

The literature search retrieved a total of 6748 publications across 11 databases. Duplicate articles ($n = 2814$) and articles not meeting the inclusion criteria ($n = 3843$) were removed after being screened by two independent reviewers to assess the title, keywords, and abstract. During the second step, the full text of 91 articles was reviewed, of which 82 articles were excluded, resulting in nine articles relating to eight different measures meeting the inclusion criteria. These nine articles related to the following measures: Assessment of Sensory Processing and Executive Function in Childhood (EPYFEI);[49] Knickerbocker Sensorimotor History Questionnaire;[54] Sensory Experiences Questionnaire-Version 3 (SEQ-3.0);[55] Sensory Processing and Self-Regulation Checklist (SPSRC);[56] Sensory Processing Measure-Home (SPM-H);[57] Sensory Processing Measure-Preschool (SPM-P);[58] Sensory Processing Scale Inventory;[50] and the Short Sensory Profile 2 (SSP2).[59]

A manual search yielded an additional 11 publications; eight were peer-reviewed journal articles and three were measurement manuals. These 11 publications related to four additional measures: the Participation and Sensory Environment Questionnaire-Home (PSEQ-H);[60–62] Participation and Sensory Environment Questionnaire-Community;[60,63] Sensory Behavior Questionnaire;[64] and the Child Sensory Profile 2 (CSP2) (Figure S2).[47]

The number of manually searched articles retrieved can be accounted for according to the following reasons: (1) measurement manuals (the SPM, SPM-P, and CSP2) would not be retrieved through the searched databases; (2) two publications were released after the search date;[62,65] and (3) cultural studies were published in journals not affiliated with the databases searched. All three publications pertaining to the PSEQ (Home and Community)[60,61,63] were only retrieved through manual searching.

In total, 20 publications (17 articles and three manuals) (Table S6) were included in this study pertaining to 12 different sensory-based measures (Table S7). All measures retrieved were proxy-reported; no child-reported questionnaires were identified. Fourteen measures were excluded due to one of the following reasons: (1) the measure was superseded by either updated versions of the same measure or by the development of a new measure ($n = 5$);[48,66–70] (2) because there were no publications relating to either the development or psychometrics of the measure ($n = 6$); or (3) the age range of the target population the measure was designed for or the psychometric studies relating to the measure were outside the scope of this systematic review and subgroup analysis was not possible ($n = 3$)[71–73] although data were requested (Table S8).

Many of the included studies were developed for children diagnosed with ASD. Twenty-three per cent of publications were studies involving typically developing children and no clinical sample was included. Three were studies conducted with children receiving occupational therapy interventions and three engaged children with a range of neurodevelopmental disorders. Twelve studies reported on study samples greater than 100. Eight studies recruited fewer than 100 study participants, ranging from 20 to 70 study participants.

Half of the included measures were published within a 3-year period from 2017 to 2019. Five of the included studies were cross-cultural studies whereby the measures had been translated into another language (the SPM-Hong Kong Chinese version,[74] SPM-Malay version,[65] the CSP2 Spanish

version,[47] and the SSP2 Polish version).[59] One study used the English version of a measure in a cross-cultural study (i.e. the SPM-P administered to English-speaking Saudi participants).[75]

Evaluation of the measurement properties for content validity (step 1)

Of the 12 reviewed measures, three had an associated peer-reviewed published report of measure development and content validity. Those three measures were the EPYFEI,[49] PSEQ (which relates to both the home and community scales of this measure),[60] and the CSP2.[47]

For these three measures, the conceptual framework to define the construct being measured was well described for both the PSEQ and the CSP2. Although all three measures consulted with professionals in item generation and measure development, only the PSEQ (Home and Community) included patient involvement in measure design.

Patient involvement consisted of semi-structured interviews with 34 parents/caregivers; 35 items were generated. For this reason, the PSEQ (Home and Community) scored adequately for the quality of measure development, whereas the EPYFEI and CSP2 scored inadequately (Table S9). The developers of the CSP2 tested the measure to ensure grade 6 reading ability of the measure using the Flesch–Kincaid Grade Level index. However, the comprehensibility of both the CSP2 and EPYFEI measures was not tested with patients.[47,49]

Of the three measures, the PSEQ was the only measure to have moderate quality of evidence for 'sufficient' (+) overall content validity. The PSEQ was also the only measure that ensured comprehensibility. There was moderate quality of evidence for sufficient relevance and comprehensiveness and high quality of evidence for comprehensibility for the PSEQ (Table 1).

The graded evidence for both the EPYFEI and CSP2 was low (Table 1) as they scored within an inadequate range for the COSMIN Risk of Bias checklist (Table S9). There were also inconsistencies with scores in terms of criteria for quality of evidence for overall measure development, content validity, and rating of reviewer scores (Table 1).

Results for the psychometric properties of measures for internal structure (step 2)

To determine the methodological quality of all 12 measures, data were extracted and evaluated for all but one publication.[60] This single study[60] reported only on measure development and not psychometric properties.

The methodological quality ratings of the studies using the COSMIN Risk of Bias checklist is reported in Table 2. Table 3 summarizes the quality of the psychometric properties of the studies pertaining to the 12 measures based on the COSMIN quality criteria[37] (Table S3) and provides an overall

TABLE 1 Content validity ratings for relevance, comprehensiveness, and comprehensibility and quality of evidence using the COSMIN adaptation of GRADE for the three measures where content validity could be evaluated

| Measurement rating | EPYFEI[49] | | | | PSEQ[60,61] | | | | CSP2[47] | | | |
	Measure development	Content validity study	Reviewer rating	Overall rating	Level of evidence	Measure development	Content validity study	Reviewer rating	Overall rating	Level of evidence	Measure development	Content validity study	Reviewer rating	Overall rating	Level of evidence
Relevance rating	±	?	?	±	Low	+	?	+	+	Moderate	±	?	?	±	Low
Comprehensiveness rating	?	?	?	?	Low	+	?	+	+	Moderate	?	?	+	+	Low
Comprehensibility rating	?	?	+	+	Low	+	+	+	+	High	?	−	+	±	Low
Overall content validity rating	?	?	?	?	Low	+	?	+	+	Moderate	?	?	+	+	Low

Ratings are provided for (1) measurement development, (2) the content validity study of specific measures, and (3) the reviewer's own rating of the content of the measures. These ratings were then synthesized into an overall rating score. COSMIN quality rating: the results are rated as either sufficient (+), whereby good measurement properties are met, insufficient (−), inconsistent (±), or indeterminate (?) when the reviewers were unable to rate the quality of evidence due to inadequate information.[39] Reviewer ratings are based on the following criteria: +, ≥85% of the items of the measure (or subscale) fulfilled the criterion; ?, no or not enough information available or quality of (part of) the study was inadequate. Quality of evidence using the COSMIN adaptation of GRADE: high evidence: there are multiple studies of at least adequate quality or there is one study of very good quality available; moderate evidence: there are multiple studies of doubtful quality available or there is only one study of adequate quality; low evidence: there are multiple studies of inadequate quality or there is only one study of doubtful quality available; very low: there is only one study of inadequate quality available.[36–39] Abbreviations: COSMIN, Consensus-based Standards for the selection of health Measurement Instruments; CSP2, Child Sensory Profile 2; EPYFEI, Assessment of Sensory Processing and Executive Function in Childhood; GRADE, Grading of Recommendations, Assessment, Development and Evaluations; PSEQ, Participation and Sensory Environment Questionnaire.

TABLE 2 The COSMIN Risk of Bias ratings for each of the studies for the 12 different measures across all psychometric measurement properties

Measure	Study	Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity	
								Convergent validity	Discriminant validity
EPYFEI	Romero-Ayuso et al.[49]	A	V	NR	V	NR	V	V	I
CSP2	Dunn[47]	NR	V	I	D	D	V	A	A
KSHQ	Carrasco[54]	NR	I	NR	NR	NR	NR	NR	NR
PSEQ-H	Pfeiffer et al.[60,61]	NR	V	NR	A	NR	NR	NR	V
	Bevans et al.[62]	V	V	NR	A	A	V	V	A
PSEQ-C	Pfeiffer et al.[63]	NR	V	NR	D	NR	NR	NR	D
SBQ	Neil et al.[64]	NR	D	NR	D	NR	V	V	V
SEQ-3.0	Ausderau et al.[55]	V	NR	NR	NR	NR	NR	NR	A
SPM	Parham and Ecker[57]	A	V	NR	D	A	V	A	V
	Dugas et al.[78]	NR	NR	NR	NR	NR	NR	V	NR
	Brown et al.[76]	NR	NR	NR	NR	NR	NR	V	NR
	Brown et al.[79]	NR	V	NR	D	NR	NR	V	NR
	Lai et al.[74]	NR	NR	A	NR	NR	NR	NR	NR
	Ahmad et al.[65]	NR	NR	I	NR	NR	NR	NR	NR
SPM-P	Miller Kuhaneck et al.[58]	A	V	NR	D	A	V	A	V
	Alkhalifah[75]	NR	NR	I	NR	NR	NR	NR	NR
Sensory Processing Scale Inventory	Schoen et al.[50]	A	V	NR	NR	NR	NR	NR	A
SPSRC	Lai et al.[56]	A	V	NR	A	NR	NR	A	A
SSP2	Dunn[47]	NR[a]	V	NR	D[a]	D[a]	V	A[a]	A[a]
	Chojnicka and Pisula[59]	NR	NR	V	NR	NR	NR	NR	NR

[a]Same data and sample reported on for the CSP2 and the SSP2. Two studies reporting on the psychometric properties of the PSEQ-H scale used the same data and sample.[55,56] The results report on the methodological quality of the measures using the COSMIN Risk of Bias checklist. A 4-point rating scale determines quality: very good (V), adequate (A), doubtful (D), or inadequate (I). An overall score is determined by taking the lowest score across all items scored in each domain.[36–39] Where no data were available or reported, the study was scored as not rated (NR). Abbreviations: CSP2, Child Sensory Profile 2; COSMIN, Consensus-based Standards for the selection of health Measurement INstruments; EPYFEI, Assessment of Sensory Processing and Executive Function in Childhood; KSHQ, Knickerbocker Sensorimotor History Questionnaire; PSEQ-C, Participation and Sensory Environment Questionnaire-Community; PSEQ-H, Participation and Sensory Environment Questionnaire-Home; SBQ, Sensory Behavior Questionnaire; SEQ-3.0, Sensory Experiences Questionnaire-Version 3; SPM, Sensory Processing Measure; SPM-P, Sensory Processing Measure-Preschool; SPSRC, Sensory Processing and Self-Regulation Checklist; SSP2, Short Sensory Profile 2.

psychometric quality rating for each psychometric property using the COSMIN adaptation of GRADE (Table S5).

Structural validity

Of the seven measures that had studies reporting on structural validity (EPYFEI, PSEQ-H, SEQ-3.0, SPSRC, SPM, SPM-P, Sensory Processing Scale Inventory), only three measures had a high level of evidence for sufficient quality of evidence for this measurement property (EPYFEI, PSEQ-H, and SEQ-3.0) (Table 3). Of the measures that conducted content validity studies, the EPYFEI had one study of adequate quality[49] (Table 2); therefore, there was moderate quality of evidence for indeterminate structural validity (factor loadings for items = 0.487–0.800). The PSEQ-H had one study of very good quality and no inconsistencies[62] (Table 2), which resulted in high quality of evidence for sufficient structural validity (confirmatory factor analysis scores = 0.71–0.91) (Table 3). The SEQ-3.0 had one study of very good quality[55] (Table 2); therefore, it had high quality of evidence for structural validity (Table 3). The other four measures had moderate evidence for sufficient quality of evidence (Sensory Processing Scale Inventory), insufficient quality of evidence (SPM, SPM-P), or indeterminate quality of evidence (SPSRC). The sample sizes in these studies ranged between 407 and 1732 (Table 3).

Internal consistency

Thirteen of the studies in this review[41,47,49,50,54,56–58,61–64,76] reported on the internal consistency rating for 11 of the 12 measures (SEQ-3.0 excluded), indicating that internal consistency is the measurement property most commonly reported. Of the 11 measures, nine (EPYFEI, CSP2, PSEQ-H, Participation and Sensory Environment Questionnaire-Community Scales, SPM, SPM-P, Sensory Processing Scale Inventory, SPSRC, and SSP2) had studies of very good quality. Therefore, all of these measures had high quality evidence for sufficient internal consistency (Tables 2, 3, and S10). Apart from the PSEQ-H, all measures only had one study reporting on internal consistency for each measure. The two studies reporting on the PSEQ-H[61,62] reported on the same study sample; thus, when pooling the summary of the results, the reviewers did not double the study sample. Therefore, since there were only single studies for each measure, the summary of pooled results can be found in Table S6.

Cross-cultural validity

Five studies addressed cross-cultural validity. Both the Polish version of the SSP2 (n = 1230)[59] and the SPM-Hong Kong Chinese version (n = 642)[74] had adequate methodological quality in terms of the process of translation and sample size for pilot testing (Table 2). There was a sufficient quality

of evidence (one study of very good quality)[59] for the cross-cultural validity of the Polish version of the SSP2.

The SPM had low quality of evidence for inconsistent cross-cultural validity because one study had adequate quality[74] and one study, the Malay version of the SPM,[65] had inadequate quality as the sample size in each study was 30.

The methodological quality of the SPM-P administered to the English-speaking Saudi participants[75] (n = 56) and the CSP2 translated into Spanish[47] (n = 67) were inadequate because the study sample sizes were under the recommended COSMIN criteria (i.e. n = 100) (Table 2).

Psychometric properties of measures for other measurement properties (step 3)

Reliability

Only one study addressed the interrater reliability of a measure,[76] whereas the other reliability studies addressed the test–retest reliability of measures. The test–retest period for all studies was between 2 and 3 weeks, except for the CSP2, with 7 to 121 days between retest periods. No study mentioned if study participants were stable during the test–retest period. However, reviewers assumed that they were stable across all studies due to the target population being either typically developing or consisting of children with neurodevelopmental disorders in the community. One of the eight measures had high quality of evidence for reliability using the COSMIN adaptation of GRADE. The EPYFEI had high quality of evidence for sufficient reliability with one study of very good quality,[49] a sample size of 1394, and intraclass correlation coefficient scores between 0.75 and 0.93 (Table 4). The PSEQ-H had moderate quality of evidence of insufficient reliability due to inconsistencies between two studies of very good quality,[61,62] resulting in the quality of evidence being downgraded by one level. For the summary of pooled results for the PSEQ-H, intraclass correlation coefficient scores ranged between 0.5 and 0.75. Because there were scores below 0.7, the study results were insufficient. There was one study of adequate quality[56] for the reliability of the SPSRC. This study reported intraclass correlation coefficient scores of 0.91 (emotional regulation), 0.95 (sensory processing), and 0.94 for the overall score (Table 3).

Measurement error

Only five of the studies addressed measurement error. All the studies reporting on measurement error except for the CSP2 had adequate methodological quality (Table 2). The difference in time frame length between the test and retest period (7 and 121 days) resulted in a doubtful rating for this measure (Table 2). Three of the measures (PSEQ-H, SPM, and SPM-P) had moderate quality of evidence for sufficient measurement error and all measures had one study of adequate quality for the quality of the measurement properties

TABLE 3 The overall quality score per psychometric measurement property for the 12 measures is reported for (1) overall quality of ratings and (2) for the synthesis and grading of the overall quality of evidence based on the COSMIN adaptation of GRADE[37–40]

Measure	Structural validity		Internal consistency		Cross-cultural validity		Reliability	
	Overall rating[a]	Level of evidence	Overall rating	Level of evidence	Overall rating	Level of evidence	Overall rating	Level of evidence
EPYFEI	?	Moderate	+	High	NR	NR	+	High
CSP2	NR	NR	+	High	–	Very low	+	Low
KSHQ	NR	NR	–	Very low	NR	NR	NR	NR
PSEQ-H	+	High	+	High	NR	NR	–	Moderate
PSEQ-C	NR	NR	+	High	NR	NR	–	Low
SBQ	NR	NR	+	Low	NR	NR	NR	NR
SEQ-3.0	+	High	NR	NR	NR	NR	NR	NR
SPM	–	Moderate	+	High	±	Low	?	Low
SPM-P	–	Moderate	+	High	–	Very low	?	Low
Sensory Processing Scale Inventory	+	Moderate	+	High	NR	NR	NR	NR
SPSRC	?	Moderate	+	High	NR	NR	+	Moderate
SSP2	NR	NR[b]	+	High	+	High	+	Low[b]

Abbreviations: CSP2, Child Sensory Profile 2; COSMIN, Consensus-based Standards for the selection of health Measurement INstruments; EPYFEI, Assessment of Sensory Processing and Executive Function in Childhood; KSHQ, Knickerbocker Sensorimotor History Questionnaire; PSEQ-C, Participation and Sensory Environment Questionnaire-Community; PSEQ-H, Participation and Sensory Environment Questionnaire-Home; SBQ, Sensory Behavior Questionnaire; SEQ-3.0, Sensory Experiences Questionnaire version 3; SPM, Sensory Processing Measure; SPM-P, Sensory Processing Measure-Preschool; SPSRC, Sensory Processing and Self-Regulation Checklist; SSP2, Short Sensory Profile 2.

[a]Overall rating using COSMIN quality rating: quality criteria ratings.

[b]These ratings for the SSP2 are the same as the CSP2 since the same data and information were provided in the manual for both measures. The results are rated as either sufficient (+), whereby the good measurement properties are met, insufficient (–), inconsistent (±), or indeterminate (?) when the reviewers were unable to rate the quality of evidence due to inadequate information.[39] Quality of evidence based on the COSMIN adaptation of GRADE:[36–39] high evidence: there are multiple studies of at least adequate quality or there is one study of very good quality available; moderate evidence: there are multiple studies of doubtful quality available or there is only one study of adequate quality; low evidence: there are multiple studies of inadequate quality or there is only one study of doubtful quality available; very low: there was only one study of inadequate quality available. Where the study reviewers were not able to retrieve data on the psychometric properties of a measure, not reported (NR) was used.
Abbreviations: CSP2, Child Sensory Profile 2; COSMIN, Consensus-based Standards for the selection of health Measurement INstruments; EPYFEI, Assessment of Sensory Processing and Executive Function in Childhood; KSHQ, Knickerbocker Sensorimotor History Questionnaire; PSEQ-C, Participation and Sensory Environment Questionnaire-Community; PSEQ-H, Participation and Sensory Environment Questionnaire-Home; SBQ, Sensory Behavior Questionnaire; SEQ-3.0, Sensory Experiences Questionnaire version 3; SPM, Sensory Processing Measure; SPM-P, Sensory Processing Measure-Preschool; SPSRC, Sensory Processing and Self-Regulation Checklist; SSP2, Short Sensory Profile 2.

(Tables 2 and S10). No measure had high overall quality when the COSMIN adaptation of GRADE was applied. Although three measures scored within moderate quality of evidence for sufficient measurement error (PSEQ-H, SPM, and SPM-P), only the PSEQ-H reported on content validity; therefore, it is the only measure that should be considered by clinicians and researchers when selecting a measure with regard to this measurement property.

Criterion validity

All three measures that had studies on content validity (EPYFEI, PSEQ-H, and CSP2) had a single study of very good quality, resulting in high quality of evidence using the COSMIN adaptation of GRADE. The EPYFEI study[49] was downgraded one level from high to a moderate level of evidence because of inconsistencies. The EPYFEI was evaluated as having moderate quality of evidence for insufficient criterion validity. The insufficient rating came from the correlation between the SSP2 and EPYFEI ranging between 0.02 and 0.80; therefore, the results were below the expected score of 0.70 to be regarded as sufficient. The correlation between the PSEQ-H and Caregiver Strain Questionnaire was 0.7, resulting in this measure having sufficient criterion validity. The CSP2 was correlated with the Sensory Profile (0.47–0.86), Behavior Assessment for Children, Second Edition (0.28–0.82), and the Social Skills Improvement Rating Scales (−0.10 to −0.38). These pooled results meant that the overall rating for the CSP2 was indeterminate. Although the SPM and SSP2 both had high quality of evidence for sufficient criterion validity, these measures did not report on content validity; therefore, they should be used at the discretion of the clinician or researcher.

Measurement error		Criterion validity		Hypothesis testing for construct validity			
				Convergent validity		Discriminant validity	
Overall rating	Level of evidence	Overall rating	Level of evidence	Overall rating	Level of evidence	Overall rating	Level of evidence
NR	NR	–	Moderate	–	Very low	–	Very low
+	Low	?	High	+	Moderate	+	Moderate
NR	NR	NR	NR	NR	NR	NR	NR
+	Moderate	+	High	+	High	+	High
NR	NR	NR	NR	NR	NR	+	High
NR	NR	+	Moderate	+	High	+	High
NR	NR	NR	NR	NR	NR	+	Moderate
+	Moderate	+	High	±	Moderate	+	Moderate
+	Moderate	–	Moderate	+	Moderate	+	Moderate
NR	NR	NR	NR	NR	NR	?	Moderate
NR	NR	NR	NR	+	Moderate	+	Moderate
+	Low[b]	+	High	+	Moderate[b]	+	Moderate[b]

The quality of evidence for this measurement property for the SPM-H (sufficient) and SPM-P (insufficient) was not the same for criterion validity, although these two measures performed the same across all other psychometric properties (Table 2). The correlation between the SPM-H and SSP[48] was 0.72. In contrast, the same correlation between the SPM-P and SSP2 resulted in a correlation of 0.62, below the required quality criterion of 0.7.

Hypothesis testing for construct validity: convergent validity

All the convergent validity studies either used the CSP2 or SSP2[48] to compare their sensory measures against, except in two studies. Bevan et al.[62] correlated the PSEQ-H scores against the Caregiver Strain Questionnaire.[77] The PSEQ-H focuses on participation and assesses parent perspectives concerning the impact of the sensory environment on participation in daily activities for young children with ASD rather than sensory integration. Therefore, researchers used a measure other than a sensory measure as a criterion standard comparator measure to determine convergent validity. The EPYFEI[49] was the only study to use the updated version of the SSP2 rather than the original version of the measure in the study.

The convergent validity studies ranged between very good (EPYFEI,[49] PSEQ-H,[62] Sensory Behavior Questionnaire,[64] SPM[76,78,79]) and adequate (SPSRC,[56] SPM,[57] CSP2,[47] SSP2[47]) for the methodological quality of these studies (Tables 2 and S10). With this said, when using the COSMIN adaptation of GRADE to determine the overall quality of evidence for these measures, the PSEQ-H and Sensory Behavior Questionnaire both had high-quality evidence for sufficient convergent validity (Table 3).

T A B L E 4   Feasibility of the three measures evaluated for content validity

Feasibility aspects	EPYFEI[49]	PSEQ[60,61]	CSP2[47]
Patient's comprehensibility	Patient's comprehensibility not tested	Pilot-tested comprehensibility through pilot testing with patients	Grade 6 reading ability of measure using the Flesch–Kincaid Grade Level index
Clinician's comprehensibility	Consulted with five occupational therapists and neuropsychologists	Content experts developed and reviewed an initial set of items for the tool	Consulted with six occupational therapists
Type and ease of administration	Completed by parent/caregiver	Completed by parent/caregiver	Completed by parent/caregiver
Length of the instrument	34 items	15 items	86 items
Completion time	15 minutes	Completion time not stated	15–20 minutes to complete
Patient's required mental and physical ability level	Reading and writing ability required	Reading and writing ability required	Reading and writing ability required
Ease of standardization	5-point Likert scale to score	5-point Likert scale to score	5-point Likert scale to score
Ease of score calculation	Manual scoring	Manual scoring	Manual scoring/can purchase Q Global for computerized administration, scoring, and reporting
Copyright	Not stated	Beth Pfeiffer[60–62]	PsychCorp, Pearson Clinical Assessment
Cost of an instrument	No cost	No cost	• SP2 Administration Manual: A$158.00 • SP2 Child Record Form 3:00–14:11 (25 pack) A$115.00 • Q Global Sensory Profile 2: Unlimited use scoring 1-year subscription: A$45
Required equipment	Writing implement and EPYFEI questionnaire	Writing implement and PSEQ questionnaire	Writing implement and SP2 questionnaire If uses Q Global, will need access to computer and Internet
Availability in different settings	Measure can be completed in different settings	Measure can be completed in different settings PSEQ-H and PSEQ-C measures Teacher questionnaire available[a]	Measure can be completed in different settings School companion questionnaire available[a]
Regulatory agency's requirement for approval	Not stated	Clinician	Speech pathologist, allied health, special education and human resources professionals, medical practitioner

[a]School/teacher versions of measures are outside the scope of this systematic review; they have been listed but their psychometric properties have not been evaluated.
Abbreviations: CSP2, Child Sensory Profile 2; EPYFEI, Assessment of Sensory Processing and Executive Function in Childhood; PSEQ, Participation and Sensory Environment Questionnaire; PSEQ-C, Participation and Sensory Environment Questionnaire-Community; PSEQ-H, Participation and Sensory Environment Questionnaire-Home; SP2, Sensory Profile 2.

The EPYFEI had a very low quality of evidence because the single hypothesis was not confirmed in the study.[49] The EPYFEI was hypothesized to correlate most strongly with the sensory processing scale of the SSP2 and least strongly with the behaviour scale of the SSP2. This was not proven since the total score obtained for the EPYFEI had a high positive correlation with the SSP2 behavioural subscale (0.80, $p < 0.001$) and the SSP2 sensory subscale (0.68, $p = 0.008$).

Discriminant validity

Twelve studies reported on discriminant validity for the different measures. Subgroup analysis was conducted between children with ASD (including Asperger syndrome, ASD, and pervasive developmental disorder-not otherwise specified) and neurotypical children. The study for the EPYFEI reported on a clinical sample that included study participants with various neurodevelopmental disorders (i.e. ADHD [$n = 95$, 5.5%];

ASD [$n = 84$, 4.8%]; language-specific disorders [$n = 106$, 6.1%]; developmental delay [$n = 15$, 0.9%]; and other neurodevelopmental disorders [$n = 83$, 4.8%]). For both the SPM and SPM-P parent questionnaires, the comparator group consisted of children receiving occupational therapy.

Although 10 different subgroups were reported on for discriminant validity for the CSP2, and it scored adequately for the quality of discriminate validity, each of these groups had small sample sizes (i.e. developmental delay [$n = 11$]; ASD [$n = 78$]; ADHD [$n = 96$]; dual diagnosis of ASD and ADHD [$n = 24$]; learning disability [$n = 45$]; gifted and talented [$n = 18$]; intellectual disability [$n = 9$]; Down syndrome [$n = 9$]; English as a second language [$n = 7$]; and other [$n = 62$]).

For the Sensory Processing Scale Inventory, the statistical method used to determine discriminant validity between subgroups was appropriate for hypothesis testing. Researchers attempted to match the typically developing cohort with the clinical sample.[50] However, there was still a significant difference in age ($z = 5.25$, $p < 0.1$), with typically

developing participants being slightly older (mean = 8 years 2 months, SD = 2 years 5 months) than the clinical sample (mean = 6 years 8 months, SD = 2 years 5 months). Although the effect size was small, the two groups differed in sex distribution, with proportionally more males in the clinical sample than in the typically developing group (φ = 0.15). Ethnicity and socioeconomical statistical data were also not reported.[50]

Responsiveness

To determine the responsiveness of the measures, the term 'responsiveness' was included in the search strategy, yet none of the studies reported on this measurement property for any measure. The SP2, SSP2, and EPYFEI were all screening measures and not used as pre-/post-test measures; therefore, it is not appropriate for these measures to report on responsiveness.

Feasibility

Information on the feasibility of implementing the three measures that provided information on content validity is provided in Table 4. All three measures can be feasibly implemented by clinicians and researchers. The measures vary in the number of items (the EPYFEI has 34 items, the PSEQ has 15, and the SP2 has 86) and cost (the EPYFEI and PSEQ are freely available, the SP2 requires the user to purchase the administration manual and record forms). All three measures use a 5-point Likert scale.

DISCUSSION

To our knowledge, this is the first systematic review evaluating multiple sensory dysregulation measures for children and adolescents using current best practice measurement standards according to COSMIN.

Twelve measures were assessed across 20 publications that provided validation data; they included three manuals. Of the 12 measures reviewed, only three (EPYFEI, PSEQ, CSP2) provided information on the development of the measure and content validity. Although the EPYFEI, PSEQ, and CSP2 were all designed through cooperation with professional experts, only the PSEQ included consumer involvement through qualitative interviews and pilot testing. This is despite the essential nature of cooperation and consumer engagement in developing items that constitute a measure to ensure relevance, comprehensiveness, and comprehensibility to the patients completing them.[36–39]

The other nine measures did not describe measurement design, nor were content validity studies conducted. In the absence of evidence that the measures are relevant, comprehensive, and comprehensible, clinicians and researchers should question the usefulness of these measures.[36–39]

When studies reporting on the PSEQ-H are evaluated using the COSMIN methodology, it is the most comprehensive, comprehensible, relevant, and psychometrically robust measure of the 12 measures evaluated and is recommended for children aged 2 and 7 years. For children older than 7 years, two measures reported on measure development and content validity, that is, the EPYFEI (designed for Spanish children aged 3–11 years) and the CSP2 (intended for children aged 3–14 years 11 months). Although content validity studies were provided for the EPYFEI and CSP2 measures, the quality of evidence was low for relevance, comprehensiveness, and comprehensibility. Of these two measures, the EPYFEI had the better quality of evidence across the psychometric measurement properties, but cross-cultural studies must be conducted to use this measure with an English-speaking population. The SEQ-3.0 (for ages 2–12 years), SPM-H (5–12 years), and Sensory Processing Scale Inventory (4–18 years) were designed to be used with an older age range. However, no measure development and content validity studies have been published for these measures; therefore, they should be used at the clinician's discretion.

Seven measures (EPYFEI, PSEQ-H, SEQ-3.0, SPM, SPM-P, Sensory Processing Scale Inventory, SPSRC) had evidence of structural validity. To determine structural validity, newer psychometric methods, such as item response theory and Rasch modelling, are recommended;[80,81] however, uptake of these methods across the studies was limited. Reasons for this include its computational complexity and limited availability of user-friendly analytical software.[82] Of the discriminant validity studies, nine were conducted with children with ASD. Understandably, the focus has been on testing these measures with children with ASD because of the high prevalence of sensory dysregulation in this cohort.[19] However, there needs to be a focus on developing measures for a broad range of neurodevelopmental disorders.[20–24]

Measures from the same suite of tools, such as the SPM and the SPM-P (used with different age ranges), PSEQ-H, Participation and Sensory Environment Questionnaire-Community Scales (used for different environmental settings), and the CSP2 and SSP2 (full version and abbreviated version of the questionnaire) did not have the same quality of evidence across measurement properties. Therefore, clinicians and researchers may consider the evidence for each measure since all measures from the same suite of tools have variable quality of evidence. In addition to assessing quality, measure selection needs to consider age group, target populations, and environment to ensure that measures are fit for purpose.

None of the studies reported on responsiveness for any of the 12 measures included in this review. Three of the measures were designed as screening tools (EPYFEI, SP2, and SSP2). When selecting a measure as a pre-/post-test measure, clinicians and researchers ought to ensure that the measure is designed as an outcome measure and not as a screening tool; there is no evidence, in terms of these studies, on responsiveness on any of the other nine measures.

It is interesting to note that half of the measures[49,50,56,60–64] included in this review were published in the past 6 years (between 2017 and 2020). This indicates that there is a growing interest and need for the development of new sensory-based measures.

Limitations: evidence

Studies in this review did not state if participants had commenced or received any previous sensory-based intervention or if study participants were stable at the time of recruitment or test–retest administration of questionnaires. Parents/caregivers whose children attended sensory-based interventions may have a more sophisticated understanding of sensory processing issues and heightened sensitivity to the behaviours associated with sensory input. Also, their children's behaviour may change due to therapeutic interventions.[66] This raises potential bias in reporting study results.

Seven measures[47,50,54,55,57,58,64] were developed before 2018. However, modern psychometric measurement development has evolved since, such as the development of COSMIN standards for measurement development in 2018.[36–39] Since COSMIN emphasizes the need for adequate content validity of a measure, among other psychometric standards, these measures no longer meet the current standard for measure development.[36–39] Thus, highlighting the importance of ongoing evaluation of existing measures against the continuously improving criteria for measure development is needed to ensure that measures meet current standards.

Only two measures (SPM and SPM-P) had studies reporting on all of the 10 measurement properties. Most measures (*n* = 8) had only a single publication reporting on psychometric properties.

Limitations: review process

When developers elect to partner with a publishing company as part of test development, it limits the ability to publish in peer-reviewed journals. This creates a challenge when conducting systematic reviews since measurement manuals were not identified in a literature search across databases. We overcame this challenge by searching all known publishers and distributors of assessment measures. This study identified six measures that were excluded from the review because there were no published measure development, content validity, or psychometric studies for these measures (Table S8). Although these measures may potentially be psychometrically sound, they could not be included or evaluated in this review. Therefore, these measures should be used with caution due to the lack of evidence pertaining to content validity and psychometric measurement properties.

Although the term 'responsiveness' was used in the search strategy, no studies on responsiveness were retrieved for any of the measures. Since the PSEQ was the only measure to have adequate content validity, we recommend that an additional systematic review be conducted specifically to identify all studies that have used this measure in intervention studies. Meta-analysis on the pre-/post-test data to determine the responsiveness of this measure should be conducted.

Conclusion

To assess treatment efficacy, validated, sensitive, and reliable proxy-reported sensory-based measures are necessary. This review provides a guide for clinicians and researchers to aid the selection of these sensory measures.

It is imperative that, as part of measurement development, content validity studies are included to ensure comprehensibility, comprehensiveness, and relevance. Of the 12 measures included in this review, only three (EPYFEI, PSEQ, and SP2) provided studies on content validity. Of these three measures, the PSEQ-H had moderate quality for content validity studies but also had high-to-moderate quality evidence for sufficient psychometric properties that were tested. Although the other measures varied in quality across the other measurement properties, these should be used at the discretion of the clinician since measures without content validity cannot be recommended for use.[36–38]

This review highlights the importance of consumer involvement in the development of measures. Clinicians and researchers should consider content validity and psychometric measurement properties to ensure measures are fit for purpose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Nicolette Soler https://orcid.org/0000-0001-7328-7063
Russell C. Dale https://orcid.org/0000-0002-2495-1826
Paula Bray https://orcid.org/0000-0002-5874-1667

REFERENCES

1. Lane SJ, Miller LJ, Hanft BE. Toward a consensus in terminology in sensory integration theory and practice. II: Sensory integration patterns of function and dysfunction. Sensory Integration Special Interest Section Quarterly. 2000; 23:1-3.
2. Koomar JA, Bundy AC. The art and science of creating direct intervention from theory. In Fisher AG, Murray EA, Bundy AC editors. Sensory integration theory and practice. Philadelphia: F. A. Davis. 1991; p. 251-317.
3. Orth M, Munchau A. Transcranial magnetic stimulation studies of sensorimotor networks in Tourette syndrome. Behav Neurol. 2013;27(1):57-64.
4. Weisman H, Parush S, Apter A, Fennig S, Benaroya-Milshtein N, Steinberg T. A study of sensory dysregulation in children with tic disorders. J Neural Transm. 2018;125(7):1077-85.
5. Tomchek SD, Dunn W. Sensory processing in children with and without autism: a comparative study using the short sensory profile. Am J Occup Ther. 2007;61(2):190-200.
6. Bijlenga D, Tjon-Ka-Jie JYM, Schuijers F, Kooij JJS. Atypical sensory profiles as core features of adult ADHD, irrespective of autistic symptoms. Euro Psychiatry. 2017;43:51-7.
7. Gouze KR, Hopkins J, LeBailly SA, Lavigne JV. Re-examining the epidemiology of sensory regulation dysfunction and comorbid psychopathology. J Abnorm Child Psychol. 2009;37(8):1077-87.
8. Abele-Webster LA, Magill-Evans JE, Pei JR. Sensory processing and ADHD in children with fetal alcohol spectrum disorder. Can J Occup Ther. 2012;79(1):60-3.
9. Adams JN, Feldman HM, Huffman LC, Loe IM. Sensory processing in preterm preschoolers and its association with executive function. Early Hum Dev. 2015;91(3):227-33.
10. Baranek GT, Chin YH, Hess LM, Yankee JG, Hatton DD, Hooper SR. Sensory processing correlates of occupational performance in children with fragile X syndrome: preliminary findings. Am J Occup Ther. 2002;56(5):538-46.
11. Bruni M, Cameron D, Dua S, Noy S. Reported sensory processing of children with Down syndrome. Phys Occup Ther Pediatr. 2010;30(4):280-93.
12. Pavao SL, Rocha N. Sensory processing disorders in children with cerebral palsy. Infant Behav Dev. 2017;46:1-6.
13. Shimizu VT, Bueno OF, Miranda MC. Sensory processing abilities of children with ADHD. Braz J Phys Ther. 2014;18(4):343-52.
14. Nieto C, Lopez B, Gandia H. Relationships between atypical sensory processing patterns, maladaptive behaviour and maternal stress in Spanish children with autism spectrum disorder. J Intellect Disabil Res. 2017;61(12):1140-50.
15. Siper PM, Kolevzon A, Wang AT, Buxbaum JD, Tavassoli T. A clinician-administered observation and corresponding caregiver interview capturing DSM-5 sensory reactivity symptoms in children with ASD. Autism Res. 2017;10(6):1133-40.
16. Van Etten HM, Kaur M, Srinivasan SM, Cohen SJ, Bhat A, Dobkins KR. Increased Prevalence of Unusual Sensory Behaviors in Infants at Risk for, and Teens with, Autism Spectrum Disorder. J Autism Dev Disord. 2017;47(11):3431-45.
17. Green D, Chandler S, Charman T, Simonoff E, Baird G. Brief Report: DSM-5 Sensory Behaviours in Children With and Without an Autism Spectrum Disorder. J Autism Dev Disord. 2016;46(11):3597-606.
18. Brockevelt BL, Nissen R, Schweinle WE, Kurtz E, Larson KJ. A comparison of the Sensory Profile scores of children with autism and an age- and gender-matched sample. S D Med. 2013;66(11):459, 61, 63-5.
19. Thye MD, Bednarz HM, Herringshaw AJ, Sartin EB, Kana RK. The impact of atypical sensory processing on social impairments in autism spectrum disorder. Dev Cogn Neurosci. 2018;29:151-67.
20. Soler N, Hardwick C, Perkes IE, Mohammad SS, Dossetor D, Nunn K, et al. Sensory dysregulation in tic disorders is associated with executive dysfunction and comorbidities. Mov Disord. 2019;34(12):1901-9.
21. da Silva Prado H, do Rosário MC, Lee J, Hounie AG, Shavitt RG, Miguel EC. Sensory phenomena in obsessive-compulsive disorder and tic disorders: a review of the literature. CNS spectrums. 2008;13(05):425-32.
22. Miguel EC, Prado H, Rauch S, Coffey B, Baer L, Savage C, et al. Sensory phenomena in obsessive-compulsive disorder and Tourette's disorder. J Clin psychiatry. 2000;61(2):150-6; quiz 7.
23. Russo M, Naro A, Mastroeni C, Morgante F, Terranova C, Muscatello M, et al. Obsessive-compulsive disorder: a "sensory-motor" problem? Int J Psychophysiol. 2014;92(2):74-8.
24. Bijlenga D, Tjon-Ka-Jie J, Schuijers F, Kooij J. Atypical sensory profiles as core features of adult ADHD, irrespective of autistic symptoms. Eur Psychiatry. 2017;43:51-7.
25. Bar-Shalita T, Vatine JJ, Parush S. Sensory modulation disorder: a risk factor for participation in daily life activities. Dev Med Child Neurol. 2008;50(12):932-7.
26. Chien CW, Rodger S, Copley J, Branjerdporn G, Taggart C. Sensory Processing and Its Relationship with Children's Daily Life Participation. Phys Occup Ther Pediatr. 2016;36(1):73-87.
27. Dunn W. Supporting children to participate successfully in everyday life by using sensory processing knowledge. Infants Young Child. 2007;20(2):84-101.
28. Baranek GT, David FJ, Poe MD, Stone WL, Watson LR. Sensory Experiences Questionnaire: discriminating sensory features in young children with autism, developmental delays, and typical development. J Child Psychol Psychiatry. 2006;47(6):591-601.
29. Lin C-L, Min Y-F, Chou L-W, Lin C-K. Effectiveness of sensory processing strategies on activity level in inclusive preschool classrooms. Neuropsychiatr Dis Treat. 2012;8:475.
30. Pfeiffer BA, Koenig K, Kinnealey M, Sheppard M, Henderson L. Effectiveness of sensory integration interventions in children with autism spectrum disorders: A pilot study. Am J Occup Ther. 2011;65(1):76-85.
31. Singh J, Santosh P. Psychopharmacology of neurodevelopmental disorders in children. Child and Adolescent Psychiatry: Springer; 2016. p. 325-62.
32. Benfer KA, Weir KA, Boyd RN. Clinimetrics of measures of oropharyngeal dysphagia for preschool children with cerebral palsy and neurodevelopmental disabilities: a systematic review. Dev Med Child Neurol. 2012;54(9):784-95.
33. Eeles AL, Spittle AJ, Anderson PJ, Brown N, Lee KJ, Boyd RN, et al. Assessments of sensory processing in infants: a systematic review. Dev Med Child Neurol. 2013;55(4):314-26.
34. Jorquera-Cabrera S, Romero-Ayuso D, Rodriguez-Gil G, Triviño-Juárez J-M. Assessment of Sensory Processing Characteristics in Children between 3 and 11 Years Old: A Systematic Review. Front pediatr. 2017;5:57-.
35. Licciardi L, Brown T. An overview & critical review of the sensory profile - second edition. Scand J Occup Ther. 2021:1-13.
36. Mokkink LB, De Vet HC, Prinsen CA, 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.
37. Terwee CB, Prinsen CA, Chiaretto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.
38. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-57.

39. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.

40. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10(1):1-8.

41. Schaaf RC, Lane AE. Toward a best-practice protocol for assessment of sensory features in ASD. J Autism Dev Disord. 2015;45(5):1380-95.

42. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Reprint—preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Phys Ther. 2009;89(9):873-80.

43. 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. Syst Rev 2021;10(1), 1-11.

44. Researchsoft T. EndNote 9.0. 1998-2005.

45. Eeles AL, Spittle AJ, Anderson PJ, Brown N, Lee KJ, Boyd RN, et al. Assessments of sensory processing in infants: a systematic review. Dev Med Child Neurol. 2013;55(4):314-26.

46. Dunn W, Westman K. The sensory profile: the performance of a national sample of children without disabilities. Am J Occup Ther. 1997;51(1):25-34.

47. Dunn W. Sensory Profile 2: User's Manual: Psych Corporation; 2014.

48. Dunn W. The sensory profile manual. San Antonio, TX: Psychological Corporation. 1999.

49. Romero-Ayuso D, Jorquera-Cabrera S, Segura-Fragoso A, Toledano-Gonzalez A, Rodriguez-Martinez MC, Trivino-Juarez JM. Assessment of Sensory Processing and Executive Functions in Childhood: Development, Reliability, and Validity of the EPYFEI. Front Pediatr. 2018;6:71.

50. Schoen SA, Miller LJ, Sullivan J. The development and psychometric properties of the Sensory Processing Scale Inventory: A report measure of sensory modulation. J Intellect Dev Dis. 2017;42(1):12-21.

51. Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter L, de Vet HC, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User manual. 2018;78(1).

52. Schünemann H, Brożek J, Guyatt G, Oxman A. editors Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. 2013 [cited 2021 July 1]. gdt.guidelinedevelopment.org/app/handbook/handbook.html

53. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-6.

54. Carrasco RC. Reliability of the Knickerbocker Sensorimotor History Questionnaire. OTJR. 1990;10(5):280-2.

55. Ausderau K, Sideris J, Furlong M, Little L, Bulluck J, Baranek G. National Survey of Sensory Features in Children with ASD: Factor Structure of the Sensory Experience Questionnaire (3.0). J Autism Dev Disord. 2014;44(4):915-25.

56. Lai CYY, Yung TWK, Gomez INB, Siu AMH. Psychometric Properties of Sensory Processing and Self-Regulation Checklist (SPSRC). Occup Ther Int. 2019;2019:8796042.

57. Parham LD, Ecker C. Sensory processing measure (SPM): Western Psychological Services; Los Angeles, CA: Western Psychological Services;2007.

58. Miller Kuhaneck H, Ecker C, Parham L, Henry D, Glennon T. Sensory processing measure-preschool (SPM-P): Manual. Los Angeles, CA: Western Psychological Services. 2010.

59. Chojnicka I, Pisula E. Adaptation and psychometric properties of the Polish version of the Short Sensory Profile 2. Medicine. 2019;98(44):e17689.

60. Pfeiffer B, Coster W, Tucker C, Piller A. Development and Content Validity of the Participation and Sensory Environment Questionnaire. Occup Ther Ment Health. 2018;34(2):105-21.

61. Pfeiffer B, Piller A, Slugg L, Shiu C. Brief Report: Reliability of the Participation and Sensory Environment Questionnaire: Home Scales. J Autism Dev Dis. 2018;48(7):2567-76.

62. Bevans KB, Piller A, Pfeiffer B. Psychometric Evaluation of the Participation and Sensory Environment Questionnaire–Home Scale (PSEQ–H). Am J Occup Ther. 2020;74(3):1-9.

63. Pfeiffer B, Piller A, Bevans K, Shiu C. Reliability of the participation and sensory environment questionnaire: Community scales. ResAutism Spec Dis. 2019;64:84-93.

64. Neil L, Green D, Pellicano E. The psychometric properties of a new measure of sensory behaviors in autistic children. J Autism Dev Dis. 2017;47(4):1261-8.

65. AHMAD NM, KADAR M, CHUI CS, RASDI HFM. Adaptation, Validation and Reliability Testing of Sensory Processing Measure Home Form Malay version for Children with Autism. Jurnal Sains Kesihatan Malaysia. 2020;18(1):37-45.

66. Johnson-Ecker CL, Parham LD. The evaluation of sensory processing: a validity study using contrasting groups. Am J Occup Thery. 2000;54(5):494-503.

67. Baranek GT, David FJ, Poe MD, Stone WL, Watson LR. Sensory Experiences Questionnaire: Discriminating sensory features in young children with autism, developmental delays, and typical development. J Child Psychol Psychiatry. 2006;47(6):591-601.

68. Little LM, Freuler AC, Houser MB, Guckian L, Carbine K, David FJ, et al. Psychometric validation of the sensory experiences questionnaire. Am J Occup Ther. 2011;65(2):207-10.

69. Schoen SA, Miller LJ, Green KE. Pilot study of the sensory over-responsivity scales: Assessment and inventory. Am Jo Occup Ther. 2008;62(4):393-406.

70. Williams ZJ, Failla MD, Gotham KO, Woynaroski TG, Cascio C. Psychometric evaluation of the Short Sensory Profile in youth with autism spectrum disorder. J Autism Dev Dis. 2018;48(12):4231-49.

71. Kolacz J, Raspa M, Heilman KJ, Porges SW. Evaluating Sensory Processing in Fragile X Syndrome: Psychometric Analysis of the Brain Body Center Sensory Scales (BBCSS). J Autism Dev Dis. 2018;48(6):2187-202.

72. Tavassoli T, Hoekstra RA, Baron-Cohen S. The Sensory Perception Quotient (SPQ): development and validation of a new sensory questionnaire for adults with and without autism. Mol Autism. 2014;5:29.

73. Brown C, Dunn, W. Adolescent / Adult Sensory Profile. San Antonio, TX: Pearson. 2002.

74. Lai CY, Chung JC, Chan CC, Li-Tsang CW. Sensory processing measure-HK Chinese version: psychometric properties and pattern of response across environments. Res Dev Dis. 2011;32(6):2636-43.

75. Alkhalifah S. Psychometric Properties of the Sensory Processing Measure Preschool-Home among Saudi Children with Autism Spectrum Disorder: Pilot Study. J Occup Ther, Schools, and Early Interven. 2019;12(4):401-16.

76. Brown T, Morrison IC, Stagnitti K. The reliability of two sensory processing scales used with school-age children: Comparing the response consistency of mothers, fathers, and classroom teachers rating the same child. J Occup Ther, Schools, & Early Interven. 2010;3(4):331-47.

77. Brannan AM, Heflinger CA, Bickman L. The Caregiver Strain Questionnaire: Measuring the impact on the family of living with a child with serious emotional disturbance. J Emot Beh Dis. 1997;5(4):212-22.

78. Dugas C, Simard M, Fombonne E, Couture M. Comparison of two tools to assess sensory features in children with autism spectrum disorder. Am J Occup Ther. 2018;72(1):1-9.

79. Brown T, Morrison IC, Stagnitti K. The convergent validity of two sensory processing scales used with school-age children: comparing the Sensory Profile and the Sensory Processing Measure. N Z J Occup Ther. 2010;57(2):56-65.

80. Thomas ML. Advances in applications of item response theory to clinical assessment. Psychol Assess. 2019;31(12):1442.

81. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehab. 2000;81:S15-S20.

82. Doucette A, Wolf AW. Questioning the measurement precision of psychotherapy research. Psychother Res. 2009;19(4-5):374-89.

SUPPORTING INFORMATION

The following additional material may be found online:

Appendix S1: Prisma-P Protocol.

Figure S1: Flow diagram of our approach to using COSMIN for conducting a systematic review of measures and COSMIN taxonomy and definitions.

Figure S2: PRISMA 2020 flowchart including all databases, registers, and other sources.

Table S1: PRISMA 2020 checklist.

Table S2: Example of search strategy for Medline via Ovid.

Table S3: Search strategies for all 11 databases used in this systematic review.

Table S4: COSMIN criteria for good measurement properties.

Table S5: Definitions of the four different ratings for the quality of evidence.

Table S6: Description of all included studies evaluated in this systematic review.

Table S7: Details the characteristics of the 12 included measures.

Table S8: List of excluded sensory measures from this systematic review and rationale for exclusion.

Table S9: Quality of measure development results using the COSMIN Risk of Bias checklist.

Table S10: Quality of measurement properties per study based on COSMIN quality criteria.