



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Child Abuse & Neglect

journal homepage: www.elsevier.com/locate/chiabuneg

Systematic review and critical appraisal of Childhood Trauma Questionnaire — Short Form (CTQ-SF)

Sylvia Georgieva^a, Jose M. Tomas^a, José Javier Navarro-Pérez^{b,*}

^a Faculty of Psychology, University of Valencia, Av. Blasco Ibáñez, 21, Valencia 46010, Spain

^b Faculty of Social Work, University of Valencia, Av. Dels Tarongers, 4b, Valencia 46022, Spain

ARTICLE INFO

Keywords:

CTQ-SF
Childhood maltreatment
Assessment
Strength of evidence
Psychometric properties

ABSTRACT

Background: Child maltreatment is a complex and multidimensional construct that encompasses a great number of risk factors. The Childhood Trauma Questionnaire — Short Form, one of the most widely used and validated instruments to assess childhood maltreatment in the past ten years, is a retrospective instrument that assesses several types of childhood abuse and maltreatment which is divided into five dimensions.

Objective: The objectives of this systematic review are to critically appraise, compare, and summarize the methodological quality and psychometric properties of published research articles validating the Childhood Trauma Questionnaire — Short Form utilizing the COSMIN checklist.

Method: Articles published in English or Spanish, in the past ten years in the databases of Scopus, Web of Science, and ProQuest, and which, directly or indirectly analyzed psychometric properties of the CTQ-SF were screened, examined, and assessed utilizing the COSMIN checklist.

Results: Main results indicate that there is a general pattern of assessing the same three psychometric properties (internal consistency, structural validity, and hypothesis testing) in a variety of samples, but leaving unassessed the rest of properties examined by the COSMIN checklist. Additionally, there are some problems with the internal consistency of several factors.

Implications and conclusions: While replicability and internal consistency are good psychometric indicators of the CTQ-SF, there is a big scientific gap of information regarding some psychometric properties. It is suggested that future research should address the remaining psychometric properties, reliability, measurement error, content validity, cross cultural and criterion validity, as well as re-examining internal consistency of some dimensions, in order to advance in the knowledge on childhood maltreatment assessment.

1. Introduction

According to the World Health Organization (WHO, 2020), child maltreatment is conceptualized as any type of maltreatment, neglect, physical or emotional mistreatment, sexual abuse, negligence, exposure to violent scenes, and/or any commercial or other exploitation performed to an underaged child or teenager, which results in an actual or potential harm to the child or adolescent's life, health, development, or dignity (WHO, 2020). In the United Nations Convention of the Rights of the Child (UNCRC) of 1989, children's civil, political, economic, and social rights were stated, thus, they became subjects of social rights, as well as the rights to be protected

* Corresponding author.

E-mail addresses: sylvia.georgieva@uv.es (S. Georgieva), tomasjm@uv.es (J.M. Tomas), j.javier.navarro@uv.es (J.J. Navarro-Pérez).

<https://doi.org/10.1016/j.chiabu.2021.105223>

Received 11 February 2021; Received in revised form 3 June 2021; Accepted 22 July 2021

Available online 2 August 2021

0145-2134/© 2021 Elsevier Ltd. All rights reserved.

and provided with what they need to fulfil their developmental needs (Ramiro, 2016). It is a duty of all nations integrated in the UNCRG to promote and preserve the positive development of children and adolescents as well as their wellbeing.

Nowadays, levels of children and adolescent's maltreatment have been internationally reported as having reached epidemic levels (Evans, Garner, & Honig, 2014). It is estimated that there are 41,000 homicide deaths of children under 15 years of age every year, 25% of adults confirm having been physically maltreated as children (WHO, 2020). These data highlight the seriousness of this global problem.

Children and adolescents are more subject to risks due to the inherent characteristics of that period in life (Fineman, 2017). Thorough history, children have suffered acts of violence by parents, care providers, and others. This fact places them in a vulnerable collective, with higher risk of developing psychosocial problems, if not properly nurtured and protected. Risk can have its origin in their social and family conditions, emotional deprivation, neglect, or maltreatment (Evans, Garner, & Honig, 2014).

Child maltreatment is a complex and multidimensional construct that encompasses a great number of risk factors. This fact has hindered its study, due to the differences in definition, factors, and types of reports which include it. There is a number of risk factors associated with the probability of child maltreatment. The World Health Organization (WHO, 2020) divides these factors in: child characteristics, caregivers or parents' relationships in the family network, and community and social factors. Caregivers' characteristics are prevalent in scientific literature, and they have often been associated to child and adolescent maltreatment (Choi, 2012; Laslett, Ferris, Dietze, & Room, 2012; McMillin et al., 2016). However, children and adolescents' characteristics, socioeconomic factors and family network factors are also relevant predictors of child and adolescent's maltreatment (Belsky, Bell, Bradley, Stallard, & Stewart-Brown, 2007; Evans, Garner, & Honig, 2014; WHO, 2020).

Child maltreatment causes long term suffering to children and families, and leads to dramatic consequences in the development and later functioning of children (WHO, 2020). Being a victim of maltreatment as a child or adolescent increases the probability of depression, anxiety and other externalizing and internalizing behaviors in further years (Hanson, Knodt, Brigidi, & Hariri, 2015; Kisely et al., 2020). A systematic review and meta-analysis performed by Norman, De Byambaa, Butchart, Scott, and Vos (2012) concluded that childhood maltreatment is associated to a range of mental disorders, drug abuse, suicide attempts, sexually transmitted infections, and risky sexual behavior. Additionally, the relation between child maltreatment and post-traumatic stress disorder (PTSD), is well known, as stated in the systematic review of Messman-Moore and Bhuptani (2017) which indicated that all forms of child maltreatment are associated with increased risk of PTSD in adulthood, especially when it comes to sexual or emotional abuse; or in the study of Hsieh et al. (2020) who explored the mediational effect of PTSD between child maltreatment and dysfunctional family environment, and children's bullying perpetration.

1.1. Detection and measurement of child maltreatment

The detection and assessment of child maltreatment is paramount for the decision-making process in child welfare, to define the magnitude of the problem (prevalence) and to track changes over time, as well as to assess the effectiveness of intervention and prevention programs, laws, and/or policies. Despite the relevance of assessing risk in children and adolescents, there is a generalized lack of comprehensive instruments to assess risk of maltreatment as a whole. Saini, Hoffmann, Pantelis, Everall, and Bousman (2019) performed a systematic review of instruments measuring child maltreatment by assessing 52 different instruments examining the following criteria: internal consistency, measurement error, content validity, structural validity, hypothesis testing, cross-cultural validity, criterion validity, responsiveness, and interpretability. The study concluded that only 8 of the 52 studies had strong to moderate evidence for at least 3 of the criteria of methodological quality assessed. They concluded that no instrument had acceptable levels for all criteria, and no criterion was met by all evaluated instruments. These results imply that it can't be concluded that there is one superior instrument regarding psychometric properties. However, the three instruments that appeared to have the strongest psychometric evidence were The Childhood Trauma Questionnaire (CTQ; Bernstein & Fink, 1998), the short form of CTQ (CTQ-SF; Bernstein et al., 2003), and the Maltreatment and Abuse Chronology of Exposure (MACE; Teicher & Parigger, 2015).

The Childhood Trauma Questionnaire (CTQ; Bernstein & Fink, 1998) is a retrospective instrument that assesses several types of childhood abuse and maltreatment. It is composed of 70 items divided into five dimensions: 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). Five years after its release, Bernstein et al. (2003) developed a short form of the CTQ after computing exploratory and confirmatory factor analyses. They first performed an Exploratory Factor Analysis estimated with maximum likelihood and with direct quartimin rotation, in which the five best items for each previously defined factor were retained, in other words, the five items that loaded the highest on their factor (loadings >0.50) and the lowest on the other factors, ensuring that each of the five factors of the scale would be distinguishable. Secondly, they tested the new composition of items via Confirmatory Factor Analysis in four different samples. A final version of 25 items was obtained, plus 3 items of minimization and denial, with 5 items for each original dimension of the scale. This scale was also tested for measurement invariance across four groups (adult substance abusing patients, adolescent psychiatric inpatients, adult substance abusers in the community, and a normative community sample of adults) and for criterion validity with good psychometric results (Bernstein et al., 2003). Since then, the CTQ-SF (Bernstein et al., 2003) has been widely validated across a variety of samples: adolescents, substance users, psychiatric inpatients, (Bernstein et al., 2003), street youth (Forde, Baron, Scher, & Stein, 2012), prisoners (Dovran et al., 2013), juvenile delinquents (Kongerslev et al., 2019), and pregnant women (Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020a).

Additionally, 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 considered types of maltreatment are the following: 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 towards siblings. This scale was developed using Item Response Theory (IRT). It is composed by 52 items divided into ten Rasch's scales measured by binary answers (yes/no). Authors of this scale have examined convergent validity with the Childhood Trauma Questionnaire (CTQ; Bernstein et al., 2003) and Adverse Childhood Experiences Questionnaire (ACE) (Felitti et al., 1998). They also proved predictive validity with adult psychopathologies, and test-retest within a 6 months period (Teicher & Parigger, 2015). This scale has also been validated in a Brazilian-Portuguese sample (Kluwe-Schiavon, Viola, & Grassi-Oliveira, 2016), as well as in Norwegian (Fosse et al., 2020).

There are several other frequently used scales in the scientific literature as the Adverse Childhood Experiences Questionnaire (ACE) (CDC, 1997; Felitti et al., 1998) which is an instrument to assess child abuse and exposure to violence, composed by 10 items: five of them are related to child maltreatment while the other five are related to parental incapacities (Felitti et al., 1998). Additionally, the Child Abuse Potential Inventory (CAP; Milner, 1986) is another instrument designed to assess caregiver's likelihood of mistreating or abusing a child. It is a self-report measure composed by 77 items divided into six maltreatment factors, three validity scales, and responses distortion indexes. Another scale is the Identification of Parents At Risk for child Abuse and Neglect (IPRAN; Bouwmeester-Landweer, 2006), which is a 37-item, three-pages self-report instrument developed to identify parents of newborns who are at risk of experiencing parenting stress which can later develop in potential abuse or neglect. It is composed by three main factors: child and family characteristics, parental developmental history and personality, and characteristics of the social context.

1.2. The Childhood Trauma Questionnaire — Short Form (CTQ-SF)

This study will focus specifically on the short form of Childhood Trauma Questionnaire (CTQ-SF; Bernstein et al., 2003). There are several reasons to select this specific instrument given the above-mentioned variety of tools to measure child maltreatment: The CTQ-SF is one of the instruments which has undergone most examinations of validity in the past 10 years of research, and it is one of the most widely used instruments to assess childhood maltreatment in the world (Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020a). Second, it is one of the easiest instruments to administer since it is composed by only 28 items administered in a Likert-type scale, which contributes to the usability of this tool in research. Third, it is the most recent and author-recommended version of the Childhood Trauma Questionnaire, and it has become a leading measure in assessment of childhood maltreatment nowadays (Baker & Maiorino, 2010). And fourth, Saini, Hoffmann, Pantelis, Everall, and Bousman (2019) concluded that the CTQ-SF is one of the three instruments with strongest psychometric properties. Because of these reasons, it is crucial to examine and summarize the main psychometric information regarding one of the most used instruments to assess child maltreatment.

However, no study has conducted a systematic evaluation of the methodological quality, quality of the instrument, and measurement properties of the studies validating the CTQ-SF (Bernstein et al., 2003). Taking into account that it is one of the most widely used and validated instruments to assess childhood maltreatment (Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020b), it is key to examine if the studies that examine this instrument possess indeed the methodological quality and strength of evidence to support the psychometric properties of the CTQ-SF, and examine if this instrument has good psychometric characteristics according to the updated criteria for good measurement properties shown in the COSMIN manual (Mokkink et al., 2018).

1.3. The current study

The aims of this systematic review were to critically appraise, compare, and summarize the methodological quality and psychometric properties of published research articles validating the Childhood Trauma Questionnaire — Short Form (Bernstein et al., 2003) utilizing the COnsensus based Standards for the selection of health status Measurement INstruments (COSMIN) for systematic reviews of Patient-Reported Outcome Measures (PROMs) (Mokkink et al., 2018). To fulfil this aim, a number of articles which examine the validity of the CTQ-SF (Bernstein et al., 2003) were collected considering the following information: authors, sample characteristics, country, design quality of the studies, and measurement properties indicated in the studies.

2. Method

2.1. Eligibility criteria

Articles which directly or indirectly analyzed psychometric properties of the CTQ-SF (Bernstein et al., 2003) were included. 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 2010 to 2020, and 3) Articles published in scientific journals indexed in the Journal Citation Reports (JCR) or Scopus, assessed by Scimago Journal and Country Rank. Given the specific nature of the CTQ-SF as a retrospective measure, studies utilizing adult samples were not excluded, since the questionnaire is still assessing events regarding childhood and adolescence. Additionally, the search was performed with that specific time margin because the CTQ-SF has been a widely explored instrument, however, psychometric properties might change across generations, and we wanted to explore the behavior of the scale during the last years of research.

Exclusion criteria were: 1) Articles considered gray literature (doctoral thesis, conference papers, press articles, etc.); 2) Articles presenting results about the CTQ but without psychometric information; and 3) Systematic reviews of articles.

2.2. Search strategy

A systematic search of the electronic databases of ProQuest, Scopus, and Web of Science was performed. All articles in English and Spanish were included. Four groups of keywords were identified: Children or adolescents, risk, instruments or questionnaires, and Childhood Trauma Questionnaire — Short Form (CTQ-SF). The first group focused on the collective of interest: children and adolescents, since although the CTQ-SF is a retrospective measure, meaning that it can be filled up by adults, it still assesses maltreatment in childhood. The second group had its focus on risk of maltreatment, the third group focused on the search for a tool or instrument assessing risk, while the fourth group, specified the examined instrument: Childhood Trauma Questionnaire — Short Form. Keywords are shown in [Table 1](#). The search was performed during the months of July and August in 2020.

2.3. Data collection process

Two independent judges screened the titles and abstracts identified with the aforementioned search strategy and assessed the selected full texts (flowchart is shown in [Fig. 1](#)). In case of disagreement, it was discussed with a third judge, as recommended by [Bekhet and Zauszniewski \(2012\)](#). There were disagreements in two studies. Conflicts were solved by a full read of the article by the three judges and a debate until reaching an agreement regarding the inclusion of the article.

Information extracted among eligible studies was the following: Names of authors of the studies, characteristics of the samples in which the instrument was applied, language of administration, and country, methodological quality of the studies, and quality of the PROM (patient-reported outcome measure).

2.4. Study selection and data extraction

This systematic review followed the ‘Preferred Reporting Items for Systematic Review and Meta-analysis Protocols’ (PRISMA-P; [Moher et al., 2015](#)). The results of the searches were transferred into an electronic reference manager (Mendeley Desktop, version 1.17.9), which eliminated duplicate entries. A total of 153 studies were identified from which 10 were selected for this review (see [Fig. 1](#)).

2.5. The COSMIN checklist for systematic reviews of PROMs

Appraisal of the methodological quality of studies was based on the COSMIN checklist for systematic reviews of PROMs ([Mokkink et al., 2018](#)). This instrument aims to assess the methodological quality of measurement instruments. The COSMIN guide distinguishes between “standards” and “criteria”, standards being considered to refer to the research design requirements of every study, meaning that it mostly indicates the quality of the study itself, while “criteria” refers to what constitutes a good measurement property (the quality of the PROM) ([Mokkink et al., 2018](#)). The COSMIN manual provides a comprehensive guideline divided into three steps and 10 boxes. The first part refers to standard procedures performing systematic reviews in agreement with existing guidelines as Cochrane ([Higgins et al., 2019](#)) and PRISMA ([Moher et al., 2015](#)), the second part concerns the evaluation of the measurement properties of them PROM itself, and the third part consists of evaluation of the interpretability and feasibility of the PROM, as well as formulating and reporting the systematic review. Authors suggest to use the COSMIN checklist as a modular tool, in the way that only the boxes of interest might be completed.

Evaluation was performed in 3 steps: first, the methodological quality of the included studies was evaluated (the above mentioned “standards”); second, criteria for good measurement properties were applied (the above mentioned “criteria”); third, evidence utilizing the modified GRADE (Grades of Recommendation, Assessment, Development and Evaluation) approach was summarized.

2.6. Assessment of methodological quality, quality of the PROM and strength of evidence

Methodological quality was assessed in three main groups: content validity, internal structure, and remaining measurement properties. Content validity is composed by box 1 (PROM development) and box 2 (Content validity), authors argument that although

Table 1
Terms used for the Boolean search.

Children or adolescent		Risk of maltreatment		Assessment		CTQ-SF
child OR	AND	*risk* OR	AND	*assessment* OR	AND	*Childhood Trauma Questionnaire*
children OR		*adversity* OR		*tool* OR		OR
childhood OR		*maltreatment* OR		*instrument* OR		* Childhood Trauma Questionnaire — Short Form*
adolescence OR		*trauma* OR		*scale* OR		OR
teenager		*victimization* OR		*questionnaire* OR		*CTQ-SF*
OR		*neglect*		*battery* OR		
early life OR				*inventory* OR		
				structured OR		
				recognition OR		
				index		

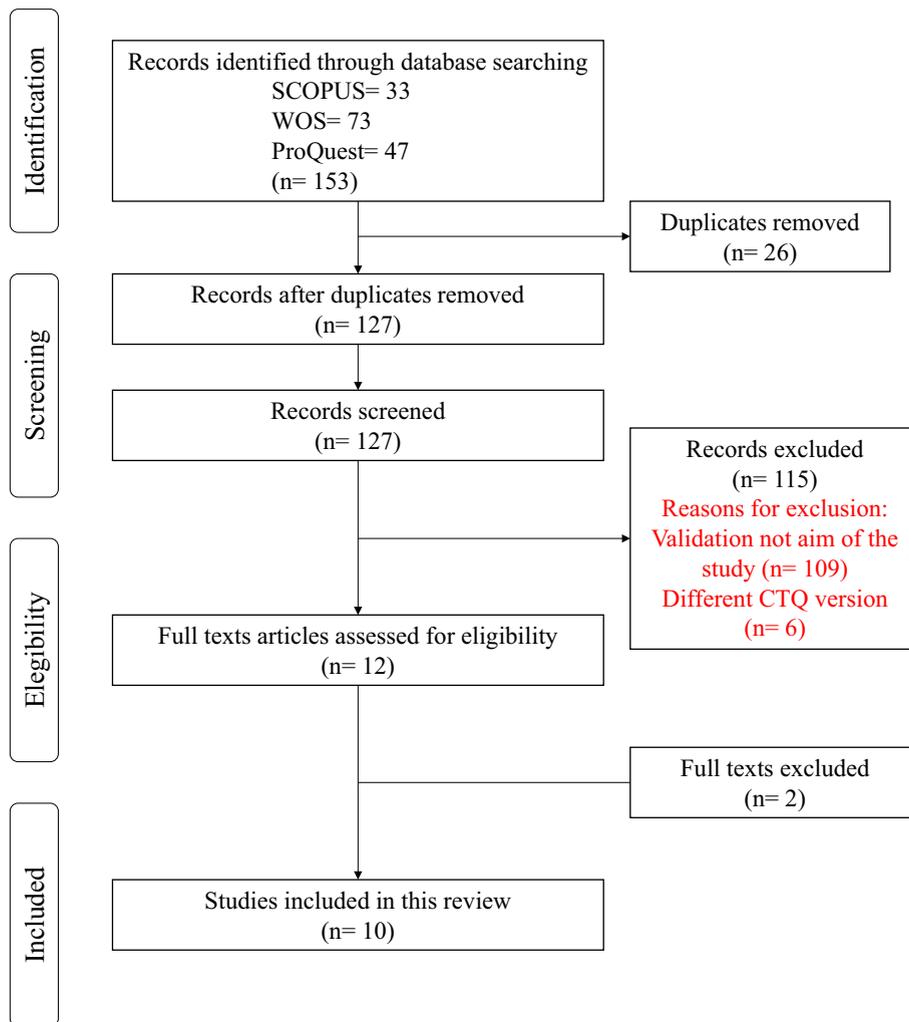


Fig. 1. Flowchart of study selection

development is not a measurement property, must be taken into account when evaluating content validity. Internal structure is composed by structural validity (box 3), internal consistency (box 4), and cross-cultural validity/measurement invariance (box 5). Finally, remaining measurement properties are reliability (box 6), measurement error (box 7), criterion validity (box 8), hypothesis testing for construct validity (box 9), and responsiveness (box 10). All items in the checklist were scored on a 4-point scale (excellent, good, fair and poor), the general score of each box was determined by the lowest score (1–4) on an item in the box (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2012). In this specific study, boxes 1 and 2 regarding content validity were not assessed since the instrument of interest has already been developed in 2003 (Bernstein et al., 2003) and assessed studies contained no information regarding its development.

Quality of the PROM was assessed with the updated criteria for good measurement properties shown in the COSMIN manual (Mokkink et al., 2018) and based on Prinsen et al. (2016) and Terwee et al. (2007) indications. These criteria were evaluated in a three-point scale: sufficient (+), insufficient (–), and indeterminate (?).

Strength of evidence assessment was performed based on the modified GRADE (Grades of Recommendation, Assessment, Development and Evaluation) approach, which determines an overall assessment of each property ranging from strong evidence, to moderate, limited, conflicting, and unknown; based on the methodological quality and consistency of results for each study (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.

3. Results

A total of 153 research articles were identified through the search of three databases (SCOPUS, Web of Science, and ProQuest), 26 were duplicated and therefore removed, leaving 127 articles to be assessed based on titles and abstracts. 115 articles were excluded since they did not meet the aforementioned criteria, leaving 12 articles to be assessed by examining their full text, from which 2 articles were excluded: 1 because it was a systematic review and another because it gave no psychometric information of the instrument. Systematic reviews were excluded since the COSMIN checklist is not designed to assess data from this type of articles.

3.1. Characteristics of the studies

All validations used Classical Test Theory. Sample sizes ranged from 77 to 3648 participants. There were 5 studies performed with adolescent or young samples, while 5 were performed with adults. Samples are described in detail in Table 2. However, there were two samples with only female participants (Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020a; Spies, Kidd, & Seedat, 2019). Ages of participants ranged from 14.9 years old (Charak, de Jong, Berckmoes, Ndayisaba, & Reis, 2017) to 41.64 years old (Karos, Niederstrasser, Abidi, Bernstein, & Bader, 2014). Most of the participants, however, were around 20 years old, this was mainly due to the fact that the CTQ-SF is a retrospective tool. Regarding countries, studies were performed in the following countries: Canada (Forde, Baron, Scher, & Stein, 2012), Norway (Dovran et al., 2013), Spain (Hernández et al., 2013), Switzerland (Karos, Niederstrasser, Abidi, Bernstein, & Bader, 2014), Burundi (Charak, de Jong, Berckmoes, Ndayisaba, & Reis, 2017), China (He, Zhong, Gao, Xiong, & Yao, 2019), Denmark (Kongerslev et al., 2019), South Africa (Spies, Kidd, & Seedat, 2019), Nigeria (Aloba, Opakunle, & Ogunrinu, 2020), and the United States (Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020a). From the 10 studies, the CTQ-SF was administered in English language in 2 of them, in one study, it was administered in both, English and Afrikaans, in another study, it was administered in Spanish. The rest of languages of administration were Norwegian, German, French, Chinese, and Danish. Detailed information about samples is presented in Table 2.

3.2. Methodological and measurement quality of the studies

Methodological quality of eligible studies is summarized in Table 3. None of the articles presented information regarding all eight COSMIN criteria. Particularly, no study assessed test-retest reliability, measurement error, content validity, or responsiveness, and only one study assessed criterion validity. Most of the studies focused on internal consistency (80%) and structural validity (70%). Half of the included studies performed different forms of hypothesis testing. However, many of them did not clearly state a priori hypotheses. Three studies tested for measurement invariance, from which only one was considered a good measurement invariance routine (He, Zhong, Gao, Xiong, & Yao, 2019). Finally, only one study presented information regarding criterion validity (Hernández et al., 2013).

3.3. Criteria of measurement — quality of the CTQ-SF

Quality of the CTQ-SF was examined according to the updated criteria for good measurement properties shown in the COSMIN manual (Mokkink et al., 2018) (see Table 4).

Examining structural validity, we could appreciate that eight of the ten studies performed structural validity analyses, they were all

Table 2
Samples of studies validating the CTQ-SF.

Article	N	Characteristics
Forde, Baron, Scher, and Stein (2012)	397	Street youth in Vancouver, British Columbia. Mean age 20 years old (SD = 2.6), 34% female and 66% male.
Dovran et al. (2013)	517	In Norway. Substance abusers (n = 126) mean age 30.9 years (SD = 11.6), 60% males; psychiatric patients (n = 210), mean age 25.5 (SD = 10.5), 30% males; prisoners (n = 109), mean age 32.2 years (SD = 9.9), 100% males; and adolescents in out-of-home placements (n = 72), mean age 19.1 years (SD = 3.5), 44% males.
Hernández et al. (2013)	185	Spanish female sample from mental health centers. Mean age 41.64 years (SD = 11.83).
Karos, Niederstrasser, Abidi, Bernstein, and Bader (2014)	661	Swiss sample of clinical (n = 565, mean age 36.64 years, SD = 11.79) and non-clinical (n = 96, mean age 25.7, SD = 5.43) participants.
Charak, de Jong, Berckmoes, Ndayisaba, and Reis (2017)	321	Adolescents and youth from Burundi, mean age 14.9 years, SD = 1.99, 58.4% females.
He, Zhong, Gao, Xiong, and Yao (2019)	3648	Chinese sample of 3431 undergraduates (mean age 19.6 years, SD = 1.56, 56.6% males) and 217 depressive patients (mean age 22.6 years, SD = 5.7, 54.4% males)
Kongerslev et al. (2019)	393	Danish clinical sample. Patients enrolled as part of routine assessment (n = 142, mean age 29 years, SD = 8.4, 68% female); patients commencing treatment for psychosis (n = 101, mean age 23 years, SD = 3.4, 26%); patients diagnosed with depression (n = 71, mean age 35 years, SD = 11.6, 78% women); and juvenile delinquent boys (n = 80, mean age 17 years, SD = 0.8)
Spies, Kidd, and Seedat (2019)	314	Female sample in Cape Town. 170 HIV uninfected, 144 HIV infected. Mean age 30.7 years, SD = 8.
Aloba, Opakunle, and Ogunrinu (2020)	1337	High school adolescents in Nigeria. Mean age 15.24 years, SD = 1.43. 45.3% males and 54.7% females.
Schmidt et al. (2020)	77	Ethnically diverse pregnant women in San Francisco (US), mean age 29.19 years, SD = 6.55. 34% Latina, 24% White, 22% African American, 14% Biracial/Multiracial, 5% Asian/ Pacific Islander, 1% Native American, 18% Monolingual Spanish-speaking, and 26% foreign born.

Table 3
COSMIN results of the standards of measurement (methodological quality) of the articles assessed.

Psychometric property	Articles	Psychometric property	Articles
<i>Structural validity</i>		<i>Measurement error</i>	
Excellent	1; 3; 5; 6; 7; 8; 9	Excellent	
Good		Good	
Fair	4	Fair	
Poor	2	Poor	
Unknown/NA	10	Unknown/NA	1; 2; 3; 4; 5; 6; 7; 8; 9; 10
<i>Internal consistency</i>		<i>Criterion validity</i>	
Excellent	1; 2; 3; 4; 5; 6; 8; 9	Excellent	3;
Good		Good	
Fair		Fair	
Poor	7	Poor	
Unknown/NA	10	Unknown/NA	1; 2; 4; 5; 6; 7; 9; 10
<i>Cross-cultural validity/measurement invariance</i>		<i>Hypothesis testing for construct validity</i>	
Excellent	6	Excellent	3; 4; 5; 6; 7
Good		Good	
Fair	1; 9	Fair	9
Poor		Poor	10
Unknown/NA	2; 3; 4; 5; 7; 8; 10	Unknown/NA	1; 2; 8
<i>Reliability</i>		<i>Responsiveness</i>	
Excellent		Excellent	4, 5, 6, 7
Good		Good	3, 8
Fair		Fair	1, 9
Poor		Poor	10
Unknown/NA	1; 2; 3; 4; 5; 6; 7; 8; 9; 10	Unknown/NA	2

Notes: 1: Forde, Baron, Scher, and Stein (2012); 2: Dovran et al. (2013); 3: Hernández et al. (2013); 4: Karos, Niederstrasser, Abidi, Bernstein, and Bader (2014); 5: Charak, de Jong, Berckmoes, Ndayisaba, and Reis (2017); 6: He, Zhong, Gao, Xiong, and Yao (2019); 7: Kongerslev et al. (2019); 8: Spies, Kidd, and Seedat (2019); 9: Aloba, Opakunle, and Ogunrinu (2020); 10: Schmidt et al. (2020). NA: not applicable.

Confirmatory Factor Analyses with satisfactory results ($CFI > 0.95$ or $RMSEA < 0.06$ or $SRMR < 0.08$). All the studies that performed CFA explored the classic five-factor structure of the CTQ-SF, although some added some modifications, as Charak, de Jong, Berckmoes, Ndayisaba, and Reis (2017) who eliminated item 24 from the sexual abuse factor. Aloba, Opakunle, and Ogunrinu (2020), on the other hand, correlated the residual error of items 11 and 12 (from physical abuse factor), and items 7 and 13 (from emotional neglect). Kongerslev et al. (2019), who tested eight structural models, concluded again, that the 5-factor structure had the best fit, however they also correlated three residuals.

Regarding internal consistency, eight of the studies estimated Cronbach's alpha as a standard, one study utilized McDonald's omega (Aloba, Opakunle, & Ogunrinu, 2020), and one study did not report any internal consistency information (Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020a). From studies which examined internal consistency, there were some consistent problems with some factors of the scale: seven studies had an alpha or omega below 0.7 for the physical neglect factor, one study had an alpha below 0.7 for emotional neglect, two studies had problems with the emotional abuse factor, and two others with the alpha of physical abuse.

Test-retest reliability assessed with intraclass correlation coefficients or weighted kappa was not reported in any of the studies. The same occurred with measurement error assessed with SDC (smallest detectable change), LoA (limits of agreement), and MIC (minimal important change).

Hypothesis testing for construct validity was performed in four of the studies. No study examined cross-cultural validity. However, four studies examined measurement invariance across different groups with acceptable results.

Finally, there was no information regarding criterion validity, as stated by the COSMIN updated criteria for good measurement properties, since there is no agreed gold standard to compare the instrument with. There is no information regarding responsiveness as stated by these criteria, as well, since no study reported AUC (Area Under the Curve) information, and there were not enough studies testing hypothesis in order to calculate the percentage of agreement, as stated by the COSMIN manual.

3.4. Strength of evidence

The strength of evidence for each article is specified in Table 5. From the percentages indicated in Table 5, it can be determined that none of the studies provided information for all nine criteria assessed by the COSMIN checklist. According to the modified GRADE criteria (Prinsen et al., 2018), it can be concluded that the CTQ-SF instrument had moderate evidence with 44.4% of the COSMIN criteria considered as moderate to strong.

4. Implications and conclusions

The main aim of this systematic review was to critically appraise, compare, and summarize the methodological quality and psychometric properties of published research articles which examined the validity of the Childhood Trauma Questionnaire — Short Form (Bernstein et al., 2003) between 2010 and 2020. The CTQ-SF has been one of the most employed instruments to assess childhood

Table 4
COSMIN results of the criteria of measurement (quality of the PROM) for the CTQ-SF in the articles assessed.

	Structural validity	Internal consistency	Reliability	Measurement error	Hypothesis testing for construct validity	Cross-cultural validity/invariance	Criterion validity	Responsiveness
Forde et al. (2012)	+	- ¹	?	?	+	+	?	?
Dovran et al. (2013)	+	+	?	?	?	+	?	?
Hernández et al. (2013)	+	- ²	?	?	+	?	?	?
Karos et al. (2014)	+	- ²	?	?	?	?	?	?
Charak et al. (2017)	?	- ²	?	?	+	?	?	?
He et al. (2019)	+	- ^{2,3}	?	?	?	+	?	?
Kongerslev et al. (2019)	+	- ^{2,4}	?	?	?	?	?	?
Spies et al. (2019)	+	- ^{2,3}	?	?	?	?	?	?
Aloba et al. (2020)	+	- ^{2,3,4*}	?	?	?	+	?	?
Schmidt et al. (2020)	?	?	?	?	+	?	?	?
Sufficient	80%	10%	0%	0%	40%	40%	0%	0%
Insufficient	0%	80%	0%	0%	0%	0%	0%	0%
Indeterminate	20%	10%	100%	100%	60%	60%	100%	100%

+	Sufficient
-	Insufficient
?	Indeterminate

¹Emotional neglect $\alpha < 0.7$.

²Physical neglect $\alpha < 0.7$.

³Emotional abuse $\alpha < 0.7$.

⁴Physical abuse $\alpha < 0.7$.

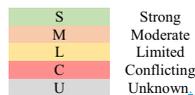
*McDonald's omega calculated.

maltreatment in recent research (Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020a; Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020b).

The results of this study were partially congruent with the results from Saini, Hoffmann, Pantelis, Everall, and Bousman (2019), who examined the methodological quality and measurement properties of published child abuse measures, in the way that both results indicate that there is a big scientific gap of information regarding some specific psychometric properties such as reliability,

Table 5
Strength of evidence of each study.

	Structural validity	Internal consistency	Cross-cultural validity/MI ¹	Reliability	ME ²	Criterion validity	Hypothesis testing for content validity	Responsiveness	% strong-moderate evidence
Forde et al. (2012)	S	S	L	U	U	U	U	U	22.2%
Dovran et al. (2013)	C	S	U	U	U	U	U	U	11.1%
Hernández et al. (2013)	S	S	U	U	U	S	S	U	44.4%
Karos et al. (2014)	L	S	U	U	U	U	S	U	22.2%
Charak et al. (2017)	S	S	U	U	U	U	S	U	33.3%
He et al. (2019)	S	S	S	U	U	U	S	U	44.4%
Kongerslev et al. (2019)	S	C	U	U	U	U	S	U	33.3%
Spies, Kidd and Seedat (2019)	S	S	U	U	U	U	U	U	22.2%
Aloba, Opakunle and Ogunrinu (2020)	S	S	L	U	U	U	L	U	22.2%
Schmidt et al. (2020)	U	U	U	U	U	U	C	U	0%
% strong-moderate evidence	70%	80%	10%	0%	0%	10%	50%	0%	
% limited conflicting evidence	20%	10%	20%	0%	0%	0%	20%	0%	
% unknown evidence	10%	10%	70%	100%	100%	90%	30%	100%	



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.

measurement error, content validity, and responsiveness. Specifically, in this study, none of the ten examined articles provided any information regarding these properties. Likewise, in [Saini, Hoffmann, Pantelis, Everall, and Bousman \(2019\)](#) results, only one study out of 68 reported information regarding measurement error, only three studies tested cross cultural validity, and no study had any information regarding responsiveness, out of the 68 articles included in the review. This information implies that deeper research is needed in these areas in order to test a scale in all its characteristics.

In this study, there was a general pattern in reporting psychometric properties, as virtually all studies reported information regarding internal consistency and structural validity (90% of studies), and almost all did hypotheses testing (70%). However, many of them did not clearly state the hypotheses a priori ([Aloba, Opakunle, & Ogunrinu, 2020](#); [Forde, Baron, Scher, & Stein, 2012](#); [He, Zhong, Gao, Xiong, & Yao, 2019](#); [Karos, Niederstrasser, Abidi, Bernstein, & Bader, 2014](#)). Three studies computed measurement invariance ([Aloba, Opakunle, & Ogunrinu, 2020](#); [Forde, Baron, Scher, & Stein, 2012](#); [He, Zhong, Gao, Xiong, & Yao, 2019](#)) from which two had limited strength of evidence. These results showed a trend in the procedures of instruments validations in the past ten years, in which the general tendency has been to examine internal consistency, structural validity, and hypotheses testing, but leave unexamined the psychometric properties related to reliability, measurement error, content validity, cross-cultural validity, criterion validity, and responsiveness interpretations, in the way that we can conclude that we do not dispose of enough information to assess all the psychometric properties of the Childhood Trauma Questionnaire — Short Form (CTQ-SF).

None of the studies examined cross-cultural validity, although the studies have been performed in a variety of countries as Canada ([Forde, Baron, Scher, & Stein, 2012](#)), Norway ([Dovran et al., 2013](#)), Spain ([Hernández et al., 2013](#)); Sweeden ([Karos, Niederstrasser, Abidi, Bernstein, & Bader, 2014](#)), Burundi ([Charak, de Jong, Berckmoes, Ndayisaba, & Reis, 2017](#)), China ([He, Zhong, Gao, Xiong, & Yao, 2019](#)), Denmark ([Kongerslev et al., 2019](#)), South Africa ([Spies, Kidd, & Seedat, 2019](#)), Nigeria ([Aloba, Opakunle, & Ogunrinu, 2020](#)), and the United States ([Schmidt, Narayan, Atzl, Rivera, & Lieberman, 2020a](#)). The most probable reason for this lack of cross-cultural information could be that none of these studies disposed of two culturally different samples to perform cross-cultural validity analyses. Therefore, this is a topic that remains unexplored. Languages of administration were diverse (English, Spanish, Norwegian, German, French, Chinese, Danish and Afrikaans). However, no language appeared more frequently than the others, which means that in the past ten years of research, this instrument has been validated only once in each language, as far as our results indicate. Therefore, generalizability within language of psychometric properties is not available.

Regarding the quality of the Childhood Trauma Questionnaire — Short Form ([Bernstein et al., 2003](#)), we can conclude that although it is one of the most well-established instruments to asses child maltreatment, further research is needed regarding its test-retest reliability, measurement error, and criterion validity, since no study examined such properties. Regarding factor structure, it seemed to be quite well established that the questionnaire is measuring five distinguishable dimensions, and although some studies introduced slight modifications, the 5-factor structure seemed to be consistent. However, when it comes to internal consistency, there

were some factors which consistently seemed to have problems, especially, the physical neglect factor. In the light of these results, we suggest a re-examination of the items of this factor in order to introduce some further modifications to improve its internal consistency. Finally, this instrument appeared to have good results for hypothesis testing for construct validity, and measurement invariance across different groups.

There are some limitations in this systematic review that should be taken into consideration. Firstly, only studies published in Spanish and English were considered, resulting in the exclusion of validations of the CTQ-SF that were published in any other languages, which could have resulted in missing valuable information on its psychometric properties. Secondly, the use of the COSMIN checklist (Mokkink et al., 2010), which, although well composed and validated, still allows some subjectivity in the interpretation of the information provided from the studies to the guidelines the COSMIN manual (Prinsen et al., 2018). For example, in some characteristics as in box 6 (reliability) 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. Another example can be box 10b (construct approach) where measurement properties regarding relations with other constructs/measures have to be assessed as “clear or sufficient”. This decision, again, can largely depend on the knowledge, expertise, and discipline of the researcher assessing the study.

All considered, this is the first study, to our knowledge, that examined and encapsulated the main psychometric properties of the Childhood Trauma Questionnaire — Short Form (CTQ-SF). The main results of this study showed that there was a consistent pattern in the validation studies of the CTQ-SF indicating that the same psychometric properties have been tested over and over again. Replicability and consistency in results are valuable indicators of the quality of this instrument. However, future research should also address the remaining properties as reliability, measurement error, content validity, cross cultural and criterion validity, and responsiveness, as well as re-examining some problematic dimensions when it comes to internal consistency, in order to address this gap and advance in the knowledge on childhood maltreatment measurement.

Funding

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).

Declaration of competing interest

None.

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*, Article 104357.
- Baker, A. J. L., & Maiorino, E. (2010). Assessments of emotional abuse and neglect with the CTQ: Issues and estimates. *Children and Youth Services Review*, *32*, 740–748.
- Bekhet, A. K., & Zauszniewski, J. A. (2012). Methodological triangulation: An approach to understanding data. *Nurse Researcher*, *20*(2), 40–43.
- Belsky, J., Bell, B., Bradley, R. H., Stallard, N., & Stewart-Brown, S. L. (2007). Socioeconomic risk, parenting during the preschool years and child health age 6 years. *European Journal of Public Health*, *17*(5), 508–513.
- Bernstein, D. P., & Fink, L. (1998). *Childhood trauma questionnaire: A retrospective self-report manual*. San Antonio, TX: The Psychological Corporation.
- Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., ... Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Maltreatment & Neglect*, *27*(2), 169–190.
- Bouwmeester-Landweer, M. B. R. (2006). *Early home visitation in families at risk for child maltreatment*. Doctoral thesis. 2006. Leiden: Leiden University Medical Center (LUMC), Leiden University.
- 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*, 383–392.
- Choi, S. (2012). Family drug courts in child welfare. *Child and Adolescent Social Work Journal*, *29*(6), 447–461.
- Dovran, A., Winje, D., Øverland, S. N., Breivik, K., Arefjord, K., Dalsbø, A. S., ... 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.
- 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.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., & Marks, J. S. (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 Preventive Medicine*, *14*(4), 245–258.
- Fineman, M. A. (2017). Vulnerability and inevitable inequality. *Oslo Law Review*, *4*(3), 133–149.
- 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.
- Fosse, R., Skjelstad, D. V., Schalinski, I., Thekkumthala, D., Elbert, T., Aanonsen, C. M., ... 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.
- Hanson, J. L., Knodt, A. R., Brigid, 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*, 1611–1619.
- 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*, 102–108.
- Hernández, A., Gallardo-Pujol, D., Pereda, N., Arntz, A., Bernstein, D. P., Gaviria, A. M., ... 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.
- 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.

- Hsieh, Y. P., Shen, A. C. T., Hwa, H. L., Wei, H. S., Feng, J. Y., & Huang, S. C. Y. (2020). Associations between child maltreatment, dysfunctional family environment, post-traumatic stress disorder and children's bullying perpetration in a national representative sample in Taiwan. *Journal of Family Violence*, *36*, 27–36.
- 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.
- Kisely, S., Abajobir, A. A., Mills, R., Strathearn, L., Clavarino, A., Gartner, C., & Najman, J. M. (2020). Child maltreatment and persistent smoking from adolescence into adulthood: A birth cohort study. *Nicotine and Tobacco Research*, *22*(1), 66–73.
- 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.
- 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*, Article 104026.
- Laslett, A. M., Ferris, J., Dietze, P., & Room, R. (2012). Social demography of alcohol-related harm to children in Australia. *Addiction*, *107*(6), 1082–1089.
- McMillin, S. E., Bultas, M. W., Zander, T., Wilmott, J., Underwood, S., Broom, M. A., & Zand, D. H. (2016). The role of maternal knowledge of child development in predicting risk for child maltreatment. *Clinical Pediatrics*, *55*(4), 374–376.
- Messman-Moore, T. L., & Bhuptani, P. H. (2017). A review of the long-term impact of child maltreatment on posttraumatic stress disorder and its comorbidities: An emotion dysregulation perspective. *Clinical Psychology: Science and Practice*, *24*(2), 154–169.
- Milner, J. S. (1986). *The child abuse potential inventory manual* (2nd ed.). Webster: Psytec Inc.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1–9.
- 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.
- Norman, R. E., De Byambaa, M. 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.
- 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.
- Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., ... Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”—A practical guideline. *Trials*, *17*(1), 1–10.
- Ramiro, J. (2016). Los derechos de los niños en las políticas españolas para la infancia. *Revista de derecho político*, *95*, 117–146.
- 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*, 106–113.
- Schmidt, M. R., Narayan, A. J., Atzl, V. M., Rivera, L. M., & Lieberman, A. F. (2020a). 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.
- Schmidt, M. R., Narayan, A. J., Atzl, V. M., Rivera, L. M., & Lieberman, A. F. (2020b). 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>
- 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*, 157–166.
- 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.
- Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42.
- Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651–657.
- WHO. (2020). Child maltreatment [online]. Available at: <https://www.who.int/news-room/fact-sheets/detail/child-maltreatment>.