Evaluating the Psychometric Quality of Social Skills Measures: A Systematic Review

Reinie Cordier[1]☯*, Renée Speyer[2]☯, Yu-Wei Chen[3]☯, Sarah Wilkes-Gillan[4]☯, Ted Brown[5]‡, Helen Bourke-Taylor[6]‡, Kenji Doma[2]‡, Anthony Leicht[2]‡

1 School of Occupational Therapy and Social Work, Curtin University, Perth, WA, Australia, 2 College of Healthcare Sciences, James Cook University, Townsville, QLD, Australia, 3 Faculty of Health Sciences, The University of Sydney, Sydney, NSW, Australia, 4 School of Allied Health, Australian Catholic University, North Sydney, NSW, Australia, 5 Department of Occupational Therapy, Faculty of Medicine, Nursing and Health Sciences, Monash University–Peninsula Campus, Frankston, VIC, Australia, 6 School of Allied Health, Australian Catholic University, St Patricks Campus, Fitzroy, VIC, Australia

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* reinie.cordier@curtin.edu.au

Abstract

Introduction

Impairments in social functioning are associated with an array of adverse outcomes. Social skills measures are commonly used by health professionals to assess and plan the treatment of social skills difficulties. There is a need to comprehensively evaluate the quality of psychometric properties reported across these measures to guide assessment and treatment planning.

Objective

To conduct a systematic review of the literature on the psychometric properties of social skills and behaviours measures for both children and adults.

Methods

A systematic search was performed using four electronic databases: CINAHL, PsycINFO, Embase and Pubmed; the Health and Psychosocial Instruments database; and grey literature using PsycExtra and Google Scholar. The psychometric properties of the social skills measures were evaluated against the COSMIN taxonomy of measurement properties using pre-set psychometric criteria.

Results

Thirty-Six studies and nine manuals were included to assess the psychometric properties of thirteen social skills measures that met the inclusion criteria. Most measures obtained excellent overall methodological quality scores for internal consistency and reliability. However, eight measures did not report measurement error, nine measures did not report cross-cultural validity and eleven measures did not report criterion validity.

Conclusions

The overall quality of the psychometric properties of most measures was satisfactory. The SSBS-2, HCSBS and PKBS-2 were the three measures with the most robust evidence of sound psychometric quality in at least seven of the eight psychometric properties that were appraised. A universal working definition of social functioning as an overarching construct is recommended. There is a need for ongoing research in the area of the psychometric properties of social skills and behaviours instruments.

Introduction

Social Functioning

Most theorists agree that social functioning is a complex construct that encompasses social skills as well as social behaviour and cognition during inter-personal interactions [1]. Social functioning involves the integration of emotional, linguistic and cognitive skills, which develop from early childhood to adolescence [1]. Social functioning is foundational for the development and maintenance of meaningful relationships and community participation and is critical for both physical health and psychological well-being [2]. Impairments in social functioning manifest in approximately one in every ten children [3], with greater levels of social impairment reported in developmental disabilities [4–9]. Impairments in social functioning are associated with an array of adverse outcomes in adolescence and adulthood, such as delinquency [10], social withdrawal and isolation [11].

Theoretical Frameworks on Social Functioning

Theoretical models of social functioning are most commonly embedded within psychology; with more recent models emerging from neuroscience [1]. These models are commonly embedded within a social information processing (SIP) framework and are focused on infant development and the cognitive processes needed for social skills in adults [12]. Researchers have extended SIP to include both cognitive and affective dimensions. In their extensions of SIP, Mostow et al. [13] and Guralnick [14] included emotional, cognitive and behavioural predictors of peer related social competence.

While few comprehensive theoretical models exist, analogous perspectives and definitions are evident throughout literature [1, 15–18]. There appears to be consensus among theorists that social functioning is an overarching construct which is reliant on a range of cognitive, emotional and linguistic skills; reflecting a person's overall performance in the area of social development [1, 19, 20].

Cognitive functions, social emotional and linguistic skills. In their model of socio-cognitive integration of abilities, Beauchamp and Anderson [1] considered social skills and functions from a range of perspectives, integrating them into a model of social competence. The socio-cognitive integration of abilities model defines the core dimensions of social functioning (biological–psychological–social) and their interactions within a developmental framework founded on empirical research and clinical principles.

Cognitive and executive functions have been central to most models of social functioning [1]. In most models, cognitive function is used to reflect a range of higher cognitive processes, such as: attentional control (e.g., selective and sustained attention, response inhibition, self-monitoring, self-regulation) and skills linked to executive functioning (e.g., working memory,

planning, problem-solving, strategic behaviour). Researchers have linked deficits in these skills to poor social outcomes, including: antisocial behaviour, emotional dysregulation, delinquency and peer-rejection [21–23].

Another principal component of most social functioning models is socio-emotional skills. These skills have been reported to include: face-emotion perception, theory of mind, and empathy. Face-emotion perception is fundamental to recognising emotion, which is needed for reciprocal social interactions [24, 25]. Theory of mind is a social cognitive skill and involves understanding the emotion, intention and perceptions of others and how the knowledge or beliefs of someone else may differ from one's own [26]. Empathy involves identifying the emotional state of another, the capacity to take the perspective or role of the other, and the evocation of shared affective responses. Empathy is associated with pro-social behaviour and comprises both affective and cognitive components [27–29].

There is an array of literature highlighting the influence of communication skills on social functioning. However, linguistic skills are infrequently incorporated into models of social functioning [1, 30]. In particular, pragmatic language skills have been described as fundamental to social functioning as they are needed to: a) integrate verbal and non-verbal communication; b) detect and interpret underlying meaning in social cues; c) respond appropriately during social interactions; and d) regulate emotions [30–32].

Approaches to social development and social competence. Some theorists view social functioning from a hierarchical perspective with 'social skills,' and 'social cognition' representing different levels of social behaviour under the auspice of social functioning. However, other theorists use the constructs 'social competence', 'social skills' and 'social functioning' interchangeably. In developmental literature, there is general consensus that children may exhibit externalising or internalising behaviour when social competence is lacking [33]. Social competence has often been conceptualised to reflect effectiveness or success in social interactions. However, there is a noteworthy lack of agreement on the nature of its relationship to social functioning (if viewed as a separate construct) and how to define, measure and approach the skills attributable to social competence [33].

Researchers have adopted four approaches to social functioning and competence in the field of social development: 1) a social skills approach, focusing on specific skills and pro-social behaviours; 2) a peer status approach, focusing on sociometric status and peer acceptance or rejection; 3) a relationship approach, based on one's ability to form and maintain friendships and positive relationships with parents, teachers and intimate relationships in adulthood; and 4) an adaptive approach, that takes into account the need for individuals to adjust their social interactions and behaviour across a variety of contexts and types of social situations [16, 34]. Despite these divergent approaches to social functioning, there is consensus among theorists regarding social functioning. There is also agreement among theorists that these skills and behaviours are mediated by a range of internal and external factors.

Mediating factors of social development and social functioning. Much research has focused on factors that mediate social functioning. Mediating factors have often been categorised as external factors or internal factors which influence a person's natural predisposition when interacting with others [1]. Internal factors include brain development and integrity, personality and temperament and are often conceptualised within the domain of cognitive skills. External factors have been described to comprise environmental influences, such as: family factors [35], parent behaviours [36], socioeconomic status (SES) [37] and culture [38]. Finally there is agreement that the mediating effect of the stated internal and external factors facilitate the product of either well-developed social functioning or result in maladaptive social behaviours [1, 19]. Surprisingly, little is known about the contextual factors and contexts within which social interaction occur, which may either inhibit or facilitate social functioning.

The influence of contextual factors on social functioning. The contexts within which the social interaction occurs (e.g., school, home, work) may also influence the way in which the individual interacts with others, as well as the nature and quality of social interactions [39, 40]. Despite research having a strong focus on adaptive behaviour [16], there has been limited research into the influence of contextual factors on a person's social functioning; perhaps one of the most neglected areas in contemporary conceptual models [1, 33].

Assessments of Social Functioning

In recognition of the importance of social functioning, the assessment and treatment of social difficulties has become a focus of research over past decades; particularly in the area of developmental disabilities [19, 41, 42]. Given the array of skills subsumed within the domain of social functioning, some measures attempt to broadly address these different domains, while others specifically target the assessment of a sub-set of skills [19].

The assessments examining social functioning are mostly in the form of pen-and-paper questionnaires with the majority of identified assessments relying on child self-report, while some revert to using peers, parents and teachers in proxy reporting [19]. Reliance on self-report alone is problematic due to: poor construct and/or criterion validity (low correlations with other assessments), susceptibility to social desirability, and reliance on a child's ability to comply with instructions [19, 43]. Further, one-time parent and teacher ratings have been described as inadequate methods of measurement, as they solely rely on the objectivity of the rater and perform poorly in capturing behavioural changes over time, across contexts and between different interactants [19, 44].

In addition to the form of the assessment, it is important for the measure to be based on a sound theoretical model and for the underlying psychometric properties of the assessment to be evaluated [1]. Measures can have different prognostic and/or analytical functions. Measures can be prognostically used to: a) predict a later outcome; b) determine suitability for a particular intervention; c) report on the responsiveness to a particular intervention; or d) determine the amount of intervention required (dosage) [45]. Measures may also be used analytically to: a) explain or understand the contexts; b) classify or identify subgroups of patients; c) allow exploration of relationship between factors; d) detect within subject change or between subgroup differences; and e) enable comparison of patients to other population subgroups or norms [45]. If an outcome measure is used to evaluate changes in clients over time following a particular intervention, the quality of responsiveness becomes important. Conversely if the measure is used as a screening measure to accurately diagnose the presence or absence of a condition, identification accuracy and therefore the interpretability of the measure is of primary concern as it indicates the overall precision of making a diagnosis [46].

Reviews of Social Functioning Measures

In response to the plethora of assessments measuring social functioning, four reviews aiming to provide an overview of these measures have been conducted [9, 19, 47, 48]. While these reviews provided valuable information on some of the many assessments focusing on social functioning, they have several limitations. Two of the earlier reviews by Demaray et al. [47] and Merrell [48] only evaluated a small number of measures and the review by Matson and Wilkins [9] did not adopt a systematic design [19]. However, a more recent systematic review by Crowe et al. [19] built on the review by Matson and Wilkins [9]. Crowe et al. [19] evaluated 86 assessments under the broad domain of social functioning. The review focused on: fairly recently published assessments (i.e., 1988–2010), whether psychometric properties were reported and the popularity of the measures (i.e., number of citations) [19]. Therein lies the

main limitation of the Crowe et al. [19] review in that they only reported on whether and if authors reported their measure's psychometric properties, but the review did not report on all psychometric properties. Moreover, the review lacked rigor in evaluating the quality of the psychometric properties of the measures in a systematic and uniform manner.

Limitations of Current Assessments Measuring Social Functioning

Across the literature, several limitations to current social functioning assessments have been identified. These limitations include discrepancies regarding the definition of social functioning and subsequently a lack of connection between measures to a theoretical model [19, 49]. We add two further limitations to current social functioning assessments: 1) a lack of observational assessments, and 2) lack of uniform reporting on the psychometric properties of these assessments.

Definitions of social functioning. Many assessments lack a clear connection to a theoretical framework and clear definitions of the domains of social functioning being measured. Under the umbrella term *social functioning* are the following constructs which are often used interchangeably: pro-social behaviour [50], social adjustment [12], social cognition [51], social competence [52], social outcomes [53] and social skills [54]. The discrepancies present several challenges to scientific literature; including comparisons across studies, evaluation of the quality of assessments being used and most importantly, barriers to determining the effectiveness of treatments aiming to ameliorate difficulties surrounding social functioning [1, 49].

Lack of observational assessments. There is a near complete absence of well validated measures that assess social functioning through direct observation. Observational measures may provide numerous benefits including a social and ecologically valid approach, whereby individuals can be assessed performing the skills within the contexts they experience the difficulties. However, the few observational assessments that currently exist are study-specific, unpublished or were modified from adult or aggression measures [19].

Uniform reporting of psychometric properties. Evident within this area of research are lack of uniform reporting on both the description and psychometric properties of such measures. As noted by Crowe and colleagues [19], the literature on the psychometric properties of assessments of social functioning is continually being updated. Further reviews are needed to update and provide a comprehensive review into the psychometric properties of the available assessments.

Study Aim

The purpose of this systematic review was twofold. Firstly, this review aimed to provide an overview of information on existing assessments that measure areas of social functioning across the lifespan; highlighting current gaps in the age, type, or context in which assessments can be administered. Within the area of social functioning, we focused the review to assessments of social skills and social behaviour.

Secondly, within this review, a central aim was to comprehensively evaluate the quality of psychometric properties reported across these assessments. To guide this aim, we used the COSMIN taxonomy of measurement properties and definitions for health-related patient-reported outcomes [55].

Methods

The PRISMA statement was used to guide the methodology and reporting of this systematic review. The PRISMA statement checklist contains a total of 27 item areas that are deemed

essential for the transparent reporting of systematic reviews [56]. A completed PRISMA check-list applicable to the current review is accessible (see Table in S1 Table).

Eligibility Criteria

Eligibility criteria for studies in this review included research articles or published manuals on the psychometric properties of instruments designed to measure the social skills and behaviours of the general population. We adopted the following, widely used definition of social skills to guide our review, which comprises both skills and behavioural elements that result in positive social interactions, encompassing: 1) cooperation, 2) verbal and non-verbal communication, 3) engagement and participation, 4) empathy, and 5) self-regulation and adaptive behaviours in situations where interpersonal interaction occurs [16, 57]. Within this search, instruments measuring these skills and behaviours in both children and adults were included. For instruments to be included in this review, their main components or subscales needed to meet the definition we adopted of social skills. Instruments or published articles written in languages other than English were not eligible. As we were interested in evaluating the quality of psychometric properties of contemporary measures being used in recent research, instruments were excluded if they were published before 1994. For the purpose of this review, instruments that had an update of their psychometric properties in the last 20 years at the time of the search were regarded as contemporary. Instruments were further excluded if they were developed for a specific target population (e.g., autism spectrum disorders) rather than a normative sample. Articles were excluded if no psychometric properties were reported. Conference abstracts, reviews, case reports, student dissertations and editorials were also excluded.

Information Sources

A systematic literature review was performed using four electronic databases: CINAHL, PsycINFO, Embase and Medline. Furthermore, Ovid's Health and Psychosocial Instruments (HAPI) database was used to identify potential instruments that met the inclusion criteria. The HAPI database provides access to information on instruments relevant to health related disciplines, including the fields of social sciences, organisational behaviour, and library and information sciences. Database searches were conducted between 3/05/2014 and the 15/05/2014. Search strategies included both free text words and subject headings (see Table 1), and comprised all journal articles up to May 2014. The second author conducted the searches because of her expertise in conducting systematic reviews. The databases were accessed from the libraries of Curtin University and James Cook University.

We searched for grey literature using Google Scholar and PsycEXTRA. PsycEXTRA is the American Psychological Association's grey literature database which accompanies the PsycINFO database. It combines bibliographic records with full-text professional and lay-audience literature in the behavioural and social sciences. To be comprehensive, we also searched the websites of three major publishers of assessments in social sciences (Pearson, Acer and Western Psychological Services) to identify potential assessment not identified in earlier search strategies.

Literature Search

A total of 2,117 abstracts were retrieved, including duplicates. The total abstracts from free text words and subject headings searches across each database were: CINAHL = 574, PsycINFO = 186. Embase = 711, Medline = 646. A total of 220 duplicates across the four databases were removed. The electronic search strategy used for each database, including: subject

Table 1. Search Terms.

	Database and Search Terms	Limitations
Subject Headings	**CINAHL**: ((MH "Psychometrics") OR (MH "Measurement Issues and Assessments") OR (MH "Reliability") OR (MH"Interrater Reliability") OR (MH "Reliability and Validity") OR (MH Test-Restest Reliability") OR (MH "Intrarater Reliability") OR (MH "Criterion-Related Validity") OR (MH "Validity") OR (MH "Predictive Validity") OR (MH "Internal Validity") OR (MH "Face Validity") OR (MH "External Validity") OR (MH "Discriminant Validity") OR (MH "Consensual Validity") OR (MH "Concurrent Validity") OR (MH "Qualitative Validity") OR (MH "Construct Validity") OR (MH "Content Validity") OR (MH "Measurement Error") OR (MH"Bias (Research)")) AND ((MH "Social Skills") OR (MH "Social Skills Training") OR (MH "Social Interaction Skills (IowaNOC)") OR (MH "Social Behavior Disorder") OR (MH "Social Behavior") OR (MH "Social Adjustment") OR (MH "Social Attitudes") OR (MH "Social Participation") OR (MH "Social Problems")) AND ((MH "Outcome Assessment") OR (MH "Behavior Rating Scales") OR (MH "Social Readjustment Rating Scale") OR (MH "Psychological Tests") OR (MH "Health Screening") OR (MH "Patient Assessment") OR (MH "Self Assessment") OR (MH "Measurement Issues and Assessments") OR (MH "Scales/ED/MT/ED") OR (MH "Questionnaires/EV/MT/ED") OR (MH "Clinical Assessment Tools"))	English Language; Human
	PsycINFO: (Psychometrics/ OR statistical reliability/ OR statistical validity/ OR "error of measurement"/) AND (social skills/ OR social reinforcement/ or social responsibility/ OR social acceptance/ OR social adjustment/ OR social anxiety/ OR social approval/ OR social behavior/ OR social cognition/ OR social comparison/ OR social control/ OR social deprivation/ OR social desirability/ OR social discrimination/ OR social skills training/ or social stress/ or social structure/ or social support/ or social values/) AND (Measurement/ OR questionnaires/ or rating/ OR rating scales/ OR psychological screening inventory/ OR screening/ OR surveys/)	English language; Human
	Embase: (Validation study/ OR validity/ OR psychometry/ OR reliability/ OR measurement accuracy/ OR measurement error/ OR measurement precision/ OR measurement repeatability/) AND (social acceptance/ OR social adaptation/ OR social bonding/ OR social change/ OR social cognition/ OR social cognitive theory/ OR social competence/ OR social adaptation/ OR social learning/ OR social learning theory/ OR social life/ OR socialization/ OR social behavior/ OR social stress/ OR social support/ OR social desirability/ OR social attitude/ OR social disability/ OR social exclusion/ OR social interaction/ OR social participation/ OR social phobia/ OR social problem/ OR social rejection/ OR social validity/) AND (questionnaire/ OR rating scale/ OR psychological rating scale/ OR screening/ OR screening test/ OR outcome assessment/ OR "social and occupation functioning assessment scale"/ OR social interaction anxiety scale/ OR social interaction test/ OR social support index/ OR "named inventories, questionnaires and rating scales"/ OR social phobia scale/ OR social readjustment rating scale/ OR social recognition test/ OR social support index/)	English language; Human
	Medline: (Psychometrics/) AND (Questionnaires/ OR Treatment Outcome/ OR "Outcome Assessment (Health Care)"/ OR Neuropsychological Tests/ OR Psychological Tests/) AND (Social behavior/ OR social participation/ or Social Problems/ OR Social Behavior Disorders/ OR social adjustment/ OR Emotions/)	English language; Human
Free Text Words	**CINAHL**: (social* OR emotion*) AND (psychometric* OR reliability OR validit* OR reproducibility OR bias) AND (questionnaire* OR assessment* OR test OR tests OR evaluation*)	Publication Date: 20130401–20130631 English Language; Human
	PsycINFO: *As per CINAHL Free Text*	(No Related Terms) yr = "2014-Current" and English language and human
	Embase: *As per CINAHL Free Text*	(No Related Terms) yr = "2014-Current" and english language and human
	Medline: *As per CINAHL Free Text*	yr = "2013-Current" and English language and human

doi:10.1371/journal.pone.0132299.t001

heading, free text and limitations are reported in Table 1. Reference lists of the included articles were searched for additional literature.

The HAPI database identified 22 instruments that potentially met the inclusion criteria; thus warranting further scrutiny. Search of grey literature identified an additional 25 records;
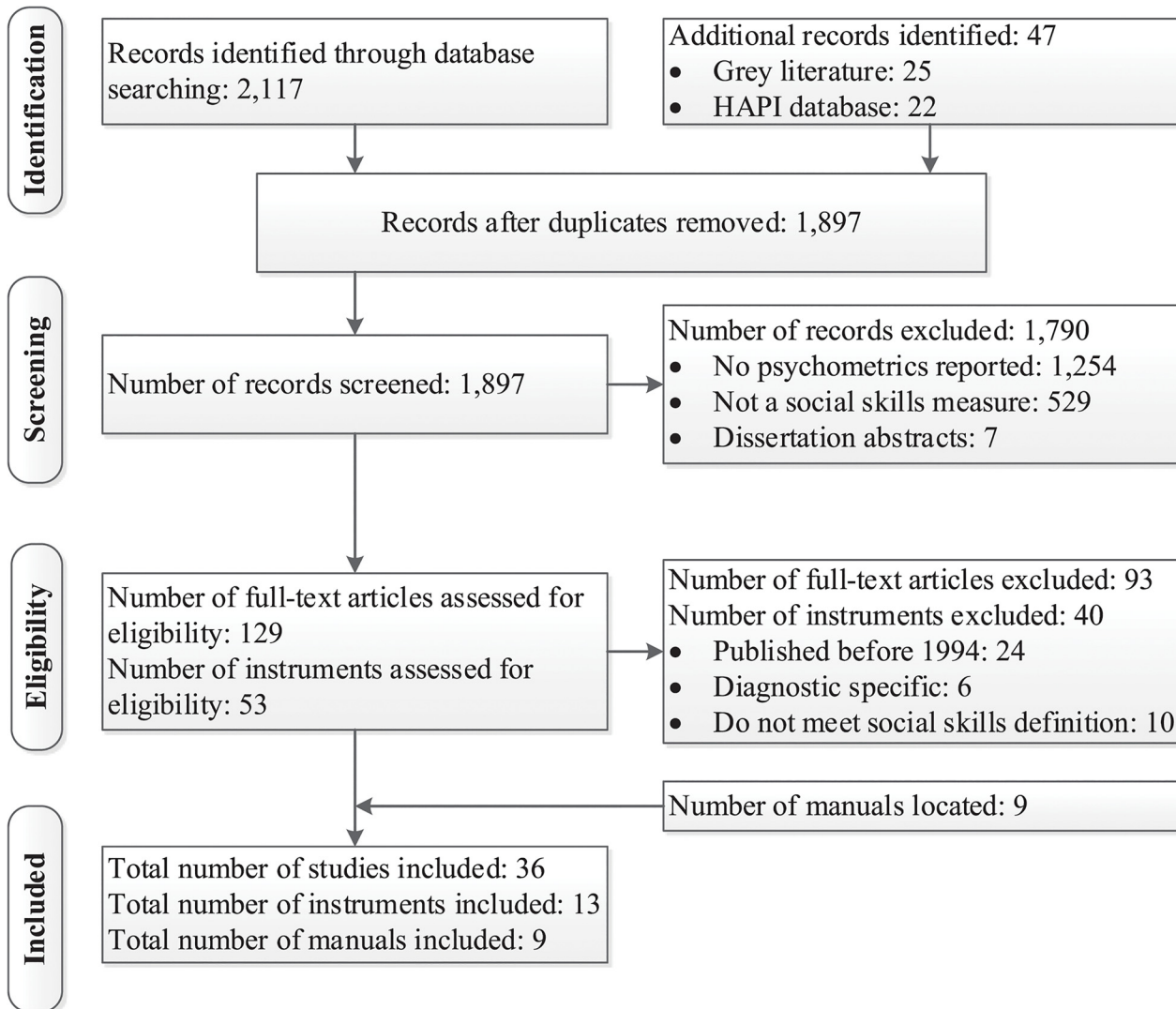
Fig 1. Flow diagram of the reviewing process according to PRISMA. Adapted from Moher et al. [58].

thus a total of 47 records were identified. Fig 1 presents the flow diagram of the reviewing process according to PRISMA [58].

Study Selection

Two independent abstract reviewers rated the abstracts on the following inclusion criteria: abstracts had to describe an instrument or outcome measure; address its psychometric measurement properties; and assess social skills and behaviours. A random sample of 40% of the abstracts was examined to determine the inter-rater reliability: Weighted Kappa = 0.79 (95% CI 0.72–0.86). To ensure all instruments measured social skills and/or social behaviour, two doctoral candidates with knowledge in the area of social skills reviewed the instruments together. At this level, the abstracts of the articles and descriptions of the assessment were located to ensure the instrument met the adopted definition of social skills [16, 57].

Data Collection Process / Data Extraction

To capture the data contained within the included studies and manuals (45) we used the Cochrane Handbook for Systematic Reviews section 7.3a [59], and the Systematic Reviews Centre for Reviews and Dissemination [60]. Data were extracted under the following headings: study design, purpose of the study, study population, age of the population, and instrument characteristics. To both capture the data and assess the methodological quality of the data the COSMIN was used [55].

Methodological Quality

The psychometric quality of the included instruments were then analysed using the COSMIN taxonomy of measurement properties and definitions for health-related patient-reported outcomes [61]. The COSMIN checklist [55] is a standardised tool for assessing the methodological quality of studies on measurement properties and consists of nine domains: internal consistency, reliability (relative measures: including test-retest reliability, inter-rater reliability and intra-rater reliability), measurement error (absolute measures), content validity (including face validity), structural validity, hypotheses testing, cross-cultural validity, and criterion validity. Responsiveness as a psychometric property was not evaluated in this review. Definitions of all the psychometric properties, as defined in the COSMIN statement, are provided in Table 2. Interpretability is not considered to be a psychometric property under the COSMIN framework and was therefore not described in this review. Each domain of the COSMIN checklist includes 5 to 18 items focussing on different aspects of study design and statistical analyses. Terwee et al. [62] proposed using a 4-point rating scale per item (excellent, good, fair, and poor), obtaining an overall methodological quality score per psychometric property by taking the lowest rating of any item in the corresponding domain. As this rating system appears to be so severe that it inhibits differentiation between more subtle psychometric qualities of instruments [63], a revised scoring was introduced. The outcome was presented as percentage of rating (Poor = 0–25.0%, Fair = 25.1%-50.0%, Good = 50.1%-75.0%, Excellent = 75.1%-100.0%). Given that some COSMIN items only have excellent and good as an option for rating, we calculated the total score for each psychometric property using the following formula to most accurately capture the quality of the psychometric properties:

$$\text{Total score for psychometric property} = \frac{(\text{Total score obtained} - \text{minimum score possible})}{(\text{Max score possible} - \text{minimum score possible})} \times 100$$

To ensure consistency of the COSMIN checklist ratings, the first author trained two independent doctoral candidates to complete the COSMIN checklist. To ensure accuracy, the two raters completed COSMIN checklists together for 10 of the 13 instruments.

Data Items, Risk of Bias and Synthesis of Results

All data items for each instrument were obtained. When an item was not reported, an 'NR' was recorded. Risk of bias was assessed at an individual study level during the rating of the COSMIN checklist through the inclusion of 'methodological limitations items'. The results were synthesised and grouped as follows: 1) development and validation of the instrument, 2) the psychometric properties of the instruments, and 3) the instrument characteristics.

Table 2. COSMIN: Definitions of Domains, Psychometric Properties, and Aspects of Psychometric Properties for Health-Related Patient-Reported Outcomes based on Mokkink et al. [61]

Psychometric property	Domain and Definition[a]
	Reliability: the degree to which the measurement is free from measurement error.
Internal consistency	The degree of the interrelatedness among the items.
Reliability	The proportion of the total variance in the measurements which is because of "true" differences among patients.
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.
	Validity: the degree to which an instrument measures the construct(s) it purports to measure.
Content validity	The degree to which the content of an instrument is an adequate reflection of the construct to be measured.
Face validity[b]	The degree to which (the items of) an instrument indeed looks as though they are an adequate reflection of the construct to be measured.
Construct validity	The degree to which the scores of an instrument are consistent with hypotheses based on the assumption that the instrument validly measures the construct to be measured.
Structural validity[c]	The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured.
Hypothesis testing[c]	Item construct validity.
Cross-cultural validity[c]	The degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument.
Criterion validity	The degree to which the scores of an instrument are an adequate reflection of a "gold standard".
Responsiveness	**Responsiveness:** the ability of an HR-PRO instrument to detect change over time in the construct to be measured.
Interpretability[d]	**Interpretability**[a]: the degree to which one can assign qualitative meaning to an instrument's quantitative scores/ score change.

Notes.
[a]Applies to Health-Related Patient-Reported Outcomes (HR-PRO) instruments.
[b]Aspect of content validity under the domain of validity.
[c]Aspects of construct validity under the domain of validity.
[d]Interpretability is not considered a psychometric property

doi:10.1371/journal.pone.0132299.t002

Results

Systematic Literature Search

After the removal of duplicate abstracts across the four databases, a total of 1,897 studies were screened for inclusion in this review. Of these studies, 129 full-text articles on 53 measures were assessed for eligibility (see Fig 1). Of these 53 measures, 13 measures met the inclusion criteria and 40 were excluded for the following reasons: 24 were published before 1994, 6 were diagnostic specific, and 10 did not meet the definition of social skills adopted for the purpose of this review. See Table 3 for an overview of the 40 social skills instruments and the reasons for exclusion. Through additional searches another 9 manuals were located. Thus, the psychometric properties were obtained for a total of 13 social skills measures which were accessed using 36 articles and 9 manuals.

Table 3. Overview of Social Skills Instrument: Reasons for Exclusion.

	Instrument	Acronym	Reason for Exclusion
1.	Aberrant Behavior Checklist [64]	ABC-ID	Published before 1994; target population: intellectual disabilities;
2.	Alarme Distress de Bebe Scale (Alarm Distress Baby scale) [65, 66]	ADBB	Did not meet social skills definition
3.	Anticipatory Social Behaviours Questionnaire [67]	ASBQ	Did not meet social skills definition
4.	Autism Behavior Checklist [68]	ABC	Target population: ASD
5.	Autism Social Skills Profile [69]	ASSP	Target population: ASD
6.	Behavioral Assessment System for Children (BASC) [70]	BASC	Published before 1994
7.	Behavioral Repertoire Self-Rating Questionnaire [71]	BRSRQ	Published before 1994; did not meet social skills definition
8.	Blushing, Trembling, and Sweating Questionnaire [72]	BTSQ	Did not meet social skills definition; Target population: Phobias
9.	Child Behavior Rating Scale [73–75]	CBRS	Published before 1994
10.	Children's Action Tendency Scale [76, 77]	CATS	Published before 1994
11.	Children's Assertive Behavior Scale [76, 78]	CABS	Published before 1994
12.	Children's Social Behavior Questionnaire [79–82]	CSBQ	Target population: ASD
13.	Community-based Social Behavior Assessment Battery [83]	CBSB	Target population: emotional and behavioral disorders
14.	Conflict Behavior Questionnaire [84]	CBQ	Published before 1994; did not meet social skills definition
15.	Coping Inventory for Stressful Situations [85]	CISS	Published before 1994; did not meet social skills definition
16.	Experienced Social Mobility Questionnaire [86]	ESMQ	Did not meet social skills definition
17.	General Health Questionnaire-12 [87]	GHQ-J	Published before 1994; did not meet social skills definition;
18.	Maternal Social Support Index [88]	MSSI	Published before 1994; did not meet social skills definition
19.	Minimal social behavior scale [89, 90]	MSBS	Did not meet social skills definition; published before 1994
20.	Personality Disorder Questionnaire—Revised [91]	PDQ-R	Published before 1994
21.	Personal and Social Performance Scale [92]	PSP	Did not meet social skills definition; target population: psychiatric population
22.	Scale of Negative Social Exchanges [93]	SNSE	Did not meet social skills definition
23.	Self-Reported Antisocial Behavior Questionnaire [94]	SRABQ	Published before 1994; target population: ADHD
24.	School Social Skills Rating Scale [47, 95]	S-3	Published before 1994
25.	Social Adjustment Rating Scale—Self-Report [96]	SAS-SR	Published before 1994
26.	Social Behavior Inventory [97]	SBI	Target population: child abuse
27.	Social Behavior Assessment Inventory [47, 98]	SBAI	Published before 1994
28.	Socially Desirable Response Set [99]	SDRS	Published before 1994; did not meet social skills definition
29.	Social Influences Scale [100]	SIS	Published before 1994; did not meet social skills definition
30.	Social Phobia and Anxiety Inventory [101]	SPAI	Published before 1994; did not meet social skills definition
31.	Social Problems Questionnaire [102]	SPQ	Published before 1994
32.	Social Readjustment Rating Scale-modified [103]	SRRS-M	Published before 1994; target population: emotional and behavioral disorders
33.	Social Role Competence Measure [104]	SRCM	Did not meet social skills definition; No psychometrics
34.	Social Skills Inventory [105, 106]	SSI	Published before 1994
35.	Subtypes of Antisocial Behavior Questionnaire [107]	STAB	Did not meet social skills definition
36.	Teenage Inventory of Social Skills [108, 109]	TISS	Published before 1994
37.	The Social Skills Q-Sort [110]	SSQ	Target population: ASD
38.	Waksman Social Skills Rating Scale [47, 111]	WSSRS	Published before 1994
39.	Walker-McConnell Scale of Social Competence and School Adjustment [47, 112]	WMS	Published before 1994
40.	Young Adult Social Behavior Scale [113]	YASB	Did not meet social skills definition

Note. ASD = Autism Spectrum Disorders; ADHD = Attention Deficit Hyperactivity Disorder.

doi:10.1371/journal.pone.0132299.t003

Included Social Skills Measures

Information on the development and validation of the 13 included social skills measures is reported in Table 4. All measures had some evidence of development and validation through the use of a normative study population/sample; without targeting a specific diagnostic group. Of the 13 measures, 12 were developed using children up to 12 years of age; with 6 of these measures also using an adolescent sample (13–18 years). No measure was developed using an adult population alone (older than 18 years) and only 1 measure (i.e., the Evaluation of Social Interaction [ESI]) was developed and validated using all three age groups (i.e., children, adolescents and adults)

The characteristics of the included measures are reported in Table 5. Of the 13 measures, 6 were published within the last 5 years (since 2009). Regarding the measure type, 9 measures used self-, parent- or teacher-report; with 3 of these measures using only teacher-report. Of the remaining measures, 3 were observation-based and only 1 used a semi-structured interview (see Table 5). Regarding the response options within the measures, 12 reported the use of Likert scales and only the ESI reported the use of a criterion-referenced rating scale. Of the 12 measures using a Likert response scale, 11 reported the use of a 3 to 5 point scale, and the Peer Social Maturity Scale (PSMS) reported the use of a 7-point scale. Additionally, the Interaction Rating Scale (IRS) reported the use of a dichotomous (yes or no) rating system for its scale.

Psychometric Properties

The quality ratings of the psychometric properties of all 13 measures, which were evaluated against the COSMIN quality criteria, are summarised in Table 6. The overall means and standard deviations of each psychometric property across all social skills measures were also calculated. S*tructural validity* was most frequently reported; the mean rating across 12 measures was 78.1 (SD = 14.8), indicating excellent quality. The least reported psychometric property was *criterion validity*; 3 measures had a mean rating of 80.7 (SD = 7.9), indicating excellent quality. Overall, 11 measures had evidence for *internal consistency*; the mean rating across these measures was 84.5 (SD = 12.9), indicating excellent quality. The mean rating of the 11 measures reporting on *reliability* and *hypothesis testing* was 74.7 (SD = 11.9) and 56.8 (SD = 17.5) respectively; indicating good quality. The mean rating of the 6 measures that reported on *measurement error* was 70.0 (SD = 12.8); indicating good quality. *Content validity* was reported by 8 measures; mean rating 74.0 (SD = 17.0), indicating good quality. Cross-cultural validity was reported by 4 measures; mean rating 54.3 (SD = 13.2), indicating good quality.

Discussion

In this systematic review, we identified and evaluated the quality of psychometric properties of instruments that measure social skills and behaviours developed after 1994. We identified 13 instruments that evaluated a component of social skills and behaviours that fit the definition we used in the review. The vast majority of instruments (11) were developed mainly for school aged children and adolescents, with two instruments solely developed for children 2–5 years (i.e., QRSH-PR and SEEC).

Alongside the validity evidence, reliability findings also need to be reported. This systematic review of social skills and behaviour instruments using the COSMIN framework provided a comprehensive summary of this. Application of the COSMIN checklist based taxonomy provided the framework for a critical evaluation of the quality and extent of psychometric evidence of the 45 research articles and manuals on the 13 social skills and behaviour instruments. Based on the COSMIN taxonomy, the social skills and behaviour instrument with the most robust psychometric properties to date was the SSBS-2, given that all eight psychometric

Table 4. Description of Studies for the Development and Validation of Instrument for the Assessment of Social Skills.

Instrument	Reference	Purpose of study	Study population	Age (range[R] and/or Mean [M] Standard deviation [SD])
Evaluation of Social Interaction (ESI)	Simmons et al. [114]	Internal scale validity, items' skill hierarchy and intended purpose, differentiation between people with and without disability	*N = 134* (*I*) 60–73y age group; (*II*) 35–59y age group; (*III*) 17–34y age group; (*IV*) 4–15y age group. Number of participants in each group is not reported; population characteristics are not pecified.	*Total sample:* R = 4–73y; M = NR; SD = NR. (*I*) R = 60–73y; M = 65.4y; SD = NR. (*II*) R = 35–59y; M = 45.2y; SD = NR. (*III*) R = 17–34y; M = 24.9y; SD = NR. (*IV*) R = 4–15y; M = 6.5y; SD = NR.
	Søndergaard and Fisher [115]	Sensitivity to differentiate between people without identified diagnosis and those with neurologic or psychiatric disorders	*N = 185* (*I*) control: *n* = 304; (*II*) neurologic: *n* = 77; (*III*) psychiatric: *n* = 104	*Total sample:* R = 16y–69y; M = NR; SD = NR. (*I*) R = NR; M = 41.9y; SD = 13.6y. (*II*) R = NR; M = 44.9y; SD = 11.9y. (*III*) R = NR; M = 35.5 SD = 11.1y.
	Griswold and Townsend [116]	Sensitivity to differentiate between children with and without disabilities	*N = 46* (*I*) with disabilities: *n* = 23; (*II*) matched without disabilities: *n* = 23	*Total sample:* R = 2y–12y; M = 7y; SD = NR. (*I*) R = NR; M = NR; SD = NR. (*II*) R = NR; M = NR; SD = NR.
	Fisher and Griswold [117]	Development and validation	*N = 6,552* (*I*) well: *n* = 4,099; (*II*) at risk/mild: *n* = 235; (*III*) psychiatric: *n* = 915; (*IV*) developmental: *n* = 320; (*V*) neurological: *n* = 300; (*VI*) other/ multiple: *n* = 682	*Total sample:* R = 2y–80+y; M = NR; SD = NR. (*I*) R = NR; M = NR; SD = NR. (*II*) R = NR; M = NR; SD = NR. (*III*) R = NR; M = NR; SD = NR. (*IV*) R = NR; M = NR; SD = NR. (*V*) R = NR; M = NR; SD = NR. (*VI*) R = NR; M = NR; SD = NR.
Home and Community Social Behavior Scales (HCSBS)	Merrell et al. [118]	Convergent and discriminative validity. *Study 1:* Comparisons with the Social Skills Rating System and Conners Parent Rating Scales; *Study 2:* Comparisons With the Child Behavior Checklist; and *Study 3:* Comparisons With the Behavior Assessment System for Children.	*N = 325 Study 1: n* = 59; *Study 2: n* = 60; *Study 3:* = 206 (*I*) *n* = 76 (*II*) *n* = 130	*Total sample:* R = NR; M = NR; SD = NR. *Study 1:* Grade 6 (R = NR; M = NR; SD = NR). *Study 2:* Grade 2–5 (R = NR; M = NR; SD = NR). *Study 3:* (*I*) R = 6–12y; M = NR; SD = NR (*II*) R = 12–18y; M = NR; SD = NR.
	Lund and Merrell [119]	Construct validity to discriminate sensitivity of group differences	*N = 180* (*I*) emotional-behavioral disorders: *n* = 60 (*II*) learning disabilities: *n* = 60 (*III*) without disabilities: *n* = 60	*Total sample:* R = 6y–12y; M = NR; SD = NR. (*I*) R = NR; M = NR; SD = NR. (*II*) R = NR; M = NR; SD = NR. (*III*) R = NR; M = NR; SD = NR.
	Merrell and Caldarella [120]	Internal consistency, correlations among subscales, discrimination between groups (at-risk vs. non-risk)	*N = 267* (*I*) non-risk: *n* = 107; (*II*) at-risk: *n* = 160	*Total sample:* R = 11y–16y; M = NR; SD = NR. (*I*) R = NR; M = NR; SD = NR. (*II*) R = NR; M = NR; SD = NR.
	Merrell and Caldarella [120]	Development and validation of HCSBS	*N = 1,562* School aged children	*Total sample:* R = 5y–18y; M = NR; SD = NR.
Interaction Rating Scale (IRS)	Anme et al. [121]	Describe the features of IRS as an evidence-based practical index of children's social skills and parenting difference in age skills	*N = 814* (*I*) *n* = 213 (18m age group); (*II*) *n* = 344 (30m age group); (*III*) *n* = 175 (42m age group); (*IV*) *n* = 82 (7y age group)	*Total sample:* R = 18m–7y; M = NR; SD = NR. (*I*) R = NR; M = 18m; SD = NR. (*II*) R = NR; M = 30m; SD = NR. (*III*) R = NR; M = 42m; SD = NR. (*IV*) R = NR; M = 7y; SD = NR.
Interaction Rating Scale Advanced (IRS–Advanced)	Anme et al. [122]	Validity of IRS-A as an evidence-based practical index of social competence	*N = 17* High school students	*Total sample:* R = 16–17y; M = NR; SD = NR.
	Anme et al. [123]	Validity and reliability of the IRS-A	*N = 50* Adults	*Total sample:* R = 18–48y; M = NR; SD = NR.
Interaction Rating Scale (IRS–BC)	Anme et al. [124]	Validate the IRS-BC.	*N = 20* Children	*Total sample:* R = 5–6y; M = NR; SD = NR.
Matson Evaluation of Social Skills with Youngsters-II (MESSY-II)	Matson et al. [125]	Internal consistency in different age group, convergent and divergent validity. *Study 1:* Assess the reliability and validity of the MESSY parent-report form using an updated sample of typically developing children *Study 2:* Examine internal consistency reliability, convergent and divergent validity.	*N = 1,138 Study 1:* (*I*) *n* = 286 (2–5y age group); (*II*) *n* = 301 (6–9y age group); (*III*) *n* = 298 (10–16y age group); *Study 2:* (*I*) *n* = 78 (2–5y age group) (*II*) *n* = 130 (6-9y age group) (*III*) *n* = 45 (10–17y age group)	*Total sample:* R = NR; M = NR; SD = NR. *Study 1:* (*I*) R = 2–5y; M = 3.9y; SD = 1.0y. (*II*) R = 6–9y; M = 7.4y; SD = 1.1y. (*III*) R = 10–16y; M = 12.3y; SD = 1.8y. *Study 2:* (*I*) R = 2–5y; M = 4.2y; SD = 0.9y. (*II*) R = 6–9y; M = 7.20y; SD = 1.2y. (*III*) R = 10–17y; M = 10.9y; SD = 1.0y.
	Matson et al. [126]	Factor structure and internal consistency	*N = 886* Typically developing children	*Total sample:* R = 2–16y; M = 7.9y; SD = 3.7y.
	Méndez et al. [127]	Psychometric properties of the Spanish translation of MESSY	*N = 634* Adolescents	*Total sample:* R = 12–17y; M = 14.3; SD = 1.7.
	Teodoro et al. [128]	Psychometric properties of the Brazilian version of MESSY	*N = 382* Children	*Total sample:* R = 7–15y; M = 10.3y; SD = 1.6y.
	Chou [129]	Psychometric properties of the Chinese translation of MESSY	*N = 191* Children	*Total sample:* R = 11–18y; M = 14.0; SD = 1.6y
	Matson et al. [130]	(*I*) internal consistency, split half reliability (*II*) inter-rater reliability	*N = 147* (*I*) children with ASD/ PDD-NOS: *n* = 114; (*II*) children with ASD/ PDD-NOS: *n* = 33 (participants a subset of the children from study I)	(*I*) R = 2–16y; M = 7.5y; SD = 3.3y. (*II*) R = 5–16y; M = 9.1y; SD = 3.1y.

(Continued)

Table 4. (Continued)

Instrument	Reference	Purpose of study	Study population	Age (range[R] and/or Mean [M] Standard deviation [SD])
Preschool and Kindergarten Behavior Scales 2 (PKBS-2)	Merrell [131]	Development and validation of the instrument PKBS-2	**N = 3,258** Community sample: n = 286 (3y age group); n = 577 (4y age group); n = 1,194 (5y age group); n = 1,201 (6y age group)	Total sample: R = 3–6y; M = NR; SD = NR.
	Merrell [132]	Relationship between social skills and internalizing problems (PKBS)	**N = 2,455** Internalizing at-risk group: n = 193 children; Non-internalizing group: n = 2262	Total sample: R = NR; M = NR; SD = NR.
	Merrell [133]	Convergent and discriminant construct validity of PKBS	**N = 295** Study 1: children who had been referred to a specific education child find program: n = 86; Study 2: developmentally delayed: n = 116; Study 3: kindergarten students: n = 46; Study 4: kindergarten students: n = 47	Total sample: R = NR; M = NR; SD = NR. Study 1: R = 3–6y; M = NR; SD = NR. Study 2: R = 3–6y; M = NR; SD = NR. Study 3: R = 5–6y; M = NR; SD = NR. Study 4: R = 5–6y; M = NR; SD = NR.
	Canivez and Rains [134]	Construct validity	**N = 124** (I) normal: n = 107; (II) disabled/at-risk: n = 17	Total sample: R = 5–6y; M = 6.2y; SD = 0.4y (I) R = NR; M = NR; SD = NR (II) R = NR; M = NR; SD = NR
	Canivez and Bordenkircher [135]	Convergent and divergent validity of the PKBS	**N = 154** (I) typical: n = 137; (II) disabled/at-risk: n = 17	Total sample: R = 5–6y; M = 5.4y; SD = 0.5y. (I) R = NR; M = NR; SD = NR. (II) R = NR; M = NR; SD = NR.
	Carney and Merrell [136]	Reliability and comparability of a Spanish-language form of the PKBS	**N = 45** Children fluent in both English and Spanish	Total sample: R = 3–6y; M = NR; SD = NR.
	Wang et al. [137]	Psychometric properties used in ASD population; detecting change in social behavior of children with ASD in an early education program (PKBS)	**N = 22** Children with ASD	Total sample: R = NR; M = NR; SD = NR. At pre-test: R = 36–76m; M = 56.5m; SD = NR. At post-test: R = 42–83m; M = 63.3m; SD = NR.
	Merrell [138]	Development of the PKBS	**N = 2,885** Standardized sample, including: (I) previously identified as being developmentally delayed: n = 239; (II) currently being identified with disabilities: n = 73	Total sample: R = 3–6y; M = 56.5m; SD = NR.
	Carney and Merrell [139]	Construct validity in differentiating ADHD and non-ADHD (PKBS)	**N = 60** (I) ADHD: n = 30; (II) comparison group: n = 30	Total sample: R = 60–83m; M = 56.5m; SD = NR. (I) R = NR; M = 74.8m; SD = 6.6m. (II) R = NR; M = 73.2m. SD = 6.0m.
	Fernández et al. [140]	Structural validity of Spanish language version of PKBS-2	**N = 1,509** Preschoolers	Total sample: R = 3–5y; M = 3.8y; SD = 0.8y.
	Al Awmleh and Woll [141]	Reliability of German language versions of PKBS	**N = 37** Preschoolers	Total sample: R = NR; M = 4.5y; SD = 0.7y.
	Edwards et al. [142]	Unidimensionality and reliability of PKBS	**N = 1,679** Children from a program designed to integrate behavioral health services for children at risk (due to low income) into early childhood education or pediatric service settings	Total sample: R = 25–73m; M = 47.7m; SD = 7.2m.
Peer Social Maturity Scale (PSMS)	Fink et al. [143]	Convergent validity, internal consistency, reliability, factor analysis. Study 1: convergent validity. Study 2: reliability, confirmatory factor analysis, convergent validity.	**N = 454** Children without developmental disabilities. Study 1: (I) Kindergarten: n = 54 (II) Grade 1: n = 44 (III) Grade 2: n = 40; Study 2: Phase 1: n = 114 Phase 2: n = 106 Phase 3: n = 96	Total sample: R = NR; M = NR; SD = NR. Study 1: Total sample: R = NR; M = 6y6m; SD = 10.4m. (I) R = NR; M = 5y7m; SD = 4.3m. (II) R = NR; M = 6y7m; SD = 4.5m. (III) R = NR; M = 7y6m; SD = 4.9m. Study 2: Phase 1: R = 55–77m; M = 5y7m; SD = 5.1m. Phase 2: R = 70–90m; M = 6y8m; SD = 4.8m. Phase 3: R = 80–102m; M = 7y8m; SD = 4.6m.
Questionario de Respostas Socialmente Habilidosas Segundo Relato de Professores (QRSH-PR)	Bolsoni-Silva et al. [144]	Content validity, internal consistency, construct validity, discrimination validity, concurrent validity, predictive validity	**N = 260** (I) behavioral problems: n = 130; (II) social skilled behaviors: n = 130	Total sample: R = NR; M = NR; SD = NR. (I) R = NR; M = NR; SD = NR. (II) R = NR; M = NR; SD = NR.
Social Competence Inventory (SCI)	Rydell et al. [145]	Internal consistency, stability across one year, inter-rater agreement, discriminate peer status, factor analysis	**N = 785** (I) pilot: n = 121; (II) Uppsala sample: n = 123; (III) provincial sample: n = 118; (IV) country sample: n = 423	Total sample: R = 7–10y; M = NR; SD = NR. (I) R = 7–10 y; M = NR; SD = NR. (II) R = NR; M = 7y11m; SD = 4m. (III) R = NR; M = 9y6m; SD = 4m. (IV) R = NR; M = 8y5m; SD = 3y3m.

(Continued)

Table 4. (*Continued*)

Instrument	Reference	Purpose of study	Study population	Age (range[R] and/or Mean [M] Standard deviation [SD])
Social-Emotional Assets and Resilience Scales (SEARS)	Merrell et al. [146]	Development and validation of SEARS (parent form)	**N = 2,356** Children and adolescents, including 10.3% of the children with a disability and 89.7% without a disability	*Total sample:* R = 5–18y; M = NR; SD = NR.
	Merrell et al. [147]	Development and validation of SEARS (teacher form)	**N = 1,673** Children and adolescents, including 17.7% of the students identified and eligible for special education services and 89.7% without a disability	*Total sample:* Kindergarten to Grade 12 (R = NR; M = NR; SD = NR).
	Romer and Merrell [148]	Temporal stability and test-retest reliability of SEARS (children, adolescent, and teacher rating)	**N = 287**(*I*) student self-report sample: *n* = 169 students from general education social studies and science classrooms (49.1% for SEARS-C; 50.9% for SEARS-A); (*II*) teacher report sample: *n* = 118 students	*Total sample:* R = NR; M = NR; SD = NR. (*I*) Grade 6-Grade 8 (R = NR; M = NR; SD = NR). (*II*) kindergarten to Grade 5 (R = NR; M = NR; SD = NR).
	Merrell [149]	Validation of the instrument	**N = 5,555** (*I*) SEARS-C normative sample: *n* = 1,224; (*II*) SEARS-A normative sample: *n* = 1,727; (*III*) SEARS-T normative sample: *n* = 1,400; (*IV*) SEARS-P normative sample: *n* = 1,204	*Total sample:* R = NR; M = NR; SD = NR. (*I*) R = 8–12y; M = 9.9y; SD = 1.2y. (*II*) R = 13–18y; M = 4.4y; SD = 1.9y. (*III*) children: R = 5–12y; M = NR; SD = NR; adolescents: R = 13–18y; M = NR; SD = NR. (*IV*) children: 5–12y; M = 8.7y; SD = 2.2y; adolescents: 13–18y; M = 15.3y; SD = 1.6y.
Vineland Social-Emotional Early Childhood Scales (SEEC)	Sparrow et al. [150]	Development and validation of the instrument	**N = 1,200** Community sample across 36 US states; *n* = 200 in each 0–1y, 1y, 2y, 3y, 4y, & 5y age cohorts	*Total sample:* R = Birth-5y11m; M = NR; SD = NR.
Social Profile (SP)	Donohue [151]	Content validity, construct validity, inter-rater reliability	**N = 187** Typically developing children	*Total sample:* R = 2–5y; M = NR; SD = NR.
	Donohue [152]	Inter-rater reliability	**N = 35**(*I*) substance abuse group: *n* = 11; (*II*) general/geriatric group: *n* = 10; (*III*) preschool group: *n* = 12; (*IV*) senior group: *n* = 2	*Total sample:* R = NR; M = NR; SD = NR.
School Social Behavior Scales-2 (SSBS-2)	Merrell [17]	Development and validation of the instrument	**N = 2,280** Students, including 10.7% with learning difficulty, 2.6% with speech or language impairment, 3.0% with intellectual disability, 4.1% with emotional disturbance, 1.9% with other disabilities, and 22.3% with all disabilities	*Total sample:* Grades K-12 (R = NR; M = NR; SD = NR).
	Raimundo et al. [153]	Analyze the psychometric properties of a Portuguese version of the social competence scale form the SSBS-2	**N = 595** *Sample 1:* *n* = 344 students; *Sample 2:* *n* = 251 students	*Total sample:* R = NR; M = NR; SD = NR. *Sample 1:* R = 6–18y; M = 12.1y; SD = 3.4y. *Sample 2:* R = 8–14y; M = 9.3y; SD = 0.8y
Social Skills Improvement System Rating Scales (SSIS)	Gresham et al. [154]	Agreement of ratings among teachers, parents/caregivers, and students in SSIS	**N = 332** Sample 1: *n* = 168 students; (*II*) *n* = 164 students	*Total sample:* R = NR; M = NR; SD = NR. (*I*) R = NR; M = 11.9y; SD = NR. (*II*) R = NR; M = 9y; SD = NR.
	Gresham and Elliott [155]	Development and validation	**N = 4,700** Children and Adolescents aged 3–18y (Spanish sample: *n* = 486)	*Total sample:* R = 3–18y; M = NR; SR = NR

Note. BC = Between Children; m = months; NR = not reported; y = years.

doi:10.1371/journal.pone.0132299.t004

Table 5. Characteristics of the Instrument for the Assessment of Social Skills.

Instrument	Purpose of instrument	Published year	Type of measure	Number of Subscales/ Forms	Number of items in total	Response Options
ESI [117]	To assess a person's performance of social interaction skills in the natural context with typical social partners during any area of occupation	2009/2014 (latest version)	Observation	1 scale	27	*4-point criterion-referenced rating scale*: 4 = socially appropriate, polite, respectful, and timely
HCSBS [156]	Assess social skills and antisocial behavior across environment	2000	Parent/ caregiver rating	2 subscales	65	*5 point scale*: 1 = never to 5 = frequently
IRS [121]	To measure child's social competence and the caregiver's child rearing competence in caregiver-child interactions	2009	Observation	5 subscales	70	*Behavior items*: 1 = yes or 0 = no. *5-point rating for impression items*: 1 = not evident at all; 2 = not evident; 3 = neutral; 4 = evident; 5 = evident at high level
IRS– Advanced [122]	An advanced version of IRS, measure social competence for individuals over 15 years old	2011	Observation	6 subscales	92	See IRS
IRS–BC [157]	A peer relationship version of the IRS, to measure child-child interaction	2012	Observation	3 subscales	43	See IRS
MESSY-II [125]	To measure both appropriate and inappropriate social behaviors for children	2010	Parent/ (teacher)/ (self-report) rating	1 scale; 3 forms	Parent/ (teacher) rating: 64	*5-point scale*: 1 = not at all to 5 = very much
PKBS-2 [131]	To measure social and emotional problems of children with significant behavioral, emotional and developmental problems	2003	Parent/teacher rating	2 scales	*Social Skill Scale*: 34 *Problem Behavior Scale*: 42	*4-point scale*: 0 = never; 1 = rarely; 2 = sometimes; 3 = often
PSMS [143]	To capitalize on teacher's experience with pupils of varying social competence	2007	Teacher rating	1 scale	7	*7-point rating*: from 1 = very much less mature than the average child this age to 7 = very much more mature than the average child this age
QRSH-PR [144]	To assess social skills in preschool children	2009	Teacher rating	1 scale	24	*3-point Likert scale*: 0 = definitely applies; 1 = applies a little; 2 = does not apply
SCI [98]	To measure social competence	1997	Parent/ teacher rating	2 subscales	29	*5-step response scale*: 1 = doesn't apply at all; 5 = applies very well to the child
SEARS [149]	To assess social and emotional characteristics of children and adolescents focuses on the affective, interpersonal, behavioral, and cognitive aspects of their adjustment	2011	Parent/ teacher/ children/ adolescent rating	4 forms	*Parent/ teacher rating*: 54 *Children/ adolescent rating*: 35	*4-point rating*: 0 = never to 3 = always
SEEC [150]	To assess usual social and emotional functioning for children from birth to 5 year 11 months	1998	Semi-structured interview	3 subscales	*Interpersonal relationship*: 44 *Play and leisure time*: 44 *Coping skills*: 34	0 = never performed; 1 = performed sometimes or with partial success; 2 = usually or habitually performed

(Continued)

Table 5. (Continued)

Instrument	Purpose of instrument	Published year	Type of measure	Number of Subscales/ Forms	Number of items in total	Response Options
SP [158]	To evaluate the social participation levels of children in activity groups	2005	Observation	3 sections/ 2 versions (children, adult / adolescent)	166	*5-point Likert scale* (for each level of social participation, 5 levels): 1 = never; 2 = rarely; 3 = sometimes; 4 = frequently; 5 = almost always
SSBS-2 [153]	To screen and assess social competence and antisocial behavior of students from K to 12 educational settings	2002	Teacher rating	2 Scales	*Social Competence* Scale: 32 *Antisocial Behavior Scale*: 33	1 = never to 5 = frequently
SSIS [155]	To measure social skills and problematic behaviors	2008	Self-rating / parent/ teacher rating	3 forms	*Teacher form*: 83 *Parent form*:79 *Student form*: 75	*4-point scale for skills*: never, seldom, often and almost always; *4-point scale for problems*: not true; a little true, a lot true; very true. *3-point scale for importance in skills*: important, not important, critical; *teacher form*: additional 5-point scale for academic competence

properties were evaluated and it had an overall rating of excellent. Three measures (HCSBS [73.9%]; PKBS-2 [67.6%]; SSIS [74.6%]) evaluated seven of the eight psychometric properties; however, their overall ratings were good (not excellent). Five of the measures had overall excellent ratings, but have evaluated only five psychometric properties (ESI [77.5%]; PSMS [84.2%]), four psychometric properties (QRSH-PR [84.2%], SCI [86.8%], or three psychometric properties (SP [80.3%]) respectively. The IRS is the measure with the least evidence of having sound psychometric properties; having rated only two psychometric properties and achieving an overall fair rating (31.9%).

Reliability and Validity

The COSMIN checklist provides information about the instruments' properties with reliability testing for internal consistency. While these aspects of reliability were not reported for a number of instruments, the current review showed good to excellent reliability for the majority of the instruments. However, only six of the thirteen instruments reported on measurement error. When selecting appropriate outcome measures for a study, consideration of the measurement error of the instruments is important as a small measurement error will allow the instrument to detect smaller treatment effects and allow for stronger conclusions to be drawn. Thus, clinical trials will require smaller sample sizes if the measurement error is small in relation to its minimal important change (MIC), compared with instruments where the opposite applies [159].

The results of the current systematic review revealed considerable variability and range of sample sizes used for the validation and development of measures. For instance, the ESI was developed and validate using a total sample size of 6,552 people classified under various diagnostic categories [117], whereas the IRS-BC [124] was validated with a sample of 20 children. Other measures which were developed and validated with small sample sizes included the IRS-BC [124], IRS-A [122, 123], SP [151, 152]. Large numbers in normative samples used for validation and development increases the generalisability of the results of measures to a population, and allows clinicians to make informed assessments about a client's functioning in

Table 6. Overview of the Psychometric Measurement Properties of Social Skills Instrument.

Instrument	Author(s)	Year	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Average
ESI	Simmons, Griswold [114]	2010[A]	Excellent (90.5)	-	-	-	Excellent (83.3)	Excellent (82.6)	-	-	NA
	Søndergaard and Fisher [115]	2012[A]	-	-	-	-	-	Excellent (82.6)	-	-	NA
	Griswold and Townsend [116]	2012[A]	-	-	-	-	-	Good (52.9)	-	-	NA
	Fisher and Griswold [117]	2014[M]	Excellent (100)	Good (62.1)	Good (61.4)	-	Excellent (100)	Excellent (90.9)	-	-	NA
	TOTAL	**NA**	**Excellent (95.2)**	**Good (62.1)**	**Good (61.4)**	**NR**	**Excellent (91.7)**	**Excellent (77.3)**	**NA**	**NR**	**77.5 (15.9)**
HCSBS	Merrell, Streeter [118]	2001[A]	-	-	-	-	-	-	-	Excellent (88.9)	NA
	Lund and Merrell [119]	2001[A]	-	-	-	-	-	Excellent (76.5)	-	-	NA
	Merrell and Caldarella [120]	1999[A]	Good (71.4)	-	-	-	Fair (41.7)	Good (70.8)	-	-	NA
	Merrell and Caldarella [156]	2000[M]	Excellent (85.9)	Good (74.2)	Excellent (86.3)	Good (57.1)	Good (75.0)	-	-	-	NA
	TOTAL	**NA**	**Excellent (78.7)**	**Good (74.2)**	**Excellent (86.3)**	**Good (57.1)**	**Good (58.3)**	**Good (73.6)**	**NR**	**Excellent (88.9)**	**73.9 (12.4)**
IRS	Anme, Shinohara [121]	2010[A]	-	-	-	-	-	Fair (29.4)	-	-	NA
	TOTAL	**NA**	**NR**	**NR**	**NR**	**NR**	**NR**	**Fair (29.4)**	**NA**	**NR**	**NA**
IRS-Advanced	Anme, Watanabe [122]	2011[A]	-	-	-	-	-	Poor (13.0)	-	-	NA
	Anme, Tokutake [123]	2014[A]	Fair (47.8)	-	-	-	-	Fair (34.8)	-	-	NA
	TOTAL	**NA**	**Fair (47.8)**	**NR**	**NR**	**NR**	**NR**	**Poor (23.9)**	**NA**	**NR**	**NA**
IRS-BC	Anme, Sugisawa [124]	2012[A]	-	-	-	-	-	Fair (30.4)	-	-	NA
TOTAL	**TOTAL**	**NA**	**NR**	**NR**	**NR**	**NR**	**NR**	**Fair (30.4)**	**NA**	**NR**	**31.9 (10.3)**
MESSY-II[b]	Matson, Neal [125]	2010[A]	Good (57.3)	-	-	-	-	Good (60.9)	-	-	NA
	Matson, Neal [126]	2012[A]	Excellent (100)	-	-	-	Excellent (83.3)	-	-	-	NA
	Teodoro, Kappler [128]	2005[A]	Excellent (85.9)	-	-	-	Good (75.0)	Good (54.3)	Fair (39.3)	-	NA
	Méndez, Hidalgo [127]	2002[A]	Excellent (100)	-	-	-	Excellent (100)	Excellent (79.7)	Good (63.5)	-	NA
	Chou [129]	1997[A]	Good (57.3)	-	-	-	Fair (41.7)	-	-	-	NA
	Matson, Horovitz [130]	2013[A]	Excellent (81.0)	-	Excellent (82.9)	-	-	-	Fair (30.4)	-	NA

(Continued)

Table 6. (Continued)

Instrument	Author(s)	Year	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Average
PKBS-2[b]	**TOTAL**	NA	**Excellent (80.3)**	**Excellent (82.9)**	**NR**	**NR**	**Good (75.0)**	**Good (68.5)**	**Fair (44.4)**	**NR**	**70.2 (15.4)**
	Merrell [131]	2003[M]	Excellent (85.9)	Good (68.9)	Fair (48.4)	Excellent (78.6)	-	-	-	-	NA
	Merrell [133]	1995[A]	-	-	-	-	-	Good (73.2)	-	-	NA
	Canivez and Rains [134]	2002[A]	-	-	-	-	-	Excellent (82.6)	-	-	NA
	Canivez and Bordenkircher [135]	2002[A]	-	-	-	-	-	Excellent (82.6)	-	-	NA
	Carney and Merrell [136]	2002[A]	-	-	-	-	-	-	Fair (48.2)	-	NA
	Wang, Sandall [137]	2001[A]	Good (66.9)	-	-	-	-	Good (57.7)	-	-	NA
	Merrell [138]	1996[A]	-	-	-	-	Good (75.0)	-	-	-	NA
	Carney and Merrell [139]	2005[A]	-	-	-	-	-	Fair (47.8)	-	-	NA
	Edwards, Whiteside-Mansell [142]	2003[A]	Excellent (85.9)	-	-	-	Good (75.0)	-	-	-	NA
	Fernández, Benítez [140]	2010[A]	-	-	-	-	Good (75.0)	-	-	-	NA
	Al Awmleh and Woll [141]	2013[A]	Good (71.4)	-	Excellent (75.7)	-	-	-	Fair (36.4)	-	NA
PSMS	**TOTAL**	NA	**Excellent (77.5)**	**Good (68.9)**	**Good (62.1)**	**Excellent (78.6)**	**Good (75.0)**	**Good (68.8)**	**Fair (42.3)**	**NR**	**67.6 (12.6)**
	Fink, Rosnay [143]	2013[A]	Excellent (100)	Excellent (96.6)	-	-	Excellent (100)	Good (51.3)	-	Good (73.2)	NA
	TOTAL	NA	**Excellent (100)**	**Excellent (96.6)**	**NR**	**NR**	**Excellent (100)**	**Good (51.3)**	**NA**	**Good (73.2)**	**84.2 (21.6)**
QRSH-PR	Bolsoni-Silva, Marturano [144]	2009[A]	Excellent (90.5)	-	-	Excellent (85.7)	Excellent (83.3)	Good (62.0)	-	-	NA
	TOTAL	NA	**Excellent (90.5)**	**NR**	**NR**	**Excellent (85.7)**	**Excellent (83.3)**	**Good (62.0)**	**NA**	**NR**	**80.4 (12.6)**
SCI	Rydell, Hagekull [145]	1997[A]	Excellent (100)	Excellent (86.3)	-	-	Excellent (100)	Good (60.9)	-	-	NA
	TOTAL	NA	**Excellent100**	**Excellent (86.3)**	**NR**	**NR**	**Excellent (100)**	**Good (60.9)**	**NA**	**NR**	**86.8 (18.4)**

(Continued)

Table 6. (Continued)

Instrument	Author(s)	Year	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Average
SEARS-P	Merrell, Felver-Gant [146]	2011[A]	Excellent (100)	Good (58.7)	-	Good (64.3)	Excellent (100)	Excellent (81.1)	-	-	NA
	Merrell [149]	2011[M]	Excellent (85.9)	Good (62.1)	-	Good (64.3)	Good (75.0)	Good (63.6)	-	-	NA
	TOTAL	**NA**	**Excellent (93.0)**	**Good (60.4)**	**NR**	**Good (64.3)**	**Excellent (87.5)**	**Good (72.3)**	**NA**	**NR**	**NA**
SEARS-T	Merrell, Cohn [147]	2011[A]	Excellent (85.9)	-	-	Fair (35.7)	Good (75.0)	Good (57.8)	-	-	NA
	Romer and Merrell [148]	2012[A]	-	Good (68.9)	-	-	-	-	-	-	NA
	Merrell [149]	2011[M]	Excellent (85.9)	Good (65.5)	-	Good (64.3)	Good (75.0)	Good (53.4)	-	-	NA
	TOTAL	**NA**	**Excellent (85.9)**	**Good (67.2)**	**NR**	**Fair (50.0)**	**Good (75.0)**	**Good (55.6)**	**NA**	**NR**	**NA**
SEARS-C	Romer and Merrell [148]	2012[A]	-	Good (72.3)	-	-	-	-	-	-	NA
	Merrell [149]	2011[M]	Excellent (85.9)	Good (62.1)	-	Good (64.3)	Good (75.0)	Fair (43.5)	-	-	NA
	TOTAL	**NA**	**Excellent (85.9)**	**Good (67.2)**	**NR**	**Good (64.3)**	**Good (75.0)**	**Fair (43.5)**	**NA**	**NR**	**NA**
SEARS-A	Romer and Merrell [148]	2012[A]	-	Good (72.3)	-	-	-	-	-	-	NA
	Merrell [149]	2011[M]	Excellent (85.9)	Good (62.1)	-	Good (64.3)	Good (75.0)	Fair (48.7)	-	-	NA
	TOTAL	**NA**	**Excellent (85.9)**	**Good (67.2)**	**NR**	**Good (64.3)**	**Good (75.0)**	**Good (56.4)**	**NA**	**NR**	**69.4 (14.2)**
SEEC	Sparrow, Balla [150]	1998[M]	Good (71.4)	Good (62.1)	Good (58.6)	Good (64.3)	Good (75.0)	-	-	-	NA
	TOTAL	**NA**	**Good (71.4)**	**Good (62.1)**	**Good (58.6)**	**Good (64.3)**	**Good (75.0)**	**NR**	**NA**	**NR**	**66.3 (6.8)**
SP	Donohue [151]	2005[A]	-	Excellent (92.8)	-	Excellent (100)	Good (58.3)	-	-	-	NA
	Donohue [152]	2007[A]	-	Good (72.3)	-	-	-	-	-	-	NA
	TOTAL	**NA**	**NR**	**Excellent (82.6)**	**NR**	**Excellent (100)**	**Good (58.3)**	**NR**	**NA**	**NR**	**80.3 (20.9)**
SSBS-2	Merrell [17]	2002[M]	Excellent (85.7)	Excellent (93.1)	Good (65.5)	Excellent (100)	Excellent (100)	Excellent (78.3)	-	Excellent (80.0)	NA
	Raimundo, Carapito [153]	2012[A]	Excellent (90.5)	-	-	-	Excellent (83.3)	-	Good (60.6)	-	NA
	TOTAL	**NA**	**Excellent (88.1)**	**Excellent (93.1)**	**Good (65.5)**	**Excellent (100)**	**Excellent (91.7)**	**Excellent (78.3)**	**Good (60.6)**	**Excellent (80.0)**	**82.2 (13.8)**

(Continued)

Table 6. (*Continued*)

Instrument	Author(s)	Year	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Average
SSIS	Gresham, Elliott [154]	2010^A	-	Excellent (79.1)	-	-	-	Good (70.8)	-	-	NA
	Gresham and Elliott [155]	2008^M	Excellent (87.4)	Excellent (79.1)	Excellent (86.3)	Excellent (85.7)	Fair (50.0)	Good (56.5)	Good (69.9)	-	NA
	TOTAL	**NA**	**Excellent (87.4)**	**Excellent (79.1)**	**Excellent (86.3)**	**Excellent (85.7)**	**Fair (50.0)**	**Good (63.7)**	**Good (69.9)**	**NR**	**74.6 (14.1)**
-	**OVERALL AVERAGE (Mean [SD])**	**NA**	84.5 (12.9)	74.7 (11.9)	70.0 (12.8)	74.0 (17.0)	78.1 (14.8)	56.8 (17.5)	54.3 (13.2)	80.7 (7.9)	NA

Notes. The measurement properties of each instrument were evaluated according to the COSMIN rating. Four-point scale was used (1 = Poor, 2 = Fair, 3 = Good, 4 = Excellent) and the outcome was presented as percentage of rating (Poor = 0–25.0%, Fair = 25.1%–50.0%, Good = 50.1%–75.0%, Excellent = 75.1%–100.0%)

[b]The psychometric properties of the previous version were examined from the studies published after 1994; BC = Between Children; NR: not reported; NA: not applicable.

doi:10.1371/journal.pone.0132299.t006

relation to a representative sample of people with similar characteristics (e.g., age, sex). In contrast, validation studies using a limited sample size are not considered adequate for reaching conclusions about the clinical findings of the measure, as the small number of participants does not allow for informed clinical assessment.

The theoretical construct being measured by an instrument must be clearly defined and then a body of evidence of the instrument's construct validity must be accrued. Within the COSMIN taxonomy, construct validity is comprised of content validity, structural validity and hypothesis testing. Assessment of content validity revealed excellent quality for the PKBS-2, QRSH-PR, SP, SSBS-2 and SISS. However, the SCI, ESI, three versions of the IRS, MESSY-II and PSMS did not provide any evidence of content validity, highlighting a need for further research. Reported structural validity revealed that the SCI, ESI, MESSY-II, four versions of the SEARS, PSMS, PKSB-2, SSBS-2 and SEEC all had published evidence of their structural validity leading to rankings of 'excellent'. Conversely, the three versions of the IRS did not have any published information in this domain; again highlighting a need for further research. The majority of the social skills and behaviour instruments (11 of 13) provided evidence of hypothesis testing ranging at the 'good' to 'excellent' level. Only the SP and the SEEC did not provide any evidence of hypothesis testing. Evidence for both cross-cultural validity (four measures: SSIS, MESSY-II, PKBS-2, and SSBS-2) and criterion validity (three measures: HCSBS, PSMS, and SSBS-2.) were the least reported psychometric property of validity.

When a scale is used without the documented measurement properties (such as construct validity) it can have potentially negative consequences, such as an error in clinical judgment or practitioners inaccurately interpreting assessment findings. Being able to investigate how well a scale measures what it claims to measure and its ability to hold its meaning across varied contexts and sample groups is vital so that it can be used with confidence in clinical settings. This systematic review of social skills and behaviour instruments provides a concise summary of the current state of play of the psychometric properties of these scales.

The importance of external or environmental influences has been emphasised by numerous theorists and researchers; however, few studies have examined the psychometric properties of an instrument in the different environmental contexts within which the social interaction occurs. It is likely that the environment (family, organisational or institutional structures, community, education, and culture) has a substantial impact on the social functioning of children and youth with and without disabilities. Therefore, future investigations of cross-cultural validity would seem valuable in the development of instruments that purport to measure social skills and behaviours.

Definition of Social Functioning as a Construct

In line with discrepancies within the theoretical models and frameworks [1, 19], there was only moderate consensus between the instruments when the stated purposes of all instruments were compared. Among the stated aims of the instruments, the stated purpose was that they measured social competence, child-child interactions, social behaviours, or social and emotional problems, adjustment, or functioning. All terms were in agreement with widely accepted and long standing definitions of the components of social functioning as involving internal or person-related factors (i.e., cognitive, affective, linguistic and personality traits) [57, 160]. However, it remains problematic that a uniform overarching definition of social functioning remains elusive within a body of research that focuses on the reliable, valid, and responsive measurement of social functioning.

This is particularly so when current conceptual models highlight the influence of external factors as well [1]. The purpose of an instrument needs to include a robust definition or

statement about the construct(s) that an instrument seeks to measure [46]. In the 13 instruments evaluated in this review, the articulated constructs may be viewed as the instrument developers' attempt to operationalise an important aspect of social functioning. Before evaluating the merits and weaknesses of the published data about multiple social interaction assessments, the first step is to situate the constructs they measure within theoretical underpinnings. This is necessary whether the instrument measures a single entity (e.g., expressive language) or part of an overarching broader construct (e.g., social functioning). Accordingly, we recommend communication and open discourse among researchers and practitioners who strive to operationalise social functioning as a universally acceptable and defined construct. In the absence of a universally agreed framework, the overlapping yet unclear differentiation of social functioning constructs, collectively described, may prove confusing for both researchers and practitioners. One way to overcome the heterogeneity of definitions is to apply a universally accepted framework with clear interrelated concepts. The International Classification of Functioning [161] has the potential to lend itself to be such a framework.

Application of the International Classification of Functioning (ICF)

The World Health Organization promotes the ICF as a potential guiding framework for professionals, organisations and governments seeking to address social and health inequalities among people with disabilities [161]. The ICF is non-discipline specific, theoretically neutral, and is based on the social model of disability. The ICF offers the advantage of being compatible with current leading psychological models and theorists described previously [1, 16]. Such theories recognise internal factors such as brain integrity and personality (ICF: body structure and function and person factors), external factors such as family, organisational/institutional places (ICF: environmental factors) and engagement and participation within the person's natural environments (ICF: activity participation). Furthermore, evidence of cross-cultural validity testing would provide evidence of the relationships between ICF person and environment factors. Social skills and behaviour, as widely defined in this review, encompasses several key aspects of the ICF including a description of body structure and function (i.e. voice and speech functions); activity and participation factors (learning and applying knowledge, communication, domestic life, interpersonal relations and interactions, and community, social and civic life); person factors (i.e. age, gender, education level, culture); and environmental factors (support and relationships, attitudes, services, systems, and policies).

Limitations

This systematic review has a number of limitations. Information published in languages other than English were not included; therefore, some research findings may have been overlooked. Not all authors who published research on the psychometric properties of social skills and behaviour instruments were directly contacted; therefore, information may have been neglected. Evaluating the quality of responsiveness as a psychometric property was outside the scope of this systematic review, due to the size of this systematic review. We are of the opinion that evaluating the responsiveness of the included instruments warrants a review in itself, given that the number of papers to be evaluated would increase exponentially. Instruments developed for specific clinical populations were outside the scope of this systematic review and they may have sound psychometric qualities for clinical use; further research is needed to evaluate this.

Implications for Practice and Future Research

A number of implications arise from the findings of this systematic review. Measuring social functioning is complex as it involves numerous distinct yet related social skills that are used

during interactions with multiple interactants with varying levels of social competence within a multitude of contexts. Therefore it is unlikely that one measure can address all the assessment needs of researchers and practitioners. As such it is not prudent to recommend one singe measure for use. The SSBS-2 is the social skills and behaviour measure with the most robust psychometric properties. Of particular strength is that all eight psychometric properties have been investigated. The SSBS-2 is recommended for use as a screening measure in educational settings. The PKBS-2 also has sound psychometric properties and is recommended for use as a diagnostic measure of social and emotional problems in children with significant behavioral, emotional and developmental problems. The SSIS has sound psychometric properties, but was clearly developed as an outcome measure; thus needing it to detect change over time following an intervention. Responsiveness of the SSIS is therefore one of the most important psychometric properties. As the review did not evaluate responsiveness, it is not possible to make a recommendation for the purpose it was developed. The HCSBS is another measure with robust psychometric properties and is recommended as a screening measure for home and community contexts. While there are other measures that were appraised in this review that show great promise in terms of the available evidence of the quality of their psychometric properties, more research is needed to evaluate the psychometric properties that have not been reported on to date.

It is important that researchers and practitioners utilise instruments with sound psychometric properties in support of evidence-based and research practices. It is recommended that practitioners collaborate with researchers to further develop the body of knowledge related to the reliability and validity of the social skills and behaviour scales. There is a need for ongoing research in the area of the psychometric properties of social skills and behaviours instruments. The body of psychometric evidence for instruments is dynamic and constantly being added to. It is strongly recommended that a universal working definition of social functioning as an overarching construct plus any related sub-constructs be generated. This would ensure that a consistent approach to evaluating the outcomes of social skills interventions is followed. In particular, it is recommended that the cross-cultural validity and criterion validity of all 13 instruments be further investigated.

There is a need to evaluate the responsiveness of the instruments and therefore to evaluate their suitability for use as an outcome measure of social skills and behaviour. Measures can be prognostically used to report on the responsiveness to a particular intervention or analytically to detect within subject change or between subgroup differences [45]. In an evidence-based practice era for all professionals who may work with clients presenting with impaired social functioning, the development of appropriate and psychometrically sound measurements is crucial to substantiate the effectiveness of interventions and programs. Consequently, instruments require statistical evaluation to determine stability over time in the absence of an intervention, as well as reliability, prior to thorough investigations of responsiveness and sensitivity to change over time. Of the included measures, a considerable number of measures had been validated within the last 10 years. Only the SEEC [150] and SCI [145] had not re-evaluated or updated their psychometric properties within the last 10 years. Furthermore, the SEEC and SCI measures had only been evaluated on a singular occasion. Future evaluation of psychometric properties is needed to determine the stability of these measures over time.

Conclusion

This systematic review presented the results of 45 studies and manuals that reported evidence of the psychometric properties of 13 social skills and behaviour instruments used with children and youth. The COSMIN taxonomy was used to rate the reliability and validity information

reported about the instruments. Three social skills and behaviour scales were found to have the strongest level of psychometric evidence reported in at least seven of the eight psychometric properties that were appraised. The authors recommend that practitioners and researchers consider using the robust SSBS-2, HCSBS and PKBS-2 with children and youth for the purposes and context for which they have been developed. It is also recommended that a more consistent definition of social skills and behaviours as a construct be generated. Only the SSBS-2 has reported on all the psychometric properties evaluated in this systematic review. There is a need for the authors of the measures included in this systematic review to evaluate and report on the quality of the psychometric properties that have not been assessed to date. The body of psychometric evidence of any scale or measure is constantly changing and evolving and it is important for practitioners to be knowledgeable of the best instruments and outcome measures for use when monitoring and assessing children's social functioning.

Supporting Information

S1 Table. PRISMA 2009 Checklist. From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097. (PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: RC RS YWC SWG TB HBT KD AL. Performed the experiments: RC RS YWC SWG TB HBT KD AL. Analyzed the data: SWG YWC. Contributed reagents/materials/analysis tools: RC RS YWC SWG TB HBT KD AL. Wrote the paper: RC RS YWC SWG TB HBT KD AL.

References

1. Beauchamp MH, Anderson V. SOCIAL: an integrative framework for the development of social skills. Psychological Bulletin. 2010; 136(1):39–64. doi: 10.1037/a0017768 PMID: 20063925

2. Cacioppo JT. Social neuroscience: Understanding the pieces fosters understanding the whole and vice versa. American Psychologist. 2002; 57(11):819–31. PMID: 12564179

3. Asher S. Recent advances in the study of peer rejection. In: Asher S, Coie J, editors. Peer rejection in childhood. New York: Cambridge University Press; 1990. p. 3–16.

4. Bagwell CL, Molina BS, Pelham WE, Hoza B. Attention-deficit hyperactivity disorder and problems in peer relations: Predictions from childhood to adolescence. Journal of the American Academy of Child and Adolescent Psychiatry. 2001; 40(11):1285–92. PMID: 11699802

5. Hilton C, Graver K, LaVesser P. Relationship between social competence and sensory processing in children with high functioning autism spectrum disorders. Research in Autism Spectrum Disorders. 2007; 1(2):164–73.

6. Lee LC, David AB, Rusyniak J, Landa R, Newschaffer CJ. Performance of the Social Communication Questionnaire in children receiving preschool special education services. Research in Autism Spectrum Disorders. 2007; 1(2):126–38.

7. Matson JL, Francis KL. Generalizing spontaneous language in developmentally delayed children via a visual cue procedure using caregivers as therapists. Behavior Modification. 1994; 18(2):186–97. PMID: 8002924

8. Matson JL, Matson ML, Rivet TT. Social-skills treatments for children with autism spectrum disorders an overview. Behavior Modification. 2007; 31(5):682–707. PMID: 17699124

9. Matson JL, Wilkins J. Psychometric testing methods for children's social skills. Research in Developmental Disabilities. 2009; 30:249–74. doi: 10.1016/j.ridd.2008.04.002 PMID: 18486441

10. Roff MF, Sells SB, Golden MM. Social adjustment and personality development in children. Minneapolis, MN: University of Minnesota Press; 1972.

11. Matson JL, Boisjoli JA. Differential diagnosis of PDD-NOS in children. Research in Autism Spectrum Disorders. 2007; 1:75–84.

12. Crick NR, Dodge KA. A review and reformulation of social information-processing mechanisms in children's social adjustment. Psychological Bulletin. 1994; 115(1):74–101.

13. Mostow AJ, Izard CE, Fine S, Trentacosta CJ. Modeling Emotional, Cognitive, and Behavioral Predictors of Peer Acceptance. Child Development. 2002; 73(6):1775–87. doi: 10.1111/1467-8624.00505 PMID: 12487493

14. Guralnick MJ. Family and child influences on the peer-related social competence of young children with developmental delays. Mental Retardation and Developmental Disabilities Research Reviews. 1999; 5(1):21–9.

15. Bellack AS, Morrison RL. Interpersonal dysfunction. In: Bellack AS, Hersen M, Kazdin AE, editors. International handbook of behavior modification and therapy. New York: Plenum; 1982. p. 717–48.

16. Gresham FM. Conceptual and definitional issues in the assessment of children's social skills: Implications for classification and training. Journal of Clinical and Child Psychology. 1986; 15(1):3–15.

17. Merrell KW. School Social Behavior Scales, Second Edition. Baltimore, MD: Paul H. Brookes Publishing; 2002.

18. Yeates KO, Bigler ED, Dennis M, Gerhardt CA, Rubin KH, Stancin T, et al. Social outcomes in childhood brain disorder: a heuristic integration of social neuroscience and developmental psychology. Psychological bulletin. 2007; 133(3):535. PMID: 17469991

19. Crowe LM, Beauchamp MH, Catroppa C, Anderson V. Social function assessment tools for children and adolescents: A systematic review from 1988 to 2010. Clinical Psychology Review. 2011; 31 (5):767–85. doi: 10.1016/j.cpr.2011.03.008 PMID: 21513693

20. Yager JA, Ehmann TS. Untangling social function and social cognition: a review of concepts and measurement. Psychiatry: Interpersonal and Biological Processes. 2006; 69(1):47–68.

21. Hill J. Biological, psychological and social processes in the conduct disorders. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2002; 43:133–64.

22. Jorge RE, Robinson RG, Moser D, Tateno A, Crespo-Facorro B, Arndt S. Major depression following traumatic brain injury. Archives of General Psychiatry. 2004; 61:42–50. PMID: 14706943

23. Greenberg MT. Promoting resilience in children and youth: Preventive interventions and their interface with neuroscience. Sciences AotNYAo, editor. New York, NY: New York Academy of Sciences; 2006.

24. McClure EB. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. Psychological Bulletin. 2000; 126:424–53. PMID: 10825784

25. Fujiki M, Spackman MP, Brinton B, Illig T. Ability of children with language impairment to understand emotion conveyed by prosody in a narrative passage. International Journal of Language and Communication Disorders. 2007; 43:330–45.

26. Carlson SM, Koenig MA, Harms MB. Theory of mind. Wiley interdisciplinary reviews Cognitive science (1939–5078). 2013; 4(4):391. doi: 10.1002/wcs.1232

27. Cordier R, Bundy A, Hocking C, Einfeld S. Empathy in the play of children with ADHD. OTJR, Occupation, Participation, and Health. 2010; 30(3):122–32.

28. Feshbach N. Empathy: The formative years: Implications for clinical practice. In: Bohart AC, Greenberg LS, editors. Empathy reconsidered: New directions in psychotherapy. Washington, DC: American Psychological Association; 1997. p. 33–51.

29. Marton I, Wiener J, Rogers M, Moore C, Tannock R. Empathy and social perspective taking in children with attention-deficit/hyperactivity disorder. Journal of Abnormal Child Psychology. 2008; 37:107–18.

30. Adams C. Practitioner review: The assessment of language pragmatics. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2002; 43: 973–87.

31. Van Der Meulen S, Janssen P, Den Os E. Prosodic abilities in children with specific language impairment. Journal of Communication Disorders. 1997; 30:155–70. PMID: 9160237

32. Cordier R, Munro N, Speyer R, Wilkes-Gillan S, Pearce W. Reliability and validity of the Pragmatics Observational Measure (POM): A new observational measure of pragmatic language for children. Research in Developmental Disabilities. 2014; 35:1588–98. doi: 10.1016/j.ridd.2014.03.050 PMID: 24769431

33. Rose-Krasnor L. The nature of social competence: A theoretical review. Social Development. 1997; 6 (1):111–35.

34. Gifford-Smith M, Brownell C. Childhood peer relationships: Social acceptance, friendships, and peer networks. Journal of School Psychology 2003; 41:235–84.

35. Masten AS, Hubbard JJ, Gest SD, Tellegen A, Garmezy N, Ramirez M. Competence in the context of adversity: Pathways to resilience and maladaptation from childhood to late adolescence. Development and Psychopathology. 1999; 11(1):143–69. PMID: 10208360

36. Healy KL, Sanders MR, Iyer A. Facilitative parenting and children's social, emotional and behavioral adjustment. Journal of Child and Family Studies, Early online. 2014:1–18.

37. McLoyd VC. Socioeconomic disadvantage and child development. American Psychologist. 1998; 53 (2):185. PMID: 9491747

38. Kirmayer LJ, Rousseau C, Lashley M. The place of culture in forensic psychiatry. Journal of the American Academy of Psychiatry and the Law Online. 2007; 35(1):98–102.

39. Conner T, Tennen H, Fleeson W, Barrett L. Experience sampling methods: A modern idiographic approach to personality research. Social and Personality Psychology Compass. 2009; 3(3):292–313. PMID: 19898679

40. Cordier R, Brown N, Chen Y-W, Wilkes-Gillan S, Falkmer T. Piloting the use of experience sampling method to investigate the everyday social experiences of children with Asperger syndrome/ high functioning autism. Developmental Neurorehabilitation. 2014;Advance online publication:1–8. doi: 10. 3109/17518423.2014.915244

41. Matson JL, Ollendick T. Enhancing children's social skills: Assessment and training. New York Pergamon Press; 1988.

42. Mrug S, Hoza B, Gerdes AC. Children with attention-deficit/hyperactivity disorder: Peer relationships and peer-oriented interventions. New Directions for Child and Adolescent Development. 2001; 91:51–78. PMID: 11280014

43. Frankel F, Feinberg D. Social problems associated with ADHD vs. ODD in children referred for friendship problems. Child Psychiatry and Human Development. 2002; 33(2):125–46. PMID: 12462351

44. Martin RP, Hooper S, Snow J. Behavior rating scale approaches to personality assessment in children and adolescents. In: Knoff H, editor. The assessment of child and adolescent personality. New York: Guilford Press; 1986. p. 309–51.

45. Wade DT. Assessment, measurement and data collection tools. Clinical Rehabilitation. 2004; 18 (3):233–7. PMID: 15137553

46. Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use 4th ed. Oxford, UK: Oxford University Press; 2008.

47. Demaray MK, Ruffalo SL, Carlson J, Busse R. Social skills assessment: A comparative evaluation of six published rating scales. School Psychology Review. 1995; 24:648–71.

48. Merrell KW. Assessment of children's social skills: Recent developments, best practices, and new directions. A Special Education Journal. 2001; 9(1–2):3–18.

49. Rao PA, Beidel DC, Murray MJ. Social skills interventions for children with Asperger's syndrome or high-functioning autism: A review and recommendations. Journal of Autism and Developmental Disorders. 2008; 38(2):353–61. PMID: 17641962

50. Scourfield J, John B, Martin N, McGuffin P. The development of prosocial behaviour in children and adolescents: A twin study. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2004; 45(5):927–35.

51. Scourfield J, Martin N, Lewis G, McGuffin P. Heritability of social cognitive skills in children and adolescents. The British Journal of Psychiatry. 1999; 175(6):559–64.

52. Waters E, Sroufe LA. Social competence as a developmental construct. Developmental Review. 1983; 3(1):79–97.

53. Priebe S. Social outcomes in schizophrenia. The British Journal of Psychiatry. 2007; 191(50):s15–s20.

54. Bedell JR. Handbook for communication and problem-solving skills training: A cognitive-behavioral approach. New York: John Wiley & Sons; 1997.

55. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. Quality of Life Research. 2010; 19:539–49. doi: 10.1007/s11136-010-9606-8 PMID: 20169472

56. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions:

Explanation and elaboration. Annals of Internal Medicine. 2009; 151(4):W1–W30. doi: 10.7326/0003-4819-151-4-200908180-00136. PubMed PMID: RS_600034819Wexplanationandelaboration.

57. Elliott S, Gresham F. Children's social skills: Assessment and classification practices. Journal of Counseling and Development. 1987; 66:96–9.

58. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta analyses: The PRISMA statement. PLoS Medicine. 2009; 6(6):e1000097.

59. Higgins JP. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley-Black-well; 2008.

60. CRD. Systematic Reviews: CRD's guidance for undertaking reviews in health care. Layerthorpe, York: CRD, University of York; 2008.

61. 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 proper-ties for health-related patient-reported outcomes. Journal of Clinical Epidemiology. 2010; 63(7):737–45. doi: 10.1016/j.jclinepi.2010.02.006 PMID: 20494804

62. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological qual-ity in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. Quality of Life Research. 2012; 21(4):651–7. doi: 10.1007/s11136-011-9960-1 PMID: 21732199

63. Speyer R, Cordier R, Kertscher B, Heijnen BJ. Psychometric properties of questionnaires on func-tional health status in oropharyngeal dysphagia: A systematic literature review. BioMed Research International. 2014. http://dx.doi.org/10.1155/2014/458678.

64. Aman MG, Singh NN, Stewart AW, Field CJ. The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. American Journal of Mental Deficiency. 1985; 89:485–91. PMID: 3993694

65. Guedeney A, Fermanian J. A validity and reliability study of assessment and screening for sustained withdrawal reaction in infancy: The Alarm Distress Baby Scale. Infant Mental Health Journal. 2001; 22:559–75.

66. Matthey S, Guedeney A, Starakis N, Barnett B. Assessing the social behavior of infants: Use of the ADBB scale and relationship to mother's mood. Infant Mental Health Journal. 2005; 26:442–58.

67. Mills AC, Grant DM, Lechner WV, Judah MR. Psychometric properties of the Anticipatory Social Behaviours Questionnaire. Journal of Psychopathology and Behavioral Assessment. 2013; 35:346–55.

68. Krug DA, Arick J, Almond P. Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. Journal of Child Psychology & Psychiatry. 1980; 21:221–9.

69. Bellini S, Hopf A. The development of the Autism Social Skills Profile: A preliminary analysis of psy-chometric properties. Focus on Autism and Other Developmental Disabilities. 2007; 22:80–7.

70. Reynolds CR, Kamphaus RW. Behavior Assessment System for Children manual. Circle Pines, MN: American Guidance Service; 1992.

71. MacPhillamy DJ, Lewinsohn PM. The Pleasant Events Schedule: Studies on reliability, validity, and scale intercorrelation. Journal of Consulting Clinical Psychology. 1982; 50:363–80.

72. Bogels SM, Reith W. Validity of two questionnaires to assess social fears: The Dutch Social Phobia and Anxiety Inventory and the Blushing, Trembling and Sweating Questionnaire. Journal of Psycho-pathology and Behavioral Assessment. 1999; 21:51–66.

73. Bronson MB, Goodson BD, Layzer JI, Love JM. Child Behavior Rating Scale. Cambridge, MA: ABT Associates; 1990.

74. Lim SM, Rodger S, Brown T. Validation of child behavior rating scale in Singapore (Part 1): Rasch analysis. Hong Kong Journal of Occupational Therapy. 2010; 20:52–62.

75. Lim SM, Rodger S, Brown T. Validation of Child Behavior Rating Scale in Singapore (Part 2): Conver-gent and discriminant validity. Hong Kong Journal of Occupational Therapy. 2011; 21:2–8.

76. Groot M, Prins P. Children's social behavior: Reliability and concurrent validity of two self-report mea-sures. Journal of Psychopathology and Behavioral Assessment. 1989; 11:195–207.

77. Deluty RH. Children's Action Tendency Scale: A self-report measure of aggressiveness, assertive-ness, and submissiveness in children. Journal of Consulting and Clinical Psychology. 1979; 47:1061–71. PMID: 512161

78. Michelson L, Wood R. Development and psychometric properties of the Children's Assertive Behavior Scale. Journal of Behavioral Assessment. 1982; 4:3–13.

79. de Bildt A, Mulder EJ, Hoekstra PJ, van Lang ND, Minderaa RB, Hartman CA. Validity of the Chil-dren's Social Behavior Questionnaire (CSBQ) in children with intellectual disability: Comparing the

CSBQ with ADI-R, ADOS, and clinical DSM-IV-TR classification. Journal of Autism and Developmental Disorders. 2009; 39:1464–70. doi: 10.1007/s10803-009-0764-x PMID: 19495951

80. Hartman CA, Luteijn E, Serra M, Minderaa R. Refinement of the children's social behavior questionnaire (CSBQ): An instrument that describes the diverse problems seen in milder forms of PDD. Journal of Autism and Developmental Disorders. 2006; 36:325–42. PMID: 16617405

81. Luteijn E, Luteijn F, Jackson S, Volkmar F, Minderaa R. The Children's Social Behavior Questionnaire for milder variants of PDD problems: Evaluation of the psychometric characteristics. Journal of Autism and Developmental Disorders. 2000; 30:317–30. PMID: 11039858

82. Luteijn E, Jackson S, Volkmar FR, Minderaa RB. The development of the Children's Social Behavior Questionnaire: Preliminary data. Journal of Autism and Developmental Disorders. 1998; 28:559–65. PMID: 9932242

83. Bullis M, Bull B, Johnson P, Johnson B. Identifying and assessing community-based social behavior of adolescents and young adults with EBD. Journal of Emotional and Behavioral Disorders. 1994; 2:173–88.

84. Prinz RJ, Foster S, Kent RN, O'Leary KD. Multivariate assessment of conflict in distressed and non-distressed mother-adolescent dyads. Journal of Applied Behavior Analysis. 1979; 12:691–700. PMID: 541311

85. Endler N, Parker JDA. Coping Inventory for Stressful Situations (CISS). Toronto, Canada: Multi-Health Systems; 1990.

86. Tougas F, Brown R, Beaton AM, St Pierre L. Neosexism among women: The role of personally experienced social mobility attempts. Personality and Social Psychology Bulletin. 1999; 25:1487–97.

87. Goldberg DP. General Health Questionnaire-12 (GHQ-12). London, United Kingdom: nferNelson; 1988.

88. Paskoe J, Ialong NS, Horn W, Reinhart MA, Perradatto D. The reliability and validity of the Maternal Social Support Index. Family Medicine. 1988; 20:271–6. PMID: 3203834

89. Lentz RJ, Paul GL, Calhoun JF. Reliability and validity of three measures of functioning with" hard-core" chronic mental patients. Journal of Abnormal Psychology. 1971; 78:69–76. PMID: 5097098

90. Farina A, Arenberg D, Guskin S. A scale for measuring minimal social behavior. Journal of Consulting Psychology. 1957; 21:265–8. PMID: 13439075

91. Hyler SE, Rieder RO. Personality Disorder Questionnaire-Revised (PDQ-R). New York, NY: New York State Psychiatric Institute; 1987.

92. Morosini P-L, Magliano L, Brambilla L, Ugolini S, Pioli R. Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. Acta Psychiatrica Scandinavica. 2000; 101:323–9. PMID: 10782554

93. Lang FR, Heckhausen J. Perceived control over development and subjective well-being: Differential benefits across adulthood. Journal of Personality and Social Psychology. 2001; 81:509–23. PMID: 11554650

94. Loeber R, Stouthamer-Loeber M, Van Kammen WB, Farnington DP. Development of a new measure of self-reported antisocial behavior for young children: Prevalence and reliability. In: Klein MW, editor. Cross-national research in self-reported crime and delinquency. Boston, MA: Kluwer; 1989. p. 203–25.

95. Brown LJ, Black DD, Downs JC. School Social Skills Rating Scale. New York: Slosson Educational Publications; 1984.

96. Weissman MM, Bothwell S. Assessment of social adjustment by patient self-report. Archives of General Psychiatry. 1976; 33:1111–5. PMID: 962494

97. Gully KJ. The Social Behavior Inventory for children in a child abuse treatment program: Development of a tool to measure interpersonal behavior. Child Maltreatment. 2001; 6:260–70. PMID: 11471633

98. Stephens TM, Arnold KD. Social Behavior Assessment Inventory: Professional manual. Odessa, FL: Psychological Assessment Resources 1992.

99. Hays RD, Hayashi T, Stewart AL. A five-item measure of Socially Desirable Response Set. Educational and Psychological Measurement. 1989; 49:629–36.

100. Ajzen I, Fishbein M. Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice Hall; 1980.

101. Turner SM, Beidel DC, Dancu CV, Stanley MA. An empiracally derived inventory to measure social fears and anxiety: The Social Phobia and Anxiety Inventory. Psychological Assessment. 1989; 1:35–40.

102. Corney RH, Clare AW. The construction, development, and testing of a self-report questionnaire to identify social problems. Psychological Medicine. 1985; 15:637–49. PMID: 4048322

103. Mullenkamp A, Gress L, Flood M. Perception of life change events by the elderly. Nursing Research. 1975; 29:109–13.

104. Bettencourt BA, Sheldon K. Social roles as mechanisms for psychological need satisfaction within social groups. Journal of Personality and Social Psychology. 2001; 81:1131–43. PMID: 11761313

105. Galeazzi A, Franceschina E, Holmes GR. Factor analysis of social skills inventory responses of Italians and Americans. Psychological Reports. 2002; 90:1115–21. PMID: 12150394

106. Riggio RE, Carney DR. Social Skills Inventory manual, 2nd edition. Menlo Park, CA: Mind Garden; 2003.

107. Burt SA, Donnellan MB. Development and validation of the Subtypes of Antisocial Behavior Questionnaire. Aggressive Behavior. 2009; 35(5):376–98. doi: 10.1002/ab.20314 PMID: 19618380

108. Inglés CJ, Hidalgo MD, Méndez FX, Inderbitzen HM. The Teenage Inventory of Social Skills: Reliability and validity of the Spanish translation. Journal of Adolescence. 2003; 26:505–10. PMID: 12887938

109. Inderbitzen HM, Foster SL. The teenage inventory of social skills: Development, reliability, and validity. Psychological Assessment. 1992; 4:451–9.

110. Locke J, Kasari C, Wood JJ. Assessing social skills in early elementary-aged children with autism spectrum disorders: The Social Skills Q-Sort. Journal of Psychoeducational Assessment. 2014; 32:62–76.

111. Waksman SA. The Waksman Social Skills Rating Scale. Portland, OR: ASIEP Education; 1985.

112. Walker HM, McConnell SR. The Walker–McConnell Scale of Social Competence and Social Adjustment: A social skills rating for teachers. Austin, TX: Pro-Ed; 1988.

113. Crothers LM, Schreiber JB, Field JE, Kolbert JB. Development and measurement through confirmatory factor analysis of the Young Adult Social Behavior Scale (YASB): An assessment of relational aggression in adolescence and young adulthood. Journal of Psychoeducational Assessment. 2008; 27:17–28.

114. Simmons CD, Griswold LA, Berg B. Evaluation of social interaction during occupational engagement. American Journal of Occupational Therapy. 2010; 64(1):10–7. PMID: 20131560

115. Søndergaard M, Fisher AG. Sensitivity of the evaluation of social interaction measures among people with and without neurologic or psychiatric disorders. American Journal of Occupational Therapy. 2012; 66:356–62. doi: 10.5014/ajot.2012.003582 PMID: 22549601

116. Griswold LA, Townsend S. Assessing the sensitivity of the Evaluation of Social Interaction: Comparing social skills in children with and without disabilities. American Journal of Occupational Therapy. 2012; 66:709–17. doi: 10.5014/ajot.2012.004051 PMID: 23106991

117. Fisher AG, Griswold LA. Evaluation of Social Interaction (3rd ed.). Fort Collins, CO: Three Star Press; 2014.

118. Merrell KW, Streeter AL, Boelter EW, Caldarella P, Gentry A. Validity of the home and community social behavior scales: Comparisons with five behavior-rating scales. Psychology in the Schools. 2001; 38(4):313–25.

119. Lund J, Merrell KW. Social and antisocial behavior of children with learning and behavioral disorders: Construct validity of the Home and Community Social Behavior Scales. Journal of Psychoeducational Assessment. 2001; 19(2):112–22.

120. Merrell KW, Caldarella P. Social-behavioral assessment of at-risk early adolescent students: Psychometric characteristics and validity of a parent report form of the School Social Behavior Scales. Journal of Psychoeducational Assessment. 1999; 17(1):36–49.

121. Anme T, Shinohara R, Sugisawa Y, Tong L, Tanaka E, Watanabe T, et al. Interaction Rating Scale (IRS) as an evidence-based practical index of children's social skills and parenting. Journal of Epidemiology. 2009; 20(Suppl 2):S419–26.

122. Anme T, Watanabe T, Tokutake K, Tomisaki E, Mochizuki Y, Tanaka E, et al. A Pilot study of social competence assessment using Interaction Rating Scale Advanced. ISRN Pediatrics. 2011; 2011:Article ID 272913. doi: 0.5402/2011/272913.

123. Anme T, Tokutake K, Tanaka E, Watanabe T, Tomisaki E, Mochizuki Y, et al. Validity and reliability of the Interaction Rating Scale Advanced (IRSA) as an index of social competence development. Public Health Research. 2014; 4(1):25–30.

124. Anme T, Sugisawa Y, Shinohara R, Matsumoto M, Watanabe T, Tokutake K, et al. Validity and reliability of the Interaction Rating Scale between Children (IRSC) by using motion capture analysis of head movement. Public Health Research. 2012; 2(6):208–12.

125. Matson JL, Neal D, Fodstad JC, Hess JA, Mahan S, Rivet TT. Reliability and validity of the Matson Evaluation of Social Skills with Youngsters. Behavior Modification. 2010; 34(6):539–58. doi: 10.1177/0145445510384844 PMID: 20935234

126. Matson JL, Neal D, Worley JA, Kozlowski AM, Fodstad JC. Factor structure of the Matson Evaluation of Social Skills with Youngsters-II (MESSY-II). Research in Developmental Disabilities. 2012; 33 (6):2067–71. doi: 10.1016/j.ridd.2010.09.026 PMID: 22750669

127. Méndez FX, Hidalgo MD, Inglés C. The Matson Evaluation of Social Skills with Youngsters: Psychometric properties of the Spanish translation in the adolescent population. European Journal of Psychological Assessment. 2002; 18:30–42.

128. Teodoro MLM, Kappler K, de Lima Rodrigues J, de Freitas PM, Haase VG. The Matson Evaluation of Social Skills with Youngsters (MESSY) and its adaptation for Brazilian children and adolescents. Interamerican Journal of Psychology. 2005; 39:239–46.

129. Chou K-L. The Matson Evaluation of Social Skills with Youngsters: Reliability and validity of a Chinese translation. Personality and Individual Differences. 1997; 22:123–5.

130. Matson JL, Horovitz M, Mahan S, Fodstad J. Reliability of the Matson Evaluation of Social Skills with Youngsters (MESSY) for children with autism spectrum disorders. Research in Autism Spectrum Disorders. 2013; 7:405–10.

131. Merrell KW. Preschool and Kindergarten Behavior Scales: Second Edition (PKBS-2). Austin, TX: Pro-ed; 2003.

132. Merrell KW. An investigation of the relationship between social skills and internalizing problems in early childhood: Construct validity of the Preschool and Kindergarten Behavior Scales. Journal of Psychoeducational Assessment. 1995; 13:230–40.

133. Merrell KW. Relationships among early childhood behavior rating scales: Convergent and discriminant construct validity of the Preschool and Kindergarten Behavior Scales. Early Education and Development. 1995; 6:253–64.

134. Canivez GL, Rains JD. Construct validity of the Adjustment Scales for Children and Adolescents and the Preschool and Kindergarten Behavior Scales: Convergent and divergent evidence. Psychology in the Schools. 2002; 39:621–33.

135. Canivez GL, Bordenkircher SE. Convergent and divergent validity of the Adjustment Scales for Children and Adolescents and the Preschool and Kindergarten Behavior Scales. Journal of Psychoeducational Assessment. 2002; 20:30–45.

136. Carney AG, Merrell KW. Reliability and comparability of a Spanish-language form of the Preschool and Kindergarten Behavior Scales. Psychology in the Schools. 2002; 39:367–73.

137. Wang H-T, Sandall SR, Davis CA, Thomas CJ. Social skills assessment in young children with autism: A comparison evaluation of the SSRS and PKBS. Journal of Autism and Developmental Disorders. 2011; 41:1487–95. doi: 10.1007/s10803-010-1175-8 PMID: 21225453

138. Merrell KW. Social-emotional assessment in early Childhood: The Preschool and Kindergarten Behavior Scales. Journal of Early Intervention. 1996; 20:132–45.

139. Carney AG, Merrell KW. Teacher ratings of young children with and without ADHD: Construct validity of two child behavior rating scales. Assessment for Effective Intervention. 2005; 30:65–75.

140. Fernández M, Benítez JL, Pichardo MC, Fernández E, Justicia F, García T, et al. Confirmatory factor analysis of the PKBS-2 subscales for assessing social skills and behavioral problems in preschool education. Electronic Journal of Research in Educational Psychology. 2010; 8(3):1229–52.

141. Al Awmleh A, Woll A. Reliability of the German language version of the Preschool and Kindergarten Behavior Scales Second Edition. Journal of Social Sciences. 2013; 9:54–8.

142. Edwards MC, Whiteside-Mansell L, Conners NA, Deere D. The unidimensionality and reliability of the preschool and kindergarten behavior scales. Journal of Psychoeducational Assessment. 2003; 21:16–31.

143. Fink E, Rosnay M, Peterson C, Slaughter V. Validation of the Peer Social Maturity Scale for assessing children's social skills. Infant and Child Development. 2013; 22:539–52.

144. Bolsoni-Silva AT, Marturano EM, Loureiro SR. Construction and validation of the Brazilian questionário de respostas socialmente habilidosas segundo relato de professores (QRSH-PR). The Spanish Journal of Psychology. 2009; 12(1):349–59. PMID: 19476246

145. Rydell A-M, Hagekull B, Bohlin G. Measurement of two social competence aspects in middle childhood. Developmental Psychology. 1997; 33(5):824–33. PMID: 9300215

146. Merrell KW, Felver-Gant JC, Tom KM. Development and validation of a parent report measure for assessing social-emotional competencies of children and adolescents. Journal of Child and Family Studies. 2011; 20(4):529–40.

147. Merrell KW, Cohn BP, Tom KM. Development and validation of a teacher report measure for assessing social-emotional strengths of children and adolescents. School Psychology Review. 2011; 40 (2):226–41.

148. Romer N, Merrell KW. Temporal stability of strength-based assessments: Test–retest reliability of student and teacher reports. Assessment for Effective Intervention. 2012; 38(3):185–91.

149. Merrell KW. SEARS: Social Emotional Assets and Resilience Scales. Lutz, FL: Psychological Assessment Resources; 2011.

150. Sparrow SS, Balla DA, Cicchetti DV. Vineland SEEC Social-Emotional Early Childhood Scales. Bloomington, MN: NCS Pearson, Inc.; 1998.

151. Donohue MV. Social profile: Assessment of validity and reliability with preschool children. Canadian Journal of Occupational Therapy. 2005; 72(3):164–75.

152. Donohue MV. Interrater reliability of the Social Profile: Assessment of community and psychiatric group participation. Australian Occupational Therapy Journal. 2007; 54(1):49–58.

153. Raimundo R, Carapito E, Pereira AI, Pinto AM, Lima ML, Ribeiro MT. School Social Behavior Scales: An adaptation study of the Portuguese version of the Social Competence Scale from SSBS-2. The Spanish Journal of Psychology. 2012; 15:1473–84. PMID: 23156949

154. Gresham FM, Elliott SN, Cook CR, Vance MJ, Kettler R. Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System—Rating Scales. Psychological Assessment. 2010; 22(1):157–66. doi: 10.1037/a0018124 PMID: 20230162

155. Gresham FM, Elliott SN. Social skills improvement system (SSIS) rating scales. Bloomington, MN: Pearson Assessments. 2008.

156. Merrell KW, Caldarella P. Home & Community Social Behavior Scales. Iowa City, IA: Assessment Intervention Resources; 2000.

157. Anme T, Tokutake K, Tanaka E, Watanabe T, Tomisaki E, Mochizuki Y, et al. Short version of the Interaction Rating Scale Advanced (IRSA-Brief) as a practical index of social competence development. International Journal of Applied Psychology. 2013; 3(6):169–73.

158. Donohue MV. Social Profile manual. Bethesda, MD: American Occupational Therapy Association; 2013.

159. Guyatt GH, Walter S, Norman GR. Measuring change over time: assessing the usefulness of evaluative instruments. Journal of Chronic Diseases. 1987; 40:171–8. PMID: 3818871

160. Gresham FM. Conceptual issues in the assessment of social competence in children. In: Strain P, Guralnick M, Walker H, editors. Cildren's social behaviour: Development, assessment, and modification. New York: Academic; 1986. p. 143–79.

161. WHO. International classification of functioning, disability, and health. Geneva, Switzerland: World Health Organization; 2001.