

Systematic Review and Critical Appraisal of Five of the Most Recurrently Validated Child Maltreatment Assessment Instruments from 2010 to 2020

Sylvia Georgieva¹, José M. Tomás¹, José J. Navarro-Pérez² , and Paula Samper-García¹

Abstract

Assessment of child maltreatment has been inconsistent across literature due to its complexity, multidimensionality, and the variety of conceptualizations of this construct. Five instruments have recurrently examined psychometric properties across the last years of research: Childhood Trauma Questionnaire, Maltreatment and Abuse Chronology of Exposure, Child Abuse Potential Inventory, Identification of Parents at Risk for child Abuse and Neglect, and Psychosocial Screening Tool. This article aims to examine and wrap up the knowledge regarding the psychometric properties of these instruments. A systematic review was performed through three of the most relevant databases in order to identify the most validated instruments to assess child maltreatment from 2010 to 2020, and 19 research articles were identified. Results indicate that there is a lack of information regarding some psychometric properties and therefore, in the light of this information, it is not possible to clearly determine if there are instruments with stronger scientific evidence for their psychometric properties, although the Maltreatment and Abuse Chronology of Exposure Scale (MACE) obtained the strongest psychometric evidence. This systematic review provided a comprehensive review on the main psychometric properties of five child maltreatment instruments in order to facilitate researchers and child welfare professionals the selection of the most suitable instrument for their specific purpose. We recommend addressing these gaps of information by further examining the psychometric properties of these instruments, and developing valid and reliable instruments for early detection in child maltreatment.

Keywords

child abuse, child sexual abuse, child neglect, child mistreatment, child maltreatment, psychometrics

Introduction

Child maltreatment is a complex, multidimensional, and heterogeneous construct that encompasses a great number of risk factors and creates a worldwide social and public health problem (Guastaferro & Lutzker, 2021; Zhang et al., 2022). According to the WHO (2020), “child maltreatment is the abuse and neglect that occurs to children under 18 years of age. It includes all types of physical and/or emotional maltreatment, sexual abuse, neglect, and any commercial or other exploitation, which results in actual or potential harm to the child’s health, survival, development, or dignity in the context of a relationship of responsibility, trust, or power.”

way to select the most reliable and valid instruments is to systematically review the literature on its psychometric properties (Terwee, et al., 2016; Yoon, et al., 2020). However, the availability of psychometrically sound and reliable assessment instruments is scarce, as stated in a recent systematic review of child maltreatment instruments performed by Saini et al., (2019) in which the psychometric quality of 52 instruments was assessed, from which, only 8 of them had strong to moderate evidence for at least 3 of the criteria of methodological quality. The study concluded that the three instruments that appeared to have the strongest psychometric

Child Maltreatment Assessment

Assessment of child maltreatment has been inconsistent across the academic literature due to the differences in its definitions, factors, and types of reports, which may include it. The best

¹Faculty of Psychology, University of Valencia, Spain

²Faculty of Social Work, University of Valencia, Spain

Corresponding Author:

José J. Navarro-Pérez, Faculty of Social Work, University of Valencia, Av. Dels Tarongers, 4b, Valencia, 46022, Spain.

Email: j.javier.navarro@uv.es

evidence were The Childhood Trauma Questionnaire (CTQ; Bernstein & Fink, 1998), the short form of CTQ (CTQ-SF; Bernstein et al., 2003), and the Abuse Chronology of Exposure Scale (MACE; Teicher & Parigger, 2015).

In order to address this problem, participatory child-protection social work requires not only better training and different instruments, but a reorganization in the value base of children's services (Wilkins & Whittaker, 2018). Institutions must stay updated with the latest results of research on child maltreatment assessment and provide professionals with the current most objective and reliable tools to support them in their decision-making process. These tools must be psychometrically sound, replicable, and restructured according to the current cultural and value changes. Failing to provide professionals with reliable instruments can hinder their work, reduce the ability to rapidly detect child maltreatment and delay any potential interventions (Saini et al., 2019); these consequences can be dramatic for children, adolescents and their families and that is why early detection must be priority in child protection services (Cowley et al., 2018; Petrowski et al., 2021). Thus, it is essential to have available scientifically rigorous and evidence-based prevention of child maltreatment.

There are several instruments that have consistently appeared in the scientific literature in the last years and their psychometric properties need to be further scrutinized.

First, one of the most validated instruments to measure child maltreatment is the short form of the Childhood Trauma Questionnaire Short Form (CTQ-SF; Bernstein et al., 2003). According to Schmidt et al., (2020) and Madison et al., (2020), the CTQ-SF is one of the most validated and most widely used instruments to assess child maltreatment for research. This questionnaire, firstly developed by Bernstein and Fink in 1998, is a retrospective instrument that assesses several types of childhood abuse and maltreatment. The first version had 70 items divided into five dimensions. Five years later, Bernstein et al. (2003) developed a short form of the CTQ after estimating exploratory and confirmatory factor analyses, obtaining a 25-item scale, with 3 additional items of minimization and denial. There are 5 items for each original dimension of the scale (emotional, physical, and sexual abuse and physical and emotional neglect). All items are rated on a 5-point Likert scale from 1 (never true), to 5 (very often true). The CTQ-SF has been validated in different samples, such as, street youth (Forde, et al., 2012), prisoners (Dovran et al., 2013), adolescents, substance users, psychiatric inpatients (Bernstein et al., 2003), juvenile delinquents (*Kongerslev et al., 2019), and pregnant women (Schmidt et al., 2020). Additionally, Saini et al., (2019) performed a systematic review of instruments in which this instrument was identified as one of the three instruments with the strongest psychometric properties of the 52 examined, obtaining strong evidence for internal consistency and structural validity, and moderate evidence for hypothesis testing and cross-cultural validity. An even more recent systematic review of the psychometric

properties of the CTQ-SF concurred about the structural validity of the CTQ-SF, although there were some concerns about its internal consistency and there was a lack of information about many psychometric properties (Georgieva et al., 2021).

Secondly, the Maltreatment and Abuse Chronology of Exposure scale (MACE; Teicher & Parigger, 2015) is a retrospective measure developed to measure the severity of exposure to ten types of child maltreatment during child development. The main innovation in this instrument is the assessment of maltreatment over time. The ten factors of the scale are: emotional neglect, non-verbal emotional abuse, parental physical maltreatment, parental verbal abuse, peer emotional abuse, peer physical bullying, physical neglect, sexual abuse, witnessing inter-parental violence, and witnessing violence towards siblings. The MACE has 52 items, and it was developed under Item Response Theory (IRT). Convergent validity has been examined by authors of the scale with the Childhood Trauma Questionnaire (CTQ; Bernstein, et al., 2003) and Adverse Childhood Experiences Questionnaire (ACE) (CDC, 1997; Felitti et al., 1998). Additionally, they proved predictive validity with adult psychopathologies, and test-retest within a 6 months period in youth and adults (Teicher & Parigger, 2015). This scale has also been validated in Norwegian in psychiatric outpatients and employees (Fosse et al., 2020) and in Brazilian-Portuguese language (Kluwe-Schiavon et al., 2016). The psychometric properties of this instrument have been examined in a recent systematic review concluding that it had strong evidence for hypothesis testing for construct validity, and moderate evidence for reliability, measurement error, and content validity (Saini et al., 2019).

In third place, the Child Abuse Potential Inventory (CAP; Milner, 1986) is a scale designed to assess caregiver's probability of physically mistreating or abusing a child, with a binary response ("agree"/"disagree"). It is a self-report measure filled up by caregivers with 77 items divided into six factors (family problems, problems with others, distress, rigidity, unhappiness, and problems with the child and with oneself) plus three validity scales (lies, random responses, and inconsistencies) plus responses distortion indexes as, for instance, faking good or bad responses. Milner (1986) obtained good psychometric results in the first validation of the scale. Years after that, Walker and Davies (2010) performed a systematic review on the psychometric qualities of the Child Abuse Potential Inventory from 1994 to 2010, in which they obtained similar results to Milner (1986). The CAP scales obtained good internal consistency estimates across a variety of samples, and similar factorial structure (Kutsal et al., 2011; Milner, 1986; Wells et al., 2011). Additionally, differential validity has been explored with different sensitivity and specificity classification rates, as well as the capacity of detecting changes in pre-post interventions in the study of Walker and Davies (2010). Their research, however, finishes suggesting deeper research on the properties of this scale. In the same line, Milner and Crouch (2012) explored the

psychometric adequacy of translated versions of this scale, concluding that the reliability and classification rates were similar to those reported for the English version. This scale has been validated in parents who have and who have not abused their children (Kutsal et al., 2011), home visited at-risk adolescent mothers and undergraduate students (Wells et al., 2011).

Additionally, the Identification of Parents at Risk for child Abuse and Neglect (IPRAN; Bouwmeester-Landweer, 2006) is a three-page self-report instrument developed to detect parents of newborns who are at risk of experiencing parenting stress which can later develop in potential abuse or neglect. This instrument is composed by 37 items with an assigned score from 0 to 2. Items are divided into three main factors: child and family characteristics, parental developmental history and personality, and characteristics of the social context. Additionally, there is a general part of 5 items which are filled by both, mothers and fathers (or other caregivers) related to the first factor: child and family characteristics; and a specific part of 16 items which is filled separately by mothers, and fathers or other caregivers (related to the two remaining factors). This scale was validated with good psychometric properties in samples of parents of newborns considered at risk versus a control group (Horrevorts et al., 2017; van der Put et al., 2017a).

Lastly, the Psychosocial Screening Tool (PAT; Pai et al., 2008) is a brief screening tool developed to evaluate psychosocial risk in families with children recently diagnosed with cancer. This instrument is based on the Pediatric Preventive Psychosocial Health Conceptual Model (PPPHM; Kazak, 2006). It is composed by a first section of demographic questions, and a second section divided into seven specific factors: family structure and resources, social support, child problems, sibling problems, caregivers stress reactions, family problems, and family beliefs. Different sections provide different response formats depending on the nature of questions (yes/no answers, frequency, amount of concern). Authors have reported good psychometric results evaluating internal consistency and test-retest reliability (Pai et al., 2008). This scale has also been validated among Canadian children newly diagnosed with cancer (Barrera, et al., 2014), Australian children and infants admitted to oncology, cardiology, or pediatric intensive care (McCarthy et al., 2016) and Dutch families of children diagnosed with cancer (Sint Nicolaas et al., 2016).

Purpose of the Current Study

Child maltreatment causes deep sorrow to children and their families and it can lead to lifelong consequences in children's physical and mental health (WHO, 2020). Stoltzenborgh et al., (2015) examined the prevalence of child abuse across the globe and concluded that child maltreatment is a widespread, global problem affecting the lives of millions of children all over the world. A systematic review and meta-analysis performed in 2020 concluded that child physical abuse, emotional

abuse, and neglect is associated to several mental disorders, drug abuse, suicide attempts, sexually transmitted infections, and risky sexual behavior (Norman et al., 2012). In the same vein, more recent research agrees that being a victim of maltreatment as a child increases the probability of emotional disorders as depression and anxiety (Hanson et al., 2015; Kisely et al., 2018); or even thinking of taking their own life (Newins et al., 2021)

According to Evans, Gardner, and Honig (2014), in the last years, child maltreatment has reached epidemic levels. According to WHO, 2020, 25% of adults confirm having been physically maltreated as children, and 20.7% report having been sexually abused as a child (WHO, 2020).

Early detection and efficient and reliable assessment of child maltreatment is essential for preventing all the personal and social negative consequences of child maltreatment (Tufford et al., 2021; *van der Put et al., 2017b). Therefore, it is crucial to constantly revise and improve the ways child maltreatment is assessed, since an early detection might greatly reduce the risk of experiencing negative long-term consequences. The main focus of this study is to critically appraise, compare, and summarize the methodological quality and psychometric properties of published research articles validating the five most recurrently validated child abuse assessment scales from 2010 to 2020 utilizing the Consensus based Standards for the selection of health status Measurement Instruments (COSMIN) checklist (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). To fulfill this aim, a systematic review was conducted, detecting the instruments that have been validated at least twice in this time period. The CTQ-SF (Bernstein, et al., 2003), the MACE (Teicher & Parigger, 2015), the CAP inventory (Milner, 1986), the IPRAN (Bouwmeester-Landweer, 2006), and the PAT (Pai et al., 2008) were analyzed considering the following information: instrument names and authors, sample characteristics, country, interpretability, generalizability, and psychometric properties.

Method

Eligibility Criteria

Articles which directly or indirectly analyzed psychometric properties of the aforementioned instruments were selected for this study. In order to guarantee the thoroughness of the data, only studies which met the following criteria were included: (1) original articles published in English or Spanish in journals with peer-reviewed process, (2) articles published from January 2010 to August 2020, (3) articles published in scientific journals indexed in the Journal Citation Reports (JCR) and/or Scimago Journal and Country Rank (SJR). On the other hand, exclusion criteria were: (1) articles considered gray literature (doctoral thesis, conference papers, press articles, etc.), thus, scientific quality was assured by including only peer reviewed articles, (2) articles without psychometric information

Table 1. Terms used for the Boolean Search.

Children or Adolescent		Risk of Maltreatment		Assessment
child OR *children* OR *childhood* OR *adolescence* OR *teenager* OR *Early life* OR	AND	*risk* OR *adversity* OR *maltreatment* OR *trauma* OR *victimization* OR *neglect*	AND	*assessment* OR *tool* OR *instrument* OR *scale* OR *questionnaire* OR *battery* OR *inventory* OR *structured* OR *recognition* OR *index*

Note. Search terms were used in both, English and Spanish language.

regarding the scales, and (3) systematic reviews of articles. Because of the diversity of different constructs measuring child maltreatment, the research team decided to include all instruments which represent any form of risk for children or adolescents, and that have been validated at least twice in the searched time period. Additionally, it is important to examine the most recent research on instruments because of the rapidly evolving contexts for children and adolescents, which means that psychometric properties might also change. Thus, it is important to stay updated and re-examine instruments in different samples, contexts, and time periods.

Information Sources and Search Strategy

A systematic search of the following databases was performed from June to September of 2020: ProQuest, Scopus, and Web of Science. Regarding search terms, 21 terms were included (see Table 1), which could be subdivided into three semantic groups: Children or adolescents (5 terms), risk (6 terms), and instruments or questionnaires (10 terms). Similar terms were included in every group in order to cover the maximum amount of information. The search was restricted to articles published from 2010 to 2020. Only instruments that measured any form of type of child maltreatment were included, in this way, instruments assessing emotional disorders were not included. Instruments that were validated at least twice in this period were retained.

Selection Process and Data Collection Process

First, two independent researchers screened the titles and abstracts of the identified articles and evaluated the selected full texts. In cases of disagreements about inclusion, they were discussed with a third judge until reaching full agreement on the inclusion of all articles, following the recommendations of Bekhet and Zauszniewski (2012), who recommend for researchers to perform the search process independently.

Eligible studies were fully read by both researchers and the following information was extracted: Information on names of instruments and authors of the study, construct to be measured and subscales, response format, target population and characteristics of the samples in which the instrument was applied, mode and language of administration, country; interpretability and generalizability of results; and psychometric properties of the instruments.

This systematic review followed the “Preferred Reporting Items for Systematic Review and Meta-analysis Protocols” (PRISMA-P; Moher et al., 2015), an established protocol for reporting this type of systematic reviews.

Risk of Bias Assessment: The Consensus Based Standards for the Selection of Health Status Measurement Instruments Checklist for Systematic Reviews of Patient-Reported Outcome Measures

Appraisal of the methodological quality of studies was based on the Consensus based Standards for the selection of health status Measurement Instruments (COSMIN) checklist for Patient-Reported Outcome Measures (PROMs) (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). COSMIN is an international initiative created by a group of multidisciplinary researchers with knowledge about the evaluation of outcome measurement instruments in the field of health care, performing systematic reviews of measurement instruments, epidemiology, psychometrics, and qualitative research. The main purpose of COSMIN is to improve the selection of outcome measurement tools in research by creating instruments for selecting the most appropriate instrument for the situation at issue. In the COSMIN manual, there is a distinguishment between “standards of measurement” and “criteria.” Standards refer to the research design requirements of every study, in other words, the quality of the study or article. Criteria, is what constitutes a good result for a measurement property (the quality of the PROM) (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). The COSMIN manual provides a comprehensive guideline divided into three steps and 10 boxes, suggested to be used as a modular tool, meaning that only boxes of interest might be assessed.

The first part is focused on standard procedures performing systematic reviews in agreement with existing guidelines as Cochrane (Higgins et al., 2019) and PRISMA-P (Moher et al., 2015), the second part consists of the evaluation of the results regarding the measurement properties of the instrument, and the third part is concerned with the interpretability and feasibility of the PROM, as well as formulating and reporting the systematic review.

In this study, assessment was performed in 3 steps: first, the methodological quality of the included studies was evaluated; second, criteria for good measurement properties were

applied; and third, evidence from utilizing the modified GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) approach was extracted to summarize results.

Step 1. Assessment of Methodological Quality of the Studies

Methodological quality of the included articles was assessed in three main groups: content validity, internal structure, and remaining measurement properties. Content validity was composed by two boxes in the instrument: PROM development and content validity, about which authors note that, although PROM development is not a measurement property, it is still relevant for the assessment of content validity. Internal structure was composed by three boxes: structural validity, internal consistency, and cross-cultural validity/measurement invariance. Finally, remaining measurement properties were reliability, measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness, assessed by one box each. All items were scored on a 4-point scale ranging from excellent to poor. General score of each property was determined by the lowest score (1–4) given to an item in every box (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018).

Step 2. Assessment of Quality of the Patient-Reported Outcome measures

Based on the criteria for good measurement, in this part the psychometric results of the instrument itself were assessed, specifically the structural validity, internal consistency, reliability, measurement error, hypothesis testing, cross-cultural validity/measurement invariance, criterion validity, and responsiveness. Quality of the PROM was assessed in a three-points scale: sufficient (+), insufficient (–), and indeterminate (?), based on the updated criteria for good measurement properties shown in the COSMIN manual (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018).

Step 3. Strength of Evidence of the Included Instruments

Finally, results present the summarized strength of evidence of each instrument, extracted based on the modified GRADE approach. This approach determines an overall assessment of each property ranging from strong evidence, to moderate, limited, conflicting, and unknown; considering both of the above-mentioned assessments together (Prinsen, et al., 2018). Strong evidence was reported when there were several methodologically good articles or one of excellent quality, moderate evidence was indicated when there are several methodologically fair articles or one of good quality, limited evidence was considered with articles with fair quality, and conflicting evidence was assigned to low quality articles. Unknown evidence was assigned to studies in which information regarding psychometric properties was missing.

Results

A total of 22,193 research articles were identified through the search of three databases, from these, the full text of 28 articles was examined, from which 19 articles were selected for inclusion in this systematic review. The full process of selection is shown in Figure 1.

Characteristics of the Included Patient-Reported Outcome Measures

Characteristics of the examined PROMS can be seen in Table 2. The final 19 articles examined the psychometric properties of five instruments assessing child maltreatment in its different manifestations: there were two studies available for the CAP scale (Milner, 1986); ten studies for the CTQ-SF (Bernstein et al., 2003); two studies for the IPRAN (Bouwmeester-Landweer, 2006); three studies for MACE scale (Teicher & Parigger, 2015); and two studies for the PAT (Pai et al., 2008).

All scales evaluated some aspect of child maltreatment, but they assessed different constructs, for instance, the CAP is especially focused on the risk of physical abuse, while the CTQ-SF assesses traumatic experiences, similar to the MACE scale, which assesses the severity of exposure to child maltreatment. Additionally, the IPRAN instrument is focused on assessing the risk of parenting stress which may result in child maltreatment, having its focus on prevention, while the PAT is focused on a very specific target population: families of children who are newly diagnosed with cancer.

Considering these differences in constructs to be measured, the target populations are still quite homogeneous, since virtually all instruments are meant to evaluate adults, most commonly parents or caregivers, and all of them are self-reports. Specifically, there are two scales which are also retrospective, these are the CTQ-SF (Bernstein et al., 2003) and the MACE scale (Teicher & Parigger, 2015).

Regarding the internal structure of the instruments, the CAP instrument has subscales intended to measure lies, random responses and inconsistencies, while the IPRAN and the PAT consider some environmental factors and the parental developmental history or the social support. The CTQ-SF is divided into the classic subscales of child maltreatment (physical, emotional, and sexual abuse, and physical and emotional neglect). Finally, the Maltreatment and Abuse Chronology of Exposure scale adds new subscales which take into account peer physical and emotional abuse, and witnessing violence at home.

It is also relevant to consider the usability of these instruments, when it comes to time costs. From this perspective, the longest scale is the Child Abuse Potential Inventory, while the shortest is the Childhood Trauma Questionnaire. The remaining three are medium length ranging from 37 to 52 items. Response formats are mostly dichotomous (yes/no or

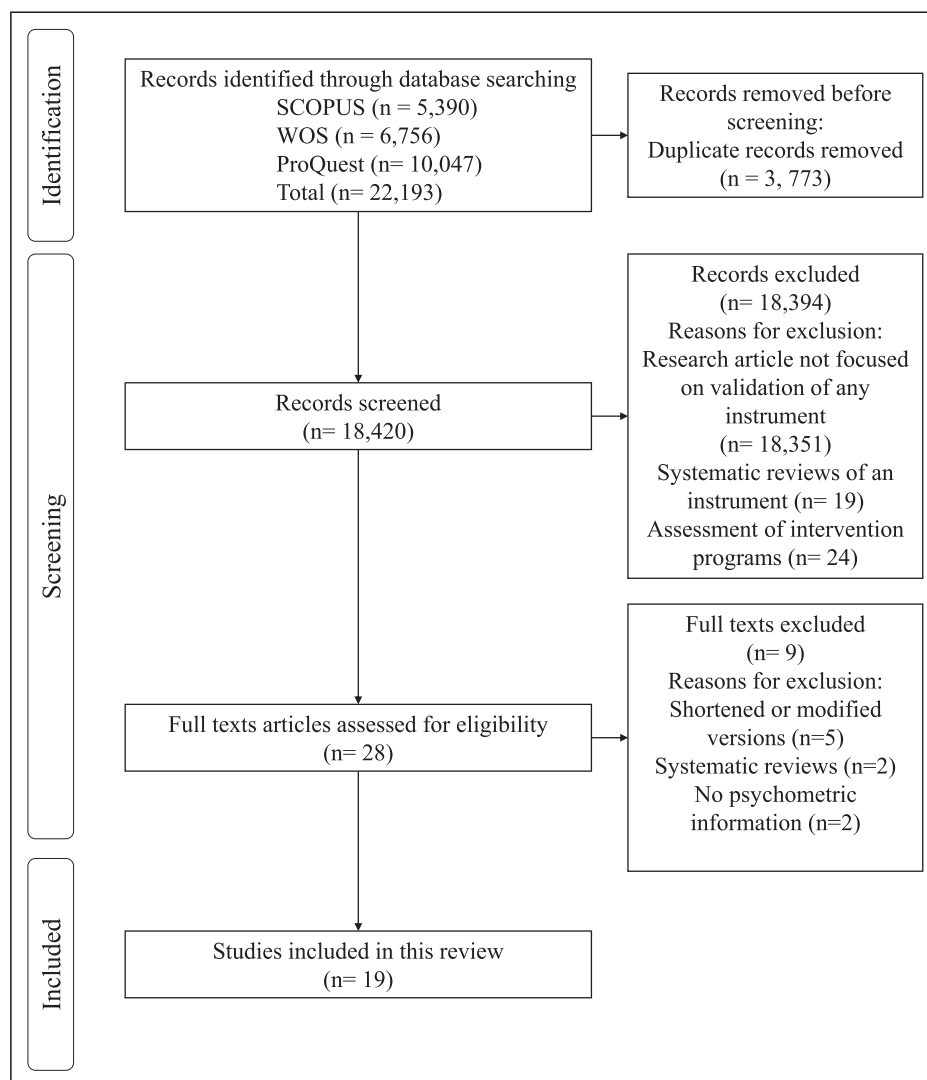


Figure 1. Flowchart of the selection process.

agree/disagree), except for the CTQ-SF and the IPRAN in which responses are measured on 4- and 5-point Likert scales, respectively.

Finally, regarding languages of validation, from the results of this search which considers the past ten years, we can confirm that all scales are available in English language. The CAP scale has been validated in Turkish. The CTQ-SF has been validated in Afrikaans, Spanish, Norwegian, German, French, Chinese, and Danish. The IPRAN has been validated in Dutch language. The MACE scale has been translated to Brazilian Portuguese and Norwegian, and the PAT is translated to Dutch, as far as we are concerned.

Characteristics of the Samples

If we focus on the samples in the examined studies, we can conclude that the vast majority of participants were youth and adults, with a mean of age from 19.1 (SD= 3.5) in [Dovran](#)

[et al., \(2013\)](#), to a mean of 47 years old (SD= 10.8) in [Fosse et al., \(2020\)](#) (see [Table 3](#)).

When it comes to the profile of participants, in [Table 3](#) we can appreciate that the CAP scale has been validated in a sample of parents who have abused and who have not abused their children; in home-visited at-risk adolescent mothers, adult mothers, and undergraduate students. The CTQ-SF has been explored in an out of home placement group, people suffering from drug addiction or from mental health problems, persons with justice system involvement, psychology students, normative adolescents and youth, young people with justice system involvement, females living and not living with HIV, and pregnant women. The IPRAN has been validated in a sample of parents at risk, parents considered to not be at risk, and families with newborns. The MACE scale has been studied in online and interviewed participants, as well as employees and outpatients. Finally, the PAT scale has been validated in children newly diagnosed with cancer, and other

Table 2. Characteristics of the included PROMs (Patient-Reported Outcome Measures).

Instrument	No of Studies Available*	Construct	Target Population	Mode of Administration	Subscales	Items	Response Format	Available Translations*
Child Abuse Potential Inventory (CAP)	2	Risk of child physical abuse	parents at high risk for child physical abuse	Self-report	Abuse scale, lie scale, random response, and inconsistency scale	160	Dichotomic (agree/disagree)	Turkish
Childhood Trauma Questionnaire-Short Form (CTQ-SF)	10	Traumatic experiences in childhood	Adults	Retrospective self-report	Physical Abuse; Emotional Abuse; Sexual Abuse, Physical Neglect; Emotional Neglect	25	5-point Likert scale (Never true, Rarely true, Sometimes true, Often true, Very often true).	Afrikaans, Spanish, Norwegian, German, French, Chinese, and Danish
Identification of Parents At Risk for child Abuse and Neglect (IPARAN)	2	Risk of parenting stress	Parents with newborn children	Self-report	Child and family characteristics, parental developmental history and personality and characteristics of the social context	37	4-point Likert scale (always, often, sometimes, never) or a dichotomic (yes/no) option	Dutch
Maltreatment and Abuse Chronology of Exposure (MACE) Scale	2	Severity of exposure to child maltreatment	Young adults	Retrospective self-report	Emotional neglect, non-verbal emotional abuse, parental physical maltreatment, parental verbal abuse, peer emotional abuse, peer physical bullying, physical neglect, sexual abuse, witnessing interparental violence and witnessing violence to siblings	52	Dichotomic (yes/no)	Brazilian Portuguese, and Norwegian
Psychosocial Assessment Tool (PAT)	3	Psychosocial risk in children who are newly diagnosed with cancer and their families	Caregivers of children newly diagnosed with cancer	Self-report	Family structure and resources; social support; child problems; sibling problems; caregiver stress reactions; family problems; and family beliefs	52	Categorical answers and dichotomic (yes/no)	Dutch

*Translations available with the specified search strategy and time period.

Table 3. Characteristics of the Samples.

Instrument	Reference	n	Age (Mean, SD)	Gender (% Female)	Country	Language
CAP	Kutsal et al., (2011)	23 Parents who have abused their children	36.6 (-)	26%	Turkey	Turkish
		47 Parents who have not abused their children	35.1 (-)	27%		
	Wells et al., (2011)	796 Home-visited at-risk adolescent mothers	*	100%	-	-
		959 Home-visited at-risk adult mothers	*	100%		
		142 Teen parent services adolescent and adult mothers	*	100%		
		1,028 Undergraduate students	*	100%		
CTQ-SF	Forde, et al., (2012)	397 Street youth who have run away or been expelled from their homes	20 (2.6)	34%	Canada	English
		72 Out-of home placement group	19.1 (3.5)	56%	Norway	Norwegian
	Dovran, et al., (2013)	126 Persons suffering from drug addiction and/or from mental health problems	30.9 (11.6)	40%		
		210 Persons seeking or receiving treatment for psychiatric symptoms	25.5 (10.7)	70%	Spain	Spanish
		109 Persons with justice system involvement	32.2 (9.9)	0%		
		185 Inpatient and outpatient females	41.64 (11.83)	100%		
	Karos et al., (2014)	565 Clinical patients	36.64 (11.79)	-	Sweedden	German
		96 Psychology students	25.7 (5.43)	-		
	*Charak et al., 2017	321 Adolescents and youth	14.9 (1.99)	58.4%	Burundi	Kirundi
	*He et al., 2019	3431 Undergraduate students	19.6 (1.0)	43.4%	China	Chinese
		217 Persons suffering from depression	22.6 (5.7)	45.6%		
	*Kongerslev et al., 2019	142 Patients enrolled as part of routine assessment	29 (8.4)	68%	Denmark	Danish
		101 Persons commencing treatment for psychosis	23 (3.4)	26%		
		71 Persons diagnosed with depression	35 (11.6)	78%		
		80 Young people with justice system involvement	17 (0.8)	0%		
	*Spies et al., 2019	314 Females (170 HIV uninfected, 144 HIV infected)	30.7 (8)	100%	South Africa	English and Afrikaans
	*Aloba et al., 2020	1337 High school adolescents	15.24 (1.43)	54.7%	Nigeria	English
		77 Pregnant women	29.19 (6.55)	100%		
IPRAN	Horrevorts et al., (2017)	177 Parents at risk	31.65 (5.65)	-	The Netherlands	Dutch
	van der Put et al., (2017)	550 Parents not at risk	31.1 (4.55)	-	The Netherlands	-
MACE	Teicher and Parigger (2015)	1051 Participants online	23.1 (1.7)	63.7%	United States	English
		407 Interviewed participants	23.1 (1.7)	61.4%		
	Fosse et al., (2020)	145 Employees	47 (10.8)	83%	Norway	Norwegian
		90 Outpatients	34.6 (10.6)	69%		
PAT	Barrera et al., (2014)	43 Children newly diagnosed with cancer	8.58 (5.26)	34.9%	Canada	English
	McCarthy et al., (2016)	177 Children admitted to oncology, cardiology, or pediatric intensive care	*	60.47%	Australia	English
	Sint Nicolaas et al., (2016)	117 Children diagnosed with cancer	*	3.5%	The Netherlands	

Note. CAP = Child Abuse Potential Inventory, CTQ-SF = Childhood Trauma Questionnaire-Short Form, IPRAN = Identification of Parents At Risk for child Abuse and Neglect, MACE = Maltreatment and Abuse Chronology of Exposure Scale, PAT = Psychosocial Assessment Tool.

Table 4. COSMIN (Consensus Based Standards for the Selection of Health Status Measurement Instruments) Results of the Standards of Measurement (Methodological Quality) of the Articles Assessed.

Psychometric Property	Articles	Psychometric Property	Articles
<i>PROM Development</i>		<i>Reliability</i>	
Excellent		Excellent	
Good		Good	17
Fair		Fair	15
Poor	15	Poor	16
Unknown/NA	1-14; 16-19;	Unknown/NA	1-14; 18-19
<i>Content Validity</i>		<i>Measurement Error</i>	
Excellent	17	Excellent	
Good		Good	
Fair	15; 16	Fair	
Poor	19	Poor	
Unknown/NA	1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 18	Unknown/NA	1-19
<i>Structural Validity</i>		<i>Criterion Validity</i>	
Excellent	3; 5; 7; 8; 9; 10; 13; 15	Excellent	1; 5; 13; 14; 15; 16; 17; 18; 19
Good		Good	
Fair	6;	Fair	
Poor	4;	Poor	
Unknown/NA	1; 2; 11; 12; 14; 16; 17; 18; 19	Unknown/NA	2; 3; 4; 6-14;
<i>Internal Consistency</i>		<i>Hypothesis testing for construct validity</i>	
Excellent	3; 4; 5; 6; 7; 8; 10; 13; 17; 18; 19	Excellent	5; 6; 7; 8; 9; 13; 14; 18; 19
Good		Good	
Fair		Fair	11; 15; 16; 17
Poor	1; 2; 9	Poor	1; 12
Unknown/NA	11; 12; 14; 15; 16	Unknown/NA	2; 3; 4; 10; 14
<i>Cross-cultural Validity/Measurement Invariance</i>		<i>Responsiveness</i>	
Excellent	8; 15	Excellent	6; 7; 8; 9; 14; 18
Good		Good	5; 10
Fair	3; 13;	Fair	1; 11; 15; 16
Poor		Poor	14
Unknown/NA	1; 2; 4; 5; 6; 7; 9; 10; 11; 12; 14; 16; 17; 18; 19	Unknown/NA	2; 3; 4; 12; 13; 17; 19

Note. 1: Kutsal et al.; (2011); 2: Wells et al.; (2011); 3: Forde; et al.; (2012); 4: Dovran; et al.; (2013); 5: Hernández; et al.; (2013); 6: Karos et al.; (2014); 7: Charak; et al.; (2017); 8: He; et al.; (2019); 9: Kongerslev; et al.; (2019); 10: Spies, Kidd & Seedat (2019); 11: Aloba, Opakunle & Ogunrinu (2020); 12: Schmidt; et al.; (2020); 13: Horrevorts et al.; (2017); 14: van der Put et al.; (2017); 15: Teicher and Parigger (2015); 16: Fosse et al.; (2020); 17: Barrera et al.; (2014); 18: McCarthy et al.; (2016); 19: Sint Nicolaas et al.; (2016).

types of pathologies related to cardiology and pediatric intensive care.

There were three studies in which all participants were female (Hernández et al., 2013; Schmidt et al., 2020; Wells et al., 2011), and two studies which did not give information regarding gender (Horrevorts et al., 2017; Karos et al., 2014).

As it can be observed in Table 3, the countries where the studies were carried out were diverse: Netherlands, United States, Norway, Canada, Turkey, Spain, Sweden, Burundi, China, Denmark, South Africa, Nigeria, Australia, and one study that did not mention information regarding country (Wells et al., 2011).

Finally, language of administration was diverse. The most common language was English, although there were translations into several other languages as Turkish (Kutsal et al., 2011), Norwegian (Dovran et al., 2013; Fosse et al., 2020), Spanish (Hernández et al., 2013), German (Karos et al., 2014), Kirundi (*Charak et al., 2017), Chinese (*He et al., 2019), Danish (*Kongerslev et al., 2019), Afrikaans (*Spies et al., 2019), and Dutch (Horrevorts et al., 2017; Sint Nicolaas et al., 2016). Wells et al. (2011) did not mention any information regarding language of administration or country, although it could be assumed it was carried out in English language.

Table 5. COSMIN (Consensus based Standards for the selection of health status Measurement Instruments) results of the criteria of measurement (quality of the PROMs (Patient-Reported Outcome Measures)).

Instrument	Article	Structural Validity	Internal Consistency	Reliability	Measurement Error	Hypothesis Testing for construct validity	Cross-cultural validity/ Measurement Invariance	Criterion Validity	Responsiveness
CAP	Kutsal et al., (2011)	?	+	?	?	+	?	+	?
	Wells et al., (2011)	?	+	?	?	?	?	?	?
CTQ-SF	Forde, et al., (2012)	+	–	?	?	+	+	?	?
	Dovran, et al., (2013)	+	+	?	?	?	+	?	?
	Hernández, et al., (2013)	+	–	?	?	+	?	?	?
	Karos et al., (2014)	+	–	?	?	?	?	?	?
	Charak, et al., (2017)	?	–	?	?	+	?	?	?
	He, et al., (2019)	+	–	?	?	?	+	?	?
	Kongerslev, et al., (2019)	+	–	?	?	?	?	?	?
	Spies, et al., (2019)	+	–	?	?	?	?	?	?
	Aloba, et al., (2020)	+	–	?	?	?	+	?	?
	Schmidt, et al., (2020)	?	?	?	?	+	?	?	?
IPRAN	Horrevorts et al., (2017)	?	?	?	?	+	?	–	?
	van der Put et al., (2017)	?	?	?	?	+	?	+	?
MACE	Teicher & Parigger (2015)	+	?	+	?	+	+	+	+
	Fosse et al., (2020)	?	?	+	?	+	?	+	+
PAT	Barrera et al., (2014)	?	+	+	?	+	?	+	?
	McCarthy et al., (2016)	?	–	?	?	–	?	+	+
	Sint Nicolaas et al., (2016)	?	+	+	?	+	?	+	?
	Sufficient	47.3%	26.3%	21%	0%	57.9%	26.3%	36.8%	15.7%
	Insufficient	0%	47.3%	0%	0%	0.5%	0%	0.5%	0%
	Indeterminate	52.6%	26.3%	79%	100%	36.8%	73.7%	62.7%	84.3%

Notes: CAP: Child Abuse Potential Inventory, CTQ-SF: Childhood Trauma Questionnaire-Short Form, IPRAN: Identification of Parents At Risk for child Abuse and Neglect, MACE: Maltreatment and Abuse Chronology of Exposure Scale, PAT: Psychosocial Assessment Tool

+	Sufficient
–	Insufficient
?	Indeterminate

Methodological and Measurement Quality of the Studies

Methodological quality of eligible studies is summarized in Table 4. None of the included studies presented information regarding all eight COSMIN criteria. Particularly, no study examined measurement error. Most of the studies analyzed structural validity, however, all of them explored the structure of the CTQ-SF scale while the structure of most the remaining scales was unexplored. Test-retest reliability was mostly examined with the MACE and the PAT instruments. Hypothesis testing for construct validity was tested among half of the included studies (52.6%), while measurement invariance was tested in 26.3% of the articles, criterion validity was tested in 6 articles, but, one of them had poor quality (Sint Nicolaas et al., 2016). There was only one study which developed the instrument itself, so PROM development was only assessed for the MACE scale with Teicher and Parigger (2015).

Criteria of Measurement – Quality of the Assessed Patient-Reported Outcome Measures

Quality of PROMS is presented in Table 5. It was performed according to the updated criteria for good measurement properties shown in the COSMIN manual (Mokkink et al.,

2018; Prinsen et al., 2018; Terwee et al., 2018). Judgments on quality and application of the COSMIN manual were performed by two researchers, and again, if disagreements occurred, they were resolved by discussing with a third judge until reaching an agreement. In this case, there were no disagreements.

Regarding the included studies, we can appreciate that the CAP has obtained sufficient scoring on both studies examining its internal consistency, while hypothesis testing and criterion validity has only been proven in one of the two studies with good results. There is a lack of information regarding the rest of psychometric properties (structural validity, test-retest reliability, measurement error, measurement invariance, and responsiveness).

Second, the CTQ-SF is by far the most validated instrument, with a total of 10 studies assessing its psychometric properties in the established search period. It can be concluded that the structural validity of this instrument is well established; obtaining good results in all eight studies that examined it. Contrariwise, the same cannot be confirmed for its internal consistency, since results were considered sufficient for only one out of nine studies, mainly because one or two factors of the scale did not reach the established alpha or omega of .7 or superior established by the criteria of good measurement properties shown in the COSMIN manual.

Table 6. Strength of Evidence of Each Study.

	Structural Validity	Internal Consistency	Reliability	Measure- ment Error	Hypothesis Testing for construct validity	Cross- cultural validity/ MI	Criterion Validity	Respons- iveness	% strong- moderate evidence
CAP	U	L	U	U	L	U	L	U	0%
CTQ-SF	S	C	U	U	M	S	U	U	37.5%
IPRAN	U	U	U	U	S	U	L	U	12.5%
MACE	S	U	S	U	S	S	S	S	75%
PAT	U	M	S	U	M	U	S	S	62.5%
% strong-moderate evidence	40%	20%	40%	0%	80%	40%	40%	40%	
% limited conflicting evidence	0%	40%	0%	0%	20%	0%	40%	0%	
% unknown evidence	60%	40%	60%	100%	0%	60%	20%	60%	

Notes: Right hand column represents the % of strong-moderate evidence for each article, bottom row indicated the strength of evidence for each psychometric characteristic evaluated by COSMIN. ¹MI: Measurement Invariance, ²ME: Measurement error

S	Strong
M	Moderate
L	Limited
C	Conflicting
U	Unknown

Additionally, hypothesis testing for construct validity and measurement invariance were examined by four studies obtaining sufficient results in all of them. However, there was a lack of information regarding test-retest reliability, measurement error, criterion validity, and responsiveness.

Additionally, the IPARAN scale is the scale with the least information available since it was examined by only two studies which reported information about only two psychometric properties: hypothesis testing for construct validity, with good results, and criterion validity, with conflicting results since they were sufficient in one study but insufficient in the other which examined it. There was a lack of information regarding all the remaining psychometric properties.

Regarding the MACE scale, the results suggest that there is evidence for sufficient test-retest reliability, construct validity proven by hypothesis testing, criterion validity and responsiveness, while only one of the two studies examined structural validity and measurement invariance obtaining good results. There was absence of information regarding only two psychometric properties: internal consistency and measurement error.

Finally, the PAT was validated by three studies, obtaining sufficient evidence for criterion validity in all three of them. Good results were also obtained in the two studies examining reliability, and the one study that examined responsiveness. However, there was conflicting evidence in the results obtained for internal consistency and hypothesis testing for construct validity since two out of three studies had good results but one study reported insufficient results based on the

updated criteria for good measurement properties (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018).

Strength of Evidence

The strength of evidence for each instrument, assessed according to the modified GRADE criteria (Prinsen, et al., 2018) is presented in Table 6. According to this approach, each instrument is assessed taking into consideration four factors: (1) risk of bias (i.e. methodological quality of the studies); (2) inconsistency (i.e. unexplained inconsistency of results across studies); (3) imprecision (i.e. total sample size of the available studies); and (4) indirectness (i.e. evidence from different populations than the population of interest in the review) in order to provide with a summary of results.

With the current information, we can conclude that the CAP has limited evidence of internal consistency, hypothesis testing, and criterion validity. It is considered limited, because although results related to the psychometric property were good, the quality of the articles, according to the COSMIN criteria, was considered poor (see Table 3). Additionally, there is a lack of information regarding its structural validity, test-retest reliability, measurement error, measurement invariance, and responsiveness.

For the CTQ-SF, we can conclude that there is strong evidence supporting its five-factor structure; however, internal consistency results are conflicting since there are consistent problems with several dimensions of the scale (physical neglect, emotional neglect, physical, and emotional abuse).

Hypothesis testing for construct validity had moderate evidence, while there was strong evidence for measurement invariance. There was no information regarding test-retest reliability, measurement error, criterion validity and responsiveness.

The Identification of Parents At Risk for child Abuse and Neglect instrument was the scale with the least information available. There were only two psychometric qualities for which we could conduct an assessment: hypothesis testing for construct validity and criterion validity. We obtained strong evidence for hypothesis testing, and low evidence for criterion validity. In criterion validity, both articles showed excellent quality, however results regarding the psychometric property were conflicting, in the way that one article supports the construct validity of the IPRAN while the results of the other indicate the contrary.

For the Maltreatment and Abuse Chronology of Exposure scale, there was strong evidence for structural validity, reliability, hypothesis testing, criterion validity, and responsiveness, while there was no information regarding internal consistency and measurement error.

Finally, the Psychosocial Screening Tool showed moderate evidence for internal consistency and hypothesis testing for construct validity, and strong evidence for test-retest reliability and responsiveness. There was a lack of information regarding structural validity, measurement error, and measurement invariance.

Discussion and Conclusions

The main objective of this study was to critically appraise, compare, and summarize the methodological quality and psychometric properties of five of the most recurrently validated child maltreatment assessment scales from 2010 to 2020 utilizing the Consensus based Standards for the selection of health status Measurement Instruments (COSMIN) checklist (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018).

Results of this study were partially congruent with previous literature. The last systematic review of child maltreatment assessment instruments to our knowledge, performed by Saini et al. (2019), indicated that there is a lack of information regarding some specific psychometric properties such as measurement error, content validity, and responsiveness. The most examined properties in Saini et al. (2019)'s results were test-retest reliability, internal consistency, content validity, and hypothesis testing, although most results were considered moderate to low evidence. They concluded that there was no single instrument that was superior to others across settings and populations, since they were missing information regarding several criteria in the COSMIN checklist.

In our results, the most examined psychometric property among all the scales was hypothesis testing for construct validity, which was tested in at least one study for each of the five examined scales. These results are slightly better than the

findings of Saini et al., (2019) in which hypothesis testing was examined for 38.4% of the examined instruments, although, there was only 19.2% which presented strong to moderate evidence, compared to 80% of strong to moderate evidence in our study. This result indicates that in the period from 2010 to 2020, hypothesis testing for construct validity has been more present in scientific literature.

The following most examined property was internal consistency, as it was examined in 73.7% of the total of the studies, however, most of them were examining the CAP and the PAT scales. This psychometric property was not examined for the IPRAN and MACE scales. This result indicates that there is a general trend in examining almost the same psychometric properties. This could be due to the fact that performing a cross-sectional study takes less research resources as time, money, and complexity than a longitudinal one. Compared to Saini et al., (2019) in which only 26.9% of scales reported information regarding internal consistency, our results are slightly better since there was information regarding 60% of the instruments examined in our study.

When considered all together, the most relevant part of our results is the huge lack of scientific rigor, conceived as accepting as valid instruments that have unknown evidence for many psychometric properties, as there is up to 100% unknown evidence for some psychometric properties: There was no information in any study regarding any instrument about measurement error, up to 84.3% of lack of information about responsiveness, 79% lack of information about test-retest reliability, and 73.7% for cross-cultural validity/measurement invariance. According to the latest COSMIN manual (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018), measurement error is "the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured," it is a relevant psychometric property to examine since it shows the "precision" of the instrument. However, as evaluated by the COSMIN manual, the sample needs to be tested at least in two different moments in as similar conditions as possible. The same occurs with test-retest reliability and responsiveness; as reliability is defined as "the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions" and responsiveness is understood as "the ability of a PROM to detect change over time in the construct to be measured" (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). These conditions make its accomplishment a harder task than carrying out a structural validity analysis, for instance, as measurements need to be undertaken in at least two different moments in time. This might be the reason of this lack of information, since many times the available time and resources a limit.

Additionally, there was a rate of 26% up to 62.7% of unknown evidence for the rest of psychometric properties (see Table 5). This lack of information creates a serious problem regarding the generalizability and decision taking based on the five childhood maltreatment instruments. Performing research

with instruments of poor or unknown quality is unethical, especially considering the nature of this topic, and can also be a waste of resources (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). In fact, if we carefully explore the limitations sections of the studies, it is sticking to discover that none of them gives a proper justification regarding why the rest of the psychometric properties were not examined. Considering the delicate nature of the construct to be measured, this is extremely dangerous, especially when more than half of the examined studies are entitled as “Validation of...” or similar. We understand that some studies’ objectives are narrower as they focus in one of two psychometric properties, indicated in their titles, however, there are very few explanations regarding why they did not examine the rest of the properties, particularly, there is only one (*Charak et al., 2017) who justifies why they did a cross-sectional study so invariance analyses over time were not possible, in the same line as Dovran et al., 2013; *Kongerslev et al., 2019, who mentioned they did not examine other psychometric properties, although they don’t give a reason for that.

It is important to distinguish the lack of information regarding a psychometric property from the limited or conflicting evidence regarding that property. In the first case, there is still a big scientific gap to examine. However, the second indicates that the instrument to which it refers should be used with extreme caution. The most common result in our study was the unexamined psychometric properties as this happened from 26.3% of studies for criterion validity, up to 100% of studies for measurement error. These are some noticeable lacks of information that strongly condition the assessment of psychometric quality of instruments.

The implications of these results in research can be uncertain conclusions in studies which utilize these instruments, especially the ones with the least available information. These uncertain conclusions might lead to possibly wrong recommendations that once applied can suppose a waste of time and resources. Additionally, we also want to highlight how paramount is to consider that besides the research purposes, social work professionals use these or similar instruments as a support for taking life changing decisions. Without the ability to effectively detect child maltreatment, interventions will be hindered or delayed (Saini et al., 2019), wrong decisions could be taken, sometimes leading to dramatic consequences in children and their families, as it could be maintaining a child in the household under extreme conditions, or the opposite, separating a child from the caregivers when there might be a possibility of a successful intervention. Our results strongly suggest that competent institutions in child protection should invest and focus in providing their professionals with updated, methodologically rigorous, and psychometrically sound assessment instruments to allow them perform their job better, faster, and more efficiently. In addition to this, we would strongly suggest for future research to examine and fill up the gaps of psychometric properties that remain unexplored in order to be able to

take informed decisions on which instrument is the most suitable in each case.

With the available information regarding the strength of evidence for the five examined instruments, it is not possible to give a categorical answer regarding which instrument has the best psychometric properties since there is a big lack of information regarding some properties and a big variety of constructs assessing different parts of child maltreatment. It is important to highlight that there is no superior instrument that can be utilized for everything, no one-fit-all solution. This systematic review aims to provide a comprehensive review on the main psychometric properties of several instruments in order to facilitate researchers the selection of the most suitable instrument for their research. Independently on the prevention or retrospective focus of the study, or the specific construct of child maltreatment that is assessed, we strongly believe that all utilized instruments should present strong psychometric evidence. Thus, it seems that in the research articles published between 2020 and 2020, the instrument which has the strongest psychometric evidence is the MACE scale (Teicher & Parigger, 2015), since it covers six of the eight psychometric properties examined by the COSMIN tool with good results. The second instrument with the strongest evidence is the PAT (Pai et al., 2008) as it covers five of the eight properties with satisfactory results; however, it is important to note that this instrument was developed for a very specific population which is parents of children newly diagnosed with cancer. This means that this instrument can be highly useful in this context but should also be examined in samples with different diseases to see if psychometric properties remain acceptable. The instrument with the weakest evidence was the CAP (Milner, 1986) since only three psychometric properties were examined, all of them with limited evidence.

Our results also showed that the CTQ-SF is by far the most validated scale of all included instruments in this study, although the most frequently utilized does not necessarily mean the one with the strongest psychometric evidence, as it has shown consistent problems with the internal consistency of the physical neglect factor and several others, these results are consistent with Georgieva et al., (2021)’s systematic review regarding the CTQ-SF psychometric quality, although not consistent with Saini et al., (2019)’s results. We would recommend a further examination of internal consistency of this instrument as well as examining reliability, measurement error, criterion validity, and responsiveness. Again, it is important to note that a lack of information does not necessarily mean poor results, but our results clearly point out that many psychometric properties need to be examined and revised in order to obtain consistent results.

Additionally, two of the five assessed instruments were retrospective, while the remaining three were utilized to prevent child maltreatment. Some of the examined studies mentioned this fact as a limitation (Aloba et al., 2020; Forde et al., 2012; Hernández et al., 2013; Karos et al., 2014). We believe that both types of instruments are necessary for

research depending on the specific research objectives and the nature, needs, characteristics, and circumstances of participants (Picornell-Lucas & López-Peláez, 2022), although utilizing research for preventing child maltreatment would potentially be more beneficial for children, their families, and institutions since an early detection avoids a number of personal and social negative consequences (Tufford et al., 2021). In fact, an early detection can suppose a better scenario for intervention and transform existent risks in intervention objectives to be fulfilled in different fields as education, health, and social services.

Furthermore, another important comment is that, while all examined instruments were self-report which is understandable since self-reports are easier to administer in research, it is important to highlight that this type of instruments present several problems when it comes to objectivity. We point out three of them: 1) there is a big subjectivity margin when it comes to self-assessing somebody's own parental abilities, being open regarding violence at home, and reporting a subjective experience (Paulhus & Vazire, 2007); 2) items interpretation can vary between participants which can condition their answers, as it has been pointed out in the study of Sprague-Jones, et al., (2020) in the development of a Protective Factors Survey; and 3) the origin and environment of participants also can condition their answers since there are some collectives and cultures which have different interpretation of a good child education and can be more allowing to introduce violent discipline (Vachon et al., 2015). Hence, we suggest the development of scales that are filled up by an impartial third person or a qualified professional.

When it comes to diversity issues, it is important to note that only two from the 19 examined studies addressed diversity: the first study noted it as a limitation since there was a lack of diversity in the selection of participants (*He et al., 2019) while the other addressed diversity as examining participants which represent the ethnic diversity in Canada (Barrera et al., 2014). The vast majority of studies were performed in the European or North American continents (specifically, 13 out of 19 studies), while one was performed in Australia, one in Asia, and three studies were performed in the African continent. No study was performed in South or Latin America. These results are relevant because the psychometric properties of child maltreatment instruments remain unexamined in a big proportion of countries and cultures which might present totally different patterns of understanding and experiencing this complex topic, in fact some collectives and cultures differ in their interpretation of a good child education and can be more supportive of violent disciplines, for example (Vachon et al., 2015). These results are in accordance with Stoltenborgh et al., (2015), who studied the prevalence of child abuse across the globe and pointed out that child maltreatment research seems to be concentrated in countries with a Western culture. Specifically, South America might be in bigger need of having objective and sound child maltreatment assessment instruments. For example, a study exploring child

maltreatment in Ecuador concluded that up to 69.6% of normalized university students have experienced child maltreatment (Jiménez-Borja et al., 2020), or in Brazil where children are used for human trafficking (Raffaelli et al., 2018). Additionally, we have to point out that gender has been roughly equally distributed except for four studies in which either all participants were male or female. No alternative gender has been considered in any of the examined studies, which opens another gap of research for child maltreatment in children and adolescents identifying with alternative genders. Out of this we can conclude that these psychometric results could be generalized to Europe and North America, but they would need further testing in the rest of geographic and hence cultural varieties.

Our results should be taken with caution since there are some present limitations in this systematic review: (1) Only studies performed in English or Spanish language were included, which might result in the loss of valuable information published in other languages. (2) Although a very valuable tool, the COSMIN checklist (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018) allows some subjectivity in this interpretation and assessment of PROMS which might largely depend on the expertise and knowledge of the researchers assessing the instruments, as for example, in box 2 Content Validity, item 13 "Was an appropriate approach used to analyze the data?" or in box 7 (measurement error) regarding test-retest reliability, it can be stated that time interval was "appropriate," However, this time interval depends on the construct to be measured, as well as the standards of the researcher evaluating the article. Contrariwise, the strengths of our research are the big number of studies being assessed, the search performed in two of the most common languages used in research, and the utilization of the latest COSMIN manual.

In summary, the results of this study provide relevant information regarding five of the most commonly validated instruments to assess child maltreatment. Main results indicate that there is a big lack of information regarding psychometric properties in a large proportion of the studies examining these five scales, creating an important conflict regarding its generalizability. Choosing an instrument to assess child maltreatment which does not have proven psychometric properties can be an extremely risky practice for a professional, given the dramatic consequences to which the decisions based on it can lead to, and an unethical practice for a researcher. For this reason, we believe future research should, on the one hand, focus on filling these informational gaps in order to take informed decisions about the utilization of child maltreatment instruments, and on the other, we suggest the development of instruments that focus on early detection and prevention of child maltreatment. Additionally, developing interdisciplinary tools in which professionals from each field can add information (as doctors, social worker, psychologists, police) could constitute a huge advance in speed and reliability of child maltreatment assessment, as although psychometrically correct, child maltreatment assessment can greatly

benefit from this multidisciplinary approach (Noll et al., 2018; Schertz et al., 2018), and being able to access a child record through a program in real time can solve mandated reporting issues; and lastly, incorporating protective factors in child maltreatment assessment can constitute a chance to detect strong points of the child and/or the family and tailor future interventions (Sprague-Jones et al., 2020).

Critical Findings

1. None of the five assessed instruments was examined in all COSMIN psychometric properties
2. There was a rate of 60% of unknown evidence for structural validity, test-retest reliability, cross cultural validity and/or measurement invariance, and responsiveness.
3. The most examined psychometric property among all the scales was hypothesis testing for construct validity, followed by criterion validity analyses.
4. It is not possible to clearly determine if there is a consistently better instrument between the five examined, although the Maltreatment and Abuse Chronology of Exposure Scale (MACE) obtained the strongest evidence

Implications of the Review for Practice, Policy, and Research

1. This systematic review provided a comprehensive review on the main psychometric properties of five child maltreatment instruments in order to facilitate researchers and child welfare professionals the selection of the most suitable instrument for their specific purpose.
2. There is a noticeable lack of information about psychometric properties of child maltreatment instruments validations, as there is up to 100% unknown evidence for some of them.
3. The implications of these results in research can be uncertain conclusions in studies which utilize these instruments, especially the ones with the least available information.
4. These uncertain conclusions might lead to possibly wrong recommendations that once applied can suppose a waste of time and resources, interventions will be hindered or delayed, wrong decisions could be taken, sometimes leading to dramatic consequences in children and their families.
5. Future research should focus on filling these informational gaps in order to take informed decisions about the utilization of child maltreatment instruments, and we suggest the development of instruments that focus on early detection and prevention of child maltreatment.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Valencian Government, in the project "Development of an instrument for the detection and assessment of risk situations in childhood and adolescence." (REF. UVEG2019/20-20163CATED CC.28509)

Grants

The author Sylvia Georgieva is currently under the finaciation of Generalitat Valenciana; Conselleria d'Educació; European Social Fund, Grant/Award Number: ACIF/2021/175.

ORCID iD

José Javier Navarro Pérez  <https://orcid.org/0000-0001-6363-7154>

References

- *Aloba, O., Opakunle, T., & Ogunrinu, O. (2020). Childhood trauma questionnaire-short form (CTQ-SF): dimensionality, validity, reliability and gender invariance among Nigerian adolescents. *Child Abuse & Neglect*, 101 (2020), 104357. <https://doi.org/10.1016/j.chiabu.2020.104357>
- *Barrera, M., Hancock, K., Rokeach, A., Cataudella, D., Atenafu, E., Johnston, D., Punnett, A., Nathan, P.C., Bartels, U., Silva, M., Cassidy, M., Jansen, P., Shama, W., & Greenberg, C. (2014). External validity and reliability of the psychosocial assessment tool (PAT) among Canadian parents of children newly diagnosed with cancer. *Pediatric Blood & Cancer*, 61(1), 165–170. <https://doi.org/10.1002/pbc.24774>
- Bekhet, A. K., & Zauszniewski, J. A. (2012). Methodological triangulation: An approach to understanding data. *Nurse Researcher*, 20(2), 40–43. <https://doi.org/10.7748/nr2012.11.20.2.40.c9442>
- Bernstein, D. P., & Fink, L. (1998). *Childhood trauma questionnaire: A retrospective self-report manual*. S. Antonio (Ed.). The Psychological Corporation.
- Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano, M., Desmond, D., & Zule, W. (2003). Development and validation of a brief screening version of the childhood trauma questionnaire. *Child Maltreatment & Neglect*, 27(2), 169–190. [https://doi.org/10.1016/S0145-2134\(02\)00541-0](https://doi.org/10.1016/S0145-2134(02)00541-0)
- Bouwmeester-Landweer, M. B. R. (2006). *Early home visitation in families at risk for child Maltreatment* (Doctoral thesis). Leiden University Medical Center (LUMC), Leiden University, Leiden.
- CDC. (1997). *Adverse childhood experiences (ACE) study* | Child maltreatment violence prevention injury center CDC. <http://www.cdc.gov/violenceprevention/acestudy/>

- *Charak, R., de Jong, J. T. V. M., Berckmoes, L. H., Ndayisaba, H., & Reis, R. (2017). Assessing the factor structure of the childhood trauma questionnaire, and cumulative effect of abuse and neglect on mental health among adolescents in conflict-affected Burundi. *Child Abuse & Neglect*, 72 (2017), 383–392. <https://doi.org/10.1016/j.chiabu.2017.09.009>
- Cowley, L. E., Maguire, S., Farewell, D. M., Quinn-Scoggins, H. D., Flynn, M. O., & Kemp, A. M. (2018). Factors influencing child protection professionals' decision-making and multidisciplinary collaboration in suspected abusive head trauma cases: A qualitative study. *Child Abuse & Neglect*, 82(2018), 178–191. <https://doi.org/10.1016/j.chiabu.2018.06.009>
- *Dovran, A., Winje, D., Øverland, S. N., Breivik, K., Arefjord, K., Dalsbø, A. S., Jentoft, M. B., Hansen, A. L., & Waage, L. (2013). Psychometric properties of the Norwegian version of the Childhood Trauma Questionnaire in high-risk groups. *Scandinavian Journal of Psychology*, 54(4), 286–291. <https://doi.org/10.1111/sjop.12052>
- Evans, R., Garner, P., & Honig, A. S. (2014). Prevention of violence, maltreatment and neglect in early childhood: A review of the literature on research, policy and practice. *Early Child Development and Care*, 184(9–10), 1295–1335. <https://doi.org/10.1080/03004430.2014.910327>
- Felitti, V., Anda, M., Nordenberg, D., Williamson, D., Spitz, A., Edwards, V., Koss, M. P., & Marks, J. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The adverse childhood experiences (ACE) study. *American Journal of Preventative Medicine*, 14(4), 245–258. [https://doi.org/10.1016/S0749-3797\(98\)00017-8](https://doi.org/10.1016/S0749-3797(98)00017-8)
- *Forde, D. R., Baron, S. W., Scher, C. D., & Stein, M. B. (2012). Factor structure and reliability of the childhood trauma questionnaire and prevalence estimates of trauma for male and female street youth. *Journal of Interpersonal Violence*, 27(2), 364–379. <https://doi.org/10.1177/0886260511416461>
- *Fosse, R., Skjelstad, D. V., Schalinski, I., Thekkumthala, D., Elbert, T., Aanondsen, C. M., Greger, H. K., & Jozefiak, T. (2020). Measuring childhood maltreatment: Psychometric properties of the Norwegian version of the maltreatment and abuse chronology of exposure (MACE) scale. *Plos One*, 15(2), Article e0229661. <https://doi.org/10.1371/journal.pone.0229661>
- Georgieva, S., Tomás, J. M., & Navarro-Pérez, J. J. (2021). Systematic review and critical appraisal of childhood trauma questionnaire—Short form (CTQ-SF). *Child Abuse & Neglect*, 120(2021), 105223. <https://doi.org/10.1016/j.chiabu.2021.105223>
- Guastaferro, K., & Lutzker, J. R. (2021). The prevention of child maltreatment: using SafeCare® to highlight successes and needs for improvement in prevention efforts. *International Journal on Child Maltreatment: Research, Policy and Practice*, 5(1), 19–29. <https://doi.org/10.1007/s42448-021-00090-w>
- Hanson, J. L., Knodt, A. R., Brigidi, B. D., & Hariri, A. R. (2015). Lower structural integrity of the uncinate fasciculus is associated with a history of child maltreatment and future psychological vulnerability to stress. *Development and Psychopathology*, 27(4pt2), 1611–1619. <https://doi.org/10.1017/S0954579415000978>
- *He, J., Zhong, X., Gao, Y., Xiong, G., & Yao, S. (2019). Psychometric properties of the Chinese version of the childhood trauma questionnaire-short form (CTQ-SF) among undergraduates and depressive patients. *Child Abuse & Neglect*, 91 (2019), 102–108. <https://doi.org/10.1016/j.chiabu.2019.03.009>
- *Hernández, A., Gallardo-Pujol, D., Pereda, N., Arntz, A., Bernstein, D. P., Gaviria, A. M., Labad, A., Valero, J., & Gutiérrez-Zotes, J. A. (2013). Initial validation of the Spanish childhood trauma questionnaire-short form: Factor structure, reliability and association with parenting. *Journal of Interpersonal Violence*, 28(7), 1498–1518. <https://doi.org/10.1177/0886260512468240>
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons. <https://doi.org/10.1002/9781119536604>
- *Horrevorts, E. M., van Grieken, A., Mieloo, C. L., Hafkamp-de Groen, E., Bannink, R., Bouwmeester-Landweer, M. B., Broeren, S., & Raat, H. (2017). Concurrent validity, discriminatory power and feasibility of the instrument for Identification of Parents at Risk for child Abuse and Neglect (IPARAN). *BMJ open*, 7(8), Article e016140. <https://doi.org/10.1136/bmjopen-2017-016140>
- Jiménez-Borja, M., Jiménez-Borja, V., Borja-Alvarez, T., & Murgueitio, J. (2020). Prevalence of child maltreatment in Ecuador using the ICAST-R. *Child Abuse & Neglect*, 99(2020), 104230. <https://doi.org/10.1016/j.chiabu.2019.104230>
- *Karos, K., Niederstrasser, N., Abidi, L., Bernstein, D. P., & Bader, K. (2014). Factor structure, reliability, and known groups validity of the German version of the Childhood Trauma Questionnaire (Short-form) in Swiss patients and nonpatients. *Journal of Child Sexual Abuse*, 23(4), 418–430. <https://doi.org/10.1080/10538712.2014.896840>
- Kazak, A. E. (2006). Pediatric psychosocial preventative health model (PPPHM): Research, practice, and collaboration in pediatric family systems medicine. *Families, Systems and Health*, 24(4), 381–395. <https://doi.org/10.1037/1091-7527.24.4.381>
- Kisely, S., Abajobir, A. A., Mills, R., Strathearn, L., Clavarino, A., & Najman, J. M. (2018). Child maltreatment and mental health problems in adulthood: Birth cohort study. *The British Journal of Psychiatry*, 213(6), 698–703. <https://doi.org/10.1192/bjp.2018.207>
- Kluwe-Schiavon, B., Viola, T. W., & Grassi-Oliveira, R. (2016). Cross-cultural adaptation of the maltreatment and abuse chronology of exposure (MACE) scale to Brazilian Portuguese. *Trends in Psychiatry and Psychotherapy*, 38(1), 33–39. <https://doi.org/10.1590/2237-6089-2015-0051>
- *Kongerslev, M. T., Bach, B., Rossi, G., Trauelsen, A. M., Ladegaard, N., Løkkegaard, S. S., & Bo, S. (2019). Psychometric validation of the childhood trauma questionnaire-short form (CTQ-SF) in a Danish clinical sample. *Child Abuse & Neglect*, 94 (2019), 104026. <https://doi.org/10.1016/j.chiabu.2019.104026>

- *Kutsal, E., Pasli, F., Isikli, S., Sahin, F., Yilmaz, G., & Beyazova, U. (2011). Preliminary validation of the child abuse potential inventory in Turkey. *Journal of Interpersonal Violence*, 26(14), 2856–2865. <https://doi.org/10.1177/0886260510390952>
- *McCarthy, M. C., Hearps, S. J., Muscara, F., Anderson, V. A., Burke, K., Hearps, S. J., & Kazak, A. E. (2016). Family psychosocial risk screening in infants and older children in the acute pediatric hospital setting using the Psychosocial Assessment Tool. *Journal of Pediatric Psychology*, 41(7), 820–829. <https://doi.org/10.1093/jpepsy/jsw055>
- Milner, J. S. (1986). *The child abuse potential inventory manual* (2nd ed.). Psytec Inc.
- Milner, J. S., & Crouch, J. L. (2012). Psychometric characteristics of translated versions of the child abuse potential inventory. *Psychology of Violence*, 2(3), 239–259. <https://doi.org/10.1037/a0026957>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L.A., PRISMA-P Group (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1–9. <https://doi.org/10.1186/2046-4053-4-1>
- Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Newins, A. R., Wilson, L. C., & Kimbrel, N. A. (2021). Child maltreatment and suicidal ideation: The role of PTSD symptoms and alcohol misuse. *Current Psychology*, 40(10), 4930–4936. <https://doi.org/10.1007/s12144-019-00436-1>
- Noll, J.G., Susman, E.J., Carmody, K.A., & Reader, J.M. (2018) The policy and practice implications of child maltreatment research: The legacy of Penelope K. Trickett. In S. Negriff. (Eds.), *Child maltreatment research, policy, and practice*. Springer. https://doi.org/10.1007/978-3-030-04561-6_6
- Norman, R. E., Byambaa, M., De, R., Butchart, A., Scott, J., & Vos, T. (2012). The long-term health consequences of child physical maltreatment, emotional maltreatment, and neglect: A systematic review and meta-analysis. *Plos Medicine*, 9(11), Article e1001349. <https://doi.org/10.1371/journal.pmed.1001349>
- Pai, A. L., Patiño-Fernández, A. M., McSherry, M., Beele, D., Alderfer, M. A., Reilly, A. T., Wei-Ting Hwang, M. D., & Kazak, A. E. (2008). The psychosocial assessment Tool (PAT2.0): Psychometric properties of a screener for psychosocial distress in families of children newly diagnosed with cancer. *Journal of Pediatric Psychology*, 33(1), 50–62. <https://doi.org/10.1093/jpepsy/jsm053>
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. Chris Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.
- Petrowski, N., Cappa, C., Pereira, A., Mason, H., & Daban, R. A. (2021). Violence against children during COVID-19: Assessing and understanding change in use of helplines. *Child Abuse & Neglect*, 116(Part 2), 104757. <https://doi.org/10.1016/j.chiabu.2020.104757>
- Picornell-Lucas, A., & López-Peláez, A. (2022). The digital citizenship of children and adolescents: Challenges for social work education. *Research in Education and Learning Innovation Archives*, 28, 32–37. <https://doi.org/10.7203/28.23001>
- Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Raffaelli, M., Santana, J. P., De Moraes, N. A., Nieto, C. J., & Koller, S. H. (2018). Adverse childhood experiences and adjustment: A longitudinal study of street-involved youth in Brazil. *Child Abuse & Neglect*, 85(2018), 91–100. <https://doi.org/10.1016/j.chiabu.2018.07.032>
- Saini, S. M., Hoffmann, C. R., Pantelis, C., Everall, I. P., & Bousman, C. A. (2019). Systematic review and critical appraisal of child maltreatment measurement instruments. *Psychiatry Research*, 272(2019), 106–113. <https://doi.org/10.1016/j.psychres.2018.12.068>
- Schertz, M., Karni-Visel, Y., Hershkowitz, I., Flaisher, M., & Hershkowitz, F. (2018). Child maltreatment identification rates in a child development center: Suggestions for improved detection. *European Journal of Pediatrics*, 177(7), 1081–1087. <https://doi.org/10.1007/s00431-018-3163-5>
- *Schmidt, M. R., Narayan, A. J., Atzl, V. M., Rivera, L. M., & Lieberman, A. F. (2020). Childhood maltreatment on the adverse childhood experiences (ACEs) scale versus the childhood trauma questionnaire (CTQ) in a perinatal sample. *Journal of Aggression, Maltreatment & Trauma*, 29(1), 38–56. <https://doi.org/10.1080/10926771.2018.1524806>
- *Sint Nicolaas, S. M., Schepers, S. A., Hoogerbrugge, P. M., Caron, H. N., Kaspers, G. J., van den Heuvel-Eibrink, M. M., Grootenhuis, M.A., & Verhaak, C. M. (2016). Screening for psychosocial risk in Dutch families of a child with cancer: reliability, validity, and usability of the psychosocial assessment tool. *Journal of Pediatric Psychology*, 41(7), 810–819. <https://doi.org/10.1093/jpepsy/jsv119>
- *Spies, G., Kidd, M., & Seedat, S. (2019). A factor analytic study of the childhood trauma questionnaire-short form in an all-female South African sample with and without HIV infection. *Child Abuse & Neglect*, 92 (2019), 157–166. <https://doi.org/10.1016/j.chiabu.2019.04.002>
- Sprague-Jones, J., Singh, P., Rousseau, M., Counts, J., & Firman, C. (2020). The protective factors survey, 2nd edition: Establishing validity and reliability of a self-report measure of protective factors against child maltreatment. *Child & Youth Services Review*, 111(2020), 104868. <https://doi.org/10.1016/j.childyouth.2020.104868>
- Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R., & van IJzendoorn, M. H. (2015). The prevalence of child maltreatment across the globe: Review of a series of meta-analyses. *Child Abuse Review*, 24(1), 37–50. <https://doi.org/10.1002/car.2353>

- *Teicher, M. H., & Parigger, A. (2015). The 'maltreatment and maltreatment chronology of exposure'(MACE) scale for the retrospective assessment of maltreatment and neglect during development. *Plos One*, 10(2), Article e0117423. <https://doi.org/10.1371/journal.pone.0117423>
- Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, 27(5), 1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>
- Terwee, C. B., Prinsen, C. A. C., Ricci Garotti, M. G., Suman, A., de Vet, H. C. W., & Mokkink, L. B. (2016). The quality of systematic reviews of health-related outcome measurement instruments. *Quality of Life Research*, 25(4), 767–779. <https://doi.org/10.1007/s11136-015-1122-4>
- Tufford, L., Lee, B., Bogo, M., Wenghofer, E., Etherington, C., Thieu, V., & Zhao, R. (2021). Decision-making and relationship competence when reporting suspected physical abuse and child neglect: An objective structured clinical evaluation. *Clinical Social Work Journal*, 49(2), 256–270. <https://doi.org/10.1007/s10615-020-00785-6>
- Vachon, D. D., Krueger, R. F., Rogosch, F. A., & Cicchetti, D. (2015). Assessment of the harmful psychiatric and behavioral effects of different forms of child maltreatment. *JAMA Psychiatry*, 72(11), 1135–1142. <https://doi.org/10.1001/jamapsychiatry.2015.1792>
- van der Put, C. E., Assink, M., & van Solinge, N. F. B. (2017a). Predicting child maltreatment: A meta-analysis of the predictive validity of risk assessment instruments. *Child Abuse & Neglect*, 73 (2017), 71–88. <https://doi.org/10.1016/j.chiabu.2017.09.016>
- *van der Put, C. E., Bouwmeester-Landweer, M. B., Landsmeer-Beker, E. A., Wit, J. M., Dekker, F. W., Kousemaker, N. P. J., & Baartman, H. E. (2017b). Screening for potential child maltreatment in parents of a newborn baby: The predictive validity of an Instrument for early identification of parents at risk for child abuse and neglect (IPARAN). *Child Abuse & Neglect*, 70 (2017), 160–168. <https://doi.org/10.1016/j.chiabu.2017.05.016>
- Walker, C. A., & Davies, J. (2010). A critical review of the psychometric evidence base of the Child Abuse Potential Inventory. *Journal of Family Violence*, 25(2), 215–227. <https://doi.org/10.1007/s10896-009-9285-9>
- *Wells, B. M., Crouch, J. L., Schubert, R., Irwin, L. M., Risser, H. J., & Skowronski, J. J. (2011). Revisiting the issue of the child abuse potential inventory's internal consistency in adolescent samples. *Journal of Adolescent Health*, 48(4), 351–357. <https://doi.org/10.1016/j.jadohealth.2010.07.023>
- WHO. (2020). *Child maltreatment*. [online]. Available at: <https://www.who.int/news-room/fact-sheets/detail/child-maltreatment> (Accessed 30 March 2021).
- Wilkins, D., & Whittaker, C. (2018). Doing child-protection social work with parents: What are the barriers in practice? *The British Journal of Social Work*, 48(7), 2003–2019. <https://doi.org/10.1093/bjsw/bcx139>
- Yoon, S., Speyer, R., Cordier, R., Aunio, P., & Hakkarainen, A. (2020). A systematic review evaluating psychometric properties of parent or caregiver report instruments on child maltreatment: Part 2: Internal consistency, reliability, measurement error, structural validity, hypothesis testing, cross-cultural validity, and criterion validity. *Trauma, Violence, & Abuse*, 22(5), 1296–1315. <https://doi.org/10.1177/1524838020915591>
- Zhang, H., Han, T., Ma, S., Qu, G., Zhao, T., Ding, X., Sun, L., Qin, Q., Chen, M., & Sun, Y. (2022). Association of child maltreatment and bullying victimization among Chinese adolescents: The mediating role of family function, resilience, and anxiety. *Journal of Affective Disorders*, 299(2022), 12–21. <https://doi.org/10.1016/j.jad.2021.11.053>

Author Biographies

Sylvia Georgieva is a junior researcher currently engaged in a PhD program in the University of Valencia financed with a grant from the Valencian Government. Her research is focused on applied structural equation models, psychometrics, educational quality, and child maltreatment and protection measurement.

José Manuel Tomás, PhD, is a Professor of Statistics in the Faculty of Psychology at University of Valencia. His research focuses on applied structural equation models and measurement issues. He is interested in applying complex statistical models to different substantive areas such as children maltreatment and neglect, educational quality and successful aging.

José Javier Navarro-Pérez, PhD, is an Associate Professor in the Faculty of Social Work and Social Services at University of Valencia. His research focuses on child maltreatment and children at risk, gender studies, juvenile offenders, juvenile unemployment and COVID, Latin gangs, and public peace-keeping policies.

Paula Samper, PhD, is an Associate Professor in the Faculty of Psychology at University of Valencia. Her research focuses on adolescents and aggressive and prosocial behavior. She is interested in analyze variables which are related with these two kinds of behavior: family, emotions, peers, school, and society.