

Instruments evaluating child outcomes used in evidence-based family support programs: A scoping review

Ana Uka^{a,1,*}, Elisabeth Stefanek^{b,1}, Daiva Skučienė^c, Carmen Schneckenteiter^d, Georg Spiel^{b,e}

^a Faculty of Technical Medical Sciences, Department of Nursing and Physiotherapy, Western Balkans University, Highway Tirana-Durres, KM 7, Kashar, Tirana, Albania

^b Pro Mente Forschung [Pro Mente Research], Pilgramgasse 8/2/29, 1050 Vienna, Austria

^c Faculty of Philosophy, Institute of Sociology and Social Work, Vilnius University, Universiteto St.9, LT-01513 Vilnius, Lithuania

^d Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

^e Pro Mente Forschung [Pro Mente Research], Pro Mente: Kinder Jugend Familie [Pro Mente: Child Youth Family], Villacherstraße 161, 9020 Klagenfurt, Austria

ARTICLE INFO

Keywords:

Scoping review
Evidence-based family support programs
Child outcome instruments
Internal consistency
Cronbach's alpha

ABSTRACT

This scoping review aims to identify instruments that measure child outcomes assessed in evidence-based family support programs and to investigate reported differences in the magnitude of Cronbach's alpha by type of study participant (i.e., community, clinical, at-risk samples). We analyzed publications extracted from three databases, which were then narrowed down to 77 articles. The most used interventions were Triple P (23 studies), Incredible Years (13), and Parent Management Training (13) conducted mainly in Europe (35) and (North) America (25). A total of 30 studies were conducted with clinical, 22 with at-risk, and 22 with community samples. The most used instruments with parents as the respondents were the Eyberg Child Behavior Inventory (32), the Strengths and Difficulties Questionnaire (25), and the Child Behavior Checklist (19). The most used instrument with children/adolescents as the respondents was the Child Depression Inventory (5), and for teachers or other professionals it was the Teacher Report Form (9). Regarding Cronbach's alpha, one-third of the studies did not report any information, one-third yielded mixed findings, and one-third reported good values. Furthermore, it became evident that information regarding Cronbach's alpha was often incomplete or missing, especially in studies conducted with clinical and at-risk samples. Further research is needed to investigate why there is a bias in reporting Cronbach's alpha. This work recommends that future studies emphasize the importance of reporting the psychometric properties of the instruments used to be able to properly compare different studies across different populations, especially when used to measure children's outcomes.

1. Introduction

Families are the most important context for children's development in all cultures. Generally speaking, most families are doing well in offering children a caring and a promoting environment and bringing up healthy and resilient children. However, due to manifold challenges, families can become unsafe places, especially for children, and such families need support to face these challenges. Researchers have developed and evaluated interventions to support families overcome such difficulties. Usually, family support interventions aim at assisting

the entire family unit. Family support programs are defined as "a wide range of activities with a clear structure, set goals, clearly defined means and methods, and a structure/guideline to assess outcomes that aim to promote the health, well-being, and development of children" (Özdemir, Vastamäki, Leijten, & Sampaio, *in press*). To ensure that these programs are beneficial for all users, there is a need for evidence-based programs, (i.e., programs with scientific evidence for their effectiveness through outcome evaluations; Hidalgo et al., 2023). One type of such programs are evidence-based parenting programs which aim to develop more effective parenting practices to assist parents to manage their child's

* Corresponding author at: Faculty of Technical Medical Sciences, Department of Nursing and Physiotherapy, Western Balkans University, Highway Tirana-Durres, KM 7, Kashar, Tirana, Albania.

E-mail addresses: ana.uka@wbu.edu.al (A. Uka), elisabeth.stefanek@promente-forschung.at (E. Stefanek), daiva.skuciene@fsf.vu.lt (D. Skučienė), carmen.schneckenreiter@univie.ac.at (C. Schneckenteiter), georg.spiel@promente-kijufa.at (G. Spiel).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.childyouth.2024.107903>

Received 2 June 2023; Received in revised form 18 August 2024; Accepted 5 September 2024

Available online 7 September 2024

0190-7409/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

behavior as well as improve the emotional and behavioral adjustment of children (Gray, Totsika & Lindsay, 2018). A number of randomized controlled trials have established the evidence-base of these programs. According to the standards of evidence criteria of the Society for Prevention Research (Flay et al., 2005; Gottfredson et al., 2015), evaluation of program effectiveness should be done using psychometrically robust measures, valid data collection procedures, and appropriate and rigorous data analyses. However, psychometric properties, especially the internal consistency of measurements, depend on the cultural context and characteristics of the sample (i.e., community, clinical and at-risk samples). Thus, it would be interesting to know what the most commonly used measures of children's behavior and adjustment in evidence-based family support programs in different samples are. Another interest is to explore the effects of these programs on children's outcomes from a multi-informant perspective. Typical informants include children, parents and teachers. A key point of a multi-informant approach is that it informs our understanding of a child's behavioral and emotional adjustment from multiple perspectives and contexts. Furthermore, we are interested in the reported differences in Cronbach's alpha, as an indicator of internal reliability of measures, by type of study sample.

Family support programs are considered to be interventions designed to improve or alter parental performance through the provision of training, support, or education focusing on fostering the well-being of the children of these parents (Morris et al., 2017; Rodrigo, Almeida, Spiel, & Koops, 2012). Programs vary in their targets, methods, and scope. Many programs have been developed to provide support for parents struggling to meet their families' needs, focusing on positive parenting and effective support for children (e.g., Sanders & Turner, 2019; Newland, 2015; Rodrigo, 2016). Furthermore, several programs exist in which the whole family is taken into account – with measurements for children and parents – focusing on child behavior and family functioning (Kumpfer, Whiteside, Greene, & Allen, 2010). Family support programs are widely used in different sociocultural contexts and implemented in different formats to help parents improve the quality of their parenting experiences. In general, programs are designed for all families (i.e., universal) or for families with specific needs (i.e., targeted). Within prevention/intervention programs that target families, three types of programs can be differentiated: 1) programs designed for everyone (i.e., universal or primary prevention programs); 2) programs designed to support families at risk (i.e., targeted or secondary prevention/intervention programs, for example, parenting in a disadvantaged neighborhood); and 3) indicated/tertiary programs focused on families that are experiencing adverse outcomes (Simeonsson, 1991).

Programs that fulfill standards of evidence as formulated by committees of prevention scientists (Corcoran, 2009; Flay et al., 2005; Gottfredson et al., 2015) are efficacious interventions with a clear rationale, a theoretical framework that describes causes and consequences, psychometrically sound outcome measures, and use multiple measures and/or sources. Furthermore, the causal theory of the intervention should be tested, and valid causal inferences should be drawn from a rigorous research design (i.e., a randomized controlled trial or quasi-experimental design). An effective intervention needs to be implemented and evaluated under real-world conditions with the specific target population, and the effects must be practically relevant.

Thus, to prove the effectiveness of programs, robust instruments are necessary and are determined by reliability, validity, and responsiveness. Reliability refers to whether an instrument reproduces similar results in different settings (populations) and the degree to which the instrument is free from measurement errors. One way reliability is explored for measures is through examination of internal consistency, that is the extent to which items within a scale are related to each other and appear to measure the same construct. A high degree of internal consistency is desirable because it "*speaks directly to the ability of the clinician or the researcher to interpret the composite score as a reflection of the test's items*" (Henson, 2001, p. 178). Despite several methodological

criticisms in the past 20 years, Cronbach's alpha (CA) is the most used reliability index in empirical studies (McNeish, 2018). One of the main criticisms of the CA is that it is a characteristic of data obtained from a particular group and not an inherent characteristic of a measure (Brennan, 2001) and should therefore be reported each time the measure is used. In particular, as programs' target groups differ with respect to their cultural or psychosocial backgrounds, and because instruments developed for the general population might work differently for clinical or at-risk populations, the CA varies among these populations (Shevlin, Miles, Davis, & Walker, 2000). Although validity i.e., whether the instrument assesses what it is intended to assess, including aspects like content, structural, construct, and criterion validity, (de Vet, Terwee, Mokkink, & Knol, 2015) is a relevant consideration, it is often not reported in effectiveness trials and is therefore not the focus of this review.

As children's well-being is the ultimate goal of interventions, it is important to review the child outcome measures and assess several perspectives of the child's well-being and adjustment, and it is necessary to include self-reports, parent reports, and other caregiver reports. Various instruments exist that are either widely used, such as the Eyberg Child Behavior Inventory (Eyberg & Pincus, 1999), newly developed, or adapted for a specific program (e.g., Strengthening Families). Furthermore, as programs are implemented in different cultural contexts, instruments are translated into the language of the country where the program is implemented, and some scales/items might be adapted for specific population groups. Although several studies have analyzed the psychometric properties of child outcome measurements used in family support programs (Bentley, Hartley, & Bucci, 2019), there is a need to explore in depth the psychometric properties of measurements used in different cultural contexts and with different populations (Abrahamse et al., 2015; Marti, Pourat, Lee, & Zima, 2022; Sorsa, Fontell, Laajasalo, & Aronen, 2019).

This scoping review aims to document the most frequently used child outcome measures of children's behavior and adjustment for evaluating the effects of evidence-based family support programs. Apart from identifying the instruments, this review aims to describe the internal consistency reported in each study and to analyze the differences in the magnitude of the CA by type of study participant. Furthermore, this analysis will help identify the most solid and methodologically unbiased instruments. Answering these questions is relevant for providing an overview of the quality of the general instrument standards needed for evidence-based programs (Bornstein, Kotler, & Lansford, 2022).

The following main questions will guide this review paper:

1. Which child outcome measurements (self-reports, parent reports, and other caregiver reports) are used to assess changes in child outcomes in evidence-based family support programs?
2. Are there differences in reported CA depending on the type of sample (i.e., community, clinical, at-risk) and type of informant (i.e., self-reports, parent reports, and other caregiver reports)?

2. Methods

This work is conducted within the framework of the "Pan-European Family Support Research Network: A bottom-up, evidence-based, and multidisciplinary approach" project (CA18123 EurofamNet), supported by the European Cooperation in Science and Technology program (<https://www.cost.eu>). EurofamNet is a collaborative initiative involving key actors in family support from across Europe aimed at providing evidence-informed responses at the European level. The current study was conducted within the deliverables of Working Group 3, "Quality standards and evidence-based programs," a subgroup compiling sound family evaluation tools with the ultimate goals of facilitating further development of evidence-based programs in accordance with the emerging European quality standards for family support and stimulating supportive policies at the European and national levels. Additionally, this work was developed in accordance with EurofamNet's

position on multimethodologies and diverse cultural approaches for the evaluation of evidence-based family support programs (Almeida, Cruz, & Canário, 2022). This study focused on evaluation tools that measure child outcomes in evidence-based family support programs. It used the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA-ScR; Tricco et al., 2018) guidelines and statements for the scoping review (Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009). Since this is a review article, ethical approval from institutional and/or national ethical review committees was not required.

2.1. Eligibility criteria

The eligibility criteria were studies written in the English language using quantitative standardized measurements with self-, teacher-, parent-, caregiver-, or peer-reported data. The final stage of the article selection consisted of reading the full text and making selections based on the following criteria: i) evidence-based family support programs; ii) interventions with child outcomes; iii) standardized instruments (i.e., whose validity has been reported in several publications; Corocan, 2009); iv) evaluation studies with pre–post testing; and v) face-to-face group intervention programs (not virtual, phone-based, internet-based, or only watching videos without experts discussing them with the parents). In this stage, three authors independently selected the final articles. The authors discussed any disagreements to reach a consensus. The total number of articles included in this study was 77.

2.2. Search strategy

The literature search was conducted in three scientific databases: SCOPUS, the Web of Science, and PsycInfo. The first literature search was performed in September 2021. Study selection was performed at all stages of the search based on the eligibility criteria. The strategy of the first search was that the potential abstracts and the keywords included at least one of the following keywords: instrument or assessment, or measure*, and evaluation, and parent* or famil* or patern* or matern* or mother or father and child* or adolescent* or teenager*, and social development or social skills or social interaction or behavior* or emotional development or educational skills or educational competence or learning or cognitive function*, and program* or intervention, and evidence*.

Next, the main disciplines searched for potential articles included psychology, sociology, social work, psychiatry, nursing, and health care. The articles included were then restricted to English-language articles published in the last 10 years (i.e., 2012–2022).

Additionally, evaluation studies from the records of evidence-based parenting programs (n = 41) were identified from the Blueprint programs² based on extra searching of other sources, and all were included based on the full-text screening. First, programs were identified by selecting “Parent Training” in the section of program specifics, yielding 25 programs. Only programs that targeted parents and their children older than 5 years were selected (e.g., Incredible Years and Triple P). Second, from these selected programs, the evaluation studies listed on the Blueprints website were chosen, resulting in 41 records.

2.3. Selection of sources of evidence

All the records identified in the database searches were exported to Endnote Web. All the publications were peer-reviewed. The initial screening was refined by excluding articles from journals that focused on medical treatment and studies of old-age patients. Then, review articles, position papers, and papers with titles including irrelevant keywords were excluded. Afterwards, the first three authors independently screened the articles based on the abstracts and methods sections.

During this process, the three authors discussed each article where they had dilemmas or disagreements, and then reached a consensus. For all eligible studies data extraction was performed independently and in duplicate by three researchers.

2.4. Data charting process

PRISMA-ScR guided the data charting and the results reporting. A data charting template was developed by three researchers to determine the variables that would be considered. Information was collected on the instruments (including the number of items and any subscale), the author(s), year of publication, type of intervention, the number of studies, study population, continent, and psychometric properties (i.e., the CA evaluation). The researchers independently charted the data from the selected articles. An essential factor discussed extensively was the level of detail provided in the articles regarding the reliability of each scale and of the whole instrument. The data charting table was collated by three researchers. The main objective of this work was to provide a complete review of the studies whose instruments could be further compared quantitatively. Some articles reported the use of up to three or more instruments, and we therefore charted the reliability of each instrument and its scales.

2.5. Data items

All instruments were coded based on the informants (i.e., self-reported (child/adolescent), parents/primary caregivers, and teachers/psychologists/other professionals, as well as whether single or multiple informants were used to measure the child outcomes. The countries of intervention were categorized on the basis of continents, i.e., (North) America, Asia, Australia, and Europe (there were no studies from Africa). The participants were categorized into community, clinical, and at-risk samples. A clinical sample is a group of people studied for public health reasons (Smyth & Arigo, 2009). Families at risk face crisis situations, stressful events, parental stress, or lack of adequate parenting skills and thus cannot adequately meet their children’s needs necessary for healthy development (Rodrigo, Byrne, & Álvarez, 2012).

2.6. Synthesis of results

Psychometric properties are the metrics used to quantitatively evaluate an instrument and compare these instruments across different studies. As most studies had sparse information on validity and reliability, we decided to assess only the reported internal consistency quality based on the CA, as this was provided in most studies. However, there were several studies that did not report any CA, several studies only reported CA of the whole instrument and other studies reported CA only for some scales. Thus, we classified the quality of the CA as follows: i) very good: the CA for every scale of the instrument is reported and is over 0.70; ii) good: the CA of every scale is reported, but one or two scales are below 0.70, or not all scales are reported but those reported are over 0.70; iii) mixed findings: only the CA of the whole instrument is reported and is over 0.70; if it is not clear whether the instrument consists of several scales; or all scales are reported, but three or more scales are below 0.70, and iv) not reported: no information regarding the CA for the current study is available.

First, we describe the studies according to their context and sample, as well as the use of single or multiple informants. Second, we examine all the instruments used in these studies separately for each respondent (i.e., children/adolescents, parents/other primary caregivers, teachers/other professionals) regarding the information on CA. Third, we examine the reported internal reliability estimates for the instruments and whether they differed with respect to the sample (i.e., community, clinical, and at-risk).

² <https://www.blueprintsprograms.org/>.

3. Results

3.1. Selection of sources of evidence

A total of 395 articles were identified in the PsycInfo database, 471 articles in SCOPUS, and 386 articles in the Web of Science database, constituting a total of 1252 articles. The latest search was conducted in December 2022 due to an expansion of PsycInfo database, resulting in an addition of 240 more articles that could be included with a new total of 1492, as new studies were published during 2022. After removing duplicates, the final number of articles identified was 920.

The screening process included several steps considering the eligibility and exclusion criteria. First, conferences, dissertations, reports, and handbooks were excluded, leaving 733 articles. The first reason is that every evidence-based family support program would probably be submitted for publication to journals, even if it was included in a dissertation, and the second reason is that we wanted to analyze results that were peer reviewed. Second, journal screening was conducted and those that were found to be out of the scope of the review were removed (e.g., palliative medicine, economic, and health policy, gerontology and geriatrics, forensics, obstetrics, schizophrenia, neonatal nursing, etc.).

The number of articles remaining was then 550. Third, the articles were screened based on their titles. As such, review articles, position papers, qualitative studies, and titles with keywords irrelevant to the search (e.g., ill patients' interventions, case studies, etc.) were excluded. As a result, 351 articles remained. These articles were independently screened based on the abstracts and methods sections by the first three authors of this publication (AU, ES, and DS). During this process, the three authors discussed each article where they had dilemmas or disagreements, and then reached a consensus. Studies describing program protocols, review papers of instruments, training of professionals, and interventions with children younger than five years old were excluded, and 126 papers eligible for full-text screening remained. The screening process was followed by selecting only articles focused on child outcomes related to cognitive functioning, educational skills, and emotional and social development. Our focus on these children's outcomes led us to exclude papers that included children younger than five years of age. Regarding educational skills, only broad concepts (e.g., learning motivation and self-regulation strategies) and not specific ones (e.g., math skills, reading, or writing) were included. When the articles did not explicitly mention child outcomes but mentioned keywords related to family outcomes they were included. As a result, 36 articles were

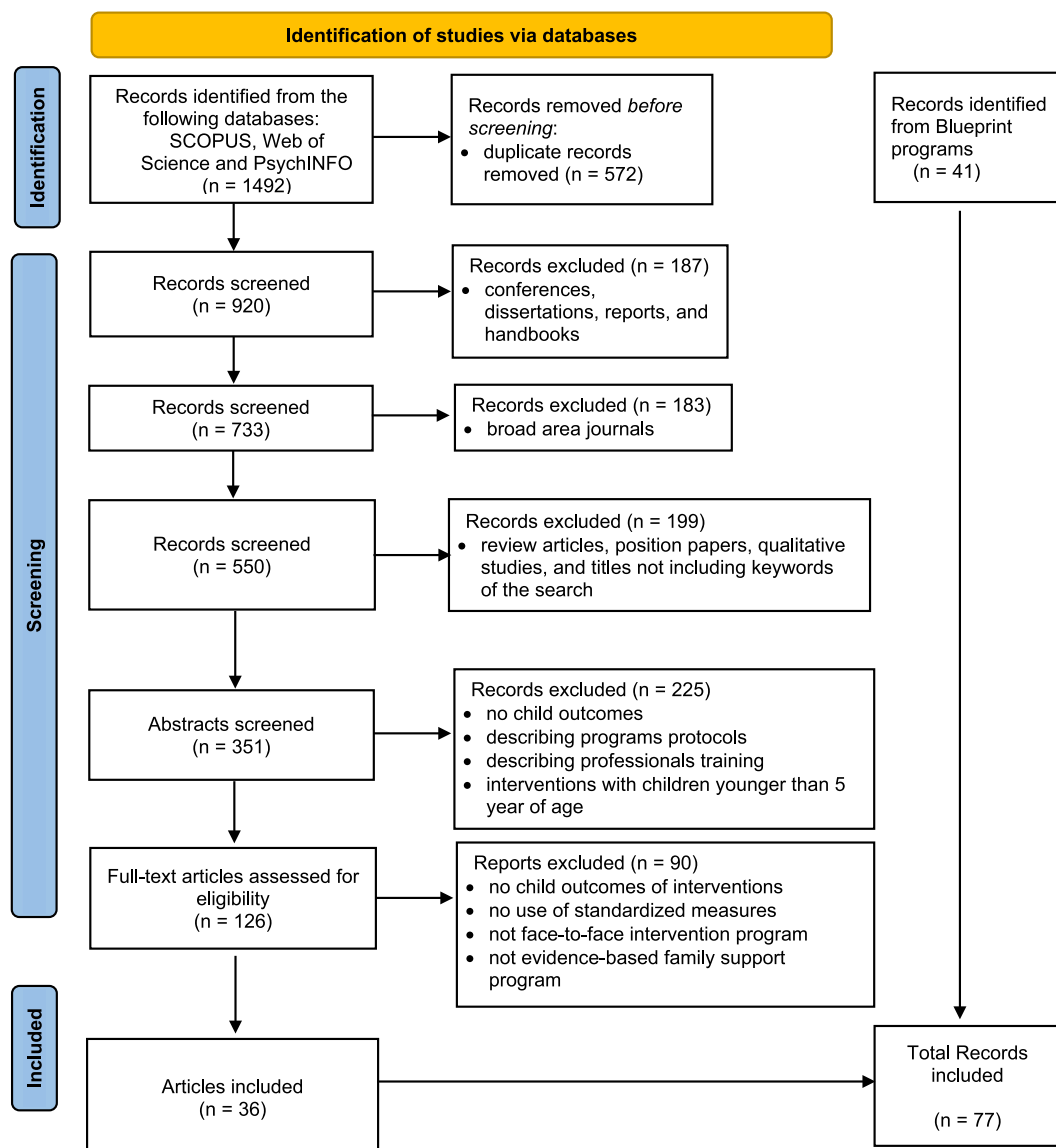


Fig. 1. Flow chart diagram for child outcome instruments.

selected. Further, the search from Blueprints resulted in an addition of 41 articles. Finally, the total number of included articles is 77. Inclusions and justifications for exclusion are provided in the flowchart in Fig. 1.

3.2. Characteristics of sources of evidence

Most studies were conducted in Europe (35 studies) and (North) America (25 studies), Australia (10 studies), and Asia (7 studies). The most used interventions were Triple P (23 studies), Incredible Years (13 studies), Parent-Management Training (13 studies), Strengthening Families (8 studies), and New Beginning (4 studies). Sixteen studies evaluated other interventions, such as Turning into Kids and Adapt (see Appendix B, Table 1).

A total of 30 studies were conducted with clinical, 22 with at-risk, 2 with clinical and at-risk samples, and 22 with community samples. In one study, community, clinical and at-risk samples were all included (Brincks et al., 2018).

Regarding informants, 46 studies used only one informant; in most of these studies, parents (n = 43) were the informants. In two studies, children/adolescents were the informants, and in one study, teachers were the informants. A total of 31 studies used more than one informant: 12 used parents and children, 11 used parents and teachers, 1 used children and teachers, and 7 used parents, children, and teachers as informants.

3.3. Results of individual sources of evidence

Overall, 187 instruments were used in the 77 studies (see Appendix A) as outcome measures of family support programs. Most measurements were related to child behavior, such as internalizing and externalizing problem behavior and prosocial behavior. The studies included few measurements of other outcomes, e.g., parent-child communication, coping, affect regulation, and quality of life.

3.4. Synthesis of results

A list of all instruments with detailed information on the study

Table 1
Most Used Instruments Measuring Child Outcomes: Parent/Other Primary Caregiver Respondents.

Instrument	First Author, Year	Items	Subscales	Studies	Intervention	Sample	Continent	CA evaluation
Eyberg Child Behavior Inventory (ECBI)	Eyberg, 1999	36	2	33	IY (12) PMTO (2) Triple P (14) Other (5)	General (6) Clinical (15) at-risk (9) Clinical and at-risk (3)	America (6) Asia (2) Australia (8) Europe (17)	Very good (14) Good (3) Mixed findings (6) Not reported (10)
Strengths and Difficulties Questionnaire (SDQ)	Goodman, 1997	25	5	26	IY (8) PMTO (1) SFP (3) Triple P (11) Other (3)	General (7) Clinical (12) at risk (5) Clinical and at-risk (2)	America (3) Asia (5) Australia (2) Europe (16)	Good (1) Mixed findings (12) Not reported (13)
Child Behavior Checklist (CBCL)	Achenbach, 1991a	14–118	2	19	IY (3) NB (3) PMTO (8) Triple P (1) Other (4)	General (4) Clinical (12) at-risk (3)	America (10) Australia (1) Europe (8)	Good (9) Mixed findings (5) Not reported (5)
Parent Daily Report (PDR)	Chamberlain, 1987	34–45	1	7	IY (1) PMTO (4) Triple P (2)	Clinical (4) at-risk (2) Clinical and at-risk (1)	Australia (2) Europe (5)	Very good (1) Mixed findings (3) Not reported (3)
Social Skills Rating System (SSRS)	Gresham, 1990	38	4	5	PMTO (3) SFP (2)	Clinical (2) at-risk (3)	America (2) Europe (3)	Good (1) Mixed findings (3) Not reported (1)
Behavioral Assessment Scale for Children (BASC)	Reynolds, 2004	25–68	4–7	3	SFP (2) Other (1)	General (2) at-risk (1)	America (1) Europe (2)	Good (2) Mixed findings (1)
Parent-Adolescent Communication Scale (PASC)	Barnes, 1982	10	1	2	NB (1) Other (1)	General (1) at-risk (1)	America (2)	Very good (1) Good (1)

Note. IY: Incredible Years; NB: New Beginnings; PMTO: Parent Management Training Oregon; SFP: Strengthening Families; Other: other evidence-based programs.

authors, samples, continent, and CA according to each informant group can be found in the supplementary section (see Appendix B, Tables 2–4).

3.5. Parents/primary caregivers as informants

The most frequently used instrument in this category were the Eyberg Child Behavior Inventory (ECBI, Eyberg & Pincus, 1999) used in 33 studies, the Strengths and Difficulties Questionnaire (SDQ, Goodman, 1999) used in 26 studies, the Child Behavior Checklist (CBCL, Achenbach, 1991a) used in 19 studies, the Parent Daily Report (PDR; Chamberlain & Reid, 1987) used in 7 studies, the Social Skills Rating System (SSRS, Gresham & Elliott, 1990) used in 5 studies, the Behavioral Assessment Scale for Children (BASC; Reynolds & Kamphaus, 2004) used in 3 studies, and the Parent-Adolescent Communication Scale (PASC; Barnes & Olson, 1982) used in 2 studies. Several other instruments were used in only one study.

The ECBI consists of 2 scales with 36 items measuring externalizing problem behavior. The intensity scale measures the frequency of disruptive behaviors, and the problem scale assesses which disruptive behaviors the parent considers to be a problem. The ECBI was used in European studies (n = 17; i.e., 49 % of the studies in this review), followed by Australia (n = 8; 70 %), (North) America (n = 6; 25 %), and Asia (n = 2; 29 %). The instrument was used in clinical (n = 15), at-risk (n = 9), and community samples (n = 6), as well as in clinical and at-risk samples (n = 3). The CA quality was categorized as very good in 14 studies, good in 3 studies, with mixed findings in 6 studies, and not reported in 10 studies.

The SDQ consists of 25 items with 5 subscales (i.e., emotional problems, conduct problems, inattention/hyperactivity problems, peer problems, and prosocial behavior). Some studies used only the subscales measuring problem behavior. The SDQ was used in European studies (n = 16; 46 %), followed by Asia (n = 5; 71 %), (North) America (n = 3; 12 %), and Australia (n = 2; 20 %). The instrument was used in clinical (n = 12), at-risk (n = 5), and community samples (n = 7), as well as in clinical and at-risk samples (n = 2). The CA quality was categorized as good in 1 study, with mixed findings in 12 studies, and was not reported in 13 studies.

Table 2
Most Used Instruments Measuring Child Outcomes: Self-Reports (Child/Adolescent Respondents).

Instrument	First Author, Year	Items	Subscales	Studies	Intervention	Sample	Continent	CA evaluation
Child Depression Inventory (CDI)	Kovacs, 1985	27	5	5	NB (2) PMTO (3)	Community (1) Clinical (2) at-risk (2)	America (4) Europe (1)	Good (4) Not reported (1)
Behavioral Assessment Scale for Children (BASC)	Reynolds, 2004	43	8	3	SFP (2) Other (1)	Community (2) at-risk (1)	America (1) Europe (2)	Very good (3)
Strengths and Difficulties Questionnaire (SDQ)	Goodman, 1997	25–68	3–8	3	PMTO (1) SFP (1) Triple P (1)	Community (1) Clinical (1) at-risk (1)	Australia (1) Europe (2)	Mixed findings (3)
Youth Self Report (YSR)	Achenbach, 1991b	30	2	3	NB (2) SFP (1)	Community (2) at-risk (1)	America (3)	Very good (1) Mixed findings (2)

Note. NB: New Beginnings; PMTO: Parent Management Training Oregon; SFP: Strengthening Families; Other: other evidence-based programs.

Table 3
Most Used Instruments Measuring Child Outcomes: Teachers/Other Professional Respondents.

Instrument	First Author, Year	Items	Subscales	Studies	Intervention	Sample	Continent	CA evaluation
Teacher Report Form (TRF)	Achenbach, 1991c	30	2	9	PMTO (9)	Community (1) Clinical (6) at-risk (2)	America (2) Europe (7)	Good (5) Mixed findings (1) Not reported (3)
Social Skills Rating System (SSRS)	Gresham, 1990	30	4	3	PMTO (3)	Clinical (2) at-risk (1)	Europe (3)	Good (1) Mixed findings (2)
Strengths and Difficulties Questionnaire (SDQ)	Goodman, 1997	25	5	2	IY (1) Triple P (1)	Clinical (1) at-risk (1)	Europe (2)	Good (1) Not reported (1)

Note. IY: Incredible Years; PMTO: Parent Management Training Oregon.

Table 4
Cronbach’s Alpha Quality Related to Sample.

	Community		Clinical		at-risk	
	n	%	n	%	n	%
Very good / good (cat 1)	24	56	23	26	21	36
Mixed / not reported (cat 2)	19	44	66	74	34	64
Whole	43	100	89	100	55	100

Table 5
Cronbach’s Alpha Quality Related to Respondents.

	Parents		Children		Teachers	
	n	%	n	%	n	%
Very good / good (cat 1)	43	34	13	39	12	43
Mixed / not reported (cat 2)	83	66	20	61	16	57
Whole	126	100	33	100	28	100

The CBCL measures externalizing problem behavior (e.g., opposition/defiance, rule breaking, and social problems) and internalizing problems (e.g., anxiety, depression, and thought problems). Reports regarding the number of items (ranging from 14 to 118 items) and scales (ranging from 1 to 4 scales) were quite heterogeneous across the studies. The CBCL was used in (North) America (n = 10; 40 %), followed by Europe (n = 8; 23 %), with one study conducted in Australia. The instrument was used in clinical (n = 12), at-risk (n = 3), and community samples (n = 4). The CA quality was categorized as good in 9 studies, with mixed findings in 5 studies, and was not reported in 5 studies.

The PDR mainly assesses the frequency of externalizing problem behavior in telephone interviews with 34 to 45 items. PDR was used in Europe (n = 5) and Australia (n = 2) in clinical (n = 4), at-risk (n = 3), as well as in a clinical and at-risk sample (n = 1). The CA quality was categorized as very good in 1 study, with mixed findings in 3 studies, and was not reported in 3 studies.

The SSRS measures prosocial behavior and consists of 38 items with 4 scales (cooperation, assertion, responsibility, and self-control; some studies assessed only 3 scales). SSRS was used in Europe (n = 3) and (North) America (n = 2) in clinical (n = 2) and at-risk (n = 3) sample.

The CA was categorized as good in 1 study, with mixed findings in 3 studies, and was not reported in 1 study.

The BASC measures internalizing and externalizing problem behavior. Reports regarding the number of items (ranging from 25 to 68 items) and scales (ranging from 3 to 8 scales) were quite heterogeneous across the studies. BASC was used in Europe (n = 2) and (North) America (n = 1) in at-risk (n = 1) and community samples (n = 2). The CA quality was categorized as good in two studies and with mixed findings in one study.

The PASC measures the quality of communication between parents and adolescents and consists of 10 items and 1 scale. It was used in (North) America in at-risk (n = 1) and community sample (n = 1). The CA quality was categorized as very good and good in one study each (see Table 1).

3.6. Children / adolescents as informants

The most frequently used instruments were the Child Depression Inventory (CDI; Kovacs, 1985) used in 5 studies, followed by the Behavioral Assessment Scale for Children (BASC; Reynolds & Kamphaus, 2004), the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999), and the Youth Self-Report (YSR; Achenbach, 1991b), all used in 3 studies each.

The CDI measuring internalizing problem behavior consists of 27 items with 5 subscales: anhedonia, ineffectiveness, interpersonal problems, negative mood, and negative self-esteem. CDI was used in (North) America (n = 4) and Europe (n = 1). The instrument was used in clinical (n = 2), at-risk (n = 2), and community samples (n = 1). The CA quality was categorized as good in four studies, as only the CA of the whole instrument was reported. One study did not report it.

BASC with children/adolescents as informants was used in Europe (n = 2) and (North) America (n = 1) in at-risk (n = 2) and community samples (n = 1). The CA quality was categorized as good in two studies and with mixed findings in one study.

The SDQ was used in European studies (n = 2) and Australia (n = 1) in a clinical sample (n = 1), at-risk (n = 1), and community sample (n = 1). The CA quality was categorized with mixed findings in three studies. Some studies used only the subscales measuring problem behavior.

The YSR measuring internalizing and externalizing problem

behavior consists of 30 items with 2 subscales (i.e., internalizing and externalizing behavior) and was used in (North) America ($n = 3$). The instrument was used in a clinical sample ($n = 1$), at-risk ($n = 1$), and community samples ($n = 1$). The CA quality was categorized as very good in 1 study and with mixed findings in 2 studies (see Table 2).

3.7. Teachers/other professionals as informants

The most often used instruments were the Teacher Report Form (TRF, Achenbach, 1991c) used in 9 studies, followed by the Social Skills Rating System (SSRS, Gresham & Elliott, 1990) used in 3 studies, and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999) used in 2 studies.

The TRF measuring internalizing and externalizing problem behavior consists of 30 items with 2 subscales: internalizing and externalizing behavior. TRF was used in Europe ($n = 7$) and (North) America ($n = 2$) in clinical ($n = 6$), at-risk ($n = 1$), and community samples ($n = 1$). The CA was categorized as good in five studies, with mixed findings in one study, and not reported in three studies.

The SSRS measures prosocial behavior and consists of 38 items with 4 scales (i.e., cooperation, assertion, responsibility, and self-control; some studies assessed only 3 scales). The SSRS was used in Europe ($n = 3$) in clinical samples ($n = 2$) and at-risk ($n = 1$). The CA quality was categorized as good in 1 study and with mixed findings in 2 studies.

The SDQ with teachers as respondents was used in Europe ($n = 2$) in a clinical ($n = 1$) and an at-risk sample ($n = 1$). The CA quality was categorized as good in one study and not reported in one study. For information regarding the items and the CA (see Table 3).

3.8. Cronbach's alpha quality related to sample

The CA quality was very good for 29 (16 %) studies, good for 39 (21 %), with mixed findings for 60 (32 %), and not reported for 59 (32 %) instruments. For further analysis these four categories were collapsed into two categories, i.e., very good and good (category 1), and mixed findings and not reported (category 2).

Chi-square tests were conducted to determine whether there was a statistically significant relationship between these two categories and the samples (i.e., community, clinical, or at-risk) as well as between these two categories and the respondents (i.e., parents/primary caregivers, children/adolescents, and teachers/other professionals).

The relation between the two categories and the samples was significant: chi-square (2, 187) = 11.37, $p = 0.003$. Significantly more instruments (56 %) used in the community samples were in category 1 (i.e., very good or good), whereas in the clinical samples, only 26 % and on the other hand, there was a significantly higher percentage of non-reporting of the CA for the instruments used in clinical samples (50 %) compared to community (16 %) and at-risk (13 %) samples. Mixed findings were mostly observed in at-risk samples (49 %) (see Table 4).

A chi-square test of independence was performed to examine the relation between the reported CA and the respondents (i.e., parents, children, and teachers). The relation between these variables was not significant: chi-square (2, 187) = 9.091, $p = 0.633$. (see Table 5).

3.9. Summary of evidence

Studies have shown that evidence-based programs should be available in communities to promote family well-being and child development (Almeida et al., 2022, Rodrigo, 2016). Effective family support programs are those designed to address parenting issues and improve children's well-being in different populations (Bornstein et al., 2022). As evidence-based family support programs vary in their targets (e.g., children and parents from different populations), methods (e.g., mixed methods), and outcomes (e.g., behavior and mental health), the main aim of this review was to identify the most frequently used instruments measuring child (ages 5–18 years) outcomes with a focus on the

emotional and social development of children/adolescents in evidence-based family support programs, as their psychosocial development is the ultimate goal of these interventions. Instruments that assess these dimensions should be robust and reliable across populations according to the standards of evidence (Flay et al., 2005). However, there is a gap in the literature whether or to what extent the psychometric properties (especially the reliability) of these instruments are reported. Thus, we searched for the most appropriate studies (i.e., evidence-based programs) as quantitative studies involving evidence-based programs are more likely to report high quality evaluation tools than any other type of studies (Asmussen, 2012; Hidalgo et al., 2023; Serene Olin et al., 2014) and compared them quantitatively with respect to their reported reliabilities (i.e., CA). In this scoping review, we identified 77 peer-reviewed studies published between 2012 and 2022 that addressed evidence-based family support programs implemented across different populations.

A multistage sampling procedure and instrument coding were used to analyze the instruments systematically. Most studies were conducted in Europe and (North) America. The most prevalent intervention programs used in these studies were Triple P, followed by Incredible Years and Parent-Management Training. Most studies were conducted with clinical, followed by those conducted with at-risk and community samples.

Most measurements were related to child behavior, such as internalizing and externalizing problem behavior, as well as prosocial behavior. Very few measurements included other outcomes, such as parent-child communication, coping, affect regulation, and quality of life. More than half of the studies included in this review used single informants, mainly parents/primary caregivers, children/adolescents, teachers, or other professionals. However, as children are key actors of family dynamics, hearing their voices by reporting about their experience with the intervention might help in improving the quality of the program (Bentley et al., 2019; Deighton et al., 2014).

The most frequently used child outcome measurement reported by parents/primary caregivers was ECBI, followed by SDQ, CBCL, PDR, SSRS, BASC, and the PASC. These findings are in line with several studies showing that the ECBI is a very practical and sensitive tool for measuring the effect of treatment on disruptive child behavior and is widely used among parents of different socioeconomic backgrounds (Abrahamse et al., 2015; Jeter, Zlomke, Shawler, & Sullivan, 2017; Weeland, van Aar, & Overbeek, 2018). The studies that used children/adolescents as informants in child outcome measurements showed that the CDI was the most frequently used measurement for child outcomes, followed by the SDQ, YSR, BASC, and LSQS. Previous evidence also indicates that the CDI is a fairly good instrument used especially with clinical samples but also with community and school samples (Kovacs, 2014; Rivera, Bernal, & Rosselló, 2005). When the studies used teachers or other professionals as the respondents for the child outcomes, the analysis revealed that the TRF was the most used measurement, followed by the SSRS, SDQ, CGISS, and GFS. The TRF was tested for its psychometric properties as a measurement used for child outcomes in various samples, and the studies indicated that this tool is useful for clinical practice, training, and research involving children and adolescents from all socioeconomic backgrounds (Bordin et al., 2013; Roussos et al., 1999; Steensma et al., 2014). Among all the measurements described for the aim of this study, the SDQ was the only child outcome measurement commonly used in all studies with different informants.

Our second research question was related to the reported instrument psychometric properties, the differences in psychometric properties, and the quality of psychometric properties with respect to the samples.

Two-thirds of the studies reported internal consistency represented by the CA of the instruments. One-third of the studies did not report any psychometric properties. The quality of the CA reports differed across the studies and instruments. Only 20 studies reported the CA of every scale, and the CA of each scale was mostly above 0.80. Of these 20 studies, 13 studies used the ECBI (Eyberg & Pincus, 1999), and the 20

studies were equally distributed among community, clinical, and at-risk samples. Twenty studies categorized as good (i.e., reported either not all the scales or the CA of the whole instrument) used the CBCL (Achenbach, 1991a) and the ECBI mostly in the community samples. The 40 studies that were categorized with mixed findings (i.e., the CA of the whole instrument was reported) used several instruments, mostly in clinical and at-risk samples. In the category of non-reported CA, 12 studies used the SDQ (Goodman, 1999), 11 used the ECBI, and 5 used the CBCL, almost all with clinical and some studies with at-risk samples. This is in line with results of previous review (Bentley et al., 2019) reports of psychometric properties being often neglected by study authors, which has implications for how study findings can be interpreted. Furthermore, even instruments that are widely used for evaluation studies, such as the CBCL, YSR, or SDQ, showed only adequate CA values (see Deighton et al., 2014, p. 9). However, when analyzing the data, it was important to keep in mind that internal consistency measures are usually sample-dependent as they express an outcome for a given sample, meaning that one measure can have good reliability in a given sample and poor reliability in another one. In such cases, further psychometric analyses are needed to determine whether the constructs are comparable across samples.

Furthermore, there is evidence of an unequal distribution of the reported CA quality between the studied samples. Although there might be several reasons why study authors do not report CA or other measures for psychometric properties, these indicators are necessary for researchers and practitioners to decide which program is the most appropriate for a specific population and assess and evaluate the results of studies with confidence. This in turn has important implications for how program selection is undertaken for specific populations and how it is evaluated in the future. Although there are several criticisms on Cronbach's alpha depending on population and due to the violation of its strict assumption, it is still the most often used reliability index. There exist several suggestions in the methodological literature on alternatives to CA (such as Omega, Coefficient H; McNeish, 2018) but more effort is needed to use these alternatives and report them in the studies.

3.10. Limitations and implications for future research

This review focused on the reliability, mostly measured with CA and quality of evaluation tools for child outcomes with respect to populations. However, the studies included in this review were mainly conducted in Europe and the (North) America and did not systematically report the reliability of the evaluation tools used for child outcomes, while in some studies the evidence was unknown. In the studies that could provide some evidence on the quality of the outcome measurements, the CA was either low or moderate (Arkan, Güvenir, Ralph, & Day, 2020; Christopher et al., 2017; Leijten, Raaijmakers, Orobio de Castro, van den Ban, & Matthys, 2017), with very few studies having CA at a high level (Gewirtz, DeGarmo, & Zamir, 2018; Thijssen, Vink, Muris, & de Ruiter, 2017), and in those cases, the evidence was limited to the reliability and validity of the instruments. This scoping review underlines the importance of establishing a comprehensive battery of child outcome measurements with high-quality psychometric properties to be considered for evidence-informed recommendations for both researchers and practitioners.

Although the current measurements used to target children's outcomes are of good quality, there is still room for improvement. This review's findings suggest that there is a need for further research to assess the psychometric properties of measurements of targeted child outcomes and provide more consistent evidence to support the use of these measurements with children older than 5 years of age (Foster & Park, 2012; Yoon, Speyer, Cordier, Aunio, & Hakkarainen, 2021). We strongly recommend that future studies emphasize the importance of reporting the psychometric properties of the evaluation tools to be able to provide stronger evidence that supports all psychometric aspects of measurements, especially when used to measure children's outcomes.

The current evidence to suggest the use of family support programs for different respondents targeting child outcomes is limited. It is essential that the programs that will be used by researchers and practitioners are strong enough to foster change within the interventions.

Many challenges to quality measurements of child outcomes in family support programs still exist. This study has several practical implications that are in accordance with the EU Strategy on the Rights of the Child and the European Child Guarantee (European Commission, 2021). An urgent need exists to increase awareness among practitioners and professionals working with children and families to use evidence-based programs consisting of measurements with strong psychometric properties to improve the quality of services (Hong et al., 2019; Wallander, Schmitt, & Koot, 2001). Next, this study recommends that such evidence-based programs should be improved and strengthened by the close cooperation of researchers who aim to develop quality evaluation tools for children's outcomes with practitioners and policymakers to help professionals working in the field with the necessary training, implementation, and evaluation of such programs. Furthermore, the development of quality evaluation tools to measure children's outcomes should be a fundamental component of evidence-based programs to increase the quality of services for children and families (Kavanagh, Adams, & Wang, 2009).

Finally, it is important to bear in mind that measurement and internal consistency are important but not the only aspects of program quality. A guiding framework for such programs could be the standards formulated by the European Family Support Network (EurofamNet, 2020), which focuses on family support policies and practices. According to this, the quality standards of a family support program can be categorized according to the phases of the program development, namely the formulation, delivery, evaluation, and dissemination phase (Özdemir et al., in press). The quality standards associated with the formulation phase relate to well-formulated theoretical models, standard delivery, responsiveness to needs, the opportunity for capacity building, appropriate participatory settings, etc. The quality standards associated with the delivery phase relate to a good implementation plan, adequate staff training, etc. Within those quality standards that qualify interventions as evidence-based programs, the standards related to program evaluation include outcome, process, and economic evaluation. Furthermore, when considering program evaluation as an essential standard, one should keep in mind that different evaluation methods exist that can be implemented within specific contexts and with specific populations (Almeida et al., 2022).

Common efforts of researchers, policy, and practice are needed to consider these aspects to further improve intervention and thus the lives of children and their families.

Funding Information

This paper is based upon work from COST Action CA18123 European Family Support Network, supported by COST (European Cooperation in Science and Technology). <https://www.cost.eu>. This study was carried out in accordance with the European Cooperation in Science and Technology Association policy on inclusiveness and excellence, as written in the Memorandum of understanding for the implementation of the COST Action "The European Family Support Network. A bottom-up, evidence-based and multidisciplinary approach" (EuroFam-Net) CA18123.

CRedit authorship contribution statement

Ana Uka: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing, Validation, Methodology, Supervision, Resources, Project administration. **Elisabeth Stefanek:** Conceptualization, Funding acquisition, Data curation, Visualization, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Validation, Methodology, Supervision, Resources, Project

administration. **Daiva Skučienė:** Conceptualization, Funding acquisition. **Carmen Schneckenreiter:** Data curation, Visualization, Investigation, Formal analysis. **Georg Spiel:** Conceptualization, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank Lucía Jimenez and John Canavan for leading our COST Action CA18123 “The European Family Support Network: A bottom-up, evidence-based and multidisciplinary approach (EurofamNet), supported by “European Cooperation in Science and Technology”, Belgium; COST, Belgium, Sonia Byrne and Ninoslava Pečnik for leading our COST Action EurofamNet WG3 and our colleague, Oriola Hamzallari for her invaluable inputs when writing this paper. The authors would like to thank the EurofamNet for supporting a Short-term scientific mission for the first author, where most of the identification and screening of articles was conducted, and a virtual mobility grant for the second author, which supported the coordination of the work conducted.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.childyouth.2024.107903>.

References

- Abrahamse, M. E., Junger, M., Leijten, P. H., Lindeboom, R., Boer, F., & Lindauer, R. J. (2015). Psychometric properties of the Dutch Eyberg Child Behavior Inventory (ECBI) in a community sample and a multi-ethnic clinical sample. *Journal of Psychopathology and Behavioral Assessment*, *37*, 679–691. <https://doi.org/10.1007/s10862-015-9482-1>
- Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4–18 and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.
- Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. (1991c). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Almeida, A., Cruz, O., & Canário, A. C. (2022). Evidence-based family and parenting support evaluation strategies: The position of EurofamNet. *EurofamNet*. https://eurofamnet.eu/sites/default/files/toolbox/academicoutputs/WG3_EurofamNetPositionEvaluationStrategies.pdf.
- Arkan, B., Güvenir, T., Ralph, A., & Day, J. (2020). The efficacy and acceptability of the Triple P: Positive Parenting Program with Turkish parents. *Journal of Child and Adolescent Psychiatric Nursing*, *33*, 148–156. <https://doi.org/10.1111/jcap.12283>
- Asmussen, K. (2012). *The evidence-based parenting practitioner's handbook*. Routledge.
- Barnes, H., & Olson, D. H. (1982). Parent-Adolescent Communication Scale. In D. H. Olson, H. I. McCubbin, H. Barnes, A. Larsen, M. Muxen, & M. Wilson (Eds.), *Family Inventories: Inventories Used in a National Survey of Families Across the Family Life Cycle* (pp. 33–48). St. Paul, MN: Family Social Science, University of Minnesota.
- Bentley, N., Hartley, S., & Bucci, S. (2019). Systematic review of self-report measures of general mental health and wellbeing in adolescent mental health. *Clinical Child and Family Psychology Review*, *22*, 225–252. <https://doi.org/10.1007/s10567-018-00273-x>
- Bordin, I. A., Rocha, M. M., Paula, C. S., Teixeira, M. C. T., Achenbach, T. M., Rescorla, L. A., & Silveira, E. F. (2013). Child Behavior Checklist (CBCL), Youth Self-Report (YSR) and Teacher's Report Form (TRF): An overview of the development of the original and Brazilian versions. *Cadernos de saúde pública*, *29*, 13–28. <https://doi.org/10.1590/s0102-311x2013000100004>
- Bornstein, M. H., Kotler, J. A., & Lansford, J. E. (2022). The future of parenting programs: An introduction. *Parenting*, *22*, 189–200. <https://doi.org/10.1080/15295192.2022.2086808>
- Brennan, R. L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, *38*, 295–317.
- Brincks, A., Perrino, T., Howe, G., Pantin, H., Prado, G., Huang, S., & Brown, C. H. (2018). Preventing youth internalizing symptoms through the Familias Unidas intervention: Examining variation in response. *Prevention Science*, *19*(Suppl 1), 49–59. <https://doi.org/10.1007/s11121-016-0666-z>
- Chamberlain, P., & Reid, J. B. (1987). Parent observation and report of child symptoms. *Behavioral Assessment*, *9*, 97–109.
- Christopher, C., Wolchik, S., Tein, J. Y., Carr, C., Mahrer, N. E., & Sandler, I. (2017). Long-term effects of a parenting preventive intervention on young adults' painful feelings about divorce. *Journal of Family Psychology*, *31*, 799–809. <https://doi.org/10.1037/fam0000325>
- Corcoran, J. (2009). Using standardized tests and instruments in family assessment. In R. A. Roberts (Ed.), *Social Work Desk Reference* (2nd ed., pp. 390–394). New York, NY: Oxford University Press.
- Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: A review of child self-report measures. *Child & Adolescent Psychiatry & Mental Health*, *8*, 1–14. <https://doi.org/10.1186/1753-2000-8-14>
- de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2015). *Measurement in Medicine*. Cambridge: Cambridge University Press.
- European Commission. (2021). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions EU strategy on the Rights of the Child (COM/2021/142 final). Available online: <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0142> (accessed on 20 November 2023).
- European Family Support Network (2020). EurofamNet principles. EurofamNet.
- Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory & Sutter-Eyberg Student Behavior Inventory-revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Flay, B. R., Biglan, A., Boruch, R. F., Gonzalez Castro, F., Gottfredson, D., Kellam, S., Moscicki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*, 151–175. <https://doi.org/10.1007/s11121-005-5553-y>
- Foster, R. L., & Park, J. H. (2012). An integrative review of literature examining psychometric properties of instruments measuring anxiety or fear in hospitalized children. *Pain Management Nursing*, *13*, 94–106. <https://doi.org/10.1016/j.pmn.2011.06.006>
- Gewirtz, A. H., DeGarmo, D. S., & Zamir, O. (2018). After deployment, adaptive parenting tools: 1-year outcomes of an evidence-based parenting program for military families following deployment. *Prevention Science*, *19*, 589–599. <https://doi.org/10.1007/s11121-017-0839-4>
- Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, *40*, 791–801.
- Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science*, *16*, 893–926. <https://doi.org/10.1007/s11121-015-0555-x>
- Gray, G. R., Totsika, V., & Lindsay, G. (2018). Sustained effectiveness of evidence-based parenting programs after the research trial ends. *Frontiers in Psychology*, *9*, Article 366935. <https://doi.org/10.3389/fpsyg.2018.02035>
- Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, *34*, 177–189.
- Hidalgo, V., Rodríguez-Ruiz, B., García Bacete, F. J., Martínez-González, R. A., López Verdugo, I., & Jiménez García, L. (2023). The evaluation of family support programmes in Spain: An analysis of their quality standards. *Psicología Educativa*, *29*, 35–43. <https://doi.org/10.5093/psed2023a9>
- Hong, S. L. S., Sabol, T. J., Burchinal, M. R., Tarullo, L., Zaslow, M., & Peisner-Feinberg, E. S. (2019). ECE quality indicators and child outcomes: Analyses of six large child care studies. *Early Childhood Research Quarterly*, *49*, 202–217. <https://doi.org/10.1016/j.ecresq.2019.06.009>
- Jeter, K., Zlomke, K., Shawler, P., & Sullivan, M. (2017). Comprehensive psychometric analysis of the Eyberg Child Behavior Inventory in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *47*, 1354–1368. <https://doi.org/10.1007/s10803-017-3048-x>
- Kavanagh, P. L., Adams, W. G., & Wang, C. J. (2009). Quality indicators and quality assessment in child health. *Archives of Disease in Childhood*, *94*, 458–463.
- Kovacs, M. (1985). Depressive disorders in childhood: II. A longitudinal study of the risk for a subsequent major depression. *Annual Progress in Child Psychiatry & Child Development*, *41*, 520–541.
- Kovacs, M. (2014). Children's depression inventory (CDI and CDI 2). *The Encyclopedia of Clinical Psychology*, 1–5. <https://doi.org/10.1002/9781118625392.wbepc419>
- Kumpfer, K. L., Whiteside, H. O., Greene, J. A., & Allen, K. C. (2010). Effectiveness outcomes of four age versions of the Strengthening Families Program in statewide field sites. *Group Dynamics: Theory, Research, and Practice*, *14*, 211–229. <https://doi.org/10.1037/a0020602>
- Leijten, P., Raaijmakers, M. A., Orobio de Castro, B., van den Ban, E., & Matthys, W. (2017). Effectiveness of the incredible years parenting program for families with socioeconomically disadvantaged and ethnic minority backgrounds. *Journal of Clinical Child & Adolescent Psychology*, *46*, 59–73. <https://doi.org/10.1080/15374416.2015.1038823>
- Marti, F. A., Pourat, N., Lee, C., & Zima, B. T. (2022). Overview of ten child mental health clinical outcome measures: Testing of psychometric properties with diverse client populations in the US. *Administration and Policy in Mental Health and Mental Health Services Research*, *49*, 197–225. <https://doi.org/10.1007/s10488-021-01157-z>

- McNeish, D. (2018). Thanks coefficient Alpha, we'll take it from here. *Psychological Methods*, 23, 412–433. <https://doi.org/10.1037/met0000144>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151, 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- Morris, A. S., Robinson, L. R., Hays-Grudo, J., Claussen, A. H., Hartwig, S. A., & Treat, A. E. (2017). Targeting parenting in early childhood: A public health approach to improve outcomes for children living in poverty. *Child Development*, 88, 388–397. <https://doi.org/10.1111/cdev.12743>
- Newland, L. A. (2015). Family well-being, parenting, and child well-being: Pathways to healthy adjustment. *Clinical psychologist*, 19(1), 3–14. <https://doi.org/10.1111/cp.12059>
- Özdemir, M., Vastamäki, S., Leijten, P., & Sampaio, F. (in press). The European Family Support Network (EurofamNet) quality standards for family support programs. EurofamNet.
- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment Scale for Children*. Bloomington: Pearson Assessments.
- Rivera, C. L., Bernal, G., & Rosselló, J. (2005). The Children Depression Inventory (CDI) and the Beck Depression Inventory (BDI): Their validity as screening measures for major depression in a group of Puerto Rican adolescents. *International Journal of Clinical and Health Psychology*, 5, 485–498.
- Rodrigo, M. J. (2016). Quality of implementation in evidence-based positive parenting programs in Spain: Introduction to the special issue. *Psychosocial Intervention*, 25, 63–68. <https://doi.org/10.1016/j.psi.2016.02.004>
- Rodrigo, M. J., Almeida, A., Spiel, C., & Koops, W. (2012). Introduction: Evidence-based parent education programmes to promote positive parenting. *European Journal of Developmental Psychology*, 9, 2–10. <https://doi.org/10.1080/17405629.2011.631282>
- Rodrigo, M. J., Byrne, S., & Álvarez Lorenzo, M. (2012). Preventing child maltreatment through parenting programs implemented at the local social services level. *European Journal of Developmental Psychology*, 9, 89–103. <https://doi.org/10.1080/17405629.2011.607340>
- Roussos, A., Karantanos, G., Richardson, C., Hartman, C., Karajiannis, D., Kyprianos, S., & Zoubou, V. (1999). Achenbach's Child Behavior Checklist and Teachers' Report Form in a normative sample of Greek children 6–12 years old. *European Child & Adolescent Psychiatry*, 8, 165–172.
- Sanders, M. R., & Turner, K. M. T. (2019). The Triple P System: Parenting support for every family. In B. H. Fiese, M. Celano, K. Deater-Deckard, E. N. Jouriles, & M. A. Whisman (Eds.), *APA handbook of contemporary family psychology: Family therapy and training* (pp. 409–424). American Psychological Association. 10.1037/0000101-025.
- Serene Olin, S., Kutash, K., Pollock, M., Burns, B. J., Kuppinger, A., Craig, N., & Hoagwood, K. E. (2014). Developing quality indicators for family support services in community team-based mental health care. *Administration and Policy in Mental Health and Mental Health Services Research*, 41, 7–20. <https://doi.org/10.1007/s10488-013-0501-9>
- Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences*, 28, 229–237.
- Simeonsson, R. J. (1991). Primary, secondary, and tertiary prevention in early intervention. *Journal of Early Intervention*, 15, 124–134. <https://doi.org/10.1177/105381519101500202>
- Smyth, J. M., & Arigo, D. (2009). Recent evidence supports emotion-regulation interventions for improving health in at-risk and clinical populations. *Current Opinion in Psychiatry*, 22, 205–210. <https://doi.org/10.1097/YCO.0b013e3283252d6d>
- Sorsa, J., Fontell, T., Laajasalo, T., & Aronen, E. T. (2019). Eyberg Child Behavior Inventory (ECBI): Normative data, psychometric properties, and associations with socioeconomic status in Finnish children. *Scandinavian Journal of Psychology*, 60, 430–439. <https://doi.org/10.1111/sjop.12541>
- Steensma, T. D., Zucker, K. J., Kreukels, B. P., VanderLaan, D. P., Wood, H., Fuentes, A., & Cohen-Kettenis, P. T. (2014). Behavioral and emotional problems on the Teacher's Report Form: A cross-national, cross-clinic comparative analysis of gender dysphoric children and adolescents. *Journal of Abnormal Child Psychology*, 42, 635–647. <https://doi.org/10.1007/s10802-013-9804-2>
- Thijssen, J., Vink, G., Muris, P., & de Ruiter, C. (2017). The effectiveness of parent management training—Oregon model in clinically referred children with externalizing behavior problems in The Netherlands. *Child Psychiatry & Human Development*, 48, 136–150. <https://doi.org/10.1007/s10578-016-0660-5>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169, 467–473.
- Wallander, J. L., Schmitt, M., & Koot, H. M. (2001). Quality of life measurement in children and adolescents: Issues, instruments, and applications. *Journal of Clinical Psychology*, 57, 571–585.
- Weeland, J., van Aar, J., & Overbeek, G. (2018). Dutch norms for the Eyberg child behavior inventory: Comparisons with other Western countries. *Journal of Psychopathology and Behavioral Assessment*, 40, 224–234. <https://doi.org/10.1007/s10862-017-9639-1>
- Yoon, S., Speyer, R., Cordier, R., Aunio, P., & Hakkarainen, A. (2021). A systematic review evaluating psychometric properties of parent or caregiver report instruments on child maltreatment: Part 1: Content validity. *Trauma, Violence, & Abuse*, 22, 1013–1031. <https://doi.org/10.1177/1524838019898456>