UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Mental health screening and assessment tools for forcibly displaced children
A systematic review

Verhagen, I.L.; Noom, M.J.; Lindauer, R.J.L.; Daams, J.G.; Hein, I.M.

REVIEW ARTICLE

🔓 OPEN ACCESS Check for updates

Mental health screening and assessment tools for forcibly displaced children: a systematic review

Ilse L. Verhagen[a], Marc J. Noom[b], Ramón J. L. Lindauer[a], Joost G. Daams[c] and Irma M. Hein[a]

[a]Department of Child and Adolescent Psychiatry, Amsterdam University Medical Center, Amsterdam, Netherlands; [b]Department of Child Development and Education, University of Amsterdam, Amsterdam, Netherlands; [c]Medical Library, Amsterdam University Medical Center, Amsterdam, Netherlands

ABSTRACT

Background: An unprecedentedly large number of people worldwide are forcibly displaced, of which more than 40 percent are under 18 years of age. Forcibly displaced children and youth have often been exposed to stressful life events and are therefore at increased risk of developing mental health issues. Hence, early screening and assessment for mental health problems is of great importance, as is research addressing this topic. However, there is a lack of evidence regarding the reliability and validity of mental health assessment tools for this population.

Objective: The aim of the present study was to synthesise the existing evidence on psychometric properties of patient reported outcome measures [PROMs] for assessing the mental health of asylum-seeking, refugee and internally displaced children and youth.

Method: Systematic searches of the literature were conducted in four electronic databases: MEDLINE, PsycINFO, Embase and Web of Science. The methodological quality of the studies was examined using the COSMIN Risk of Bias checklist. Furthermore, the COSMIN criteria for good measurement properties were used to evaluate the quality of the outcome measures.

Results: The search yielded 4842 articles, of which 27 met eligibility criteria. The reliability, internal consistency, structural validity, hypotheses testing and criterion validity of 28 PROMs were evaluated.

Conclusion: Based on the results with regard to validity and reliability, as well as feasibility, we recommend the use of several instruments to measure emotional and behavioural problems, PTSD symptoms, anxiety and depression in forcibly displaced children and youth. However, despite a call for more research on the psychometric properties of mental health assessment tools for forcibly displaced children and youth, there is still a lack of studies conducted on this topic. More research is needed in order to establish cross-cultural validity of mental health assessment tools and to provide optimal cut-off scores for this population.

HIGHLIGHTS
- Research on the psychometric properties of mental health screening and assessment tools for forcibly displaced children and youth is slowly increasing.
- However, based on the current evidence on the validity and reliability of screening and assessment tools for forcibly displaced children, we are not able to recommend a core set of instruments. Instead, we provide suggestions for best practice.
- More research of sufficient quality is important in order to establish crsoss-cultural validity and to provide optimal cut-off scores in mental health screening and assessment tools for different populations of forcibly displaced children and youth.

Herramientas de detección y evaluación de salud mental para niños desplazados por la fuerza: una revisión sistemática

Antecedentes: Un número sin precedentes de personas en todo el mundo son desplazadas por la fuerza, de las cuales más del 40 por ciento son menores de 18 años. Los niños y jóvenes desplazados por la fuerza a menudo han estado expuestos a eventos vitales estresantes y, por lo tanto, corren un mayor riesgo de desarrollar problemas de salud mental. Por lo tanto, la detección temprana y la evaluación de los problemas de salud mental son de gran importancia, al igual que la investigación que aborda este tema. Sin embargo, hay una falta de evidencia con respecto a la confiabilidad y validez de las herramientas de evaluación de la salud mental para esta población.

Objetivo: El objetivo del presente estudio fue sintetizar la evidencia existente sobre las propiedades psicométricas de los instrumentos de medición de resultado reportadas por el paciente [PROM, por sus siglas en inglés] para evaluar la salud mental de los niños y jóvenes solicitantes de asilo, refugiados y desplazados internos.

Método: Se realizaron búsquedas sistemáticas de la literatura en cuatro bases de datos electrónicas: MEDLINE, PsycINFO, Embase y Web of Science. La calidad metodológica de los estudios se examinó mediante la lista de verificación de riesgo de sesgo de COSMIN. Además, se utilizaron los criterios COSMIN de buenas propiedades de medición para evaluar la calidad de los instrumentos de medición de resultados.

Resultados: La búsqueda arrojó 4842 artículos, de los cuales 27 cumplieron con los criterios de elegibilidad. Se evaluaron la confiabilidad, consistencia interna, validez estructural, prueba de hipótesis y validez de criterio de 28 PROM.

Conclusión: En base a los resultados con respecto a la validez y confiabilidad, así como la factibilidad, recomendamos el uso de varios instrumentos para medir problemas emocionales y de conducta, síntomas de TEPT, ansiedad y depresión en niños y jóvenes desplazados por la fuerza. Sin embargo, a pesar de la petición de más investigación sobre

las propiedades psicométricas de las herramientas de evaluación de la salud mental para niños y jóvenes desplazados por la fuerza, todavía faltan estudios sobre este tema. Se necesita más investigación para establecer la validez transcultural de las herramientas de evaluación de la salud mental y proporcionar puntajes de corte óptimos para esta población.

被迫流离失所儿童的心理健康筛查和评估工具：一项系统综述

背景：全世界有史无前例的大量人口被迫流离失所，其中超过 40% 的人未满 18 岁。被迫流离失所儿童和青年经常面临应激生活事件且因此患心理健康问题的风险增加。因此，对心理健康问题进行早期筛查和评估非常重要，针对这一主题的研究也非常重要。然而，缺乏该人群心理健康评估工具可靠性和有效性的证据。
目的：本研究旨在综合评估寻求庇护、难民和国内流离失所儿童和青少年心理健康的患者报告结果测量 [PROM] 心理测量特性的现有证据。
　　方法：在四个电子数据库中对文献进行了系统搜索：MEDLINE、PsycINFO、Embase 和 Web of Science。使用 COSMIN 偏倚风险检查表检查研究的方法学质量。此外，使用良好测量特性的 COSMIN 标准评估结果测量的质量。
结果：搜索到了 4842 篇文章，其中 27 篇符合资格标准。评估了 28 个 PROM 的信度、内部一致性、结构效度、假设检验和标准效度。
结论：根据有效性和可靠性以及可行性的结果，我们建议使用多种工具来测量被迫流离失所儿童和青少年的情绪和行为问题、PTSD 症状、焦虑和抑郁。然而，尽管呼吁对被迫流离失所儿童和青少年心理健康评估工具的心理测量特性进行更多研究，此主题的研究仍然缺乏。需要更多的研究来确定心理健康评估工具的跨文化有效性，并为该人群提供最佳的划界分。

Abbreviations

APA	American Psychiatric Association
ASEBA	Achenbach System of Empirically Based Assessment
AUC	Area Under Curve
CAPS	Clinician-Administered PTSD Scale
CATS	Child and Adolescent Trauma Screen
CBCL	Child Behaviour Checklist
CES-DC	Center for Epidemiological Studies Depression Scale for Children
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CGI-s	Clinical Global Impression – severity
CIDI	Composite International Diagnostic Interview
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
CPDS	Child Psychosocial Distress Screener
CPSS	Child PTSD Symptom Scale
CRIES	Children's Revised Impact of Event Scale
DICA	Diagnostic Instrument for Children and Adolescents
DSM	Diagnostic and Statistical Manual of mental disorders
DSRS	Depression Self-Rating Scale
EFA	Exploratory Factor Analysis
EV	Explained Variance
GAPD	Global Assessment of Psychosocial Disability
HSCL	Hopkins Symptom Checklist
HTQ	Harvard Trauma Questionnaire
ICHOM	International Consortium for Health Outcomes Measurement
IES	Impact of Event Scale
Kinder-DIPS	Structured Diagnostic Interview for Mental Disorders in Children and Adolescents
K-SADS	Schedule for Affective Disorders and Schizophrenia for School-Age Children
MINI KID	Mini International Neuropsychiatric Interview for Children and Adolescents
PDS	Post-traumatic Diagnostic Scale
PHQ	Patient Health Questionnaire
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Patient-Reported Outcome Measure
PROSPERO	International Prospective Register of Systematic Reviews
PTSD	Post-Traumatic Stress Disorder
PTSDSSI	Post-traumatic Stress Disorder Semi-structured Interview
RATS	Reactions of Adolescents to Traumatic Stress questionnaire
RHS	Refugee Health Screener
RMSEA	Root Mean Square Error of Approximation
SCARED	Screen for Child Anxiety-Related Emotional Disorders
SDQ	Strengths and Difficulties Questionnaire
TRF	Teacher Report Form
UCLA	PTSD RI University of California at Los Angeles PTSD Index
UNHCR	United Nations High Commissioner for Refugees
YSR	Youth Self Report

1. Introduction

To date, an unprecedentedly large number of people are forcibly displaced. At the end of 2020, there were an estimated 82.4 million refugees, asylum seekers and internally displaced people worldwide, according to the UNHCR (2021). Syria's devastating war, an unfolding humanitarian crisis in Afghanistan, the recent war in Ukraine and several other conflicts around the globe have caused a surge of forcibly displaced in the past decade. Moreover, it is expected that the number of forcibly displaced people will continue to rise, in part as a direct and indirect result of climate change and war. Globally, children under 18 years of age account for about 42 percent of the forcibly displaced population (UNHCR, 2021).

Due to the growing population of forcibly displaced people, as well as a rise in refugees seeking protection beyond the borders of neighbouring countries in recent years, there has been an increase in research

on the mental health of forcibly displaced children and youth (Hodes, 2019). Forcibly displaced children and youth often experience many stressful life events, both pre-flight, during the flight and in the resettlement phase. Examples of stressful life events include exposure to violence, loss of loved ones, separation from parents, lack of access to basic necessities and discrimination (Fazel & Stein, 2002; Lustig et al., 2004). Stressful life events are a major risk factor for the development of mental health problems (Bean et al., 2007a; Fazel et al., 2012; Fazel & Betancourt, 2018; Heptinstall et al., 2004; Porter & Haslam, 2005; Reed et al., 2012). A recent meta-analysis showed high prevalence rates of post-traumatic stress disorder [PTSD] (23%), anxiety (16%) and depression (14%) among refugee and asylum-seeking children and adolescents (Blackmore et al., 2020). Moreover, research points towards probable long-term persistence of mental health problems in this population (Dyregrov et al., 2002; Vervliet et al., 2014). Therefore, mental health screening and assessment is of great importance to support the delivery of early interventions and treatment (Blackmore et al., 2020; Gadeberg et al., 2017; Horlings & Hein, 2018).

With this objective in mind, mental health assessment tools are used widely among both researchers and health care professionals. However, previous reviews have shown that there is a lack of research on the reliability and validity of mental health assessment tools in different refugee children and youth populations (Ehntholt & Yule, 2006; Gadeberg et al., 2017; Horlings & Hein, 2018). Hence, the aim of the present study was to synthesise the existing evidence on psychometric properties of mental health assessment tools for asylum-seeking, refugee and internally displaced children and youth.

Patient reported outcome measures [PROMs] with sufficient psychometric properties are of vital importance to perform adequate mental health screening and assessment (Mokkink et al., 2010). When constructs are of a subjective nature and therefore not directly measurable, which is the case with self- and proxy-reported questionnaires on mental health, it is even more important to ensure the reliability and validity of these tools (Mokkink et al., 2010). Moreover, reliability and validity do not represent the measurement instrument as such, but the application of that instrument within a certain population and context (Terwee et al., 2007). The majority of measurement instruments for mental health issues have been developed for adult western populations, but a number of questionnaires have been adapted or developed for children and youth. Yet, a recent systematic review on the validity of mental health measurement tools in refugee children and youth by Gadeberg et al. (2017) found only nine validation studies that met the inclusion criteria of their review. Gadeberg et al.

(2017) concluded that there is a severe lack of validated trauma and mental health assessment tools for this population. No studies have been conducted on the validity of assessment tools with refugee children under six years of age. The quality of the assessment tools was generally found to be better when assessing internalising symptoms than when assessing externalising symptoms. Nevertheless, the overall evidence was considered weak and no recommendations for best practice were provided.

When assessing mental health in forcibly displaced children and youth, it is crucial that the instrument administered is culturally valid, meaning that it is applicable and relevant to children of different cultural backgrounds (Gadeberg et al., 2017). Because of cross-cultural differences in mental health problems, such as variations in symptoms, it is possible that instruments do not measure the same construct as intended when administered to different cultural populations (Ertl et al., 2011; Kohrt et al., 2011). Additionally, cut-off scores established in particular populations and contexts could cause erroneous prevalence rates and false diagnoses when applied in other populations and contexts (Kohrt et al., 2011). Consequently, a lack of validated assessment tools could result in an overestimation or underestimation of mental health issues. Additionally, the use of non-validated assessment tools in scientific research on forcibly displaced children may lead to unreliable results (Gadeberg et al., 2017; Stolk et al., 2017). As Gadeberg et al. (2017) concluded, 'the value that can be attached to results of a study is pre-determined by the degree of reliability and validity of the tool that has been used.' (p. 445). Besides, when assessing the mental health of forcibly displaced children, it is important that the instrument is sensitive in recognising trauma and stressor-related symptoms, because of the high prevalence rates of these symptoms in this population (Gadeberg et al., 2017).

An overview of research conducted on the measurement properties of instruments assessing the mental health of forcibly displaced children is imperative. There is an urgent call for more validation studies on mental health assessment tools for refugee children, expressed in a letter by Gadeberg and Norredam (2016). The review by Gadeberg et al. (2017) needed to be updated and broadened by also including internally displaced children and youth. As Gadeberg et al. (2017) noted, these children have 'the threat of integrity in common with refugee children' (p. 445).

The aim of this systematic review was to provide a clear overview of measurement properties of PROMs in forcibly displaced children and youth, in order to provide recommendations on the most suitable instruments and to identify gaps in the current evidence on this topic.

2. Methods

A protocol for this systematic review was written using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P) checklist (Moher et al., 2015). The research protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) as number CRD42020150367, accessible at http://www.crd.yor.ac.uk/PROSPERO/. The review was conducted in accordance with the PRISMA statement (Moher et al., 2009).

The protocol for systematic reviews of PROMs that was published by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative was also used as a guideline in performing this systematic review (Mokkink et al., 2018).

2.1. Search strategy

On 6 August 2019, we conducted a systematic literature search in four electronic databases: MEDLINE, PsycINFO, Embase and Web of Science. The literature search was repeated on 8 July 2021. The search strategy was designed with the assistance of a medical information specialist (JD). No restrictions were imposed with regard to language or publication date. Search terms for the systematic search included 'child' OR 'adolescent' OR 'minor' AND 'refugee' OR 'asylum seeker' OR 'internally displaced' AND 'psychometric' OR 'validity' OR 'reliability' OR 'instrument' AND 'mental health' OR 'psychiatric' OR 'psychological'.

2.2. Selection criteria

The eligibility criteria for the selection of the full text articles to be reviewed were (1) studies on the development or evaluation of psychometric properties of a mental health measurement tool for PTSD, anxiety, depression or emotional and behavioural problems, and (2) studies where the majority of the study population was comprised of refugee, asylum-seeking children and/or internally displaced children and adolescents between 0 and 23 years of age. Excluded were (1) studies not reporting on criterion or construct validity, (2) studies on measurement tools that are administered as (semi-) structured interviews intended for diagnostic assessment, and (3) studies on assessment tools developed for a specific language or cultural group only.

2.3. Study selection

Results from the search were imported to the bibliographic database of EndNote by the medical information specialist (JD) and all duplicate studies were removed. The remaining references were uploaded into Rayyan QCRI to perform selection of the studies (Ouzzani et al., 2016). The titles and abstracts of the first 100 articles were independently screened by two review authors (IV, IH) based on the eligibility criteria. To reduce the possibility of selection bias, inter-rater reliability was measured using Cohen's Kappa statistic. When the inter-rater reliability was less than 0.80, an additional subset of 100 articles was independently assessed until the inter-rater reliability was ≥ 0.80. After screening 300 articles, the kappa was sufficiently high. The remaining 3320 titles and abstracts were then screened by one review author (IV). When the abstract contained limited information on the study, the full text was reviewed. After selection based on titles and abstracts, the full texts of potentially eligible studies were obtained and examined against predefined inclusion criteria by one review author (IV) for inclusion in the review.

The reference list of selected papers was manually searched by one author (IV) in order to identify additional relevant studies. Furthermore, manual searches of grey literature (i.e. unpublished papers, reports and conference abstracts) were performed by this author (IV).

After repeating the literature search in July 2021, the articles that were included in the first selection process were uploaded and the selection procedure was repeated using ASReview (Van de Schoot et al., 2021). ASReview is a free, open source software which overcomes time-consuming manual screening by prioritising relevant studies via machine learning algorithms. The algorithm was built on the articles that were included with the manual search. Therefore, one author (IV) only had to screen titles and abstracts of 15 percent of the papers, of which the last five percent were consecutive, irrelevant papers.

2.4. Data extraction

One research author (IV) extracted the data using a standardised form that was designed for this systematic review, while a second review author (IH) independently checked a random subset of three data extraction forms for accuracy and completeness. The characteristics retrieved included aim of the study, sample size, method of recruitment, population characteristics such as age, gender and ethnicity, setting of the study, characteristics of the measurement instrument, language adaptations, informants and measurement properties. Additionally, a table with characteristics of each measurement instrument studied in the included articles was compiled from relevant articles, websites, publications and manuals. Information in the table included reference to the

development study, construct, target population, subscales, number of items and available translations.

2.5. Measurement properties

The reliability of an instrument refers to the extent to which the instrument yields stable and consistent results. If the construct remains unchanged, the instrument should yield the same score each time it is administered. This can be tested over time (test-retest) by different persons at the same point in time (inter-rater) or by the same person at different points in time (intra-rater). Another important element of reliability is internal consistency. Internal consistency is defined as the degree to which items on an instrument are correlated and therefore measure the same construct. Validity refers to the extent to which an instrument accurately measures what it intends to measure. Validity can be divided into content validity, construct validity and criterion validity. Content validity is the degree to which the content of an instrument is relevant to and representative of the construct that it asserts to measure. Construct validity can be divided into structural validity, hypotheses testing and criterion validity. Structural validity is defined as the extent to which the scores of an instrument are an adequate reflection of the dimensionality of the construct that it intends to measure. Hypotheses testing refers to the degree to which the scores of an instrument are consistent with hypotheses. These hypotheses are often based on correlations between the score on the instrument and the score on another instrument that measures a similar (convergent) or different (divergent) construct, or on differences in scores between relevant groups. Criterion validity is defined as the extent to which an instrument reflects a 'gold standard'. A gold standard is an external criterion of the construct being measured. In most cases, a diagnostic interview is used as the gold standard. The COSMIN has defined other measurement properties, but these will not be described in this review (Mokkink et al., 2018).

2.6. Quality assessment and evidence synthesis

The COSMIN methodology was followed to assess the quality of the included studies and to evaluate the quality of the measurement properties (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). The evaluation of the measurement properties of the instruments consisted of three steps: (1) the COSMIN Risk of Bias checklist was used to assess the methodological quality of the study and rate the quality as very good, adequate, doubtful or inadequate, (2) the results on the measurement properties were rated as sufficient (+), insufficient (-) or indeterminate (?) based on the criteria for good measurement properties, and (3)

the overall evidence was summarised: the measurement properties were rated as sufficient (+), insufficient (-), inconsistent (+/-) or indeterminate (?), and the total quality of the evidence was rated as high, moderate or low.

We deviated from the COSMIN Risk of Bias checklist with regard to rating the quality of internal consistency in the study. We rated the quality of the study as 'adequate' whenever the internal consistency was reported per subscale as intended by the instrument, despite limited or inconsistent results on the structural validity of the tool in forcibly displaced children. In many studies there was only limited evidence for the structural validity, yet internal consistency provides some insight into the reliability of the instrument. We also deviated from the COSMIN guideline by only rating the total quality of evidence as 'high' whenever there were at least two studies of very good quality, due to the wide range of populations in this review and consequent limitations in generalizability.

We added additional criteria for exploratory factor analyses [EFA], since several studies did not provide evidence for structural validity by conducting a confirmatory factor analysis and the COSMIN criteria does not provide a rating system for EFAs. Thus, EFAs were rated from very poor (< 30%) to excellent (>70%). Moreover, we added additional criteria for internal consistency, internal consistency was rated from very poor (<0.30) to excellent (>0.90), as the COSMIN criteria only have the dichotomy between sufficient and insufficient (above or below 0.70). Lastly, we added additional criteria for sensitivity and specificity, both sensitivity and specificity were rated from very poor (<40%) to excellent (>90%). The COSMIN criteria have no rating system for sensitivity and specificity, but these psychometric properties give a good insight into the value of the instrument for screening purposes.

2.7. Assessment tools

Below is an overview of the mental health assessment tools that were evaluated in the studies included in this systematic review:

2.7.1. Behavioural and emotional problems
2.7.1.1. ASEBA. The Achenbach System of Empirically Based Assessment [ASEBA] consists of the Child Behaviour Checklist [CBCL], Youth Self Report [YSR] and Teacher Report Form [TRF] for parents, adolescents and teachers respectively (Achenbach, 1991; Achenbach & Rescorla, 2001). The ASEBA was developed for children between one and 18 year(s) of age. The instruments consist of around 118 items divided into eight subscales: delinquent behaviour, aggressive behaviour, withdrawn, somatic complaints, anxious/depressed, social problems, thought problems

and attention problems. The subscales delinquent behaviour and aggressive behaviour can be summarised into an externalising scale, and the subscales withdrawn, somatic complaints and anxious/depressed can be summarised into an internalising scale. Combining the internalising and externalising scales form a total score.

2.7.1.2. SDQ. The Strengths and Difficulties Questionnaire [SDQ] (Goodman, 1997; 2000) was developed for children between two and 17 years of age. There is a self-report and a caregiver-report of the instrument. The questionnaire consists of 25 items divided into five subscales: emotional symptoms, conduct problems, hyperactivity, peer problems and prosocial behaviour. The subscales emotional symptoms, conduct problems, hyperactivity and peer problems can be summarised into a total problem score.

2.7.2. Post-traumatic stress disorder [PTSD]

2.7.2.1. CPSS. The Post-traumatic Diagnostic Scale [PDS] (Foa et al., 1997) was developed to measure symptoms of PTSD in adults. The modification of the PDS for children and adolescents between eight and 18 years of age is the Child PTSD Symptom Scale [CPSS] (Foa et al., 2001; 2018). The CPSS is available in a self-report version and a caregiver-report version. Based on the DSM-IV (APA, 1994), the CPSS consists of 17 items divided into subscales, in accordance with the DSM-IV. These subscales include: intrusion, avoidance and arousal. For the CPSS-5, which is based on the DSM-5 (APA, 2013), the subscale incorporates the addition of changes in cognition and mood, resulting in a questionnaire with 20 items in total.

2.7.2.2. CRIES. The Impact of Event Scale [IES] (Horowitz et al., 1979) was developed as a screening tool for PTSD in adults. The IES-based version for children between 8 and 18 years of age is the Children's Revised Impact of Event Scale [CRIES] (Children and War Foundation, 1998). The CRIES is available as self-report and caregiver-report in both a 13-item version and an eight-item version. The items of the CRIES are based on the DSM-IV. The CRIES-8 consists of eight items divided into two subscales: intrusion and avoidance. The CRIES-13 consists of 13 items divided into three subscales, with five additional items in the subscale arousal. The CRIES has not been updated to meet the criteria of the DSM-5.

2.7.2.3. CATS. A recently developed instrument to measure PTSD symptoms in children and adolescents is called the Child and Adolescent Trauma Screen [CATS] (Sachser et al., 2017). The CATS consists of 20 items divided into four symptom clusters based on the DSM-5 and on preliminary information about PTSD criteria for the ICD-11. The CATS is

available in a caregiver-report for children between three and six years of age and as a self-report and caregiver-report for children and adolescents between seven and 17 years of age.

2.7.2.4. CBCL-PTSD. The CBCL-PTSD is a tool to measure PTSD symptoms, consisting of a selection of items derived from the original CBCL that are relevant to PTSD. Wolfe et al. (1989) used a selection of 20 items using the DSM-III (APA, 1980) as a guide. Nehring et al. (2021) created an alternative CBCL-PTSD scale consisting of 18 items. This instrument was developed by psychometrically guided selection of items with an appropriate correlation to PTSD and presence of these symptoms in more than 20% of the cases with an established PTSD-diagnosis in a sample of Syrian refugee children ($N = 61$). The CBCL-PTSD is a unidimensional scale, thus items are not divided into subscales based on the DSM-5.

2.7.2.5. UCLA PTSD RI. The University of California at Los Angeles Post Traumatic Stress Disorder Reaction Index [UCLA PTSD RI] (Pynoos et al., 1998) was developed as a caregiver-report for children and youth six years and younger, and as a self-report and caregiver-report for children and youth between seven and 21 years of age. The UCLA-PTSD Reaction Index for the DSM-IV consists of 16 items divided into three subscales, based on the DSM-IV criteria for PTSD. The instrument has been updated for the DSM-5: this updated version consists of 31 items.

2.7.2.6. HTQ. The Post Traumatic Stress Symptoms [PTSS] section of the Harvard Trauma Questionnaire [HTQ] (Mollica et al., 1992) was developed for adults. Although the HTQ has not been adapted for younger populations, the questionnaire is often used for assessing PTSD in adolescents (Jakobsen et al., 2017). The instrument consists of 16 items divided into three subscales, based on the DSM-IV criteria for PTSD. The instrument has been updated for the DSM-5 through the incorporation of nine additional items.

2.7.2.7. RATS. The Reactions of Adolescents to Traumatic Stress questionnaire [RATS] was specifically developed for refugee adolescents between 12–18 years of age (Bean et al., 2006b). The instrument consists of 22 items with three subscales, based on the DSM-IV: intrusion, avoidance and arousal. The instrument has not been updated for the DSM-5.

2.7.3. Depression and anxiety

2.7.3.1. CES-DC. The Center for Epidemiological Studies Depression Scale for Children [CES-DC] (Faulstich et al., 1986; Weissman et al., 1980) is the modification of the adult scale CES-D (Radloff, 1977) for children and adolescents between eight

and 18 years of age. The CES-DC measures depression in 20 items divided into four subscales: depressed affect, positive affect, somatic activity and interpersonal functioning (Faulstich et al., 1986; Weissman et al., 1980). The CES-DC-10 is a new 10-item version of the instrument (McEwen et al., 2020).

2.7.3.2. DSRS. The Birleson Depression Self-Rating Scale for children [DSRS] (Birleson, 1981; 1987) is an 18-item instrument that was developed to measure depression symptoms in children between eight and 14 years of age. The instrument is based on an operational definition of depressive disorder. A five-factor structure was found for the original instrument.
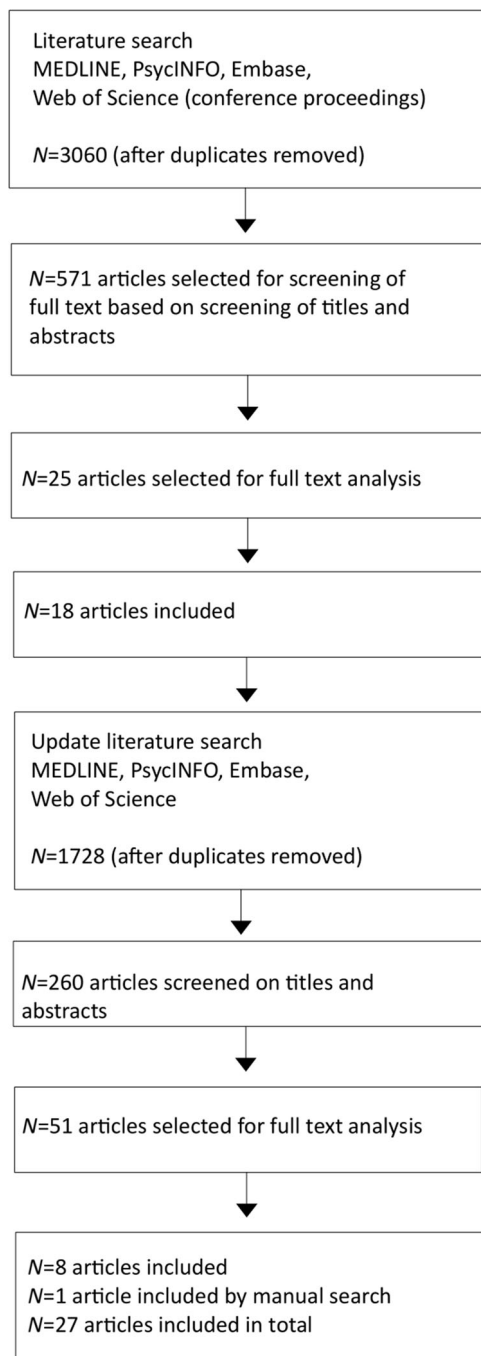
2.7.3.3. PHQ-A. The Patient Health Questionnaire is a self-report to assess anxiety, mood, eating, and substance use disorders (Spitzer et al., 1999). The PHQ-9 is the module to measure the severity of depressive symptoms with nine items reflecting the DSM-IV criteria for major depressive disorder (Kroenke et al., 2001). The Patient Health Questionnaire for Adolescents [PHQ-A] (Johnson et al., 2002) is a modification of the adult scale PHQ-9 for adolescents between 11 and 17 years of age. Like the original instrument, the PHQ-A is a unidimensional scale consisting of nine items measuring symptoms of depression.

2.7.3.4. HSCL. The Hopkins Symptom Checklist-25 was developed to assess depression and anxiety symptoms in adults (Derogatis et al., 1974). The HSCL-37A was developed specifically for refugee adolescents between 12 and 18 years of age (Bean et al., 2007b). Similar to the original instrument, the HSCL-37A consists of an anxiety and depression subscale; an additional subscale measuring externalising behaviour was also included. The HSCL-30 is a 30-item version of the HSCL developed for adults, with an added subscale on somatisation (Hoge et al., 2006). The HSCL-Y is a modification of the HSCL-30, suitable for refugee and migrant adolescents between 12 and 18 years of age. The scale consists of 16 items measuring depression, anxiety and somatisation symptoms (Khawaja et al., 2019).

2.7.3.5. SCARED. The Screen for Child Anxiety-Related Emotional Disorders [SCARED] (Birmaher et al., 1997) is an instrument developed for children and adolescents between eight and 18 years of age. The original instrument consists of 41 items divided into the following subscales: panic disorder or significant somatic symptoms, generalised anxiety disorder, separation anxiety disorder, social anxiety disorder and significant school avoidance. A new modification of the SCARED has been developed, based on a study involving Syrian refugee children in Lebanon (McEwen et al., 2020). The instrument consists of 18 items divided into the same subscales as the original tool, with the omission of the original tool's school avoidance subscale.

2.7.4. Others
2.7.4.1. CPDS. The Child Psychosocial Distress Screener [CPDS] (Jordans et al., 2008) was developed as a short measure of seven items that measure psychosocial distress in children. The instrument was developed for children between eight and 14 years of age residing in conflict-affected areas. The tool consists of the following subscales: child distress, child resilience and school context. Five items on child distress and child resilience are filled out by the child,

Literature search
MEDLINE, PsycINFO, Embase,
Web of Science (conference proceedings)

N=3060 (after duplicates removed)

↓

N=571 articles selected for screening of full text based on screening of titles and abstracts

↓

N=25 articles selected for full text analysis

↓

N=18 articles included

↓

Update literature search
MEDLINE, PsycINFO, Embase,
Web of Science

N=1728 (after duplicates removed)

↓

N=260 articles screened on titles and abstracts

↓

N=51 articles selected for full text analysis

↓

N=8 articles included
N=1 article included by manual search
N=27 articles included in total

Figure 1. Flowchart.

while two items on school context are filled out by the teacher.

2.7.4.2. RHS. The Refugee Health Screener [RHS] (Hollifield et al., 2013; 2016) was developed specifically for refugee adolescents and adults aged 14 years or older. It is a unidimensional tool consisting of 13 items measuring emotional distress, such as PTSD, depression and anxiety symptoms (Hollifield et al., 2016).

3. Results

In total, 27 articles were included in this review. An overview of the selection procedure is provided in Figure 1. The flowchart documents the number of studies remaining at each stage of the selection process. Table 1 shows the characteristics of the study populations of the included studies. Table 2 and Table 3 show the results of the studies, with regard to measurement properties.

Below are the results with regard to the quality of evidence for each separate study and the quality of the measurement properties for each PROM. Due to low evidence and indeterminate results on reliability, we have only reported these results in Table 2. Similarly, the overall evidence for hypotheses testing was low with inconsistent results. Thus, the result for this measurement property can be found in Table 3.

3.2. Behavioural and emotional problems

3.2.1. ASEBA

The original factor structure of eight subscales was not confirmed in confirmatory factor analyses [CFA]. However, there is moderate quality of evidence for a sufficient two-factor structure for both the CBCL and TRF, namely for the internalising and externalising factors (Bean et al., 2006a; 2007c). There is high quality of evidence for excellent internal consistency of the internalising and externalising scales of the CBCL (Bean et al., 2006a; Hall et al., 2014) and moderate quality of evidence for sufficient internal consistency of the internalising and externalising scales of both the YSR and TRF (Bean et al., 2007c; Hall et al., 2014). There is moderate quality of evidence for sufficient criterion validity of the CBCL, based on the AUC. However, the sensitivity and specificity were average for both the internalising and externalising scales. There is moderate quality of evidence for sufficient criterion validity for the internalising scale of the YSR, based on the AUC. However, the AUC for the externalising scale was insufficient. For the latter, no results were reported. The sensitivity and specificity for the internalising scale of the YSR was average (Hall et al., 2014).

3.2.2. SDQ

The original five-factor structure of the SDQ was not supported by the EFA of the parent-report version (Essex, 2019; Khawaja & Dhushyanthakumar, 2020; McEwen et al., 2020) or the self-report version (Essex, 2019). Since no CFA was performed, the results are considered to be indeterminate. The explained variance of the SDQ was average. The internal consistency of the subscales is inconsistent, varying from very poor to good. There is moderate quality of evidence for the internal consistency of the total scale of both the self- and parent-report. The internal consistency of the total scale of the self-report was reported to be insufficient and the results for the internal consistency of the SDQ parent-report were inconsistent. One study assessed the criterion validity of the SDQ parent-report, providing moderate evidence for sufficient AUC, good sensitivity and average specificity (McEwen et al., 2020).

3.3. Post-traumatic stress disorder

3.3.1. CPSS

For the purpose of this review, we will only discuss the results of the CPSS below, and not the PDS. The structural validity of the CPSS was only explored with EFA. Therefore, the results are considered indeterminate (Hall et al., 2014; Hasson et al., 2021; Kohrt et al., 2011; Marshall & Venta, 2021; McEwen et al., 2020; Venta & Mercado, 2019; Ventevogel et al., 2014). The three-factor structure of the CPSS for the DSM-IV was not supported for the caregiver and self-report due to moderate and low evidence respectively, but the explained variance was very good (Hall et al., 2014; Kohrt et al., 2011; Marshall & Venta, 2021; McEwen et al., 2020; Venta & Mercado, 2019; Ventevogel et al., 2014). For the self-report of the CPSS based on the DMS-5, the four-factor structure was not supported and the explained variance was low (Hasson et al., 2021). It was suggested that a two-factor structure based on one-factor consisting of criterion b, c and e and one-factor consisting of criterion d of the DSM-5 was more suitable. A one-factor structure also seemed suitable (McEwen et al., 2020). There is moderate evidence for excellent internal consistency of the CPSS self-report (Hasson et al., 2021; McEwen et al., 2020). There is high evidence for sufficient criterion validity of the CPSS self-report. The sensitivity ranged from average to very good, and the specificity from poor to good. There is a big discrepancy between the studies with regard to the recommended cut-off scores, ranging from 12 to 26 (Hall et al., 2014; Kohrt et al., 2011; McEwen et al., 2020; Ventevogel et al., 2014). Furthermore, there is moderate evidence for sufficient criterion validity of the CPSS parent-report based on the DSM-IV, with good sensitivity and specificity (Hall et al., 2014).

Table 1. Characteristics of the included studies.

Study	Screening tool	Informant	Age (in years)	Gender % female	Nationality/Ethnicity	Displacement status	Country/Setting
Al-Amer et al. (2020)	PHQ-A	Self	13-18; M = 15.9 SD = 1.5	42%	Palestinian	Refugees	Jordan
Bean et al. (2006a)	CBCL	Legal guardian	10–18; M = 15.48 SD = 1.52	28.7%	48 countries; Angola (43.9%), Iran/Afghanistan/Iraq (4.4%), Eritrea/Ethiopia (2.7%), Somalia (2.1%), Sierra Leone (7.9%), Guinea (6.7%), Other African countries (14.0%), China/Tibet (8.6%) Other countries (9.6%)	Unaccompanied refugee minors	The Netherlands
Bean et al. (2006b)	RATS	Self	8–26; M = 15.72 SD = 1.74	40.7%	Dutch URMs: 48 countries; predominantly Angola (43%), Sierra Leone (10%), and China (8%). Belgian immigrant/refugees: 111 countries, predominantly Morocco (14%), Ghana (11%) and Turkey (9%).	Unaccompanied refugee minors, immigrant/refugee adolescents, non-migrants (control group)	The Netherlands Belgium
Bean et al. (2007b)	HSCL-37A	Self	8–26; M = 15.72 SD = 1.74	40.7%	Dutch URMs: 48 countries; predominantly Angola (43%), Sierra Leone (10%), and China (8%). Belgian immigrant/refugees: 111 countries, predominantly Morocco (14%), Ghana (11%) and Turkey (9%).	Unaccompanied refugee minors, immigrant/refugee adolescents, non-migrants (control group)	The Netherlands Belgium
Bean et al. (2007a)	TRF	Teacher	9–18; M = 15.80 SD = 1.58	28.7%	48 countries; predominantly Angola (47.3%), Sierra Leone (8.2%), Guinea (7.8%), and China (8.2%).	Unaccompanied refugee minors	The Netherlands
Dyregrov et al. (1996)	IES	Self	6–15; M = 10.68 SD = 2.14	49%	Croatia and Bosnia and Herzegovina	Displaced and refugee children	Croatia
Elbert et al. (2009)	UCLA-PTSD RI	Self + parent	10–14 M = 10.5 SD = N/A	47%	Sri Lankan	War-affected and former IDPs	Sri Lanka
Ellis et al. (2006)	UCLA-PTSD RI	Self	12–19; M = 15.6 SD = 2.0	46%	Somali	Refugees (accompanied)	United States
Ertl et al. (2011)	PDS DHSCL	Self Self	12–25 M = 17.2 SD = N/A	57%	Northern Ugandans	Internally displaced	Camps for IDPs in Northern Uganda
Essex (2019)	SDQ	Self + parent	5–17 M = N/A SD = N/A	47%	Predominantly Iraqi, Afghan and Iranian	Refugees and asylum seekers	Australia
Hall et al. (2014)	CBCL YSR CPSS-I	Parent Self Self + parent	7–18; M = 11.02 SD = 2.90	58%	Somali	Refugees (with caregivers)	Refugee camps in Ethiopia
Hasson et al. (2021)	CPSS-5	Self	N/A; M = 15.2 SD = 2.6	48.4%	El Salvador, Guatemala and Honduras	Unaccompanied minors	United States
Jakobsen et al. (2017)	HSCL-25 HTQ (PTSS)	Self Self	15–18 M = 16.2 SD = N/A	0%	Predominantly Afghan and Somali	Unaccompanied Asylum-seeking adolescents	Norway
Jordans et al. (2008)	CPDS	Child + Teacher	7–17 M = 11.75 SD = 1.79	54%	Burundian	Internally displaced	Burundi

Table 1. Continued.

Study	Screening tool	Informant	Age (in years)	Gender % female	Nationality/Ethnicity	Displacement status	Country/Setting
Jordans et al. (2009)	CPDS	Child + Teacher	6–15; M = 10.37 SD = 1.42, M = 9.79 SD = 1.40, M = 10.36 SD = 1.52, M = 12.0 SD = 1.45	46.5–54.3%	Burundian (41,85%) Indonesian (16,21%) Sri Lankan (25,68%) Sudanese (16,26%)	War-affect children and internally displaced	Burundi, Indonesia, Sri Lanka and Sudan
Khawaja et al. (2019)	HSCL-Y	Self	11–18; M = 14.89 SD = 1.72	50.6%	46 different nationalities	Refugee and migrant adolescents	Australia
Khawaja & Dhushyanthakumar (2020)	SDQ	Teacher	11–18; M = 15.0 SD = 1.8	50.9%	43 different nationalities	Refugee and migrant adolescents	Australia
Kohrt et al. (2011)	CPSS DSRS	Self Self	11–14; M = N/A SD = N/A	67.9%	Nepalese	War-affected children and former child soldiers	Nepal
Marshall and Venta (2021)	CPSS	Self + Parent	15–23; M = 19 SD = 1.8	53.8%	Central America, mainly Honduras, El Salvador and Guatemala	Recently immigrated adolescents	United States
McEwen et al. (2020)	CES-DC CPSS SCARED SDQ	Self Self Self Parent	8–17; M = 11.79, SD = 2.28	53.%	Syrian	Refugee children	Lebanon
Müller et al. (2021)	CATS	Self	N/A; M = 16.8, SD = 1.54	7%	Predominantly from Afghanistan, Syria and Eritrea	(Un)accompanied refugee minors	Germany
Nehring et al. (2021)	CBCL-PTSD	Parent	4–14; M = 8.9, SD = 2.8	41%	Syrian	Refugees	Germany
Sack et al. (1998)	IES	Self	13–25; M = 20.1, SD = 3.4	48%	Cambodian (Khmer)	Refugees	United States
Salari et al. (2017)	CRIES-8	Self	9–18; M = 15.41 SD = 1.25	2.4%	Afghan (81.4%), Iranian (5.8%), Syrian (5.4%), Iraqi (2.4%), Pakistani (1%), Somali (1%), Eritrean (1%), Ethiopian (1%), Libyan (0.5%) Lebanese (0.5%)	Unaccompanied Refugee Minors	Sweden
Sarkadi et al. (2019)	RHS-13	Self	14–18; M = 16.55 SD = 1.12	24.1%	Afghan (62.1%), Indian (3.4%), Iraqi (3.4%), Sri Lankan (3.4%), Syrian (24.1%), Venezuelan (3.4%)	(Un)accompanied refugee minors	Sweden
Venta & Mercado (2019)	CPSS	Self	Sample 1 15–25; M = 19 SD = 2 Sample 2 M = 9.2 SD = N/A	40.1% 47%	Central America, mainly Honduras, El Salvador and Guatemala	Recently immigrated children and adolescents	United States
Ventevogel et al. (2014)	CPSS DSRS SCARED	Self Self Self	10–15; M = 12.8 SD = 1.3	45%	Burundian	Internally displaced	Burundi

Table 2. Results of measurement properties (structural validity, internal consistency, reliability).

Screening tool (reference)	Structural validity			Internal consistency			Reliability		
	N	Meth qual	Result (rating)	N	Meth qual	Result (rating)	N	Meth qual	Result (rating)
Behavioural and emotional problems									
CBCL (Bean et al., 2006a)	478	Adequate	Two-factor structure; CFI = 0.98 (+)	478	Very good	Total scale α = 0.94 (+) Internalising α = 0.89 (+) Externalising α = 0.90 (+)	478	Doubtful	Inter-rater r = 0.13–0.47 (?)
CBCL (Hall et al., 2014)	–	–	–	147	Very good	Internalising α = 0.92 (+) Externalising α = 0.93 (+)	147	Doubtful	Test–retest + inter-rater r = 0.54–0.60 (?)
YSR (Hall et al., 2014)	–	–	–	147	Very good	Internalising α = 0.95 (+) Externalising α = 0.92 (+)	147	Doubtful	Test–retest + inter-rater r = 0.34–0.38 (?)
TRF (Bean et al., 2007a)	461	Adequate	Two-factor structure; CFI = 0.98 (+)	461	Very good	Total scale α = 0.95 (+) Internalising α = 0.89 (+) Externalising α = 0.94 (+)	–	–	–
SDQ-P (McEwen et al., 2020)	1006	Adequate	Five-factor structure not supported; Seven-factor structure EV = 49.6% (?)	1006	Adequate	Total scale α = 0.76 (+) Emotional α = 0.66 (–) Conduct α = 0.48 (–) Hyperactivity α = 0.46 (–) Peer problem α = 0.26 (–) Prosocial α = 0.50 (–)	–	–	–
SDQ-P (Essex, 2019)	679	Adequate	Five-factor structure not supported; EV = 47,48% (?)	679	Adequate	Total scale α = 0.64 (–) Emotional α = 0.71 (+) Conduct α = 0.54 (–) Hyperactivity α = 0.60 (–) Peer problem α = 0.16 (–) Prosocial α = 0.72 (+)	–	–	–
SDQ-S (Essex, 2019)	402	Adequate	Five-factor structure not supported; EV = 42,09% (?)	402	Adequate	Total scale α = 0.66 (–) Emotional α = 0.70 (+) Conduct α = 0.47 (–) Hyperactivity α = 0.44 (–) Peer problem α = 0.29 (–) Prosocial α = 0.71 (+)	–	–	–
SDQ-T (Khawaja & Dhushyanthakumar, 2020)	175	Adequate	Five-factor structure not supported; a four-factor structure was proposed; EV = 54,86% (?)	175	Adequate	Prosocial α = 0.82 (+) Emotional α = 0.76 (+) Hyperactivity α = 0.85 (+) Behaviour α = 0.67 (–)	–	–	–
PTSD									
CATS (Müller et al., 2021)	145	Very good	Four-factor structure not supported; CFI = 0.86 (–)	145	Adequate	Total scale α = 0.84 (+) Intrusion α = 0.73 (+) Alterations in cognition and mood α = 0.66 (–) Avoidance α = 0.31 (–) Hyperarousal α = 0.59 (–)	–	–	–
CBCL-PTSD (Nehring et al., 2021) CBCL-PTSD adaptation	–	–	–	61	Doubtful	Total scale α = 0.79 (+) Total scale α = 0.89 (+)	–	–	–
CPSS-I (SR) (Hall et al., 2014)	–	–	–	147	Doubtful	Total scale α = 0.94 (+)	147	Doubtful	Test–retest + inter-rater r =.47 (?)

(*Continued*)

Table 2. Continued.

Screening tool (reference)	Structural validity			Internal consistency			Reliability		
	N	Meth qual	Result (rating)	N	Meth qual	Result (rating)	N	Meth qual	Result (rating)
CPSS-I (CR) (Hall et al., 2014)	–	–	–	147	Doubtful	Total scale α = 0.94 (+)	147	Doubtful	Test–retest + inter-rater r =.69 (?)
CPSS-5-SR (Hasson et al., 2021)	149	Adequate	Four-factor structure not supported; EV = 14.3% (?)	149	Adequate	Total scale α =.93 (+) Intrusion α = 0.86 (+) Changes in cognition and mood α = 0.84 (+) Avoidance α = 0.73 (+) Arousal and reactivity α = 0.69 (–)	–	–	–
CPSS (SR) (Kohrt et al., 2011)	–	–	–	162	Doubtful	Total scale α =.86 (+)	162	Doubtful	Test–retest r = .85 (?)
CPSS (SR) (McEwen et al., 2020)	1006	Adequate	Four-factor structure not supported; Two-factor structure EV = 59.8% One-factor also acceptable (?)	1006	Doubtful	Total scale α =.94 (+)	–	–	–
CPSS (SR + CR) (Marshall & Venta, 2021)	–	–	–	52	Adequate	Total scale α =.94 (+) Re-experiencing α =.80 (+) Avoidance α =.90 (+) Hyperarousal α =.78 (+)	–	–	–
CPSS (SR) (Venta & Mercado, 2019)	78	Inadequate	Three-factor structure not supported; EV = 64,54% (?)	78	Adequate	Total scale α =.90 (+) Re-experiencing α =.82 (+) Avoidance α =.81 (+) Hyperarousal α =.62 (–)	–	–	–
CPSS (CR) (Venta & Mercado, 2019)	103	Adequate	Three-factor structure not supported; Two-factor structure EV = 64.29%	103	Adequate	Total scale α =.95 (+) Re-experiencing α =.88 (+) Avoidance α =.89 (+) Hyperarousal α =.88 (+)			
CPSS (SR) (Ventevogel et al., 2014)	–	–	–	65	Adequate	Total scale α = 0.90 (+) Reexperiencing α = 0.84 (+) Avoidance α = 0.79 (+) Hyperarousal α = 0.77 (+)	–	–	–
PDS (Ertl et al., 2011)	–	–	–	504	Adequate	Total scale α = 0.89 (+) Reexperiencing α = 0.71 (+) Avoidance α = 0.78 (+) Hyperarousal α = 0.86 (+)	–	–	–
IES (Dyregrov et al., 1996)	1787	Adequate	Three-factor structure; EV = 47% Two-factor structure; EV = 39,8% (?)	1787	Very good	Total scale α = 0.79 (+) Intrusion α = 0.80 (+) Avoidance α = 0.73 (+) Emotional numbing α = −0.05 (–)	–	–	–
IES (Sack et al., 1998)	180	Very good	Three-factor structure; CFI = 0.987 (+)	180	Inadequate	Total scale α = 0.92 (+)	–	–	–
CRIES-8 (Salari et al., 2017)	201	Very good	Two-factor structure; CFI =.99 (+)	208	Very good	Total scale α = 0.76 (+) Intrusion α = 0.74 (+) Avoidance α = 0.65 (–)	–	–	–

(*Continued*)

Table 2. Continued.

Screening tool (reference)	Structural validity			Internal consistency			Reliability		
	N	Meth qual	Result (rating)	N	Meth qual	Result (rating)	N	Meth qual	Result (rating)
UCLA-PTSD RI (Elbert et al., 2009)	–	–	–	–	–	–	–	–	–
UCLA-PTSD RI (Ellis et al,. 2006)				76	Doubtful	Total scale α = 0.85 **(+)**			
HTQ/PTSS-16 (Jakobsen et al., 2017)	–	–	–	160	Doubtful	Total scale α = 0.89 **(+)**	–	–	–
RATS (Bean et al., 2006b)	3096	Adequate	Three-factor structure; EV = 49% **(?)**	3096	Very good	Total scale α = 0.91 **(+)** Intrusion α = 0.87 **(+)** Avoidance α = 0.81 **(+)** Hyperarousal α = 0.76 **(+)**	519	Inadequate	Test–retest r = 0.61 **(?)**
Depression and anxiety									
CES-DC-10 (McEwen et al., 2020)	1006	Adequate	One-factor structure; EV = 51% (?)	1006	Very good	Total scale α = 0.89 **(+)**	–	–	–
DSRS (Kohrt et al., 2011)	–	–	–	162	Doubtful	Total scale α = 0.67 **(–)**	162	Doubtful	Test–retest r = 0.80 **(?)**
DSRS (Ventevogel et al., 2014)	–	–	–	65	Doubtful	Total scale α = 0.85 **(+)**	–	–	–
DHSCL (Ertl et al., 2011)	–	–	–	504	Very good	Total scale α = 0.89 **(+)**	–	–	–
HSCL-25 (Jakobsen et al., 2017)	–	–	–	160	Doubtful	Total scale α = 0.94 **(+)**	–	–	–
HSCL-37A (Bean et al., 2007b)	3019	Adequate	Two-factor structure; EV = 33.1% **(?)**	3019	Very good	Total scale α = 0.90 **(+)** Internalising α = 0.92 **(+)** Externalising α = 0.75 **(+)**	519	Inadequate	Test–retest r = 0.63 **(?)**
HSCL-Y (Khawaja et al., 2019)	241	Adequate	One-factor structure; EV = 40% **(?)**	241	Very good	Total scale α = 0.91 **(+)**	–	–	–
PHQ-A (Al-Amer et al., 2020)	298	Very good	One-factor structure; CFI = 0.96 **(+)**	591	Very good	Total scale α = 0.82 **(+)**	–	–	–
SCARED-18 (McEwen et al., 2020)	1006	Adequate	Four-factor structure partially replicating original structure; EV = 53.5% **(?)**	1006	Very good	Total scale α = 0.84 **(+)** Panic/somatic α = 0.78 **(+)** Social α = 0.69 **(–)** Generalised α = 0.73 **(+)** Separation α = 0.52 **(–)**	–	–	–
SCARED-41 (Ventevogel et al., 2014)	–	–	–	65	Very good	Total scale α = 0.92 **(+)** Panic/somatic α = 0.86 **(+)** Social α = 0.76 **(+)** Generalised α = 0.71 **(+)** Separation α = 0.70 **(+)** School α = 0.49 **(–)**	–	–	–
Other									
CPDS (Jordans et al., 2008)	–	–	–	2240	Inadequate	Total scale α = 0.53 **(–)** w/probes α = 0.83 **(+)**	2240	Doubtful	Test–retest r = 0.71–0.83 **(?)**
CPDS (Jordans et al., 2009) *Burundi*	4193	Very good	Three-factor structure; RMSEA < .01 **(+)**						
CPDS (Jordans et al., 2009) *Indonesia*	1624	Very good	Three-factor structure; RMSEA < .00 **(+)**						
CPDS (Jordans et al., 2009) *Sri Lanka*	2573	Very good	Three-factor structure; RMSEA = 0.19 **(+)**						

(*Continued*)

Table 3. Results of measurement properties (criterion validity, hypotheses testing).

Screening tool (reference)	Criterion validity							Hypotheses testing		
	N	Meth quality	Criterion	Cut-off value	Result AUC (rating)	Result Sensitivity	Result Specificity	N	Meth quality	Result (rating)
Behavioural and emotional problems										
CBCL (Bean et al., 2006a)	–	–	–	–	–	–	–	478	Doubtful	Results in line with 7 hypotheses **(7+)** Result not in line with 5 hypotheses **(5−)**
CBCL (Hall et al., 2014)	159	Very good/ Adequate	Qualitative study	14 10	Internalising: AUC = 0.73 **(+)** Externalising: AUC = 0.70 **(+)**	65.82% 59.65%	65.79% 60.53%	–	–	–
YSR (Hall et al., 2014)	159	Very good/ Adequate	Qualitative study	11	Internalising: AUC = 0.70 **(+)** Externalising: Not stated **(?)**	68.35% Not stated	63.16% Not stated	–	–	–
TRF (Bean et al., 2007b)	–	–	–	–	–	–	–	461	Doubtful	Results in line with 5 hypotheses **(5+)**, results not in line with 7 hypotheses **(7−)**
SDQ-P (McEwen et al., 2020)	119	Very good	MINI KID CGI-s	17	AUC = 0.72 **(+)**	70%	66%	–	–	–
SDQ (Essex, 2019)	–	–	–	–	–	–	–	–	–	–
SDQ-T (Khawaja & Dhushyanthakumar, 2020)	–	–	–	–	–	–	–	–	–	–
PTSD										
CATS (Müller et al., 2021)	–	–	–	–	–	–	–	–	–	–
CBCL-PTSD (Nehring et al., 2021)	61	Very good	PTSDSSI Kinder-DIPS	5	AUC = 0.88 **(+)**	85%	76%	–	–	–
CBCL-PTSD adaptation				7	AUC = 0.86 **(+)**	85%	83%			
CPSS-I (Hall et al., 2014) Self-report	159	Very good/ Adequate	Qualitative study	13	AUC = 0.73 **(+)**	64.86%	81.58%	–	–	–
CPSS-I (Hall et al., 2014) Parent-report	159	Very good/ Adequate	Qualitative study	14	AUC = 0.74 **(+)**	71.62%	71.05%	–	–	–
CPSS-5-SR (Hasson et al., 2021)	–	–	–	–	–	–	–	–	–	–
CPSS (Kohrt et al., 2011)	162	Very good	K-SADS GAPD	20	AUC = 0.77 **(+)**	68%	73%	–	–	–
CPSS (McEwen et al., 2020)	119	Very good	MINI KID CGI-s	12	AUC = 0.70 **(+)**	83%	43%	–	–	–
CPSS-CR (Marshall & Venta, 2021)	–	–	–	–	–	–	–	52	Doubtful	Results in line with 1 hypothesis **(1+)**, results not in line with 2 hypotheses **(2−)**
CPSS (Venta & Mercado, 2019)	–	–	–	–	–	–	–	–	–	–
CPSS (Ventevogel et al., 2014)	65	Very good	K-SADS	26	AUC = 0.78 **(+)**	71%	83%	–	–	–

PDS (Ertl et al., 2011)	68	Very good	CAPS	16	AUC = 0.79 (+)	82% (7)	70% (7)	504	Adequate	Results in line with 2 hypotheses (2+), results not in line with 1 hypothesis (1−)
IES (Dyregrov et al., 1996)		–	–	–	–	–	–	1787	Adequate	Results not in line with 1 hypothesis (1−)
IES (Sack et al., 1998)	180	Very good	DICA	19	AUC = 0.69 (−)	66%	63%	–	–	–
CRIES-8 (Salari et al. 2017)	–	–	–	–	–	–	–	–	–	–
UCLA-PTSD Index (Elbert et al., 2009)	53	Inadequate	CIDI MINI	Not stated	Not stated (?)	62%	89%	350	Doubtful	Results in line with 1 hypothesis (1+)
UCLA-PTSD Index (Ellis et al., 2006)	–	–	–	–	–	–	–	76	Doubtful	Results in line with 1 hypothesis (1+), results not in line with 1 hypothesis (1−)
HTQ/PTSS-16 (Jakobsen et al., 2017)	160	Very good	CIDI	2.23	AUC = 0.75 (+)	80%	64%	–	–	–
RATS (Bean et al., 2006b)	–	–	–	–	–	–	–	3096	Adequate	Results in line with 9 hypotheses (9+) Results not in line with 2 hypotheses (2−)
Depression and anxiety										
CES-DC-10 (McEwen et al., 2020)	119	Very good	MINI KID CGI-s	10	AUC = 0.74 (+)	81%	56%	–	–	–
DSRS (Kohrt et al., 2011)	162	Very good	K-SADS GAPD	14	AUC = 0.82 (+)	71%	81%	–	–	–
DSRS (Ventevogel et al., 2014)	65	Very good	K-SADS	19	AUC = 0.85 (+)	64%	88%	–	–	–
DHSCL (Ertl et al., 2011)	68	Very good	MINI; depression section	2.65	AUC = 0.76 (+)	50%	83%	504	Adequate	Results in line with 4 hypotheses (4+), results not in line with 1 hypothesis (1−)
HSCL-25 (Jakobsen et al., 2017)	160	Very good	CIDI	2.17	Anxiety AUC = 0.81 (+)	92%	69%	–	–	–
				2.17	Depression AUC = 0.75 (+)	71%	66%			
HSCL-37A (Bean et al., 2007a)	–	–	–	–	–	–	–	3019	Adequate	Results in line with 16 hypotheses (16+), results not in line with 9 hypotheses (9−)
HSCL-Y (Khawaja et al., 2019)	–	–	–	–	–	–	–	241	Adequate	Results in line with 3 hypotheses (+3), results not in line with 1 hypothesis (1−)
PHQ-A (Al-Amer et al., 2020)	–	–	–	–	–	–	–	–	–	–
SCARED-18 (McEwen et al., 2020)	119	Very good	MINI KID CGI-s	12	AUC = 0.69 (−)	80%	53%	–	–	–
SCARED-41 (Ventevogel et al., 2014)	65	Very good	K-SADS	44	AUC = 0.69 (−)	55%	90%	–	–	–
Other										
CPDS (Jordans et al., 2008)	65	Very good/ Adequate	K-SADS	8	AUC = 0.81 (+)	84%	60%	–	–	–
CPDS (Jordans et al., 2009)	–	–	–	–	–	–	–	–	–	–
RHS-13 (Sarkadi et al., 2019)	–	–	–	–	–	–	–	29	Inadequate	Result in line with 2 hypotheses (2+)

3.3.2. CRIES

For the purpose of this review, we will only discuss the results of the CRIES below, and not the IES.

There is moderate evidence for sufficient structural validity of the CRIES self-report, since the two-factor structure has been confirmed by one study (Salari et al., 2017). There is also moderate evidence for internal consistency, with good internal consistency of the total scale and the subscale intrusion, but average internal consistency of the subscale avoidance (Salari et al., 2017). The criterion validity of the CRIES has not been assessed for the population of this review.

3.3.3. CATS

There is moderate evidence for insufficient structural validity of the CATS self-report, since the four-factor structure was not reproduced. The internal consistency of the subscales was inconsistent, ranging from poor to good. The internal consistency of the total scale is very good. The criterion validity of the CATS has not been assessed for the population of this review (Müller et al., 2021).

3.3.4. CBCL-PTSD

The structural validity was not assessed and there is low quality of evidence for good internal consistency of the total scale. However, there is moderate evidence for sufficient criterion validity, with very good sensitivity and good specificity (Nehring et al., 2021).

3.3.5. UCLA PTSD RI

The structural validity was not assessed and there is low evidence of a good internal consistency of the total scale, based on the DSM-IV (Ellis et al., 2006). Regarding the criterion validity, the quality of evidence is low and the results are indeterminate, since the AUC was not reported. Sensitivity was average and specificity was very good (Elbert et al., 2009).

3.3.6. HTQ

The structural validity was not assessed. Only one study reported on the internal consistency of the total scale. Thus, there is low quality of evidence for sufficient internal consistency of the instrument. Yet, there is moderate evidence for sufficient criterion validity, with very good sensitivity and average specificity (Jakobsen et al., 2017)

3.3.7. RATS

The results on the structural validity are indeterminate, since only an EFA was performed. There is moderate evidence supporting the three-factor structure of the RATS, with average explained variance. There is also moderate evidence for excellent internal consistency of the total scale and good to very good internal

consistency of the subscales. The criterion validity was not assessed (Bean et al., 2006b).

3.4. Depression and anxiety

3.4.1. CES-DC

The results for the structural validity are indeterminate, since only an EFA was performed. However, the study showed moderate evidence for a one-factor structure with good explained variance. There is moderate quality of evidence for very good internal consistency of the total scale score. With regard to the criterion validity, there is moderate quality of evidence for a sufficient AUC and very good sensitivity of the scale, but poor specificity (McEwen et al., 2020).

3.4.2. DSRS

No factor analysis was carried out. Two studies did report on the internal consistency of the total scale, with inconsistent results ranging from average to very good (Kohrt et al., 2011; Ventevogel et al., 2014). However, since no factor analysis was performed, the quality of evidence for the internal consistency is considered low. There is high quality of evidence for sufficient AUC, average to good sensitivity and very good specificity, with 14 or 19 as the recommended cut-off scores (Kohrt et al., 2011; Ventevogel et al., 2014).

3.4.3. PHQ-A

One study provided moderate evidence for sufficiency of the unidimensional scale and good internal consistency (Al-Amer et al., 2020). The criterion validity was not studied.

3.4.5. HSCL

Different versions of the Hopkins Symptoms Checklist [HSCL] were analyzed. One study assessed only the depression subscale of the HSCL [DHSCL] (Ertl et al., 2011). No studies were carried out on the structural validity of the HSCL-25 and the DHSCL. One study provided moderate evidence for the structural validity of the HSCL-37A. The results are indeterminate, since only a principal component analysis was performed. The study found a two-factor structure with the PCA, consisting of an internalising and externalising scale. The anxiety and depression subscales were not confirmed (Bean et al., 2007b). The structural validity of the HSCL-Y was only assessed with an EFA; hence, the results are considered indeterminate. There is moderate evidence for a one-factor structure of the instrument measuring psychological distress. The internal consistency of the DHSCL was very good (Ertl et al., 2011). The internal consistency of the HSCL-25 total scale was excellent (Jakobsen et al., 2017). However, since no factor analysis was performed and the internal consistency of the anxiety

and depression subscales was not reported, the quality of evidence is considered low. The internal consistency of the HSCL-Y total scale was excellent (Khawaja et al., 2019). The internal consistency of the HSCL-37 was excellent for both the total scale and the internalising scale, and good for the externalising scale (Bean et al., 2007b).

A study on the criterion validity of the DHSCL reported sufficient AUC and very good specificity, but poor sensitivity (Ertl et al., 2011). A study on the HSCL-25 showed sufficient AUC for both the anxiety and depression subscales. For the anxiety subscale, the sensitivity was excellent and the specificity was average. For the depression subscale, the sensitivity was good and the specificity was average (Jakobsen et al., 2017). The studies on the HSCL-37A and the HSCL-Y did not assess the criterion validity.

3.4.6. SCARED

The structural validity of the SCARED-41 was not analyzed. Since only an EFA was performed for the SCARED-18, the results are considered indeterminate. There is moderate evidence for a four-factor structure partially replicating the original structure, with good explained variance (McEwen et al., 2020). The internal consistency for the total score of the SCARED is very good to excellent. However, the internal consistency of the subscales ranges from poor to very good (McEwen et al., 2020; Ventevogel et al., 2014).

Regarding the criterion validity of the SCARED, both the SCARED-41 and the SCARED-18 showed insufficient AUC (McEwen et al., 2020; Ventevogel et al., 2014). The sensitivity was poor and the specificity excellent for the SCARED-41 (Ventevogel et al., 2014). Conversely, the sensitivity was very good and the specificity was poor for the SCARED-18 (McEwen et al., 2020). Thus, there is moderate evidence for insufficient criterion validity.

3.5. Others

3.5.1. CPDS

The structural validity of the CPDS was studied in four different countries and confirmed in three of them. In one of the countries, a good fit for the instrument was not found (Jordans et al., 2009). The internal consistency of the scale was poor, but when probe questions were added, the internal consistency increased to very good. However, the internal consistency of the three subscales was not reported, thus providing low quality of evidence for internal consistency (Jordans et al., 2008). Regarding the criterion validity of the CPDS, there is moderate quality of evidence for sufficient AUC, very good sensitivity and average specificity (Jordans et al., 2008).

3.5.2. RHS

No study on structural validity was performed. The internal consistency of the total scale is excellent according to one study. However, there is low quality of evidence due to a small sample size. The criterion validity was not studied (Sarkadi et al., 2019).

4. Discussion

The aim of this systematic review is to synthesise the existing evidence on psychometric properties of measurement instrument for assessing the mental health of asylum-seeking, refugee and internally displaced children and adolescents.

As noted by Gadeberg et al. (2017), internally displaced children have considerable similarities to refugee children and were therefore included in this review. The review by Gadeberg et al. (2017) was limited to studies reporting on criterion validity. We broadened the inclusion criteria by including studies reporting on any form of validity.

Based on the current evidence on psychometric properties of assessment tools for forcibly displaced children, we are not able to recommend a core set of questionnaires (Prinsen et al., 2016).

The idea that a core set of questionnaires can reliably and validly measure mental health in diverse populations can be contested (Gadeberg et al., 2017). However, PROMs are used widely and have added value in both scientific research and clinical practice (Fängström et al., 2019). Assessment tools can guide mental health screening by actively and explicitly addressing mental health issues, especially since mental health is a topic that is often not openly discussed in many cultures (Horlings & Hein, 2018). Furthermore, according to health care professionals, mental health measures can be useful tools to establish a more structured and informative overview of the mental health of forcibly displaced children (Fängström et al., 2019). Therefore, we provide suggestions based on available outcomes as well as feasibility. However, evidence is still limited and the results of the PROMs should be interpreted with caution. In a few studies the sample size was small or the methodological quality was insufficient because of other reasons, limiting the weight of the results.

Both the CBCL and SDQ are deemed acceptable instruments to screen for emotional and behavioural issues. Since the SDQ is brief and freely accessible in many languages, we suggest using this instrument for screening purposes. However, similar to previous studies, we recommend using the total problem score only, because the factor structure of the SDQ is not supported (McEwen et al., 2020; Stolk et al., 2017). Moreover, the SDQ does not measure any trauma- or stressor-related symptoms. Hence, with

forcibly displaced children, it is recommended to add an instrument measuring PTSD symptoms when assessing emotional and behavioural problems (Stolk et al., 2017). Furthermore, a follow-up interview is needed since problems may be over- or under-stated. Previous research has questioned the comprehensibility of the idioms of distress of the SDQ among children from different cultural backgrounds (Derluyn & Broekaert, 2007; Stolk et al., 2017). Copyright restrictions of the SDQ limit adaptations of the instrument to improve the reliability and validity in different populations (McEwen et al., 2020). Results should thus be interpreted with caution.

With regard to measuring PTSD symptoms, several instruments were examined. The results are inconclusive. However, we propose using the CRIES-8 for PTSD screening based on the current evidence for the structural validity and internal consistency of the scale. The CRIES is brief and widely implemented worldwide. The CRIES is also recommended as a measurement instrument for PTSD by the International Consortium for Health Outcomes Measurement [ICHOM] (Krause et al., 2021). It would be a great asset to have the CRIES available for children under eight years of age, especially since the DSM-5 has additional criteria for PTSD for children six years and younger. When an instrument is needed to provide preliminary diagnostic information on PTSD, we currently recommend using the CATS. The CATS is based on the DSM-5 whereas the CRIES is based on the DSM-IV. The CATS is an instrument with similar characteristics to the CPSS, which has been more widely studied and implemented. However, the CATS is also available for young children under eight years of age. Both the CRIES and the CATS are available for free in different languages relevant to forcibly displaced children and adolescents. However, more research is needed on the psychometric properties of these instruments in different forcibly displaced children and youth populations.

The results on instruments assessing depression and anxiety were also inconclusive. We currently recommend using the DSRS, because there is good evidence for sufficient criterion validity of the instrument. The different versions of the HSCL are also showing promising results in their usage as screening instruments to measure psychological distress. However, the separate scores on anxiety and depression items cannot be interpreted reliably. Besides, the different versions of the HSCL are only available for adolescents. The ICHOM (Krause et al., 2021) has recommended the use of the Revised Children's Anxiety and Depression Scale [RCADS] (Chorpita et al., 2000) to measure anxiety and depression symptoms. The RCADS is available as a parent-report and self-report for children and adolescents between eight and 18 years of age. The instrument has a long version consisting of 47 items and a short version consisting of 25 items. Research should address the psychometric properties of the RCADS for forcibly displaced children. These questionnaires are all freely available. It would be useful to have these questionnaires on depression and anxiety also available for children under eight years of age.

It would be highly beneficial for screening purposes and clinical practice to have a standard set of instruments that are of sufficient quality, freely available and translated into a variety of languages. Furthermore, the use of several different instruments in research can restrict the possibility of comparing interventions and treatments (Krause et al., 2021).

Moreover, solely performing translations and back-translations of instruments is not enough to capture cultural differences in mental health measurement instruments (Kohrt et al., 2011). This is also highlighted by this review, since the original factor structure of PROMs is often not replicated. A careful process of transcultural translation is necessary to achieve the semantic equivalence of questionnaires (Kohrt et al., 2011). Semantic equivalence entails that the meaning of each item is the same in each culture after translation into the language and idiom (written or oral) of each culture (Flaherty et al., 1988). Additionally, the recommended cut-off scores varied greatly, showing that not one single cut-off score can be used for different populations. Since time and effort are needed to cross-culturally adapt instruments, examine reliability and validity, as well as establish cut-off scores for different populations, focusing on a limited number of mental health measurement instruments would be beneficial.

Another interesting result is the high correlation between instruments measuring PTSD symptoms and instruments measuring anxiety and depression symptoms. Several hypotheses were rated as insufficient due to correlation scores above the selected criteria. However, this could also be an indication of overlapping symptoms between PTSD and other internalising problems. Moreover, it could demonstrate high comorbidity of PTSD and anxiety or depression in forcibly displaced children.

Lastly, a severe lack of research on content validity was identified. Content validity is one of the most important psychometric properties (Mokkink et al., 2010; Prinsen et al., 2016; Terwee et al., 2018). In order for a PROM to measure what it intends to measure, the content should have the same meaning across cultures. However, in different cultural context the meaning, clustering, and experience of symptoms may differ (Kohrt et al., 2011). Conducting qualitative research to address this issue is highly recommended. Qualitative research could help in gaining more insight into conceptions and experiences of mental health in forcibly displaced children, which can be of

assistance in developing or adapting culturally sensitive tools (Gadeberg et al., 2017).

5. Limitations

The majority of the study selection, data extraction and quality assessment was carried out by only one member of the review team (IV); this could have caused more errors and bias. However, IV regularly consulted with IH and MN. Due to time constraints and the large number of different instruments, we were not able to completely follow the COSMIN methodology for assessing the content validity of the measurement instruments. Moreover, interpretability and feasibility were not always described in depth. Several studies were identified on the reliability and validity of mental health measurement instruments for children and adolescents living in conflict areas or on the cross-cultural validity of mental health outcome measures for children and adolescents; these studies could be relevant for the populations of this review. However, this was beyond the scope of the current review. The search did not include databases from different continents, such as Latin America and Africa, which could have resulted in the inclusion of more studies. This could partially explain the lack of studies found in languages other than English, of which none met the inclusion criteria.

6. Conclusion

There is a lack of studies conducted on the reliability and validity of mental health measurement instruments for forcibly displaced children and youth, despite a call for more research on this topic. More research of sufficient quality is needed in order to establish cross-cultural validity and to provide optimal cut-off scores for this population. COSMIN provides useful guidelines to improve the quality of research on measurement properties. Special attention should be paid to studying the content validity of mental health measurement instruments utilised with forcibly displaced children and youth. Moreover, there is a scarcity of mental health assessment tools for younger children. Encouragingly, research on the psychometric properties of mental health screening and assessment tools for different populations of children and adolescents seems to be steadily growing.

Ethics statement

Institutional review board approval and informed consent are not applicable to this article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

Data sharing is not applicable to this article as no new data were created in this study.

References

Achenbach, T. M. (1991). *Manual for the Teacher's Report Form and 1991 Profile*. University of Vermont.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-age Forms & Profiles*. University of Vermont, Research Center for Children, Youth & Families.

Al-Amer, R., Maneze, D., Ramjan, L., Villarosa, A. R., Darwish, R., & Salamonson, Y. (2020). Psychometric testing of the Arabic version of the Patient Health Questionnaire among adolescent refugees living in Jordan. *International Journal of Mental Health Nursing*, *29*(4), 685–692. https://doi.org/10.1111/inm.12702

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). American Psychiatric Association.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.

Bean, T., Derluyn, I., Eurelings-Bontekoe, E., Broekaert, E., & Spinhoven, P. (2006b). Validation of the multiple language versions of the reactions of adolescents to traumatic stress questionnaire. *Journal of Traumatic Stress*, *19* (2), 241–255. https://doi.org/10.1002/jts.20093

Bean, T., Derluyn, I., Eurelings-Bontekoe, E., Broekaert, E., & Spinhoven, P. (2007b). Validation of the multiple language versions of the Hopkins Symptom Checklist-37 for refugee adolescents. *Adolescence*, *42*(165), 51–71.

Bean, T. M., Eurelings-Bontekoe, E., & Spinhoven, P. (2007a). Course and predictors of mental health of unaccompanied refugee minors in The Netherlands: One year follow-up. *Social Science & Medicine*, *64*(6), 1204–1215. https://doi.org/10.1016/j.socscimed.2006.11.010

Bean, T., Mooijart, A., Eurelings-Bontekoe, E., & Spinhoven, P. (2006a). Validation of the Child Behavior Checklist for guardians of unaccompanied refugee minors. *Children and Youth Services Review*, *28*(8), 867–887. https://doi.org/10.1016/j.childyouth.2005.09.002

Bean, T., Mooijart, A., Eurelings-Bontekoe, E., & Spinhoven, P. (2007c). Validation of the Teacher's Report Form for teachers of unaccompanied refugee minors. *Journal of Psychoeducational Assessment*, *25*(1), 53–68. https://doi.org/10.1177/0734282906293688

Birleson, P. (1981). The validity of depressive disorder in childhood and the development of a self-rating scale: A research report. *Journal of Child Psychology and Psychiatry*, *22*(1), 73–88. https://doi.org/10.1111/j.1469-7610.1981.tb00533.x

Birleson, P., Hudson, I., Buchanan, D. G., & Wolff, S. (1987). Clinical evaluation of a self-rating scale for depressive disorder in childhood (depression self-rating scale). *Journal of Child Psychology and Psychiatry*, *28*(1), 43–60. https://doi.org/10.1111/j.1469-7610.1987.tb00651.x

Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997). The screen for child anxiety related emotional disorders (SCARED): scale construction and psychometric characteristics. *Journal*

of the *American Academy of Child & Adolescent Psychiatry*, *36*(4), 545–553. https://doi.org/10.1097/00004583-199704000-00018

Blackmore, R., Gray, K. M., Boyle, J. A., Fazel, M., Ranasinha, S., Fitzgerald, G., Misso, M., & Gibson-Helm, M. (2020). Systematic review and meta-analysis: The prevalence of mental illness in child and adolescent refugees and asylum seekers. *Journal of the American Academy of Child & Adolescent Psychiatry*, *59*(6), 705–714. https://doi.org/10.1016/j.jaac.2019.11.011

Children and War Foundation. (1998). *The Children's Revised Impact of Event Scale (8): CRIES-8.* https://www.childrenandwar.org/.

Chorpita, B. F., Yim, L. M., Moffitt, C. E., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, *38*(8), 835–855. https://doi.org/10.1016/S0005-7967(99)00130-8

Derluyn, I., & Broekaert, E. (2007). Different perspectives on emotional and behavioural problems in unaccompanied refugee children and adolescents. *Ethnicity and Health*, *12*(2), 141–162. https://doi.org/10.1080/13557850601002296

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science*, *19*(1), 1–15. https://doi.org/10.1002/bs.3830190102

Dyregrov, A., Gjestad, R., & Raundalen, M. (2002). Children exposed to warfare: A longitudinal study. *Journal of Traumatic Stress*, *15*(1), 59–68. https://doi.org/10.1023/A:1014335312219

Dyregrov, A., Kuterovac, G., & Barath, A. (1996). Factor analysis of the Impact of Event Scale with children in war. *Scandinavian Journal of Psychology*, *37*(4), 339–350. https://doi.org/10.1111/j.1467-9450.1996.tb00667.x

Ehntholt, K. A., & Yule, W. (2006). Practitioner review: Assessment and treatment of refugee children and adolescents who have experienced war-related trauma. *Journal of Child Psychology and Psychiatry*, *47*(12), 1197–1210. https://doi.org/10.1111/j.1469-7610.2006.01638.x

Elbert, T., Schauer, M., Schauer, E., Huschka, B., Hirth, M., & Neuner, F. (2009). Trauma-related impairment in children: A survey in Sri Lankan provinces affected by armed conflict. *Child Abuse & Neglect*, *33*(4), 238–246. https://doi.org/10.1016/j.chiabu.2008.02.008

Ellis, H. B., Lhewa, D., Charyney, M., & Cabral, H. (2006). Screening for PTSD among Somali adolescent refugees: Psychometric properties of the UCLA PTSD Index. *Journal of Traumatic Stress*, *19*(4), 547–551. https://doi.org/10.1002/jts.20139

Ertl, V., Pfeiffer, A., Saile, R., Schauer, E., Elbert, T., & Neuner, F. (2011). Validation of a mental health assessment in an african conflict population. *International Perspectives in Psychology: Research, Practice, Consultation*, *1*(Supplement 1), 19–27. https://doi.org/10.1037/2157-3883.1.S.19

Essex, R. (2019). The psychometric properties of the Strengths and Difficulties Questionnaire for children from refugee backgrounds in Australia. *Clinical Psychologist*, *23*(3), 261–270. https://doi.org/10.1111/cp.12178

Faulstich, M., Carey, M., Ruggiero, L., Enyart, P., & Gresham, F. (1986). Assessment of depression in childhood and adolescent: An evaluation of the Center for Epidemiological Studies Depression Scale for Children (CES-DC). *American Journal of Psychiatry*, *143*(8), 1024–1027. https://doi.org/10.1176/ajp.143.8.1024

Fazel, M., & Betancourt, T. S. (2018). Preventive mental health interventions for refugee children and adolescents in high-income settings. *Lancet Child & Adolescent Health*, *2*(2), 121–132. https://doi.org/10.1016/S2352-4642(17)30147-5

Fazel, M., Reed, R. V., Panter-Brick, C., & Stein, A. (2012). Mental health of displaced and refugee children resettled in high-income countries: Risk and protective factors. *Lancet*, *379*(9812), 266–282. https://doi.org/10.1016/S0140-6736(11)60051-2

Fazel, M., & Stein, A. (2002). The mental health of refugee children. *Archives of Disease in Childhood*, *87*(5), 366–370. https://doi.org/10.1136/adc.87.5.366

Fängström, K., Dahlberg, A., Ådahl, K., Rask, H., Salari, R., Sarkadi, A., & Durbeej, N. (2019). Is the strengths and difficulties questionnaire with a trauma supplement a valuable tool in screening refugee children for mental health problems? *Journal of Refugee Studies*, *32*(1), 122–140. https://doi.org/10.1093/jrs/fey073

Flaherty, J. A., Gaviria, F. M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J. A., & Birz, S. (1988). Developing instruments for cross-cultural psychiatric research. *Journal of Nervous & Mental Disease*, *176*(5), 260–263. https://doi.org/10.1097/00005053-198805000-00001

Foa, E. B., Asnaani, A., Zang, Y., Capaldi, S., & Yeh, R. (2018). Psychometrics of the Child PTSD Symptom Scale for DSM-5 for trauma-exposed children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *47*(1), 38–46. https://doi.org/10.1080/15374416.2017.1350962

Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment*, *9*(4), 445–451. https://doi.org/10.1037/1040-3590.9.4.445

Foa, E. B., Johnson, K. M., Feeny, N. C., & Treadwell, K. R. H. (2001). The Child PTSD Symptom Scale: A preliminary examination of its psychometric properties. *Journal of Clinical Child Psychology*, *30*(3), 376–384. https://doi.org/10.1207/S15374424JCCP3003_9

Gadeberg, A. K., Montgomery, E., Frederiksen, H. W., & Norredam, M. (2017). Assessing trauma and mental health in refugee children and youth: A systematic review of validated screening and measurement tools. *European Journal of Public Health*, *27*(3), 439–446. https://doi.org/10.1093/eurpub/ckx034

Gadeberg, A. K., & Norredam, M. (2016). Urgent need for validated trauma and mental health screening tools for refugee children and youth. *European Child & Adolescent Psychiatry*, *25*(8), 929–931. https://doi.org/10.1007/s00787-016-0837-2

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry*, *9*(2), 129–134. https://doi.org/10.1007/s007870050008

Hall, B. J., Puffer, E., Murray, L. K., Ismael, A., Bass, J. K., Sim, A., & Bolton, P. A. (2014). The importance of establishing reliability and validity of assessment instruments

for mental health problems: An example from Somali children and adolescents living in three refugee camps in Ethiopia. *Psychological Injury and Law*, 7(2), 153–164. https://doi.org/10.1007/s12207-014-9188-9

Hasson IIIR. G., Easton, S. D., Díaz-Valdés Iriarte, A., O'Dwyer, L. M., Underwood, D., & Crea, T. M. (2021). Examining the psychometric properties of the Child PTSD Symptom Scale within a sample of unaccompanied immigrant children in the United States. *Journal of Loss and Trauma*, 26(4), 323–335. https://doi.org/10.1080/15325024.2020.1777760

Heptinstall, E., Sethna, V., & Taylor, E. (2004). PTSD and depression in refugee children. *European Child & Adolescent Psychiatry*, 13(6), 373–380. https://doi.org/10.1007/s00787-004-0422-y

Hodes, M. (2019). New developments in the mental health of refugee children and adolescents. *Evidence Based Mental Health*, 22(2), 72–76. https://doi.org/10.1136/ebmental-2018-300065

Hoge, C. W., Auchterlonie, J. L., & Milliken, C. S. (2006). Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. *JAMA*, 295(9), 1023–1032. https://doi.org/10.1001/jama.295.9.1023

Hollifield, M., Toolson, E. C., Verbillis-Kolp, S., Farmer, B., Yamazaki, J., Woldehaimanot, T., & Holland, A. (2016). Effective screening for emotional distress in refugees: The Refugee Health Screener. *Journal of Nervous and Mental Disease*, 204(4), 247–253. https://doi.org/10.1097/NMD.0000000000000469

Hollifield, M., Verbillis-Kolp, S., Farmer, B., Toolson, E. C., Woldehaimanot, T., Yamazaki, J., Holland, A., St. Clair, J., & Soohoo, J. (2013). The Refugee Health Screener-15 (RHS-15): development and validation of an instrument for anxiety, depression, and PTSD in refugees. *General Hospital Psychiatry*, 35(2), 202–209. https://doi.org/10.1016/j.genhosppsych.2012.12.002

Horlings, A., & Hein, I. (2018). Psychiatric screening and interventions for minor refugees in Europe: An overview of approaches and tools. *European Journal of Pediatrics*, 177(2), 163–169. https://doi.org/10.1007/s00431-017-3027-4

Horowitz, M., Wilner, N. J., & Alverez, W. (1979). Impact of events scale: A measure of subjective stress. *Psychosomatic Medicine*, 41(3), 209–218. https://doi.org/10.1097/00006842-197905000-00004

Jakobsen, M., Meyer DeMott, M. A., & Heir, T. (2017). Validity of screening for psychiatric disorders in unaccompanied minor asylum seekers: Use of computer-based assessment. *Transcultural Psychiatry*, 54(5-6), 611–625. https://doi.org/10.1177/1363461517722868

Johnson, J. G., Harris, E. S., Spitzer, R. L., & Williams, J. B. (2002). The Patient Health Questionnaire for Adolescents: Validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *Journal of Adolescent Health*, 30(3), 196–204. https://doi.org/10.1016/s1054-139x(01)00333-0

Jordans, M. J. D., Komproe, I. H., Tol, W. A., & De Jong, J. T. V. M. (2009). Screening for psychosocial distress amongst war-affected children: Cross-cultural construct validity of the CPDS. *Journal of Child Psychology and Psychiatry*, 50(4), 514–523. https://doi.org/10.1111/j.1469-7610.2008.02028.x

Jordans, M. J. D., Komproe, I. H., Ventevogel, P., Tol, W. A., & de Jong, J. T. V. M. (2008). Development and validation of the Child Psychosocial Distress Screener in Burundi. *American Journal of Orthopsychiatry*, 78(3), 290–299. https://doi.org/10.1037/a0014216

Khawaja, N. G., & Dhushyanthakumar, L. (2020). Strengths and Difficulties Questionnaire-teacher: Investigating its factor structure and utility with culturally and linguistically diverse students. *Journal of Psychologists and Counsellors in Schools*, 30(1), 43–57. https://doi.org/10.1017/jgc.2019.23

Khawaja, N. G., Pekin, C., & Schweitzer, R. D. (2019). Factor structure and psychometric properties of the Hopkins Symptom Checklist: An investigation with culturally and linguistically diverse youth in Australia. *Australian Journal of Psychology*, 71(2), 137–145. https://doi.org/10.1111/ajpy.12221

Kohrt, B. A., Jordans, M. J., Tol, W. A., Luitel, N. P., Maharjan, S. M., & Upadhaya, N. (2011). Validation of cross-cultural child mental health and psychosocial research instruments: Adapting the Depression Self-Rating Scale and Child PTSD Symptom Scale in Nepal. *BMC Psychiatry*, 11(127), 1–17. https://doi.org/10.1186/1471-244X-11-127

Krause, K. R., Chung, S., Adewuya, A. O., Albano, A. M., Babins-Wagner, R., Birkinshaw, L., Brann, P., Creswell, C., Delaney, K., Falissard, B., Forrest, C. B., Hudson, J. L., Ishikawa, S., Khatwani, M., Kieling, C., Krause, J., Malik, K., Martínez, V., Mughal, F., … Wolpert, M. (2021). International consensus on a standard set of outcome measures for child and youth anxiety, depression, obsessive-compulsive disorder, and post-traumatic stress disorder. *Lancet Psychiatry*, 8(1), 76–86. https://doi.org/10.1016/S2215-0366(20)30356-4

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lustig, S. L., Kia-Keating, M., Knight, W. G., Geltman, P., Ellis, H., Kinzie, J. D., Keane, T., & Saxe, G. N. (2004). Review of child and adolescent refugee mental health. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(1), 24–36. https://doi.org/10.1097/00004583-200401000-00012

Marshall, K., & Venta, A. (2021). Psychometric evaluation of the caregiver version of the Child PTSD Symptom Scale in a recently immigrated Spanish speaking population. *Psychiatry Research*, 301, 1–6. https://doi.org/10.1016/j.psychres.2021.113954

McEwen, F. S., Moghames, P., Bosqui, T., Kyrillos, V., Chehade, N., Saad, S., Abdul Rahman, D., Popham, C., Saab, D., Karam, G., Karam, E., & Pluess, M. (2020). *Validating screening questionnaires for internalizing and externalizing disorders against clinical interviews in 8 to 17-year-old Syrian Refugee Children*. Technical Working Paper. London: Queen Mary University of London. https://doi.org/10.31234/osf.io/6zu87.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & the PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

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

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, 19 (4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Mollica, R. F., Caspi-Yavin, Y., Bollini, P., Truong, T., Tor, S., & Lavelle, J. (1992). The Harvard Trauma Questionnaire: Validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. *Journal of Nervous and Mental Disease*, 180(2), 111–116. https://doi.org/10.1097/00005053-199202000-00008

Müller, L. R. F., Unterhitzenberger, J., Wintersohl, S., Rosner, R., & König, J. (2021). Screening for posttraumatic stress symptoms in young refugees: Comparison of questionnaire data with and without involvement of an interpreter. *International Journal of Environmental Research and Public Health*, 18(13), 6803–6809. https://doi.org/10.3390/ijerph18136803

Nehring, I., Sattel, H., Al-Hallak, M., Henningsen, P., Mall, V., & Aberl, S. (2021). The Child Behavior Checklist as a screening instrument for PTSD in refugee children. *Children*, 8(6), 521–530. https://doi.org/10.3390/children8060521

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan – a web and mobile app for systematic reviews. *Systematic Reviews*, 5(210), 1–10. https://doi.org/10.1186/s13643-016-0384-4

Porter, M., & Haslam, N. (2005). Predisplacement and postdisplacement factors associated with mental health of refugees and internally displaced persons: A meta-analysis. *JAMA*, 295(5), 602–612. https://doi.org/10.1001/jama.294.5.602

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & 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

Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "core outcome Set" – a practical guideline. *Trials*, 17(1), 449. https://doi.org/10.1186/s13063-016-1555-2

Pynoos, R., Rodriguez, N., Steinberg, A., Stuber, M., & Frederick, C. (1998). *UCLA PTSD Index for DSM-IV: Adolescent Version*. University of California at Los Angeles Trauma Psychiatry Service.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385–401. https://doi.org/10.1177/014662167700100306

Reed, R. V., Fazel, M., Jones, L., Panter-Brick, C., & Stein, A. (2012). Mental health of displaced and refugee children resettled in low-income and middle-income countries: Risk and protective factors. *Lancet*, 379(9812), 250–265. https://doi.org/10.1016/S0140-6736(11)60050-0

Sachser, C., Berliner, L., Holt, T., Jensen, T. K., Jungbluth, N., Risch, E., Rosner, R., & Goldbeck, L. (2017). International development and psychometric properties of the child and adolescent trauma screen (CATS). *Journal of Affective Disorders*, 210, 189–195. https://doi.org/10.1016/j.jad.2016.12.040

Sack, W. H., Seeley, J. R., Him, C., & Clarke, G. N. (1998). Psychometric properties of the Impact of Events Scale in traumatized Cambodian refugee youth. *Personality*

and Individual Differences, 25(1), 57–67. https://doi.org/10.1016/S0191-8869(98)00030-0

Salari, R., Malekian, C., Linck, L., Kristiansson, R., & Sarkadi, A. (2017). Screening for PTSD symptoms in unaccompanied refugee minors: A test of the CRIES-8 questionnaire in routine care. *Scandinavian Journal of Public Health*, 45 (6), 605–611. https://doi.org/10.1177/1403494817715516

Sarkadi, A., Bjärtå, A., Leiler, A., & Salari, R. (2019). Is the Refugee Health Screener a useful tool when screening 14- to 18-year-old refugee adolescents for emotional distress? *Journal of Refugee Studies*, 32(1), 141–150. https://doi.org/10.1093/jrs/fey072

Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA*, 282(18), 1737–1744. https://doi.org/10.1001/jama.282.18.1737

Stolk, Y., Kaplan, I., & Szwarc, J. (2017). Review of the Strengths and Difficulties Questionnaire translated into languages spoken by children and adolescents of refugee background. *International Journal of Methods in Psychiatric Research*, 26 (4), e1568. https://doi.org/10.1002/mpr.1568

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Terwee, C. B., Prinsen, C. A. C., 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

United Nations High Commissioner for Refugees [UNHRC]. (2021). *Global trends: Forced displacement in 2020*. https://www.unhcr.org/flagship-reports/globaltrends/.

Van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, 3(2), 125–133. https://doi.org/10.1038/s42256-020-00287-7

Venta, A. C., & Mercado, A. (2019). Trauma screening in recently immigrated youth: Data from two Spanish-speaking samples. *Journal of Child and Family Studies*, 28(1), 84–90. https://doi.org/10.1007/s10826-018-1252-8

Ventevogel, P., Komproe, I. H., Jordans, M. J., Feo, P., & De Jong, J. T. V. M. (2014). Validation of the Kirundi versions of brief self-rating scales for common mental disorders among children in Burundi. *BMC Psychiatry*, 14 (1), 36. https://doi.org/10.1186/1471-244X-14-36

Vervliet, M., Lammertyn, J., Broekaert, E., & Derluyn, I. (2014). Longitudinal follow-up of the mental health of unaccompanied refugee minors. *European Child & Adolescent Psychiatry*, 23(5), 337–346. https://doi.org/10.1007/s00787-013-0463-1

Weissman, M. M., Orvaschel, H., & Padian, N. (1980). Children's symptom and social functioning self-report scales. Comparison of mothers' and children's reports. *Journal of Nervous and Mental Disease*, 168(12), 736–740. https://doi.org/10.1097/00005053-198012000-00005

Wolfe, V. V., Gentile, C., & Wolfe, D. A. (1989). The impact of sexual abuse on children: A PTSD formulation. *Behavior Therapy*, 20(2), 215–228. https://doi.org/10.1016/S0005-7894(89)80070-X