

# What Measures are Effective in Trauma Screening for Young Males in Custody? A COSMIN Systematic Review

TRAUMA, VIOLENCE, &amp; ABUSE

1–14

© The Author(s) 2023



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/15248380231219251

[journals.sagepub.com/home/tva](https://journals.sagepub.com/home/tva)

Rachel O'Rourke<sup>1</sup>, Mike Marriott<sup>2</sup> , and Richard Trigg<sup>2</sup>

## Abstract

Despite the available evidence identifying the high prevalence rates of potentially traumatic experiences in forensic populations, there is still a lack of evidence supporting the use of suitable assessment tools, especially for young males in custody. For services to identify, support, and offer trauma interventions to this cohort, practitioners require reliable and valid assessment tools. This systematic review (Open Science Framework registration: <https://osf.io/r6hbk>) identifies those tools able to provide valid, reliable, and comparable data for this cohort. Five electronic databases and gray literature were searched to identify relevant measures. Inclusion criteria: studies of tools to assess for trauma with males aged between 12 and 25 years-old in a custodial setting, any year of publication, and available in English. Exclusion criteria: studies that did not measure psychological trauma or include a standalone trauma scale, or report primary data. A three-step quality assessment method was used to evaluate the methodological quality and psychometric properties of the measures. Fourteen studies were selected for review (which included 12 measures). The studies sampled a total of approximately 1,768 male participants and an age range of 12 to 25 years. The studies reported on various types of psychometric evidence and due to the lack of homogeneity, a narrative synthesis was used to discuss, interpret, and evaluate each measure. The overall quality of the psychometric properties of the measures in this review showed that the currently available instruments for the assessment of trauma with young males in custody is limited but promising.

## Keywords

trauma, assessment, young men, prison, systematic review

In recent years, a large amount of research has documented that young people exposed to multiple forms of adversity are more likely to evidence higher levels of delinquent and externalizing behavior during adolescence than those exposed to less adversity (Connolly & Kavish, 2019; Kretschmar et al., 2017). As a result, many researchers argue that adversity experienced during childhood may exert a long-term influence on developmental patterns of offending across adolescent development (Bonner et al., 2020; Farrell & Zimmerman, 2017), thus making childhood adversity a potential causal risk factor for adolescent delinquency. Previous studies have also demonstrated what appears to be a “dose–response” relationship between the number of times a young person was exposed to trauma and the number of later difficulties (Copeland et al., 2007; Hodges et al., 2013). Therefore, there is good evidence to support the expectation that the negative effects of trauma may be amplified in the context of multiple traumas. However, a recent study by Daniunaite et al. (2021) found that this was not always the case, and that perhaps it is the type of trauma (e.g., interpersonal trauma) that is most likely to leave its mark. Most studies examine the impact of child maltreatment adverse

childhood experiences, particularly childhood physical abuse, sexual abuse, and exposure to intimate partner violence on adult outcomes. However, less is known about the relationship between other types of childhood maltreatment, for example, emotional maltreatment or neglect on violence in adulthood.

The physical and emotional consequences of childhood abuse and neglect appear to have a profound effect on development through childhood, adolescence, and adulthood, with a range of outcomes which go some way to explain criminal justice system (CJS) involvement for many. Young men involved in the CJS represent one of the most pervasively traumatized populations. Research evidence demonstrates effects on emotional regulation, unpredictable behavior, and a lack of trust and connection within

<sup>1</sup>His Majesty's Prison and Probation Service, Milton Keynes, UK

<sup>2</sup>Nottingham Trent University, UK

## Corresponding Author:

Mike Marriott, NTU Psychology, Nottingham Trent University, 50 Shakespeare Street, Nottingham NG1 4FQ, UK.

Email: [mike.marriott@ntu.ac.uk](mailto:mike.marriott@ntu.ac.uk)

relationships (Webermann & Murphy, 2019). Wright and Little's review of the key research (2016) also showed how early child maltreatment affected emotional control, finding key developmental differences in young adults with a trauma history, for example behaving recklessly and reacting aggressively to provocation.

Therefore, effective and accurate assessment of trauma is needed with young men in custody. Trauma experienced during childhood and adolescence is significantly impactful in how it can elicit changes that may negatively impact an individual's development, changing the neurobiology of the stress response systems (Agorastos et al., 2018) and increasing later risk for psychopathologies (Pechtel & Pizzagalli, 2011). In various studies, a direct relationship has been shown between the severity of the trauma and the impact on the individual (Ben-Zion et al., 2019). Post-Traumatic Stress Disorder (PTSD) symptoms have been described in children who have survived sexual, physical abuse, and other traumatic events, indicating the impact of cumulative negative childhood experiences. It is therefore important to this field that assessment can take account of a range of traumatic experiences, expressions of difficulties, and the ways in which young males express those difficulties.

Despite early identification being a key component of preventing or mitigating the effects of traumatic experiences (Issakidis et al., 2004), practitioners are hampered by the lack of psychometrically sound trauma screening methods when working with young people (Eklund et al., 2018). The theoretical underpinnings to explain the development of a stress disorder in traumatized adolescents are also limited, compared with adults and PTSD. There is growing agreement that the lack of understanding so far is attributable to the adolescent expression of trauma exposure being more complex than the presentation of symptoms as defined by PTSD. There is some prevalence rate research (Chitsabesan et al., 2006; Moore et al., 2013) specifically targeted at young male offenders, looking at both how trauma presents and best practice in working with this cohort, although limitations are apparent. These research studies have often used mixed age groups in their methodology, studying ages across different developmental stages together and with much of the published material focused on the younger age ranges. It is also the case that more research has taken place with young offenders on probation or remand than those in custody or serving longer sentences (Cruise et al., 2008; Rogers, 1994).

There is a lack of research and evidence on what would count as trauma-specific assessment for males aged 12 to 25 years-old in the UK justice system. Arguably, there is not a direct transfer between assessing trauma with non-offending or community-based young males to those in custody. Not only is it clear that the latter group report higher rates of child maltreatment (Lader et al., 2000), but their experiences may also vary in that they have been exposed to and/or witnessed or perpetrated significant harm on others, often involving weapons and resulting in injury or death. However, as the

concept of trauma-informed services and practice becomes more popular in health and custodial services (McCartan, 2020), an evidence-based understanding of trauma assessment with this cohort is necessary.

As the clinical understanding of trauma experiences and expression have developed, a range of evaluation tools have been developed within the clinical sphere. As such, a systematic review is required to allow practitioners to make a more evidence-based decision in their selection and use of such assessment measures and tools. An agreed focus for consideration of research into the use of measures is the need to establish the level of common psychometric properties such as reliability and validity in such tools. Both domains must be established for a tool to be considered robust and fit for purpose, and both break down into various measurement properties for full evaluation purposes. Without evidence of a psychometric tool's reliability and validity, it is not advisable to adopt the tool in practice. It is critical therefore that the empirical literature is informed by studies offering robust evidence around the most appropriate tools for particular cohorts and settings.

This systematic review had four aims: (1) to establish what measures are available to screen for trauma in custody with young males, (2) to examine how trauma has been conceptualized within such measures, (3) to evaluate the psychometric properties of these measures, and (4) to appraise the availability of these measures for application in a custodial context.

## Method

### *Literature Search*

This systematic review followed the Consensus-based Standards for the selection of health Measurement Instrument (COSMIN) Guideline for Systematic Reviews of Patient Reported Outcome Measures (PROM) (Mokkink et al., 2018). The protocol for this review was registered on the Open Science Framework (<https://osf.io/r6hbk>) on December 28, 2020.

Following scoping searches to establish the appropriate databases and hone the search terms, the following electronic databases were searched in July 2020, using the specific terms identified within the protocol (see Supplemental Appendix A): The Cochrane Library, ProQuest, PsycINFO, ProQuest Dissertations and Theses, PsycTests, and PTSDpubs. The search terms combined terms for the following five concepts: adolescents, custody, trauma, males, and screening tools. The search strategy used for each database is reported in Supplemental Appendix A. Additional searches included those within key research databases in relevant health, government, and forensic organizations and professional society's websites from the United Kingdom, United States, and Canada. Additional gray literature searches were conducted in Electronic Theses Online System, and OAIster.

Further hand searches were conducted from the references of all final screened articles and use of Scopus (due to its wide health-related coverage) and Google Scholar to check for later citations, bringing one more text into scope. Finally, contact was made with all the authors from the final in-scope texts as well as others noted as experts in this field from Scopus. Responses from three of those authors brought two further texts into scope.

### Selection Criteria

Once the final set of eligible studies was established, the titles and abstracts were reviewed against the inclusion criteria: (1) must include participants aged 12 to 25 from a custody setting; (2) must include an empirical study using a trauma assessment tool; (3) must include an empirical study reporting data relevant to the psychometric properties of a trauma tool; (4) published in English; and (5) must have full text available. Studies were excluded if they (1) were not about trauma or were about another category of trauma only, for example, brain injury; (2) the text did not report any primary data; (3) were not specifically about assessment of trauma, for example, general mental health screen; or (4) did not validate a specific tool for assessing for trauma, or have a standalone trauma scale. A validation study was defined according to the COSMIN guidelines on page 20 of the COSMIN manual (2018), in that the aim of the study should be the evaluation of one or more measurement properties, the development of a PROM or the evaluation of the interpretability of the PROMs of interest. The COSMIN guidelines also stipulate that each PROM subscale must be rated separately. Studies reporting on measures that were multi-dimensional with validated sub-scales that could be scored as separate constructs were therefore included.

A cross check on the proposed criteria was conducted with a second rater using a sample of 26 texts. This resulted in 100% agreement and indicated there was consistency in the application of the criteria to the set of texts. Therefore, it was decided that there was high enough agreement to continue with single-reviewer application of the inclusion criteria, with little risk of unjustified exclusion.

### Data Extraction

The data extraction of potentially eligible literature was carried out with the following extracted data: author, year of publication, country, the study title, population, and types of psychometric properties tested. Specific details were also extracted from each study based on the psychometric property investigated, after which, quality assessment of the included studies was conducted for each study (see Supplemental Appendix B for full details).

### Overview of Methodological and Measurement Quality Assessment Process

The COSMIN methodology aims to improve the quality of studies on measurement properties by developing methodology and practical tools for assessing measurement properties (Mokkink et al., 2018). The COSMIN guidelines recommend a series of steps when evaluating the methodological and quality assessment. First, each study was evaluated for the methodological quality of the content validity and methodological quality of studies conducted on the psychometric properties of each instrument using the COSMIN Risk of Bias Checklist (Mokkink et al., 2018). Second, each study was evaluated for the quality of content validation procedures and each of the measure's psychometric properties as they were presented in each study. These were organized using COSMIN criteria that relate to internal structure (structural validity, internal consistency, and cross-cultural validity) and other measurement properties (reliability, measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness). Third, we applied the modified Grading of Recommendations Assessment, Development and Evaluation principles to examine the quality of the overall body of evidence for each instrument (Guyatt et al., 2011).

**Step 1: Methodological Quality Assessment.** To assess for risk of bias in the included studies, a robust approach was needed to assess the methodological quality of each study and to gauge the reliability of the reported results. The COSMIN methodology for systematic reviews of patient-reported outcome measures (Mokkink et al., 2018), and the COSMIN Risk of Bias Checklist (Terwee et al., 2018) were adopted to evaluate the methodological quality of the included studies (see Supplemental Appendix C for full details). The COSMIN checklist evaluates nine domains relevant to measurement properties and for the purpose of this study five of those domains were applicable and able to be evaluated: "structural validity," "internal consistency," "reliability," "measurement error" and "hypotheses testing for construct validity."

Content validity was not evaluated at this first stage as the COSMIN Risk of Bias Checklist requires a subjective judgment of whether the study asked patients and professionals about the relevance, comprehensiveness, and comprehensibility of the PROM. The studies in this review were not seeking to establish or explore content validity by this definition and so the COSMIN manual recommends that a PROM not be further considered. Therefore, to ensure full consideration of each study, the three-step method used in this systematic review allowed for the psychometric properties of the measure in the study to be reviewed at the second step, using criteria laid out in Terwee et al. (2007) and Cordier et al. (2017). Cross-cultural validity was not evaluated as the instruments

reviewed were developed and published in English, and interpretability is not considered to be a psychometric property under the COSMIN framework and was therefore not described in this review. Responsiveness was outside the scope of this review, as it was deemed not relevant where measures were being used as proxy diagnostic instruments rather than measures of change over time. Criterion validity was also not evaluated due to the absence of an agreed “gold standard” measure of the symptoms required in order to diagnose the PTSD disorder or of a measure that assesses exposure to trauma among children and adolescents.

Each of the measurement properties has a range of standards that are rated using a four-point rating system. Each standard was rated on a five-point scale “Very Good,” “Adequate,” “Doubtful” or “Inadequate.” The response option “NA” for some standards is available. The overall rating of the quality of the study is based on the lowest rating of any standard, that is, if any item is scored “inadequate” then the worst score counts (Mokkink et al., 2018). The items rate the quality of study design and the robustness of statistical analyses conducted in each of the studies.

**Step 2: Quality Assessment of Psychometric Properties.** Following the assessment of methodological quality of each study, the quality of the psychometric properties of the 12 measures themselves was then rated in step two. This gave an overall methodological quality score to each study, based on the quality of the psychometric properties in that study, and following an alternative, later approach to that proposed by the authors of COSMIN. The certainty of the evidence was classified using the method specified by Terwee et al. (2018) who suggested taking the lowest rating of any item in a checklist domain as the final quality rating for that domain. Cordier et al. (2017) however noted that subtle differences in the methodological quality between studies are difficult to detect via this method of scoring, so their revised scoring procedure was used in this systematic review. In brief, for each of the seven measurement properties at this step (not including “cross-cultural validity” or “responsiveness” from the COSMIN Risk of Bias Checklist), a criterion was defined for a positive, negative, or indeterminate rating, depending on the design, methods, and outcomes of the validation study. The results of each study were evaluated by the first author using the criteria described in Cordier et al. (2017) and Terwee et al. (2007). Supplemental Appendix D provides a summary of these criteria and the levels of evidence used to report when there was more than one study reporting findings about a measure.

**Step 3: Overall Quality of Psychometric Properties.** To create an overall quality rating, the measurement property for each measure was given an overall quality score using the criteria in Schellingerhout et al. (2012). This approach combined the scores of study quality with the psychometric quality ratings to give the overall rating. A description of this process is in Supplemental Appendix E, but in brief is a combination of

the ratings of both the methodological quality of the studies judged by the COSMIN checklist, plus the quality criteria for the psychometric properties of assessments, with an overall criterion given that was based on Terwee et al. (2007)

## Results

### *Descriptive Summary of Included Studies*

The search retrieved an initial total of 2,954 records. Searches in the five databases found 2,689 records with the following breakdown: PsycINFO=2,160, PTSDpubs=224, PsycTests=13, Cochrane=45, and ProQuest Dissertations and Theses=224. Further gray literature searching identified an additional 265 records from database, register, citation searching, and author contact. The initial search results were reduced to 2,930 after removing duplicates. From additional searching, two more records were included from contact with key authors, and one further record included from the Google Scholar citation search with no new texts found from reference mining. Screening of the title and abstracts of the remaining texts resulted in the removal of 2,896 records, and full copies of the remaining 34 texts were obtained initially.

At the full text review stage, 16 texts were excluded for the following reasons: five were not about trauma or were about another category of trauma only, for example, brain injury; three were not reporting on any specific tool for assessing for trauma; six were not specifically about the assessment of trauma; one did not seek to validate the tool in question and one did not include males from the age range targeted. At the data extraction stage, four more texts were excluded, due to not being able to be evaluated with the COSMIN methodology. This was because they either did not fully represent the construct in question, were not solely about trauma, or did not have standalone scales.

The final set of articles which met the inclusion criteria totaled 14 texts included in the current review, with a range of publication dates from 2000 to 2020. All but two of those studies originate from the United States and the others from European countries. Two doctoral theses were included as part of the final set of 14 texts; Zito (2016) and Flaherty (2017). These were considered suitable for inclusion as each had been peer reviewed by each doctoral candidate’s academic supervision team and doctoral viva committee. The studies sampled a total of approximately 1,768 male participants and an age range of 12 to 18 years. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, 2023) flow diagram in Figure 1 below illustrates the screening, overall data collection, and results of screening outcomes at the various stages. Key information about the studies can be found in Supplemental Appendix B.

### *Descriptive Summary of the Trauma Measures*

A full description of each of the measures is available in Supplemental Appendix F. Of the 14 studies in this review,

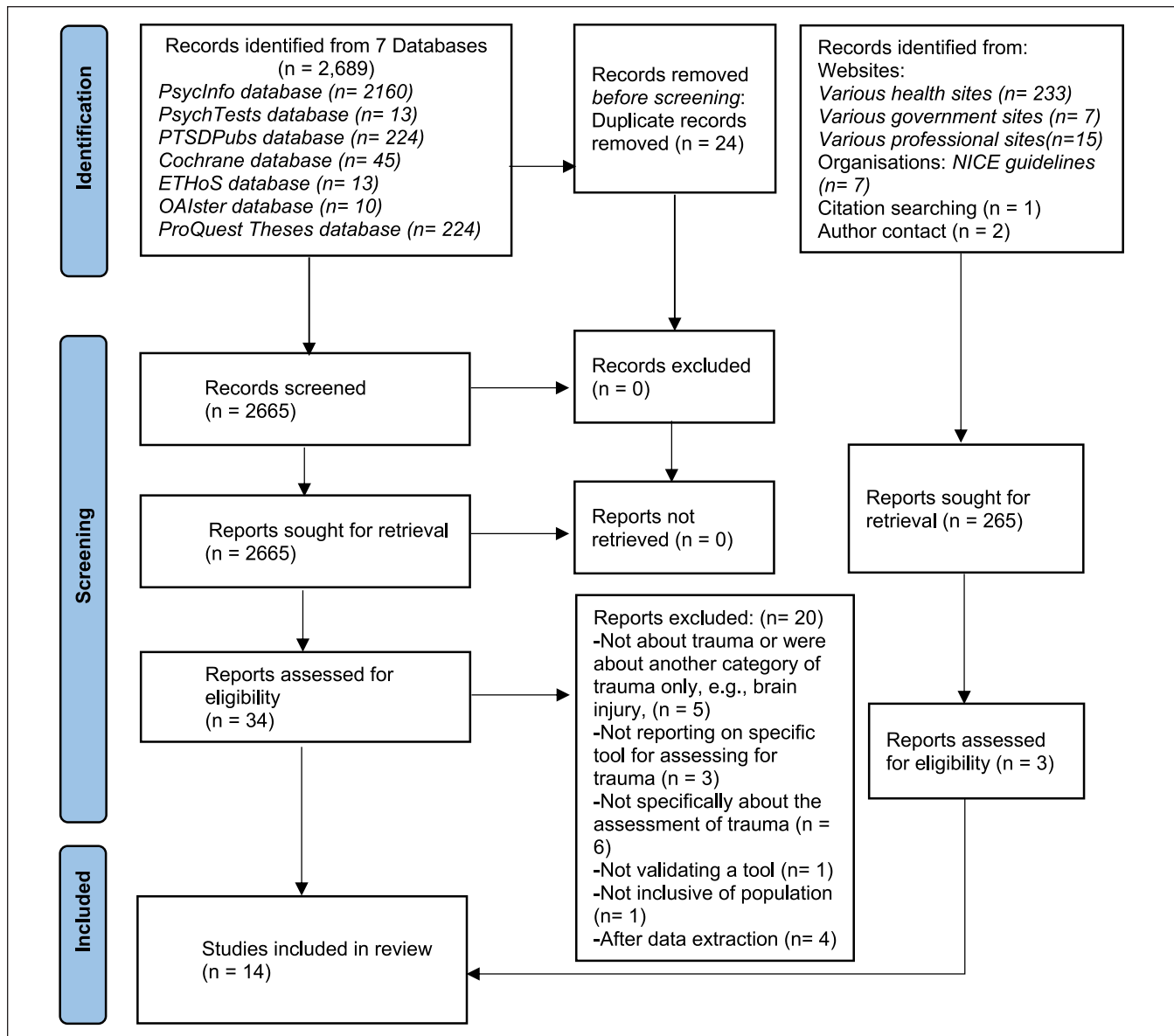


Figure 1. PRISMA flow diagram of search strategy.

the psychometric evidence of 12 measures of trauma used with the 12 to 25 years male offender population was reported. Those measures differed significantly to each other on length (ranging from 10 to 478 items). Nine of the measures were unidimensional measures of trauma, some only being validated for the first time in the reporting study. The other three measures were multiscale measures with a relevant scale assessing for trauma exposure and/or symptomatology (The Behavior Assessment System for Children, Second Edition, Adolescent Version-Self Report [BASC-2 SRP-A], Minnesota Multiphasic Personality Inventory-Adolescent [MMPI-A], The Massachusetts Youth Screening Instrument [MAYSI-2]). These studies were able to be included as the subscales fully represented the trauma construct or were standalone sub scales.

### Psychometric Properties of Trauma Measures

This stage of the review process involved an assessment of the methodology of each of the empirical studies so that they could then be weighed according to the reliability of the results. The studies included here report on various types of psychometric evidence, and due to the lack of homogeneity in methodologies and theoretical constructs a narrative synthesis of these factors was undertaken. Table 1 shows the summary version of the psychometric evidence reported in the studies. The overall rating is based on the lowest rating of any standard, that is, if any item is scored “inadequate” then the worst score counts (Mokkink et al., 2018). The items rate the quality of study design and the robustness of statistical analyses conducted in each of the studies.

**Table 1.** Psychometric Evidence Summary of the Studies.

Measures N= 12	Title Abbreviation	Tests of Factor Structure	Tests of Internal Consistency	Tests of Reliability	Tests of Validity
The Behavior Assessment System for Children, Second Edition, Adolescent Version-Self-Report	BASC-2 SRP-A	No	No	Yes	Yes
The Clinician-Administered PTSD Scale for Children and Adolescents	CAPS	No	Yes	Yes	Yes
Child Report of Posttraumatic Symptoms	CROPS	Yes (Edner et al., 2017; Edner, Piegore et al., 2020)	Yes	Yes (Edner et al., 2017; Edner, Piegore et al., 2020)	Yes (Edner et al., 2017; Edner, Piegore et al., 2020)
The Childhood Trauma Questionnaire-Short Form	CTQ-SF	Yes	Yes (Dudeck et al., 2015)	Yes (Dovran et al., 2013)	No
Dimensions of Violence Exposure Rapid Screen	DVERS	No	Yes	Yes	Yes
Minnesota Multiphasic Personality Inventory-Adolescent	MMPI-A	No	Yes (Cashel et al., 2000)	Yes (Cashel et al., 2000)	Yes
The Massachusetts Youth Screening Instrument	MAYSI-2	No	Yes	Yes	Yes
The Structured Trauma-Related Experiences and Symptoms Screener	STRESS	Yes	Yes	Yes	Yes
The Trauma-Related Symptoms and Impairment Rapid Screen	TSIRS	No	Yes	Yes	Yes
Trauma Symptom Checklist for Children	TSCC	No	Yes	No	No
Trauma Scale for Juvenile Offenders	TSJO	No	Yes	No	Yes
The Trauma Checklist	TC	Yes	Yes	Yes	Yes

### Step 1: Methodological Quality Assessment

After the data extraction phase, the methodological quality ratings of the studies were reviewed using the “worst score counts” principle (see Supplemental Appendix G for the measurement property criteria). Of the 16 studies included for review (two texts undertook two studies each, the rest were single studies), 13 were assessed as “Inadequate” and three as “doubtful.” Reflecting the exploratory nature of many of these studies, they were often robust in hypothesis testing and use of comparator instruments where applicable. Analysis was often done to test the internal consistency of the scales in question, with reliability and validity work done too, although the latter did not often meet the standard of the COSMIN benchmark. Supplemental Appendix H summarizes the ratings for each COSMIN category.

### Step 2: Psychometric Quality Assessment

In this next step, the measures themselves were evaluated. This was conducted by assessing the quality of the psychometric properties of each of the twelve measures using the criteria set out in Cordier et al. (2017) and Terwee et al. (2007) (see Supplemental Appendix I).

**Internal Consistency.** Internal consistency was rated for 11 of the 12 measures with the BASC-2 not evaluated on this property in the study by Zito (2016). Conflicting results were rated for three measures: the Trauma Symptom Checklist for Children (TSCC), the Child Report of Posttraumatic Symptoms (CROPS), and the MMPI-A. The CROPS had different methodological ratings across studies, with some reporting an EFA along with acceptable Cronbach values in two studies by Edner et al. (2017) and Edner, Glaser et al. (2020), Edner, Piegore et al. (2020), and a combination of good alpha values, but a lack of factor analysis in another study (Flaherty, 2017). Similarly, the MMPI-A was reported by Cashel et al. (2000) to have a good Cronbach alpha for the PK subscale, but no available factor analysis and Murray et al. (2013) also reported no factor analysis.

Six measures, the Trauma Checklist (TC), The Trauma-Related Symptoms and Impairment Rapid Screen (TSIRS), Dimensions of Violence Exposure Rapid Screen (DVERS), MAYSI-2, the Clinician-Administered PTSD Scale for Children and Adolescents (CAPS-CA), and Trauma Scale for Juvenile Offenders (TSJO), were all rated as “negative” as factor analysis was not performed (due to structural validity not being the stated aim of the study). Internal consistency was rated as “positive” for two measures: the Structured

Trauma-Related Experiences and Symptoms Screener (STRESS) and the Childhood Trauma Questionnaire-Short Form (CTQ-SF), which rated “positive” for having a confirmatory factor analysis, reported along with acceptable Cronbach alphas on an appropriately sized sample.

**Reliability.** The psychometric concept of reliability at this stage is defined by the COSMIN guidelines as “the extent to which patients can be distinguished from each other, despite measurement errors” and so also includes inter-rater reliability. Test–retest reliability was not a feature of these studies; this reflects the typical practice of the early stages of test development, where resources are limited and therefore tend to be focused on validation in the first instance. Reliability was rated for 10 of the 12 measures with the BASC-2 and the TSJO not evaluated on this property in the reporting studies. Conflicting results were reported for the TSIRS and the DIVERS with good intraclass correlation coefficients (ICCs) reported but low kappa coefficient results.

Reliability was rated as “negative” for 6 of the 12 measures. Four of those six were rated “negative” due to a low reported ICC and those measures were: the TC, the MAYSI-2, the STRESS and the MMPI-A. The CROPS and the TSCC were rated as “negative” as no statistical information was reported for reliability as evaluated by the COSMIN criteria. Reliability was rated as “positive” for two measures, the CAPS-CA which had a good kappa value and the CTQ-SF as the reported reliability coefficients were all greater than .70.

**Content Validity.** Content validity was rated for 10 of the 12 measures with the TSJO not evaluated according to COSMIN criteria. The TC was rated as “indeterminate” as while there was clear description of the measurement aims, target population, and concepts being measured, only some participants completed all the measures (some did not complete the CTQ) which was not explained.

Content validity was only rated “negative” for the CTQ-SF as although there were clear aims, the target population for this systematic review was not clearly delineated. Content validity was rated as “positive” for 8 of the 12 measures with all the following deemed to have sufficiently clear measurements aims, concept descriptions, and target population and clear item selection: the TSIRS, the DIVERS, the MAYSI-2, the MMPI-A, the TSCC, the CROPS (for all three reporting studies), the STRESS, and the CAPS-CA.

**Construct Validity.** Construct validity was rated for only 4 of the 12 measures with the BASC-2 and the TSJO not able to be evaluated according to COSMIN criteria. Three measures were not able to be assessed as no information was found on this property for them: the TSIRS, the DIVERS, and the TSCC. There were indeterminate ratings for three other measures as in all cases there were differently reported results: the CROPS, the CTQ-SF, and the MMPI-A. In all cases, it was due to having one study report no results for construct

validity and another report positive results explaining more than 50% of the variance. The TC was again rated as “indeterminate” due to the many measures cited in the articles being used as proxies for clinical variables rather than the TC being validated against them. Data are reported for a correlation with one measure, but that was only for a much smaller section of the original sample, for reasons not clear.

Construct validity was rated as “negative” for two of the measures, that is, the STRESS and the MAYSI-2 as the former confirmed less than 75% of the study’s hypothesis and the latter’s multiple regressions proved less than 75% of the variance. Construct validity was rated as “positive” for the CAPS-CA as there were clear hypotheses and the results explained more than 50% of the variance.

### Step 3: Overall Psychometric Quality

The overall psychometric quality for each of the measures is presented in Table 2. The process adopted to reach a conclusion about the overall psychometric quality follows the process and criteria set out in Schellingerhout et al. (2012) and Cordier et al. (2017). The judgment was reached by combining the ratings of both the methodological quality of the studies using the COSMIN checklist and the quality criteria for the psychometric properties of assessments (see Table 2). To assess whether the results of the measurement properties were positive, negative, or indeterminate, we used criteria based on Terwee et al. (2017).

When judging the overall psychometric quality of the twelve measures of trauma, no measure was rated as “moderate” or strong evidence found for either “negative” or “positive” results. This indicates the low methodological ratings achieved according to the COSMIN quality assessment in step one combined with the low psychometric quality ratings achieved in step two using the criteria described by Terwee et al. (2007) and Cordier et al. (2017).

Sixty ratings were reported using the five psychometric properties relevant to this review from Schellingerhout et al. (2012) and Cordier et al. (2017), and judged for each of the 12 measures. When judging the overall psychometric quality of the measures:

- 16 of the 60 ratings were classified as “Not Reported.”
- 17 of the 60 ratings were classified as “indeterminate.”
- 17 of the 60 ratings were classified as “limited positive” (the highest rating achieved in this study).
- 8 of the 60 ratings were classified as “limited negative.”

When judging the overall quality of internal consistency, eight measures were rated as “limited positive” due to the reported good alpha values and the scales being unidimensional. Only three of those same measures rated as “limited positive evidence” for overall construct validity, as so few

**Table 2.** Overall Psychometric Quality of the Trauma Measures.

Measures N= 12	Reliability Property: Internal Consistency	Reliability Property: Measurement Error	Reliability Property: Reliability <sup>a</sup>	Validity Property: Construct Validity	Validity Property: Hypothesis Testing <sup>b</sup>
BASC-2 SRP-A	evidence NR <sup>c</sup>	evidence NR	evidence NR	evidence NR	Indeterminate evidence
CAPS	Limited positive evidence	Indeterminate evidence	Limited positive evidence	Limited positive evidence	Limited positive evidence
CTQ-SF	Limited positive evidence	evidence NR	Limited positive evidence	Limited positive evidence	evidence NR
CROPS	Limited positive evidence	evidence NR	Indeterminate evidence	Limited negative evidence	Limited positive evidence
DVERS	Limited positive evidence	Indeterminate evidence	Indeterminate evidence	Indeterminate evidence	Indeterminate evidence
MMPI-A	Limited positive evidence	evidence NR	Limited negative evidence	Limited positive evidence	Limited negative evidence
MAYSI-2	Indeterminate evidence	Indeterminate evidence	Limited negative evidence	Limited negative evidence	Limited negative evidence
TSIRS	Limited positive evidence	Indeterminate evidence	Indeterminate evidence	Indeterminate evidence	Indeterminate evidence
TSCC	Limited positive evidence	evidence NR	Indeterminate evidence	Indeterminate evidence	evidence NR
TSJO	Limited negative evidence	evidence NR	evidence NR	evidence NR	Limited positive evidence
TC	Indeterminate evidence	evidence NR	Indeterminate evidence	Indeterminate evidence	evidence NR
STRESS	Limited positive evidence	evidence NR	Limited negative evidence	Limited negative evidence	Limited positive evidence

Note. BASC-2 SRP-A=The Behavior Assessment System for Children, Second Edition, Adolescent Version-Self Report; CAPS=The Clinician-Administered PTSD Scale for Children and Adolescents; CROPS=Child Report of Posttraumatic Symptoms; CTQ-SF=The Childhood Trauma Questionnaire-Short Form; DVERS=Dimensions of Violence Exposure Rapid Screen; MAYSI-2=The Massachusetts Youth Screening Instrument; MMPI-A=Minnesota Multiphasic Personality Inventory-Adolescent; STRESS=The Structured Trauma-Related Experiences and Symptoms Screener; TC=The Trauma Checklist; TSCC=Trauma Symptom Checklist for Children; TSIRS=The Trauma-Related Symptoms and Impairment Rapid Screen; TSJO=Trauma Scale for Juvenile Offenders.

<sup>a</sup>Definition of Reliability here: the proportion of the total variance in the measurements which is because of “true” differences among patients.

<sup>b</sup>Definition of Hypothesis Testing here: where predefined questions about expected correlations exist and at least 75% of the results were in accordance with the hypotheses.

<sup>c</sup>NR=not reported.

studies were able to explain the reported variance found in their results. Only two of those same measures were also able to be rated as having “limited positive evidence” for reliability overall. The only measure with a majority of “positive” ratings (albeit limited due to being only one study) was the CAPS. That study by Harrington (2008) reported good reliability with regard to internal consistency and inter-rater agreement. It also reported good validity results with regard to how the CAPS-CA corresponded to other measures. Notably, this was an entirely male and incarcerated sample with a good sample size.

## Discussion

Application of the inclusion criteria to the results of the various searches identified 14 empirical studies for inclusion in this review. The pilot stages of the search strategy, with additional hand searching and reference mining of the included papers, along with author contact, allows confidence in

concluding that all relevant research available at the time has been included. Therefore, the conclusions below are based on a synthesis of all available evidence at that time.

A mixed picture emerged from the findings, with the measures taking a different theoretical approach to the evaluation of trauma in this age group, and the studies taking different approaches to reporting the psychometric quality of them. Many of the measures were studied only once in this review, and the lack of replication studies and study quality issues for a number of tools limit conclusions regarding their application. Data from this review therefore support the conclusion drawn from previous research that empirical evidence is still very limited within this field. However, it is clear from the recent nature of many of the studies that this work is developing at pace and although caution is needed in drawing any firm conclusions about measures at this point, findings are beginning to emerge about how some tools could be used in a custodial context with this age group.



**Table 3.** Critical Findings.

---

There is not yet sufficient, robust methodological and psychometric validation work done on measures of trauma with young adult men in custody.

None of the measures are able to account for exposure to trauma both as a victim and as a perpetrator.

None of the measures are able to account for the potentially traumatic experiences of being imprisoned at a young age and for a long time.

Many of the measures still align with a tradition PTSD model of trauma, despite the potentially limiting aspect of this model with regard to the expression of trauma-related difficulties in this age group.

---

**Table 4.** Practice, Policy, and Research implications.

---

This is the first known study to investigate the methodological and psychometric properties of the available measures of trauma described. This is despite the increasing emphasis on the need for screening for trauma-related difficulties in this cohort.

More research work on the validity and reliability of such measures is needed, with particular attention paid to the range of diversity in this cohort.

Any further work should be informed by how cumulative trauma is worthy of different assessment and intervention than single incident, post-traumatic stress. All of which is also likely affected by the developmental stage of both the age when experienced and the age of the young male prisoner at time of screening.

---

As the findings reported in Supplemental Appendix I indicate, no measure was able to be rated by the quality assessment method as any higher than having limited “positive” or “negative” evidence. This is due in part to the number of single studies looking at different measures, as well as the low methodological and psychometric quality ratings achieved in the previous steps of the quality assessment process. The only measure with a majority of positive ratings was the CAPS, which was an unpublished thesis by Harrington (2008) and so not peer-reviewed. However, it used a robust psychometric approach, and a large, male, custodial sample to evaluate the tool and so was able to report good reliability with regard to internal consistency and inter-rater agreement, and good validity results with regard to how the CAPS-CA corresponded to other measures.

### Overview of Quality Assessment Findings

A total of 14 texts met this review’s inclusion criteria and reported on the psychometric properties of 12 measures of trauma for use with young males in a custodial setting. Three measures were evaluated in more than one study: the MMPI-A, the CROPS, and the CTQ-SF, but this did not lead to any consistent findings of reliability or validity for these measures. For the other nine measures, only single studies were identified reporting on one or more of the psychometric properties. Furthermore, most studies only addressed a few (a range of one to five) of the various measurement properties being evaluated in this review.

Using the COSMIN methodology to assess the methodological quality ratings of the 16 studies contained in the 14 texts resulted in the methodology of 13 studies being assessed as “inadequate” and three studies as “doubtful” (Mokkink et al., 2018). In the final step of determining the overall psychometric quality of the measures of trauma, many ratings

were not able to be determined due to the poor methodological or psychometric quality ratings. This insufficient evidence should be interpreted with caution by the reader, especially if making clinical decisions about trauma screening with this cohort. The mixed evidence reflected how some studies did not conduct sufficient psychometric evaluation that could be assessed for its quality. For example, the lack of factor analysis to establish internal consistency does not necessarily mean that those tools should be discounted from use. It does however suggest there are psychometric qualities that warrant further and more rigorous analysis. Due to this being an emerging field at the point of this review, studies often focused on establishing criterion validity, cut off points, or inter-rater reliability. Future studies will consider developing the evidence base with regard to internal consistency and possible measurement error, as well as further exploration of construct and content validity to ensure accurate evaluation of the construct being measured.

### Reliability Evaluation

Evaluation of the reliability of the measures was reported in 10 of the 16 studies. Internal consistency was the most frequently reported psychometric domain for 11 of the 12 measures, with 11 of the studies achieving a “very good” rating at this final step (only the BASC-2 measure did not report internal consistency). Evaluation of the factor structure was conducted in six studies for four of the measures, with the CTQ-SF and the STRESS achieving a “very good” rating, the TC achieving an “adequate” rating and the CROPS achieving an “inadequate” rating.

When judging the psychometric quality of the internal consistency of the measures, there was a mixed picture with three measures rated as “indeterminate,” two as “conflicting results” and four as “negative,” all affected by the lack of

factor analysis attempted in these studies (see Supplemental Appendix I). Only two measures had “positive” internal consistency ratings: the CTQ-SF and the STRESS. However, when judging the overall quality of internal consistency (see Table 2 above), eight measures were rated as “limited positive” due to the reported good alpha values and the scales being unidimensional. These findings suggest that the evidence base is not yet strong enough to reach firm conclusions about the overall reliability of the measures in question. While many studies had reported some evaluation of reliability, they tended to rely on Pearson’s correlation (which perhaps is less relevant to clinical samples where normal distribution is not expected), or Cronbach’s alpha which is a function of the number of items in the measure, meaning that high values cannot be assumed to be indicative necessarily of high internal consistency. As the measures were not sufficiently robust according to the COSMIN criteria, the researcher or practitioner seeking evidence on trauma measures for this group of people should exercise caution in using tools currently in use for the wider age group, rather than for this custodial cohort. It is also important to note that there has been no evaluation of change over time for any of these measures, and so a practitioner cautiously using any of these measures as part of an initial assessment should be even more cautious before considering using them to track any change over time for any reason.

### **Validity Evaluation**

Evaluation of the validity of the measures with regard to hypothesis testing was carried out in most of the studies (12 of the 16 studies) and for all the 12 measures. Ten of the 12 measures were judged to have “very good” properties with regard to “hypothesis testing” (see Supplemental Appendix H). With regard to construct validity only 4 of the 12 measures had conducted factor analysis, with a “very good” rating achieved for the CTQ-SF and the STRESS, and an “adequate” rating for the CROPS and TC.

When judging the psychometric quality of the validity of the measures, both the CAPS and the MMPI-A achieved “positive” ratings for both content and construct validity (see Supplemental Appendix I). All the others were either a mixed picture of “indeterminate” ratings or “negative” ratings. One measure, the TSJO, had neither content nor construct validity reported, due to the aim of that study being to examine the predictive utility of the newly developed measure.

Finally, when judging the overall quality of each psychometric property per measure (see Table 2), 16 of the 60 ratings were classified as “not reported” and 17 of the 60 ratings were classified as “indeterminate.” This then contributed to an inconclusive outcome for being able to judge the measures. The field of trauma assessment with young males in custody is nascent at this time; therefore, many of the studies in this review focused their statistical work on demonstrating a level of criterion-related validity or more precisely, predictive

validity, often in terms of sensitivity and specificity or in reporting receiver operating characteristics. The availability of validated instruments is paramount to research and practice work in this field. While this work is necessary to correctly classify those with trauma exposure and associated symptomatology, it is arguably not so important to a practitioner in such settings where trauma exposure within the cohort can often be safely assumed. However, without the empirical support for particular tools, any clinician runs the risk of substandard assessment and intervention, which in a criminal justice context has implications for legal challenge and misguided risk reduction work.

### **Trauma Assessment with Young Males in Custody**

To date, the majority of research on trauma and juvenile offenders focuses on assessing whether or not juvenile offenders meet the criteria for PTSD (Ford et al., 2008; Kerig et al., 2009, 2012; Perkins et al., 2016). While the prevalence of diagnosed PTSD is higher than community samples (Wolpaw & Ford, 2004), assessing for trauma using measures with a PTSD lens can limit the breadth of both the exposure to trauma and assessment of the trauma symptoms that would be reported otherwise. Of the 12 measures in this systematic review, eight were aligned with a PTSD model of trauma, that is, symptoms observed as a result of exposure to trauma, and the other four were not aligned with any diagnostic model.

In addition to the potential difficulty of measures being predicated on particular diagnostic models, some instruments failed to capture both the range of traumatic events as well as the range of symptoms that would be reported. For example, there are some that only measure dissociation and depersonalization symptoms rather than assess for all trauma symptomatology. Given the need to screen for the range of symptoms arising from trauma in this population, and to ensure an accurate understanding of the impact on the boy or young man who is not hinged solely on an adult understanding of PTSD, it is imperative that practitioners are equipped with measures that are valid, reliable, easy to use, and appropriate for their setting and population of interest. Considering the high prevalence of intellectual difficulty and acquired brain injury experienced by this group, tools must also be able to be easily accessible for them too.

### **Considerations Relating to Diversity**

Due to the lack of validation work available so far for the measures included in this study, little is known about the suitability of the measures for the kind of diverse populations found in the young adult male cohort in UK prisons. It would be useful to have data to gauge the suitability of the measures for the protected characteristics such as age, ethnicity, disability, and sexual orientation, but also more widely for language ability, or socioeconomic status for example. Further

research work is required to both develop further the theoretical understanding of trauma as well as the measurement of it, in order to develop measures suitable for all.

It will be important to ensure that future research design and methodology is able to explore for the different experiences of all the participants, with particular regard to culture and ethnicity. While some of the studies in this review reported an ethnicity breakdown, this was not always done and when done sometimes used ethnic breakdowns specific to the country of origin. As the evidence continues to demonstrate the overrepresentation of young men from black and minority ethnic backgrounds in the UK CJS, measures assessing for trauma exposure and symptoms must evidence their ability to account for how different groups of people respond to such measures. While gender is already determined to play its part in how people report previous trauma or current symptoms, it remains to be determined to what extent cultural factors also play their part. Measures that do not take account of collective racial trauma may well also fall short of fully describing how this cohort experience both single traumatic incidents and intersecting traumas.

### *Strengths and Limitation of the Review*

This systematic review appears to be the first such review to collate and synthesize the available literature regarding appropriate measures of trauma for this cohort, and to do so with a robust risk of bias checklist that is relevant to service user reported outcomes. It is likely therefore that this work will be of value to both future research work in this area and to practitioners searching the evidence base for guidance in this field. Other strengths also add to the robustness of this review and so the confidence with which others may use it. For example, it is based on a comprehensive search strategy to ensure that both a variety of databases and non-database options were searched. It also uses the most widespread and comprehensive quality assessment tool to assess measurement properties of health instruments designed for an evaluative purpose, as part of a three-step quality assessment method. This again offers the researcher or practitioner confidence in the objective benchmark being applied.

There are limitations to the review. The COSMIN methodology meant that any studies that looked at indicative items on subscales had to be excluded due to not having a sufficiently one-dimensional construct available to be analyzed with the COSMIN domains. Those studies tended to be ones that sought to develop or validate a screen for institutions to identify risk of PTSD, that is, the presence of symptoms following exposure to trauma, rather than a more exploratory understanding of the type and nature and extent of the experiences of trauma. Some of the included systematic reviews poorly reported the review process, outcomes, and conclusions, and this fact may have led to the loss of some data. So, the strict application of the COSMIN methodology meant that there was a reduction in reporting of the

quality criteria according to COSMIN, whereby at the final stage of overall psychometric quality evaluation, only three of the possible seven criteria were evaluated for each study overall. While this reflects the studies in scope, and the early stage of research into such measures, a more robust and inclusive evaluation of all the criteria would have been preferred, although was not possible at this point.

With regard to limitations, a methodological limitation is that a sole researcher undertook the risk of bias quality assessment work, rather than having a second researcher cross-check the texts for a more robust quality assessment process. To mitigate against any possible bias introduced because of this, a search strategy and review protocol was agreed before data collection started to reduce the chances of bias.

### *Implications for Practice and Research*

This review of the available research of trauma measures for this cohort allows for an overview of those measures and their psychometric properties. The studies all argue for the need for measures to be developed, normed, validated, and standardized for adolescents in custody. There is general agreement now of the dangers of using adult measures to assess for strict diagnostic frameworks of trauma, for example, PTSD. While the studies in this review do not warrant recommendation yet for use within practice, the direction of travel is clear. Researchers are now looking to avoid measures simply seeking to establish trauma exposure (as that can be safely assumed for most in this cohort) and to avoid a traditional PTSD model of trauma (as adolescents tend to express their trauma symptomatology in diverse and complex ways) as well as understand the need to use measures that are developmentally appropriate and suitable for males and females (due to the diversity of trauma expression by the genders). As that work progresses, this will need replicating with young males in custody too, in order to ensure the findings hold true for this cohort. This is particularly important as this cohort continues to change and reflect sentencing practices, for example, longer sentences at younger ages, a higher representation of males from an ethnic minority in the UK.

When evaluating the overall psychometric quality of the 12 measures of trauma, none could be assessed as demonstrating an overall strong "positive" or "negative" psychometric quality. The highest rating achieved in this review for the measures was only "limited positive" and then for only 17 of the 60 ratings and "limited negative" for only 8 of the 60 ratings. Suffice to say, these findings evidence that trauma measures for use with young males in custody is not yet sufficiently well validated, and so practitioners should be wary of using such measures at this point in time, particularly if considering tools for group level screening or administration.

Forensic practitioners are familiar with embedding psychometric measures in an assessment approach, which complements wider clinical work such as the use of risk, ability,

or strengths-based tools. The quality assessment undertaken for the tools evaluated by the studies in this review offers practitioners more detail on how to make best use of particular tools. This provides an evidence-based foundation for the use of such tools and approaches, which should continue to be set in the context of a wider case formulation approach.

When considering practice implications arising from this review, it is of note that when reviewing the studies, most authors tended to focus on the conceptual and statistical properties of their measures, and issues of the practicality of the instrument are rarely discussed, for example, administrative or respondent burden. However, this is of key importance to readers looking for tools to use in clinical practice. While an overall quality score is helpful but not necessary (as it assumes all measurement properties are equal), detailed examination of each tool's particular psychometric features will be helpful to the practitioner in this regard.

Moreover, the measures also now need to account for the other types of categories of what can be experienced as traumatic, that pushes the frontier beyond the usual ACE10 approach (Felitti et al., 1998). This cohort of boys and young men in custody are identified as more likely to have committed serious offenses, more likely to be sentenced to lengthy sentences of 20 years or more, and to be sentenced as part of joint enterprise legislation; this all points to the need for measures to be able to account not only for previous trauma but also for ongoing trauma exposure resulting from custody and length of sentence.

Ongoing research is needed to take forward the problem of a lack of robust measures to screen for trauma with young males in custody. This will need to be informed by the ongoing empirical studies which are updating our understanding of how trauma is experienced and then expressed in the various developmental stages in the under 25 years age range. This should also be informed by how cumulative trauma is worthy of different assessment and intervention than single incident, post-traumatic stress. Ensuring that such research includes the person centered and phenomenological approach is vital to capturing the nuances in this field.

## Conclusions

This systematic scoping review is the first of its kind to synthesize the research regarding the psychometric properties of measures available for use to assess for trauma exposure and symptomatology with adolescent males in custody. It reports evidence of the quality of psychometric properties of 12 instruments used to measure trauma with the targeted population. The COSMIN taxonomy (Mokkink et al., 2018) was used to rate the reliability and validity information reported about the instruments. A varying degree of evidence was reported for the psychometric properties of the trauma measures, which led to an inability to recommend any measure for use in practice. As trauma measures for this cohort are not yet well validated, there is scope for further empirical work

to inform this field. Such measures will need to adapt to the changing type of trauma exposure experience by the boys and young men in this sample, as well as the complexity of expression of that trauma at various developmental stages.

## Acknowledgments

We are grateful to all the young people who took part in the original studies included in this review. We are also grateful to those authors who responded to queries and provided unpublished studies on included measures.

## Declaration of Conflicting Interests

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

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was conducted as part of Rachel's doctoral studies at Nottingham Trent University. The doctoral study program was funded by HMPPS, Psychology Services.

## ORCID iD

Mike Marriott  <https://orcid.org/0000-0001-7743-5262>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Agorastos, A., Pervanidou, P., Chrousos, G. P., & Kolaitis, G. (2018). Early life stress and trauma: Developmental neuroendocrine aspects of prolonged stress system dysregulation. *Hormones, 17*(4), 507–520.
- Ben-Zion, Z., Fine, N. B., Keynan, N. J., Admon, R., Halpern, P., Liberzon, I., Hendler, T., & Shalev, A. Y. (2019). Neurobehavioral moderators of post-traumatic stress disorder (PTSD) trajectories: Study protocol of a prospective MRI study of recent trauma survivors. *European Journal of Psychotraumatology, 10*(1), 1683941.
- Bonner, T., DeLisi, M., Jones-Johnson, G., Caudill, J. W., & Trulson, C. (2020). Chaotic homes, adverse childhood experiences, and serious delinquency: Differential effects by race and ethnicity. *Justice Quarterly, 37*(4), 697–714.
- Cashel, M. L., Ovaert, L., & Holliman, N. G. (2000). Evaluating PTSD in incarcerated male juveniles with the MMPI-A: An exploratory analysis. *Journal of Clinical Psychology, 56*(12), 1535–1549.
- Chitsabesan, P., Krull, L., Bailey, S., Kenning, C., Sneider, S., MacDonald, W., & Theodosiou, L. (2006). Mental health needs of young offenders in custody and in the community. *The British Journal of Psychiatry, 188*(6), 534–540.
- Connolly, E. J., & Kavish, N. (2019). The causal relationship between childhood adversity and developmental trajectories of delinquency: A consideration of genetic and environmental confounds. *Journal of Youth and Adolescence, 48*, 199–211.

- Copeland, W. E., Keeler, G., Angold, A., & Costello, E. J. (2007). Traumatic events and posttraumatic stress in childhood. *Archives of General Psychiatry*, *64*(5), 577–584.
- Cordier, R., Milbourn, B., Martin, R., Buchanan, A., Chung, D., & Speyer, R. (2017). A systematic review evaluating the psychometric properties of measures of social inclusion. *PLoS One*, *12*(6), e0179109.
- Cruise, K. R., Dandreaux, D. M., Marsee, M. A., & DePrato, D. K. (2008). Identification of critical items on the Massachusetts Youth Screening Instrument–2 (MAYSI-2) in incarcerated youth. *International Journal of Forensic Mental Health*, *7*(2), 121–132.
- Daniunaite, I., Cloitre, M., Karatzias, T., Shevlin, M., Thoresen, S., Zelviene, P., & Kazlauskas, E. (2021). PTSD and complex PTSD in adolescence: Discriminating factors in a population-based cross-sectional study. *European Journal of Psychotraumatology*, *12*(1), 1890937.
- Dovran, A., Winje, D., Øverland, S. N., Breivik, K., Arefjord, K., Dalsbø, A. S., Jentoft, M. B., Hansen, A. L., & Waage, L. (2013). Psychometric properties of the Norwegian version of the childhood trauma questionnaire in high-risk groups. *Scandinavian Journal of Psychology*, *54*(4), 286–291.
- Dudeck, M., Vasic, N., Otte, S., Streb, J., Wingenfeld, K., Grabe, H. J., Freyberger, H. J., Schroeder, T., Von Schoenfeld, C. E., Driessen, M., & Barnow, S. (2015). Factorial validity of the short form of the childhood trauma questionnaire (CTQ-SF) in German psychiatric patients, inmates, and university students. *Psychological Reports*, *116*(3), 685–703.
- Edner, B. J., Glaser, B. A., & Calhoun, G. B. (2017). Predictive accuracy and factor structure of the child report of post-traumatic symptoms (CROPS) among adjudicated youth. *Psychological Trauma: Theory, Research, Practice, and Policy*, *9*(6), 706.
- Edner, B. J., Glaser, B. A., Calhoun, G. B., Dukes-Murray, H., Khan, Z., & Donaldson, K. (2020). Efficacy of the MMPI-A in detecting trauma among delinquent youth: Cross-validation of the trauma scale for juvenile offenders (TSJO). *The Journal of Forensic Psychiatry & Psychology*, *31*(1), 1–16.
- Edner, B. J., Piegore, A. L., Glaser, B. A., & Calhoun, G. B. (2020). Preliminary cross-sectional validation of the CROPS 14-item for juvenile offenders. *International Journal of Offender Therapy and Comparative Criminology*, *64*(12), 1258–1274.
- Eklund, K., Rossen, E., Koriakin, T., Chafouleas, S. M., & Resnick, C. (2018). A systematic review of trauma screening measures for children and adolescents. *School Psychology Quarterly*, *33*(1), 30.
- Farrell, C., & Zimmerman, G. M. (2017). Does offending intensify as exposure to violence aggregates? Reconsidering the effects of repeat victimization, types of exposure to violence, and poly-victimization on property crime, violent offending, and substance use. *Journal of Criminal Justice*, *53*, 25–33.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*, *14*(4), 245–258.
- Flaherty, C. (2017). *Assessing trauma in the juvenile offender population: An exploration of the TSCC and the CROPS*. (Doctoral dissertation, University of Georgia).
- Ford, J. D., Chapman, J. F., Pearson, G., Borum, R., & Wolpaw, J. M. (2008). Psychometric status and clinical utility of the MAYSI-2 with girls and boys in juvenile detention. *Journal of Psychopathology and Behavioral Assessment*, *30*(2), 87–99. <https://doi.org/10.1007/s10862-007-9058-9>
- Grasso, D. J., Felton, J. W., & Reid-Quinones, K. (2015). The structured trauma-related experiences and symptoms screener (STRESS): Development and preliminary psychometrics. *Child Maltreat*, *20*(3), 214–220. <https://doi.org/10.1177/1077559515588131>
- Guyatt, G. H., Oxman, A. D., Schünemann, H. J., Tugwell, P., & Knottnerus, A. (2011). GRADE guidelines: A new series of articles in the journal of clinical epidemiology. *Journal of Clinical Epidemiology*, *64*(4), 380–382.
- Harrington, T. (2008). *The clinician administered PTSD scale for children and adolescents: A validation study*. ProQuest.
- Hodges, M., Godbout, N., Briere, J., Lanktree, C., Gilbert, A., & Kletzka, N. T. (2013). Cumulative trauma and symptom complexity in children: A path analysis. *Child Abuse & Neglect*, *37*(11), 891–898.
- Issakidis, C., Sanderson, K., Corry, J., Andrews, G., & Lapsley, H. (2004). Modelling the population cost-effectiveness of current and evidence-based optimal treatment for anxiety disorders. *Psychological Medicine*, *34*(1), 19–35.
- Kerig, P. K., Bennett, D. C., Thompson, M., & Becker, S. P. (2012). “Nothing really matters”: Emotional numbing as a link between trauma exposure and callousness in delinquent youth. John Wiley & Sons. [https://search.proquest.com/docview/1030207083?accountid=14693&bdid=1370&\\_bd=xld6eh0sU8a4fEkTgA8tQxcP%2B2U%3D](https://ntu.idm.oclc.org/login?url=https://search.proquest.com/docview/1030207083?accountid=14693&bdid=1370&_bd=xld6eh0sU8a4fEkTgA8tQxcP%2B2U%3D)
- Kerig, P. K., Moeddel, M. A., & Becker, S. P. (2011). Assessing the sensitivity and specificity of the MAYSI-2 for detecting trauma among youth in juvenile detention. *Child & Youth Care Forum*, *40*(5), 345–362.
- Kerig, P. K., Ward, R. M., Vanderzee, K. L., & Moeddel, M. A. (2009). Posttraumatic stress as a mediator of the relationship between trauma and mental health problems among juvenile delinquents. *Journal of Youth and Adolescence*, *38*(9), 1214–1225.
- Kretschmar, J. M., Tossone, K., Butcher, F., & Flannery, D. J. (2017). Patterns of poly-victimization in a sample of at-risk youth. *Journal of Child & Adolescent Trauma*, *10*(4), 363–375.
- Lader, D., Singleton, N., & Meltzer, H. (2000). *Psychiatric morbidity among young offenders in England and Wales*. Office for National Statistics.
- McCartan, K. F. (2020). *Trauma-informed practice: HM Inspectorate of Probation Academic Insights*. 2020/05. HM Inspectorate of Probation.
- 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.
- Moore, E., Gaskin, C., & Indig, D. (2013). Childhood maltreatment and post-traumatic stress disorder among incarcerated young offenders. *Child Abuse & Neglect*, *37*(10), 861–870.
- Murray, H. D., Glaser, B., & Calhoun, G. (2013). Using the MMPI-A in identifying trauma symptoms among juvenile offenders. *Journal of Forensic Psychiatry & Psychology*, *24*(3), 337–352.

- Pechtel, P., & Pizzagalli, D. A. (2011). Effects of early life stress on cognitive and affective function: An integrated review of human literature. *Psychopharmacology*, *214*(1), 55–70.
- Perkins, A., Calhoun, G. B., Glaser, B. A., & Kunemund, A. J. (2016). Identifying trauma symptoms among adjudicated youth using the BASC-2 SRP and CROPS. *The Journal of Forensic Psychiatry & Psychology*, *27*(5), 665–683.
- PRISMA (n.d.) <http://www.prisma-statement.org/>
- Rogers, K. L. (1994). *Posttraumatic stress disorder in a sample of conduct disordered youth* [Doctoral dissertation, Theses, Dept. of Psychology/Simon Fraser University].
- Schellingerhout, J. M., Verhagen, A. P., Heymans, M. W., Koes, B. W., Henrica, C., & Terwee, C. B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: A systematic review. *Quality of Life Research*, *21*(4), 659–670.
- Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42.
- Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., De Vet, H. C., & 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.
- Webermann, A. R., & Murphy, C. M. (2019). Childhood trauma and dissociative intimate partner violence. *Violence Against Women*, *25*(2), 148–166.
- Wolpaw, J. M., & Ford, J. D. (2004). *Assessing exposure to psychological trauma and post-traumatic stress in the juvenile justice population*. National Child Traumatic Stress Network.
- Wright, S., Liddle, M., & Goodfellow, P. (2016). *Young offenders and trauma: Experience and impact: A practitioner's guide*. Beyond Youth Custody.
- Zito, A. E. (2016). *Assessing trauma symptomatology among juvenile offenders: An exploration of the BASC-2 and TSCC* [Doctoral dissertation, University of Georgia].

### Author Biographies

**Rachel O'Rourke**, DPsych, Rachel is a forensic psychologist working for HMPPS and her research and clinical work is focused on young adult men in custody. She is specifically interested in this cohort as their numbers increase in UK prisons and we seek to respond better to their needs and the needs of the staff working with them.

**Mike Marriott**, is a Principal Lecturer in Psychology at Nottingham Trent University, and a UK registered clinical psychologist with particular expertise in child and adolescent mental health. His academic work is focused on knowledge exchange and the transfer of research into practice settings, with methodological expertise in qualitative research.

**Richard Trigg**, is a Principal Lecturer in Psychology at Nottingham Trent University. His research work focuses mainly within the field of health psychology and the wellbeing of people with chronic conditions. He has methodological expertise in the use and development of psychometric scales.