RUNNING HEAD: ASSESSING CHILD ADOLESCENT SEXUAL EXPLOITATION

Advancing the Assessment of Child Sexual Exploitation: A Rapid Evidence Review and Preliminary Validation of the CASEE Structured Professional Judgment Tool

Carol A. Ireland [ad], Jane L. Ireland [bc], Svenja Wittrock[d], Teresa Carrasco Montero[d], Annalena Schmid & Phoebe Sands[d]

[a]School of Psychology, University of Greater Manchester, Bolton, Greater Manchester, UK,

[b]School of Psychology, Liverpool John Moores, Liverpool, UK, [c]Ashworth Research Centre, Mersey Care NHS Foundation Trust, Merseyside, Liverpool UK

[d]Coastal Child and Adult Therapeutic Services (CCATS), Poulton-le-Fylde, Lancashire, UK

Conflict of interest statement: CI and JI acknowledge an interest in the development and dissemination of the CASEE tool with further projects under development. The authors report there are no further competing interests to declare

Word count including supplemental material: 12,614

Word count excluding supplemental material: 10,562

Abstract

Purpose: This paper explores the quality of assessments available where a child is felt to be vulnerable to child sexual exploitation (CSE). In particular it considers the preliminary reliability and validity of the Child Adolescent Sexual Exploitation Evaluation (CASEE, Ireland et al, 2025), which aims to consider vulnerability and protective factors in a child felt to be exposed to CSE concerns.

Design/Methodology/Approach: It comprises of a Rapid Evidence Assessment (REA), followed by a study of 94 professionals who completed a two-day CASEE validation event (3 events in total).

Findings: The REA captured a continued absence of theoretical and empirical background to currently available CSE assessments for children, and a continued failure to consider a SPJ (Structured Professional Judgement) approach. Taking a SPJ approach to assessing CSE using the CASEE, this initial study demonstrated a good level of inter-rater reliability overall, with limited disagreements between raters.

Originality: This paper argues the continued lack of CSE assessment tools that are empirically and theoretically developed, that take advantage of the SPJ robust approach, in addition to considerations of reliability and validity. The CASEE is outlined as an emerging example of an empirically informed SPJ tool able to begin to consider some of these gaps.

Keywords: Child Sexual Exploitation, CASEE, Vulnerability, CSE risk, CSE protective factors, SPJ

Introduction

Child Sexual Exploitation (CSE) is noted in the literature as an example of child sexual abuse where the victim 'exchanges' sexual activities for some kind of 'reward', as a result of adult perpetrator coercion, manipulation and/or grooming. This 'reward' for the young person may be tangible (e.g., money, food, drugs, gifts) and/or intangible (e.g., affection, protection). The term 'exchange' has been previously applied but correctly challenged as wrongly implying a level of agency by the victim, when this is indeed an abusive encounter where the perpetrator(s) holds the power (Laird et al., 2022). Perpetrators take advantage of and/ormanipulate a child's vulnerabilities. Such vulnerabilities can include varying forms of child abuse and neglect, mental health problems, substance abuse, lower socioeconomic status, and a greater risk of re-victimisation across the lifespan (Astridge et al., 2023; Petruccelli et al., 2019). Poly-victimisation, the exposure to more than one type of victimisation in childhood, is further found to increase the vulnerability of exposure to CSE (De Vries et al, 2019). Exposure to traumatic experiences such as these, especially frequently, can lead to poor coping and a distorted view of self (Alderson, 2019), with self-denigration further increasing in young people exposed to CSE (Alderson et al, 2022).

The prevalence of CSE and consequent risk to young people however is not yet fully understood, yet accepted to be a problem of sizeable concern (Alderson, 2019). For example, a 12-month period of review from 2021 to 2022 indicated 16,330 completed assessments by local authorities in England that identified concerns of CSE (Department for Education, 2022). Alderson et al. (2022) further suggested from their study that every second child under the age of 16 may be approached sexually by an adult. Of those who are initially approached, one in four children may become subject to sexual exploitation. Although their study was cross-sectional and self-report, it nonetheless highlighted CSE as a significant issue worthy of increased interest. Indeed, Alderson et al. (2022) argue that figures obtained by local

authorities likely under-represent the problem, since many victims do not come to the attention of professionals.

In addition, the gender of victims may impact, with male CSE victims appearing less likely to be identified than females, with research suggesting they are less likely to be believed than female victims (Hill & Diaz, 2021). Signs of CSE in boys have arguably been misinterpreted as signs of dissocial behaviour, whereas females can be recognised more quickly, and in comparison, as CSE victims (Hill & Diaz, 2021). Underpinning the failure to recognise CSE is arguably a lack of professional knowledge. Mason-Jones & Loggie (2019), for example, analysed serious case reviews and concluded that most professionals lacked an understanding as to what factors make a child vulnerable to CSE. This then makes assessment of such concerns difficult.

The need to therefore capture the vulnerability of potential and actual CSE victims is paramount. CSE has further become an important topic of contemporary academic interest, following high-profile CSE cases in recent years, where child protective services have failed to recognise and protect children subject to CSE (Independent Office for Police Conduct, 2022; Jay, 2014). A lack of awareness and poor understanding of CSE among professionals has been identified as a key contributor to these failings, with the need therefore for good-quality, child-focused risk assessments (Jay, 2014; Independent Office for Police Conduct, 2022) that consider vulnerabilities in potential or actual victims of CSE. Such assessments are a key component to determining a child's vulnerability to CSE. This had led however to the swift development of assessment tools, arguably of limited validity. For instance, in their exploratory study among 42 practitioners, Brown et al. (2017) identified at least 19 different screening and assessment tools used across England; none of the tools were empirically and/or theoretically developed, but largely based on practitioners' own clinical experiences. Consequently, these tools include many risk factors not supported by research evidence,

further risking the omittance of empirically validated risk, as well as protective factors (Brown et al., 2016; Brown et al., 2017; Franklin et al., 2018) that can reduce or buffer against risk. In addition is the notion of 'risk' in victims of CSE when assessing concerns, and where such language can drive a potential attribution of blame toward victims. Instead, terms such as vulnerability and protective factors may be more helpful in assessment, such as that noted in Alderson et al's (2022) *Protect and Vulnerability Model against CSE* (PVM:CSE), detailed below.

Whilst the use of language in such assessment tools has been queried, additional concerns have focused on their assessment methods, where assessors are required only to tick boxes and compute a total score, therefore reducing meaningful application. Most survey respondents in Brown et al. (2017) felt this lead to tensions between professional judgements and the use of these tools, where practitioners felt unable to add case-specific factors or adjust arbitrary risk categorisations. Further, tools varied considerably in how the information of the assessment is combined and used to aid decision-making for professionals (Brown et al., 2016; Brown et al., 2017; Franklin et al., 2018). Indeed, whilst developed with good intent, the validity and reliability of several of these CSE assessment tools require focus and improvement in order to support their overall utility. This is important, as most practitioners expressed finding these tools valuable and useful when considering the CSE concerns for victims, but noted a need for clearer guidance in such CSE assessment, and in order to improve practice (Brown et al., 2017).

Further, and when considering currently available CSE assessment tools that look at determining the vulnerabilities of potential CSE victims, several continue to be developed for child/youth services that appear to exclude the scientific literature. Such exclusions include not only the research and theoretical literature concerning CSE, but also how such assessment tools should be reasonably developed in terms of their construction. Whilst the aim of the

current paper is not to discuss in detail all available measures, as this would extend past the papers specific purpose, a summary of the methodological and construction difficulties in currently available CSE victim assessment tools is noted in supplementary table S1. However, to illustrate by example would be the Canadian Sexual Exploitation among Youth (SEY) risk assessment framework screener (Pennsylvania State University: Panlilio et al., 2022). This tool considers young people aged 11 to 17 years, assessing 26 items across seven domains in regard to CSE concerns. Item generation was based on a national survey, incorporating a yes/no criteria, generating a final cut off score that indicates a low and high-risk rating. The SEY focuses solely on 'risk' factors, omitting any theoretical underpinning. It further fails to address protective factors that could then be utilised in the reduction of vulnerability in a child, alongside a failure in capturing inter-rater or test-retest considerations. It lacks individualised clinical formulations of understanding and potential future scenarios of vulnerability (i.e., scenario planning), leading to a difficulty in comprehensively and meaningfully understanding CSE concerns in potential victims, including support in conceptualising strategies of intervention and/or prevention. By some contrast, a more robustly developed CSE victim assessment tool would be the tool for the detection of risk of sexual exploitation in childhood and adolescence (EDR-ESIA) (Institut Balear de la Dona: Benavente et al., 2022). This is available in several languages, capturing young people aged 11 to 18 years, using 88 items across six domains, with some evidence for validation. Item generation was, importantly, based on a review of the literature, further complemented by Delphi expert consultations (Benavente et al., 2021) and expert review. However, the EDR-ESIA focuses on risk items, and where we would argue in this paper that the term 'vulnerability' is a better fit, with no attention to protective factors, formulation, scenario planning or inter-rater/test-retest considerations. Although showing promise when compared to the SEY, the EDR-ESIA continues to suffer from a lack of clear theoretical

underpinning. This is alongside a lack of attention to the approach adopted by Structured Professional Judgement (SPJ) tools when developing tools of assessment.

The latter is an important issue. There are clear, long-standing, evidenced developments in offender risk assessment over the last 30 years that attend to the concept of Structured Professional Judgements (SPJ) in tool construction (e.g. Olver, Stockdale, Helmus,Woods, Termeer and Prince, 2024). Whilst such criminological and forensic assessments consider offender risk, such as sexual and aggressive risk, the focus of assessment tools for victims of CSE would suggest clear value of the SPJ approach being applied, aiding to their construction and development. Such a union has yet to take place. Furthemore, research has long established the superiority of the SPJ approach in terms of predictive accuracy, reliability, fairness, accountability and utility over reliance on unstructured clinical judgment (e.g. Viljoen and Vinent, 2020). Unstructured clinical judgement is overly reliant on the clinicians view of concerns without clear theoretical and empirical guidance, and where notably the majority of current available CSE assessment tools would be placed. In addition, the SPJ approach argues for a move away from reliance on computing a single risk score, the previous actuarial approach, focusing instead on the integration of risk and protective factors alongside their critical relevance, and in a more dynamic way. The end result is a qualitative and individualised understanding of a client's risk, using the clinician's knowledge in a structured way. This includes future risk scenarios of concern that can be communicated to manage and prevent risk materialising and is far more than a simply summing up of risk factors (de Vogel, de Beuf, Shepherd & Schneider, 2022). SPJ approaches further avoid the reductionist approach of actuarial assessments that aim to reduce risk to a scored value but fail to account for context and the importance of dynamic change in risk and/or protective factors. Context and change are arguably the

cornerstone of effective risk management and, consequently, should be reflected in the CSE assessments being developed, and with reference to risk in terms of vulnerability.

The further importance of incorporating theory in the SPJ approach amplifies the large neglect of this in current CSE victim assessment tools. For example, theories of adverse childhood experiences are certainly evidenced but seemingly not routinely *integrated* into the current CSE assessment tool process. For example, according to *Ecological Systems Theory* (Bronfenbrenner, 1977, 2005; Campbell et al., 2009), a victim's well-being is actively shaped by several factors across different ecological systems. These factors include negative social reactions to disclosure, self-blame, and poly-victimisation, all frequently experienced by victims of CSE (Alderson, 2019, Alderson et al., 2022) and further exacerbating the trauma and detrimental impact on victims (Campbell et al., 2009). Thus, inclusion of such factors in a CSE vulnerability assessment tools become of paramount importance. More recently, ecological understandings have been incorporated into integrated theory that does attempt to capture CSE as a more unique adverse experience, referencing the value of vulnerability and protective factors, as opposed to risk. For example, the *Protect and Vulnerability Model against CSE* (PVM:CSE; Alderson et al., 2022), identifies vulnerability and protective factors associated with child victims of CSE, as well as the mechanisms through which vulnerability can be aggravated. PVM:CSE, for example, describes the factors that place a child on the 'vulnerability pathway', noting how strain can be caused for the child and proposing how pre-existing vulnerability interacts with situational factors to increase or decrease the risk of a child's exposure to CSE. Protective factors are further incorporated, with these serving to reduce or buffer against CSE risk and the longer-term negative impacts of CSE exposure. Such theories underpin the need to integrate our understanding of CSE into a workable formulation where vulnerability can be understood, managed and prevented. As such, this paper introduces preliminary reliability and validity of a new CSE victim assessment tool, the

CASEE (Child Adolescent Sexual Exploitation Evaluation, Ireland et al, 2025, Ireland and Ireland 2024), which is proposed as a new SPJ tool that begins to address the current difficulties in the area of CSE assessment. It is outside the focus of this paper to comprehensively present the theoretical and empirical construction of the CASEE, alongside the detailed development of the items in line with such theoretical and empirical data; readers are instead directed to Ireland and Ireland (2024) for a detailed overview on the CASEE development and its origins in a theoretical underpinning. Instead, this paper focuses on some preliminary considerations of its use solely in regard to initial reliability and validity. .

This paper therefore aims to address research gaps by initially providing a review of currently available CSE assessment tools using a Rapid Evidence Assessment (Part A), to efficiently yet robustly determine the nature and approaches taken. This is then followed by a study of 94 professionals (Part B) completing case studies using the CASEE (Ireland et al, 2025) as part of initial reliability and validation of the tool, using the findings of the REA to help drive Part B focus. This paper therefore aims to provide a comprehensive outline of current approaches to CSE victim vulnerability assessment, identifying benefits or otherwise of adopting an SPJ approach. As this is a groundwork study of some initial reliability and validation of a new assessment tool (the CASEE), further informed by a Rapid Evidence Assessment, no predictions are offered. Yet, this paper is presented as an introductory reliability and validation of the first available SPJ tool for CSE victim assessment, the CASEE.

Part A: Rapid Evidence Assessment

To examine the recent state of the field, a Rapid Evidence Assessment (REA) was conducted using academic and grey literature on currently available CSE victim assessment tools, considering their empirical and theoretical quality, validation and overall utility. The REA was conducted following the Authority/NERC guidelines for the production of Quick

Scoping Reviews and Rapid Evidence Assessments (Collins et al., 2015). An REA was decided as it can offer trustworthy evidence summaries rapidly and efficiently, and whilst less rigorous than a systematic review, it is more rigorous than more ad-hoc searches, and less time-consuming. The research team felt this was more efficient to identify the current and more limited state of knowledge on the topic area, and in order to keep the research timely. This REA was not used to develop the CASEE, as this was subject to much more detailed development, and as summarised in Ireland and Ireland (2024). Moreover, the REA was utilised to assist with the more current understanding of the available CSE victim assessment tools.

Method

Scope and search keywords

Inclusion criteria: Studies were considered eligible if they (a) covered the topic of CSE assessment; (b) focused on assessment of victims or potential victims of CSE; (c) were published in the English language; (d) were published since January 1st 2000; and (e) were available in full text. Specifically the keywords (search strings) were child sexual exploitation OR child sexual abuse and exploitation AND tool OR toolkit OR checklist OR protocol OR questionnaire OR instrument OR test OR inventory OR scale OR risk OR risk assessment OR assessment OR assess OR evaluation OR evaluate OR screening OR screen. Search engines Google and Google Scholar were used, with academic databases PsychArticles, PsychINFO, PubMed, Scopus and Web of Science. 18 organisational websites were also accessed.

Screening Process and Quality Assessment

Articles were screened for relevance. All papers included were examined, with a list of CSE assessment tools generated. When possible, the databases were set to pick up articles containing the keywords related to CSE in the title (i.e., 'child sexual exploitation' or 'child sexual abuse and exploitation') and the keywords related to assessment (e.g., 'tool',

'protocol', 'inventory'…) in the text of the article. If this was not allowed in the database, then the database was set to pick up articles containing the keywords in the title and/or abstract (and when possible, in both).

To assess the quality of each of the identified CSE tools, the classification system developed by Benavente et al. (2023) was employed. This system considers if a tool has been validated and categorises the strength of the evidence into four levels. *Solid* evidence refers to multiple pieces of evidence on different types of validity and reliability, or a systematic review with consistent adequate psychometric properties. *Moderate* evidence refers to one or two pieces of evidence concerning the validity and reliability of the tool. *Limited* evidence refers to only qualitative evidence of content validity being available. Finally, *No evidence* indicates that no data is provided regarding the sources of evidence of validity or the psychometric properties of the tool (Benavente et al., 2023).

Results

In total, 479 articles were identified through online searches across databases. After removing duplicates, the sample was reduced to 291 papers. Titles and abstracts were screened, leading to 30 papers eligible for inclusion. After full-text assessment, five papers were excluded for being relevant to CSE assessment but not covering assessment tools. A scrutiny of the bibliographic references led to the addition of two more papers. Twenty-seven papers were thus included in the REA, comprising 19 CSE tools. Figure 1 illustrates the screening process. Table S1 summarises the 19 CSE tools in terms of their characteristics and quality rating based on Benavente et al (2023), and as noted in the introduction.

<Insert Figure 1 here>

Current CSE Assessment Tools

Nineteen distinct CSE assessment tools were identified, all available in English (with the addition of the EDR-ESIA also available in Spanish and Catalan). Eight out of the 19 tools referred to the assessment of CSE as the assessment of Commercial Child Sexual Exploitation (CSEC) and/or sex trafficking; these terms were used interchangeably (Benavente et al., 2023). None of the tools provided a theoretical basis, neither was it possible to find references to theory behind the tools' development in further searches. As for the empirical foundation guiding the development of the tools items, only six of the tools derived items from scientific studies (Andretta et al., 2016; Baird et al., 2020; Benavente et al., 2022; Clutton & Coles, 2007; Greenbaum et al., 2018; Panlilio et al., 2022). For two tools, interviews and focus groups with stakeholders were carried out to gather information for item development (Asian Health Services and Banteay Srei: Chang et al., 2015; WestCoast Children's Clinic, 2016). For the remaining 11 tools, the item generation procedure was not mentioned in the tool itself and could not be discovered through further investigation.

Content of the CSE Assessment Tools

Sixteen of the 19 tools did not consider the inclusion of protective factors. Two considered protective factors within the professional judgment section in the form of an open question with no guidance. One did not consider protective factors in a designated section, but the presence of protective factors could be identified in the rating of risk factors. None of the 19 reviewed tools included reference to the use of case formulations and/or scenario planning when considering concerns in the child. All tools provided guidance in concluding whether the child or young person subject to assessment was at risk of CSE, yet the criteria to determine CSE risk varied across tools. While some tools established a specific threshold to

determine the presence of CSE risk (i.e., cut-off score), others distinguished between risk levels based on different scores. In some tools, the presence of one single indicator was considered sufficient to conclude that immediate action was required.

Evaluation of the CSE Assessment Tools and application to SPJ

In the context of reviewing these 19 tools, validation refers to the process of providing empirical evidence supporting the quality and usefulness of an assessment tool for its intended purposes (Sireci & Padilla, 2014). Out of the 19 tools reviewed, seven had undergone some form of validation. However, none reached a *solid* level of evidence (defined as having multiple pieces of evidence on different types of validity and reliability, or a systematic review with consistent adequate psychometric properties, Benavente et al., 2023). All tools lacked a theoretical background, failing to consider the use of case formulations and scenario planning to help conceptualise the concerns for the child and then develop a plan of support. None therefore followed the expectations of an SPJ approach. Furthermore, the majority of tools (84%) failed to consider protective factors that can reduce or buffer against vulnerabilities in the assessment of CSE. Finally, 58 per cent of tools included no evidence as to how items had been developed.

Summary conclusion

Despite evidence for a wide number of CSE tools having been developed, there was clear indication of these having been constructed outside the expectations of SPJ. In addition, there was a lack of empirical and theoretical basis to these tools, with little or no attention to validation. The CASEE (Child Adolescent Sexual Exploitation Evaluation) is therefore presented as a newly developed tool that can begin to bridge these gaps. The CASEE follows the expectations of SPJ, drawing from empirical and theoretical developments, with the value of this being reaffirmed by the findings of the REA. The CASEE has, until now, not been considered in relation to reliability, with the REA indicating a tendency for tools to neglect

this aspect. Consequently, Part B below builds on the REA by examining the initial reliability of the CASEE tool, alongside thematic components of the SPJ approach.

Part B: Inter-rater reliability of the CASEE items via professionals scores and thematic analysis of case formulation

Part B aims to examine the reliability of the tool, as noted by the REA, alongside a thematic analysis of the formulation SPJ elements neglected from the other CSE tools.

Method

Participants

One hundred and twenty five individuals who had experience of working with children expressed an interest in participating in a two-day live online taught validation training event that was delivered on three occasions over a three month period. 112 completed a consent form and received event details. Eighteen individuals did not attend, resulting in a final sample of 94 participants. Ninety-one provided information on their demographics; 85.7 per cent reported their gender as female, with 11% male and 3.3% not disclosing. A third (34%) were forensic psychologists, with 11% therapists and 9.9% social workers. A third (34%) identified as 'other', but all professional disciplines. Only a small proportion (4.4%) had worked for less than a year, with the most frequently reported period between one and five years (30.8%). Only 11 per cent did not complete risk assessments, with 69.2 per cent completing reports on CSE. A full summary of the participant demographics is presented in Table S2.

Materials

Child Adolescent Sexual Exploitation Evaluation (CASEE): The CASEE (Child Adolescent Sexual Exploitation Evaluation, Ireland et al. 2025) is an SPJ tool for assessing children who are at considered vulnerable to CSE, including suspected to be, or have been exposed to CSE. The aim of the CASEE is to assist practitioners who work with this

population by offering an empirically and theoretically driven structured assessment guide

that focuses on key areas of interest to help in understanding a child's vulnerability, alongside

what may reduce or buffer against such vulnerabilities. It encourages assessors to exercise

their professional clinical judgement in understanding and formulating CSE concerns,

aligning to the SPJ approach. By emphasising the dynamic nature of vulnerability and

protective factors, the CASEE provides guidance on how to manage a child's CSE

vulnerability. As noted earlier, a more detailed summary of the CASEE is noted in Ireland

and Ireland (2024). Yet, in brief, the CASEE includes 18 vulnerability items with 44 sub-

items. The vulnerability items are divided into three domains: *general vulnerability items,*

strain for the child items, and *situational factor items.* It also includes eight protective factors.

Each factor is rated as 'clearly present', 'somewhat present', 'no evidence' (or 'absent' for

the protective items), with detail as to how each factor should be rated. Each item contains a

description offering brief context and theory. Each vulnerability item is rated on the last six

months, and each protective factor item projecting the next 6 months, to illustrate the

dynamic nature of exploitation. Once the CASEE ratings are completed for the young person,

the professional is encouraged to complete a formulation and scenario plan, which then

directs consideration of management strategies to support the child felt to be at risk (Ireland

et al., 2025, Ireland and Ireland, 2024). The aim is to understand and reduce vulnerability, to

increase protective factors, and to develop an individualised understanding of concerns for

the child. This is all in order to maximise the potential for successful intervention and

management.

All participants were provided with a digital copy of the CASEE manual (version 1),

which included background on the tool, the CASEE items and scoring. An abridged example

of a CASEE vulnerability and protective factor is presented in Figure 2.

<Insert Figure 2 here>

Case Vignettes: Three fictitious case vignettes were used as part of the training, allowing sufficient detail to score each of the CASEE items, and designed to resemble actual cases and reflect the diversity of CSE. The order in which these cases vignettes were presented were counterbalanced across each event, to minimise practice effects. A summary of the profiles of the three cases is presented in Table 1.

<Insert Table 1 here>

SORC formulation Assessment: All participants completed psychological case formulations using a SORC formulation template based on the presented vignettes. This is a general formulation approach (Lee-Evans, 1994; Ireland, 2018), with the SORC comprising of four sections: Setting conditions (antecedents), Organism variables (background factors, beliefs), Response (the CSE) and Consequences (positive and negative reinforcement). The SORC was used to determine participants' views regarding potential triggers of the child's vulnerability contributing to the CSE concerns, the child's background and predisposing factors contributing to their vulnerability, the nature and circumstances of the CSE behaviour taking place, and consequences in regard to the positive and negative reinforcement elements that could maintain vulnerability.

Procedure

Ethical approval was granted by the Coastal Child and Adult Therapeutic Services (CCATS) ethics review committee. Prior to the training events, five members of the research team involved in data collection (CI, SW, TCM, AS, PS) independently scored the case studies, before meeting as a group and calibrating their answers. These final ratings were then used as

a comparison with participants scoring. In order to try to reduce a potential for vested interest, only one out of the five members of the research team who collected data was involved in the development of the CASEE as an author. This member of the research team purposely did not lead on any data analysis.

The CASEE validation online training events were open to professionals who worked with children and/or who had clinical knowledge in the CSE area (e.g. psychologists, social workers, therapists, teachers). Following receipt of a consent form, their place on one of the three events was confirmed. Each training event consisted of an introduction to the CSE literature and the CASEE, before progressing to score each of the three case vignettes independently. Finally, participants completed psychological formulations and a scenario plan for each vignette in small groups. These were then compared to the research team ratings, formulation and scenario plans.

Results

This section will present the inter-rater reliability, agreement with prior consensus ratings, levels of agreement and overall thematic analysis of the case study formulations. There was no missing data.

Inter-rater reliability

Inter-rater reliability (IRR) was initially examined by means of the intraclass correlation coefficient (ICC). A single-measure, two-way random effects model and absolute agreement type was used ($ICC_{A,1}$; McGraw & Wong, 1996). This was initially selected as it is the most common statistic for examining IRR. However, the ICC suffers from the limitation that it requires substantial variance in cases in order to obtain acceptable values (Gwet, 2021). Since this study used a smaller set of case vignettes, and results showed little variance in cases, an alternative reliability coefficient, *Finn's r* (Finn, 1970), was used. This

is conceptually different from the ICC, but not affected by the amount of variance in the sample. Since the coefficient was designed for categorical data, analyses were conducted for individual items only. Consistent with previous studies (e.g., Selby, 2018; Sutherland, 2012), raters who omitted an item (missing a rating) for at least one of the three cases were excluded on an item-by-item basis. Therefore, coefficients were calculated on a different number of raters for each item. Benchmarks by Fleiss (1986) were used for interpreting Finn's coefficient: $r \geq .75$ = excellent; $.60 \leq r < .75$ = good; $.40 \leq r < .60$ = moderate; $r < .40$ = poor. Finn's coefficients were computed using R (v4.3.3) and the irr package (v0.84.1; Gamer et al., 2019).

Overall, few raters chose to omit an item. Out of 7,254 ratings in total, 84 (1.16%) were 'omit' ratings. Item 18 ('other considerations') was the item that was omitted most often by participants, with 18% of ratings being omitted across all cases. Among the other items, the percentage of omitted ratings did not exceed 4%.

Table 2 shows Finn's coefficients for the individual CASEE items. Finn's coefficients ranged from 'moderate' to 'excellent' IRR (r = .45 to .87); 23% of items achieved 'excellent' reliability, 54% 'good' reliability, and 23% 'moderate' reliability. The mean Finn's r across all items was .65. This suggests that overall, CASEE items achieve IRR in *at least* the 'good' range.

<Insert Table 2 here>

Agreement with prior consensus rating

The level of agreement between participants' ratings and the consensus ratings already determined by the research team prior to the events (research team prior consensus rating) were considered. Whilst there are no agreed-upon benchmarks regarding what level of

agreement was acceptable, for this current study, absolute agreement of 60% or higher was considered good as this was well above chance agreement. However, and whilst percentage agreement is informative of absolute agreement, it does not take the level of *dis*agreement into account, e.g. whether participants scored 'no evidence' or 'somewhat present' instead of 'clearly present' when compared to the research team ratings. In order to examine the degree of disagreement, the mean of ratings which were not in absolute agreement with the research team ratings were calculated.

One sample *t*-tests were then conducted to examine whether participant mean total and domain scores for each case and across all cases differed significantly from the research team. The average level of absolute agreement with the research team prior consensus rating was 63.7% across all cases. At the item level, absolute agreement ranged from 30.1% to 93.2%. 18 items (69%) of the 26 CASEE items achieved agreement above 60%. This shows that participants achieved a good level of agreement with the research teams consensus rating for two-thirds of the CASEE items. Items achieving the highest level of agreement ($\geq 80\%$) were items 1 (physical isolation), 9 (social attachment with peers), 13 (instability in self-perception), 22 (supportive peer relationships), and 25 (positive view of self). Items with the lowest level of agreement ($\leq 40\%$) were items 15 (lack of control), 17 (normalisation of harmful experiences), and 18 (other considerations). Items 17 and 18 achieved 32.6% and 30.1% agreement, which is less than what would be expected by chance.

As a supplementary analysis, Spearman's rho correlation was computed to examine the association between average percentage agreement values and Finn's reliability coefficients. There was a strong, positive correlation between these measures, $r_s(24) = .82$ ($p < .001$). This suggests that the level of agreement with the research teams consensus rating is strongly associated with the IRR of an item.

To examine the level of *dis*agreement for items that received a research team rating of 2 ('clearly present') or 0 ('no evidence'), the mean difference between participant scores and the research team score was calculated. Across all cases, the mean difference was 1.07, which ranged from 1.00 to 1.43 for individual items. This demonstrated that the vast majority of participant ratings that did *not* match the research team rating, deviated by only one category. For example, when an item was rated as 'clearly present' by the research team, participants mostly decided between 'somewhat present' and 'clearly present', as opposed to 'no evidence' and 'clearly present'.

Thematic analysis of CASEE informed psychological case formulations

Reflexive Thematic Analysis (RTA: Braun & Clarke, 2006; Percy et al., 2015) was employed to identify prevalent themes in the participants case formulations for two of the case studies (case studies 'Jessica' and 'Richard'), and to compare these to themes derived from the research team. To ensure scientific rigor, the themes were reviewed by another researcher blind to the aims of the study, to ensure reliability of the themes. Any disagreement was resolved through discussion and alterations were made when necessary. The theme identification was completed based on the areas highlighted in the SORC framework: setting conditions, organism variables, response, consequences, and hypotheses.

Assessment of agreement: Prevalence calculation

Assessment of agreement was conducted between the Jessica and Richard case studies SORC formulations completed by participants and compared to the research team using a prevalence calculation. Collective trends across participants formulations (12 formulations in total as completed in small groups) were considered. The prevalence of a formulation theme was calculated as a percentage, indicating the number of formulations (out of 12) in which a theme was identified. For example, if a theme was identified in 6 out of 12 formulations, its prevalence would be 50%.

Levels of agreement

Themes identified in the research team's formulation were compared to the prevalence of the same theme in the same section of the participants formulations. Percentage agreement in qualitative research has no universally agreed-upon set of guidelines, and so reference to Cohen's Kappa (McHugh, 2012) was innovatively used as a basis for the assessment of agreement. This classification system was used to categorize levels of agreement between both groups using the labels of very low (1 – 25% agreement), low (26 – 42% agreement), moderate (43 – 59% agreement), high (60 - 75% agreement), very high (76 – 99% agreement) and perfect agreement (100%). The detailed criteria for assessing the level of agreement in noted in S3. A minimum level of high agreement (60 – 75%) was used as the threshold for agreement of themes. Table S3 notes the set criteria for assessing the level of agreement.

Upon completing the thematic analyses, both concordance and discrepancies emerged between the research team and participants formulations. Regarding concordances, because the thematic analyses of both group's formulations were completed separately, themes covering the same topic were often named differently. However, reading the supportive evidence revealed a complete overlap. For instance, the theme 'Impaired and negative sense of self' in the research teams formulation corresponded to the theme 'Negative and unstable self-image' in the participants formulation. Discrepancies included themes present in research teams formulations but absent in participants, and the reverse. It is not possible in the restrictions of this paper to offer all themes that were noted, yet examples are offered in each section below.

Thematic results

SORC: Setting Conditions (Antecedents leading up to the CSE concerns)

Across the setting conditions section, the research team considered 17 themes in total. Of these 17 themes, participants correctly considered eight in the setting conditions' section, two in other sections of the formulation, and failed to consider seven. Furthermore, two were considered by participants as setting conditions that the research team did not include in this section, but did consider in other sections of the formulation. Example themes of agreement included 'distrust in interpersonal relationships and absence of support systems in the child's life', 'maladaptive coping' and 'negative and unstable sense of self'.

SORC: Organism Variables (Background history, past experiences the young person already bringing to this CSE concern situation)

Across the organism variables section, the research team considered 19 themes in total. Out of these 19, participants correctly considered 11 of these in the organism variables section, five of these in other sections of the formulation, and failed to consider three. Example themes of agreement included 'direct and indirect exposure to abuse', 'unstable and neglectful home environment and disinterested parenting' and 'insecure attachment within the family'.

SORC: Response Behaviours (Nature of the CSE risk victim exposed to and their response)

Across the response behaviour section, the research team considered two themes in total. Out of these two themes, participants correctly considered one of these as a response behaviour and considered the other in another section of the formulation. Four themes were considered by participants as response behaviours that the research team did not include in this section. Yet, two of these were considered by the research team in other sections of the formulation, and the other two were not considered at all by the research team. Example

themes of agreement included 'perception of attachment to abusive adults' and 'exploitative dynamics in relationships with older males'

SORC: Consequences (Positive and negative reinforcement of the CSE risk, the needs being manipulated by the perpetrator)

Across the section of consequences, the research team considered 11 themes in total. Out of these 11 themes, participants correctly considered eight of these as consequences, yet failed to consider the remaining three consequences. Example themes of agreement included 'gains in sense of belonging', 'provision of material goods' and 'gains in psychological comfort and positive mood, and removal of distress'.

It was observed across the formulations of the vignette's that participants tended to misunderstand the conceptual intent of the 'consequences' section. Consequences here are to reflect the needs of the victim that are being manipulated/taken advantage of by the perpetrator. This misinterpretation occurred with a prevalence of 75% for one vignette (in nine out of the 12 formulations), and of 66.7% for the second vignette (in eight out of the 12 formulations).

SORC: Hypotheses (Determining what created the CSE vulnerability for the victim)

Eleven themes in total were considered across the vignettes. Out of these 11 themes, participants correctly considered four of these in the hypotheses section, considered six in other sections of the formulation, and failed to consider only one theme. Four themes were considered by participants as hypotheses that the research team did not include in this section, but did consider in other sections of the formulation. Example themes of agreement included 'alleviation of distress caused by loneliness and insecure attachments', 'obtaining emotional security and a sense of attachment' and regulation of self-identity'.

Professional confidence

Finally, practice with rating the 3 case vignettes over the 2-day events did not improve rating accuracy overall, yet completion of vignettes across the 2 days event markedly increased practitioners' confidence in their knowledge and ability around CSE risk assessment.

Discussion

The Rapid Evidence Assessment (REA) noted a lack of comprehensive CSE assessments that possess theoretical and empirical underpinning, informed by Structured Professional Judgement (SPJ) approaches. Further, 16 of the 19 tools assessed failed to consider protective factors in the assessment of CSE, with 12 out of the 19 tools failing to have undergone any validation. This amplifies the importance of a theoretically and empirically driven SPJ informed CSE assessment tool, such as the CASEE, which is then subject to some level of reliability assessment.

Whilst the CASEE is in its infancy regarding reliability and validation , the overall inter-rater reliability (IRR) in this study for the tool was at least 'good', with the highest individual ratings agreement of over 80% (excellent agreement) for five items, namely physical isolation, social attachment with peers, instability in self-perception, supportive peer relationships and positive view of self. One of the lowest items for agreement was that of 'other considerations'. This is not surprising as this is a fluid variable and a more open item aimed to capture any additional components in the child's vulnerability not already considered in the tool's ratings (and therefore not captured in the literature to date). As such, this item is more encouraging of a professional's own clinical opinion of what may be missing, but felt important to consider.

Whilst not in full agreement across every item, the CASEE presented with good inter-rater reliability overall; the level of agreement with the research team's consensus rating was strongly associated with the inter-rater reliability of the item. Any rater disagreements tended

to be within one category difference of rating, such as 'clearly present' and 'somewhat present'. The level of IRR evidenced for the CASEE in this paper is therefore comparable to findings of other SPJ youth risk assessment tools, such as Sutherland et al. (2012) and Wilson (2013); when evaluating the accuracy of ratings for an assessment tool on sexual aggression, there were mean levels of agreement of 49% and 76%, respectively. For the CASEE, the mean level of agreement was 64%, and where all participant ratings were in absolute agreement with the research team consensus rating. This demonstrates emerging comparability with other SPJ tools, such as Sutherland et al (2012) and Wilson (2013). This does of course mean the remainder were not in agreement. Yet, disagreements tended to be within one category of difference, as noted above. SPJ tools welcome the professionals clinical view, and therefore some level of disagreement as opposed to absolute agreement are expected, but with an expected majority consensus of agreement overall. This is certainly noted for the CASEE, which demonstrates an emerging potential in what is a new CSE assessment tool.

Formulation of the child's vulnerability using the scoring of the CASEE items was a valuable component, showcasing the need to capture this as part of the assessment process. This is a dynamic process where the professional aims to weave together the vulnerability considerations that have emerged from these items, gathering this into a coherent story of understanding using a guided formulation structure, in this instance the SORC. As such, exact replication of themes across raters would not be expected, even more so when considering complex concerns such as CSE. Despite this, out of 60 themes identified, 46 noted by participants matched those themes also identified by the research team, albeit with 14 of these themes not falling within the same section of the SORC assessment. Overall, this suggests at least similar areas of observation in developing these formulations, with more similarity between the research team and participants than dissimilarity.

The findings overall note the developing importance of an empirically and theoretically informed SPJ assessment tool when considering emerging reliability. This allows the advancement of awareness and understanding of CSE in professionals, that is theoretically and research grounded, previously noted as lacking via independent inquiries (Jay, 2014). This is especially where it has been noted that most professionals lack an understanding of what factors make a child vulnerable to CSE (Mason-Jones & Loggie, 2019). Overall, this paper presents the CASEE as a new and emerging SPJ assessment tool that can begin to provide support for professionals attempting to understand the vulnerability and protective factors of a child where concerns of CSE are raised, and to do so with a level of reliability in a tool that has empirical and theoretical underpinning.

In addition, the delivery of the validation events, served to increase confidence of understanding for professionals in a complex area. This is arguably of importance, as serious case reviews have indicated the need for confidence to be developed in professionals to better serve their understanding and then management of CSE concerns. This ultimately can help reduce the risk of further serious case review inquiries. This is especially pertinent when concerns have been raised that professionals may not recognise CSE concerns or may blame victims for their abuse (Alderson et al., 2019). Yet, even outside of the professional's completion of various case studies, the CASEE (as an example of a SPJ tool for CSE) ameliorates such concerns via its comprehensive presentation of vulnerability factors and use of language that ensures victim blame is not actively used or encouraged.

Overall, this paper argues the importance of the CASEE as an emerging example of an empirically developed tool that is in line with the value of individualising, understanding and explaining a child's CSE concerns, rather than an arbitrary summing up factors (de Vogel1, de Beuf, Shepherd & Schneider, 2022).

Strengths and limitations

The research is not without limitations, and this is only a preliminary study of reliability on a new SPJ tool in the area of CSE. The REA demonstrated clear evidence of limitations in the development of currently available tools, yet this does not mean these CSE tools are without value, but rather the articulation of evidence supporting their content (validation) and reliability is generally not crystalised. Regardless, it is accepted that there will be value within these tools, hopefully further determined by future research. However, the absence of an already developed SPJ informed tool that could be compared to the CASEE to determine concurrent reliability is an unavoidable limitation that is fully acknowledged. Indeed, the CASEE is presented as a SPJ informed tool, which is newly developed and designed to stimulate future research. Whilst this paper notes initial emerging promise for the CASEE as a new tool of CSE assessment for victims of CSE, more is needed to strength the validity and reliability of the tool. For instance, lower inter-rater agreement could be further explored with larger data sets, to determine whether disagreements are random, systematic, or arising from conceptual ambiguity in the tool items. This is alongside a need for larger volume of case examples. Particularly welcome would be validations undertaken by researchers who do not have an investment in the tool's success, reducing potential risk of bias. Whilst efforts were made to reduce bias in this study, one of the five members of the research team involved in data collection was an author of the CASEE. Invariably that, alongside authors involvement in paper write-up, can always introduce the potential for bias.

However, a strength in the current research lies in the integration of a REA to help inform the direction of reliability considerations that were utilised in this study. This further allows an initiation in addressing some of the limitations noted in the field when evaluating CSE tools. The REA was able to directly inform, therefore, the approach to considering reliability. This allowed for adoption of a novel approach of investigating IRR alongside a good sized sample of professional raters. More traditional IRR studies have often relied on only two

trained raters. This study, however, recruited close to one hundred professionals working in the field to rate. Whilst less case vignettes were used across these two-day validation events, this arguably assisted in minimising rater fatigue and ensured more controlled conditions by which the data was collected. However, by choosing a smaller sample of cases, there is the limitation that case characteristics may have impacted the level of IRR. Increasing the number of case vignettes, with varying degrees of vulnerability and wider scales of rating items, could certainly be a focus for future research.

As such, and due to the smaller sample of case vignettes, this study did not achieve adequate statistical power for analyses with the Intraclass Correlation Coefficient (ICC) and therefore adopted an alternative IRR statistic, Finn's coefficient. We argue however that this is a much more novel and advantageous approach, as the gain of Finn's coefficient, unlike traditional reliability indices, is the fact it does not require a large and diverse sample of cases. This does of course then make comparisons with previous studies of other SPJ assessment tools who have used ICC more restricted but, equally, could offer an alternative approach when utilising harder to obtain samples.

Recommendations for future research

Further IRR studies of the CASEE should be undertaken, focusing on larger and more diverse samples of cases, considering the use of ICC for closer comparison. Future studies could also build on the REA conducted as part of this paper, to compare ICCs of the CASEE to those found with other SPJ tools. This would represent an important next step in the continued evaluation of SPJ informed CSE tools, such as the CASEE. Future research should also focus on gaining a better understanding as to how professionals integrate case formulation into vulnerability assessment overall.

Conclusion

This study demonstrated that professionals across a two-day validation training event are able to achieve at least good to excellent IRR for CSE cases applied to a SPJ informed CSE assessment of victim vulnerability. Arguably, such tools could further promote an understanding of what can lead to CSE vulnerability in a child, in documenting the decision-making that a professional engages in to understand and manage the risk, what professionals can do to interrupt the vulnerability concerns, and how protective factors can be maximised. What is clearly evident, however, is how under-developed this area of research is with regard to the refinement of assessment tools and the need therefore to invest in tools, we would argue, that follow SPJ principles. Thus, the current research argues a call for more research into this area, with a more highly developed and empirically/theoretically informed approach to assessment.

References

Alderson, K. (2019). *The nature and prevalence of sexual exploitation in young people: developing an explanatory model of vulnerability in a normative sample* [Doctoral dissertation, University of Central Lancashire]. https://clok.uclan.ac.uk/30783/

Alderson, K., Ireland, C. A., Khan, R., Ireland, J. L., & Lewis, M. (2022). Child sexual exploitation, poly-victimisation and resilience. *Journal of Criminological Research, Policy and Practice*, *8*(1), 53–74. https://doi.org/10.1108/jcrpp-04-2021-0017

Alderson, K, & Ireland, C. A. (2020). Child Sexual Exploitation: Definition and the importance of language. Abuse: An International Impact Journal, *1,* 48-59. https://threequayspublishing.com/product/child-sexual-exploitation-definition-and-the-importance-of-language/#:~:text=This%20discussion%20paper%20aims%20to%20consider%20the%20issues%2C,defining%20CSE.%20This%20includes%20a%20suggested%20revised%20definition.

Andretta, J. R., Woodland, M. H., Watkins, K. M., & Barnes, M. E. (2016). Towards the Discreet identification of Commercial Sexual Exploitation of Children (CSEC) victims and Individualized Interventions: Science to Practice. *Psychology, Public Policy and Law*, *22*(3), 260-270. https://doi.org/10.1037/law0000087

Astridge, B., Li, W. W., McDermott, B., & Longhitano, C. (2023). A systematic review and meta-analysis on adverse childhood experiences: Prevalence in youth offenders and their effects on youth recidivism. *Child Abuse & Neglect*, *140*, 106055. https://doi.org/10.1016/j.chiabu.2023.106055

Baird, N., Dagon, D., Henderson, G., Kirkman, M., Kurlus, I., & Parry, R. (2020). Sexual exploitation of children involved in the Children's Hearings System: A research report by the Scottish Children's Reporter Administration and Barnardo's Scotland. Retrieved from https://cms.barnardos.org.uk/sites/default/files/2020-10/CSE%20research%20report%20-%20Sexual%20exploitation%20of%20children%20involved%20in%20the%20Children's%20Hearings%20System'%20-%20October%202020.pdf

Benavente, B. T., Brage, L. B., Solé, J., & Beltrán, N. P. (2021). Risk factors for commercial sexual exploitation of children and adolescents: results of an International Delphi Panel. *PubMed*. https://doi.org/10.7334/psicothema2020.480

Benavente, B., Ballester, L., Pich, J., & Pereda, N. (2022). Detection of child sexual exploitation through the evaluation of risk indicators in Spain. *Papeles del psicólogo*, *43*(2), 90-95. https://doi.org/10.23923/pap.psicol.2993

Benavente, B., Bully, P., Ballester, L. (2023). Instruments for the Identification of Child Sexual Exploitation: A Systematic Review. Trauma, Violence, and Abuse. SAGE Publications. https://doi.org/10.1177/15248380231207898

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. http://dx.doi.org/10.1191/1478088706qp063oa

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist,* 513-531. https://doi.org/10.1037/0003-066X.32.7.513

Bronfenbrenner, U. (2005). The bioecological theory of human development. In U. Bronfenbrenner (Ed.), *Making human beings human: Bioecological perspectives on human development* (pp. 3–15). Sage Publications Ltd.

Brown, S.J., Brady, G., Franklin, A., Bradley, L., Kerrigan, N., & Sealey, C. (2016). *Child Sexual Abuse and Exploitation: Understanding risk and vulnerability.* Early Intervention Foundation. https://www.eif.org.uk/report/child-sexual-abuse-and-exploitation-understanding-risk-and-vulnerability#:~:text=This%20review%2C%20looking%20at%20the%20risk%20indicators%20and,based%20on%20a%20rapid%20evidence%20assessment%20and%20consideratio

Brown, S., Brady, G., Franklin, A., & Crookes, R. (2017). *The use of tools and checklists to assess risk of child sexual exploitation: An exploratory study*. Centre of Expertise in Child Sexual Abuse. https://www.csacentre.org.uk/research-publications/cse-risk-tools/exploratory-study-on-the-use-of-tools-and-checklists-to-assess-risk-of-child-sexual-exploitation/

Campbell, R., Dworkin, E., & Cabral, G. (2009). An ecological model of the impact of sexual assault on women's Mental Health. *Trauma, Violence, & Abuse*, *10*(3), 225–246. https://doi.org/10.1177/1524838009334456

Chang, K. S. G., Lee, K., Park, T., Sy, E., & Quach, T. (2015). Using a Clinic-based

Screening Tool for Primary Care Providers to Identify Commercially Sexually Exploited Children. *The Journal of Applied Research on Children*, *6*(1). https://doi.org/10.58464/2155-5834.1235

Collins, A.M., Coughlin, D., Miller, J., & Kirk, S. (2015). *The production of quick scoping reviews and rapid evidence assessments: A how to guide*. https://www.gov.uk/government/publications/the-production-of-quick-scoping-reviews-and-rapid-evidence-assessments

Clutton, S., & Coles, J. (2007). Sexual exploitation risk assessment framework: A pilot study. Barnardo's Cymru. Retrieved from https://www.iicsa.org.uk/keydocuments/26534/view/BRD000284.pdf

Department for Education (2022). *Characteristics of children in need: 2021 to 2022.* https://www.gov.uk/government/statistics/characteristics-of-children-in-need-2021-to-2022. [Accessed 8 November 2023.]

de Vogel, V., De Beuf, T., Shepherd, S & Schneider, R.D. (2022) Violence Risk Assessment with the HCR-20 V3 in Legal Contexts: A Critical Reflection, Journal of Personality Assessment, 104:2, 252-264, https://doi.org/10.1080/00223891.2021.2021925

De Vries, I., Kafafian, M., Goggin, K., Bouchard, E., Goldfarb, S., & Farrell, A. (2019). Enhancing the identification of commercial sexual exploitation among a population of High-Risk youths using predictive regularization models. *Child Maltreatment*, *25*(3), 318-327. https://doi.org/10.1177/1077559519889178

Finn, R. H. (1970). A note on estimating the reliability of categorical data. *Educational and Psychological Measurement*, *30*(1), 71–76. https://doi.org/10.1177/001316447003000106

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. John Wiley & Sons.

Franklin, A., Brown, S., & Brady, G. (2018). The use of tools and checklists to assess the risk of child sexual exploitation: Lessons from UK practice. *Journal of Child Sexual Abuse*, *27*(8), 978–997. https://doi.org/10.1080/10538712.2018.1510453

Gamer, M., Lemon, J., Fellows, I., & Puspendra, S. (2019). *Various coefficients of interrater reliability and agreement*. https://cran.r-project.org/package=irr

Greenbaum, V. J., Dodd, M., & McCracken, C. (2018). A short screening tool to identify victims of child sex trafficking in the health care setting. *Pediatric Emergency Care*, *34*(1), 33-37. https://doi.org/10.1097/pec.0000000000000602

Hill, L., & Diaz, C. (2021). An exploration of how gender stereotypes influence how practitioners identify and respond to victims (or those at risk) of child sexual exploitation. *Child & Family Social Work*, *26*(4), 642–651. https://doi.org/10.1111/cfs.12845

Independent Office for Police Conduct. (2022). *The Operation Linden report*. https://www.policeconduct.gov.uk/publications/operation-linden-report-june-2022

Ireland, J. L. (2018). Individual assessments of aggression: Accounting for core factors. In J. L. Ireland, C. A. Ireland, & P. Birch (Eds.), *Violent and sexual offenders: Assessment, treatment and management* (2nd ed., pp. 203-227). Routledge.

Ireland, C. A., Lewis, M., & Ireland J. L. (2025) *CASEE - Child Adolescent Sexual Exploitation Evaluation*. UK. https://threequayspublishing.com/product/casee-child-adolescent-sexual-exploitation-evaluation-manual/

Ireland, C. A., Ireland, J. L., (2024). CASEE - Introducing a new assessment tool for supporting victims of Child Sexual Exploitation. *Abuse: An International Impact Journal,* 5(1), 15 – 25. https://threequayspublishing.com/product/casee-introducing-a-new-assessment-tool-for-supporting-victims-of-child-sexual-exploitation/

Jay, A. (2014). *Independent inquiry into child sexual exploitation in Rotherham*. https://www.rotherham.gov.uk/downloads/download/31/independent-inquiry-into-child-sexual-exploitation-in-rotherham-1997---2013

Laird, J. J., Klettke, B., Hall, K., & Hallford, D. (2022). Toward a global definition and understanding of child sexual exploitation: The development of a conceptual model. *Trauma, Violence, & Abuse*, *24*(4), 2243–2264. https://doi.org/10.1177/15248380221090980

Lee-Evans, J. M. (1994). Background to behavioural analysis. In M. McMurran, & J. Hodge (Eds.), *The assessment of criminal behaviours of clients in secure settings* (pp. 6-33). Jessica Kingsley Publishers.

Mason-Jones, A. J., & Loggie, J. (2019). Child sexual exploitation. An analysis of serious case reviews in England: Poor communication, incorrect assumptions and adolescent neglect. *Journal of Public Health, 42*(1), 62-68. https://doi.org/10.1093/pubmed/fdy227

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 276-282. https://doi.org/10.11613/bm.2012.031

Olver ME, Stockdale KC, Helmus LM, Woods P, Termeer J, Prince J (2024). Too risky to use, or too risky not to? Lessons learned from over 30 years of research on forensic risk assessment with Indigenous persons. Psychol Bull. May;150(5):487-553. doi: 10.1037/bul0000414. Epub 2024 Feb 15. PMID: 38358684.

Panlilio, C. C., Dierkhising, C. B., Richardson, J., & Runner, J. (2022). Evaluating and validating the classification accuracy of a screening instrument to assess risk for

commercial sexual exploitation of Child Welfare–Involved children and adolescents. *Public Health Reports*, *137*(1_suppl), 73S-82S. https://doi.org/10.1177/00333549211065523

Percy, W. H., Kostere, K., & Kostere, S. (2015). Generic Qualitative Research in Psychology. *The qualitative report*. https://doi.org/10.46743/2160-3715/2015.2097

Petruccelli, K., Davis, J., & Berman, T. (2019). Adverse childhood experiences and associated health outcomes: A systematic review and meta-analysis. *Child Abuse & Neglect, 97*, 104127. https://doi.org/10.1016/j.chiabu.2019.104127

Selby, S. E. (2018). *Inter-rater reliability of the Structured Assessment of Violence Risk in Youth (SAVRY) amongst mental health professionals* [Doctoral dissertation, University of Glasgow]. https://theses.gla.ac.uk/id/eprint/30869

Sutherland, A. A., Johnstone, L., Davidson, K. M., Hart, S. D., Cooke, D. J., Kropp, P. R., Logan, C., Michie, C., & Stocks, R. (2012). Sexual violence risk assessment: An investigation of the interrater reliability of professional judgments made using the risk for sexual violence protocol. *International Journal of Forensic Mental Health*, *11*(2), 119–133. https://doi.org/10.1080/14999013.2012.690020

Sireci, S. G., & Padilla, J. L. (2014). Validating assessments: Introduction to the Special Section. *PubMed*, *26*(1), 97-99. https://doi.org/10.7334/psicothema2013.255

WestCoast Children's Clinic (2016). *Identifying Commercially Sexually Exploited Children: Guidelines for administering the commercial sexual exploitation identification tool (CSE-IT) Version 2.0.* Retrieved from http://www.westcoastcc.org/wp-content/uploads/2016/09/CSE-IT2.0user-manual_9.15.16.pdf.

Wilson, C. M. (2013). *Reliability and consistency of risk formulations in assessments of sexual violence risk* [Doctoral dissertation, Simon Fraser University].

Viljoen, J. L., & Vincent, G. M. (2020). Risk assessments for violence and reoffending:

Implementation and impact on risk management. Clinical Psychology: Science and

Practice, e12378. https://doi.org/10. 1111/cpsp.12378

Figure 1. Flowchart for the selection of papers to be included in the REA.

Source: Authors own work

| Identification | 479 records identified through initial database searching
Google scholar (n=38), PsychARTICLES (n=6), PsychINFO (n=129), PubMed (n=95), Scopus (n=86), Web of Science (n=125) |

Duplicates removed (n=188)

Records remaining (n=291)

Records excluded by screening title and abstract (n=261)

Records remaining (n=30)

Records removed by screening full text for inclusion criteria, did not cover CSE assessment tools or checklists (n=5)

Records remaining (n=25)

Additional records included through second search process (n=2)
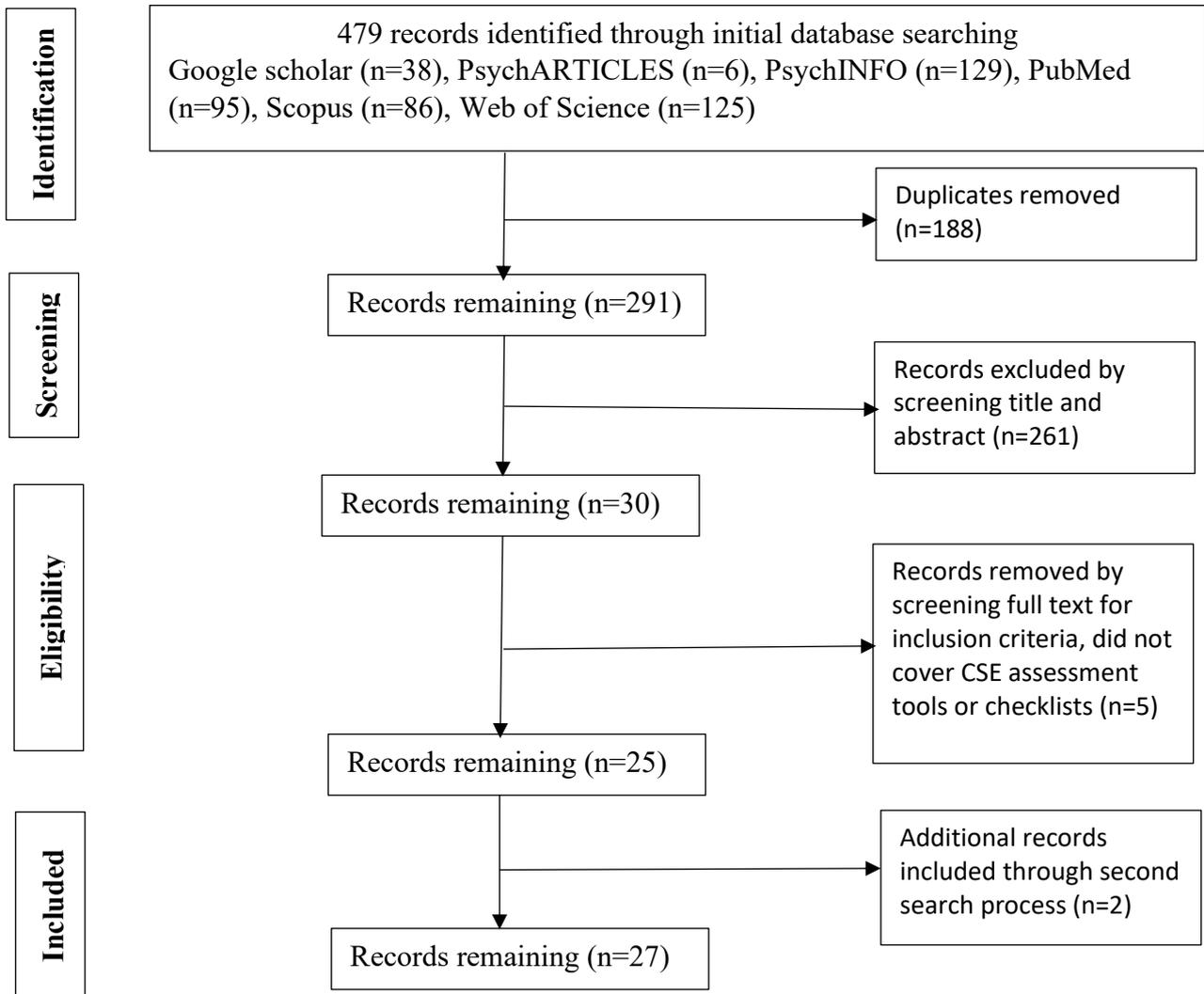
Records remaining (n=27)

Screening

Eligibility

Included

Figure 2. Abridged examples of a CASEE vulnerability and protective factor

Source: Authors own work

Vulnerability factor

Item 3: Strain on the primary caregiver

This item attends to strain on the primary caregiver, to include stress occurring through financial difficulties, unemployment, mental health concerns, addiction, and a lack of social support. All of these factors, particularly when considered collectively, hinder a primary caregiver's ability to protect and provide the basic needs for a young person.

When scoring this item, the following is considered:

Poverty; Primary caregiver's mental health; Criminal Justice involvement (including domestic or intimate partner abuse); Lack of social support and Use of substances

Relevant timeframe: Scoring focuses on the past six months (including the day of assessment).

Scoring:

Clearly present: Consistently present within the young person's life (all of the above item components are met).

Somewhat present: Defined as only somewhat present. That is, some, but not all, of the above items are met.

No evidence: Defined as not being present at all, that is, no issues of strain on the primary caregiver.

Protective factor

Item 23: Appropriate supervision by a protective adult

This item looks to determine if the young person has supervision in place by a protective adult, fitting the notion of guardianship (i.e., a protective adult who is physically available for the young person, where the young person feels able to approach them when fearful/distressed, with this adult being suitably protective of them). This protective adult may or may not be a caregiver, but could also be someone significant in the young person's life (e.g. extended family members, teachers, professionals).

Relevant timeframe: Scoring focuses on a time scale of where they are now and projecting into the next six months. This can be just the one rating unless it is felt that where they are now is likely to be substantially different to where they are likely to be within the next six months.

Scoring:

Clearly Present: Consistently present for the young person.

Somewhat present: Somewhat present. That is, there is only partial supervision by a protective adult.

Absent: Defined as not being present at all, in that there is no evidence of appropriate supervision.

Table 1
Profiles of the vignettes

Case	Victim gender	Victim age	Victim's primary caregiver	(Suspected) abuser(s) and general circumstances
Jessica	Female	15	Local authority (LA)	*Adult male perpetrators.* Residing with LA due to caregiver neglect. Jessica would leave the LA residence, being found with unknown adult males, staying overnight in their home.
Richard	Male	13	Foster placement	*Adult female perpetrators.* Regularly going missing. Concerns of engaging with unknown criminal adult males, but also adult females, whom Richard stated would engage in sexual behaviours (abuse) with him; seeing this in positive terms.
Casey	Female	14	Maternal aunt	*Adult males.* Concerns of engaging with unknown criminal adult males, where police have been involved regarding criminal sexual exploitation concerns, and where Casey was discovered by police at the home of an adult male.

Table 2

Finn's Coefficients for Individual CASEE Items Across All Cases

CASEE item	N	Finn's r	F
General vulnerability items			
1 Physical isolation	91	.77	4.39
2 Social isolation	91	.65	2.85
3 Strain on the primary caregiver	90	.62	2.64
4 Education	91	.73	3.69
5 Lack of healthy sexual knowledge, and social pressures	87	.61	2.53
6 Traumatic experiences	91	.60	2.51
Strain for the child items			
7 Unmet basic needs	91	.52	2.06
8 Unmet psychological needs in the familial/caregiver setting	90	.62	2.62
9 Social attachment with peers	91	.81	5.34
10 Difficulty trusting others	90	.61	2.58
11 Lack of a caring adult	90	.64	2.77
12 Unhelpful coping and emotional instability	90	.77	4.34
13 Instability in self-perception	91	.74	3.87
14 Perception of negative evaluation by others	88	.57	2.35
15 Lack of control	82	.48	1.9
Situational factor items			
16 Lack of risk awareness	91	.79	4.84
17 Normalisation of harmful experiences	90	.45	1.83
18 Other considerations	63	.47	1.89
Protective factor items			

19	Basic needs met	91	.65	2.84
20	Stable education	91	.65	2.83
21	Positive attachment with caregiver	89	.62	2.66
22	Supportive peer relationships	90	.87	7.84
23	Appropriate supervision by a protective adult	91	.55	2.23
24	Adaptive coping	86	.72	3.56
25	Positive view of self	91	.82	5.40
26	Positive engagement with support services	88	.65	2.86

Note. All Finn's coefficients were significant at the .01 level.

Table S1. Characteristics and quality rating of the 19 CSE Assessment Tools (using Benavente et al, 2023 benchmark for level of evidence)

Instrument/ Tool	Validated	Quality rating	Demographics	Theoretical Background	Item number and Generation Procedure	Use of case formulations/ scenario planning	Considered protective factors	Scoring criteria
Six-Item Screening Questionnaire (CSEC/CST) / Child Sex Trafficking Screening Tool *(Greenbaum et al, 2018)*	Yes	Moderate evidence	Minors up to the age of 18. Currently used in the United States.	Not mentioned/ not found.	6 items. Based on characteristics of child patients aged 12 to 18 years in a health care setting who had been victims of commercial sexual exploitation of children/child sex trafficking (CSEC/CST CSEC. Cross-sectional study of 25 CSEC victims compared to 83 alleged victims of acute sexual assault/sexual abuse (ASA) without evidence of CSEC/CST. Groups compared on 16 variables. After univariate analysis, 6 screening items were identified.	No	Not considered	Yes (1) No (0)

Sex Trafficking Assessment Review (STAR) *(Washington D.C. Commercial Sexual Exploitation of Children Subcommittee Task Force, Andretta et al., 2016)*	Yes	Moderate evidence	Minors in the juvenile justice system. Currently being used in the state of Illinois in the United States.	Not mentioned / not found	12 items. Empirically guided checklist. Items developed by identifying risk factors reported by Salisbury and Dabney (2013), Salisbury et al. (2015), Grace et al. (2012), and Leitch and Snow (2010) .	No	Not considered	No Risk (0) Moderate Risk (1) High Risk (2)
Commercial Sexual Exploitation Identification Tool (CSE-IT Version 2) *(West Coast Children's Clinic, 2016)*	Yes	Moderate evidence	Youth ages 10 and over. Currently being used in the United States.	Not mentioned / not found	46 items across 8 domains. Approach based in Grounded Theory: methodology based on observations and principles. Information gathered from focus groups and interviews with users and other stakeholders, aiming to incorporate the lived experience of victims of exploitation and of professionals working with them. Some items were not represented in	No	Not considered	No Information (0) No Concern (0) Possible Concern (1) Clear Concern (2)

				the literature, but were identified as important (Basson, 2017).				
Herramienta de detección del riesgo de explotación sexual en la infancia y adolescencia (EDR-ESIA) / Tool for the detection of risk of sexual exploitation in childhood and adolescence *(Institut Balear de la Dona, Benavente et al., 2022)*	Yes	Moderate evidence	Youth ages 11 to 18. Currently being used in Spain, more specifically in the Balearic Islands.	Not mentioned / not found	88 items across 6 domains. Review of scientific literature regarding CSE identification tools (Benavente et al., 2021), complemented by consultations to a Delphi group of experts to find relevant items (Benavente et al., 2021). Once items were selected, workgroups of professionals involved in the care of minors provided an updated view on CSE. There was a final review of four experts that reviewed the tool.	No	Not considered	CSE target indicators: Mild (1) Moderate (2) Serious (3) Vulnerabilities: Present Not present Unkown
Sexual Exploitation Risk Assessment Framework (SERAF) *(Newport Children and*	Not mentioned / not found	No evidence	Youth ages 11 to 22. Currently being used in the United Kingdom.	Not mentioned / not found	31 items across 3 domains. List of vulnerabilities and risk indicators developed drawing on research and practice evidence, such as: Scott (2001), Cusick	No	Not considered	Indicators of vulnerabilities and of moderate risk (items 1-23): No (0) Yes (1)

Family Services, Clutton & Coles, 2007)					(2002), and Chase and Statham (2005).			Target indicators (items 24-31): No (0) Present in the past (1) Currently present (5) Present before and now (6)
San Luis Obispo Commercially Sexually Exploited Children (CSEC) Screening Tool *(San Luis Obispo County Collaborative Response Team, 2014)*	Not mentioned / not found	No evidence	Youth ages 11 to 18. Currently being used in the United States.	Not mentioned / not found	22 items. Generation procedure not mentioned / not found	No	Not considered	No (0) Yes (1)
Asian Health Services and Banteay Srei's CSEC Screening Protocol	Yes	Limited evidence	Women ages 13 to 23 (Asian and from other ethnicities).	Not mentioned / not found	10 items. Methodological approach of focused interviews with stakeholders, including	No	Not considered	No (0) Yes (1)

(Asian Health Services and Banteay Srei, 2008).			Currently being used in the United States.		CSEC victims, health care providers, youth program health educators, behavioural health staff etc Furthermore, the domestic violence screening and referral protocol provided guidance throughout the development of the tool.			
Local Safeguarding Children Board (LSCB) CSE Risk Screening Tool *(Sutton Local Safeguarding Children Partnership, 2018)*	Not mentioned / not found	No evidence	Youth ages 11 to 17. Currently being used in the United States.	Not mentioned / not found	50 items across 5 domains. Generation procedure not mentioned / not found.	No	Not considered	No Yes Unknown
Sexual Exploitation among Youth (SEY) Risk Assessment Framework Screener *(Pennsylvania State*	Yes	Moderate evidence	Youth ages 11 to 17. Currently being used in Canada.	Not mentioned / not found	26 items across 7 domains. Items based on the National Survey of Child and Adolescent Well-Being (NSCAW) study.	No	Not considered	No (0) Yes (1)

University, Panlilio et al., 2022)								
OSCB Child Sexual Exploitation Screening Tool *(Oxfordshire Safeguarding Children Board, OSCB)*	Not mentioned / not found.	No evidence	Children and young people aged 10 plus. Currently being used in the United Kingdom.	Not mentioned / not found	85 items across 6 domains. Generation procedure not mentioned / not found.	No	Not considered	Present: No Yes Possible Each item has a suggested risk level: Low Medium High
Pan Lancashire Child Sexual Exploitation Standard Operating Protocol *(Children's Safeguarding Assurance Partnership, Blackburn with*	Not mentioned / not found.	No evidence	Children and young people up to the age of 18. Currently being used in the United Kingdom.	Not mentioned / not found	10 items. Generation procedure not mentioned / not found.	No	Not considered	Possible scores: 0 point 1 point 2 points

Darwen & Blackpool Safeguarding Children Boards)								
Project Phoenix Child Sexual Exploitation Measurement Tool *(Project Phoenix)*	Not mentioned / not found.	No evidence	Children referred to social services showing indicators of CSE. Currently being used in the United Kingdom.	Not mentioned / not found	10 items plus one category for professional judgment. Generation procedure not mentioned / not found.	No	Protective factors are not considered in a designated section. However, the presence of some protective factors could be registered in the rating of risk factors.	Scores for items 1 to 10: 1 point 2 points 3 points 4 points 5 points Score of professional judgment: 1-10
Rotherham Council's Child Sexual Exploitation Screening Tool *(Rotherham Safeguarding Children Partnership)*	Not mentioned / not found	No evidence	Children under the age of 18 years old. Currently being used in the United Kingdom.	Not mentioned / not found	71 items across 4 domains. Generation procedure not mentioned / not found.	No	Not considered.	Yes No

West Midlands Metropolitan Regional Child Sexual Exploitation (CSE) Screening Tool *(Birmingham Children's Trust)*	Not mentioned / not found	No evidence	Children and young people 13 years or older. Currently being used in the United Kingdom.	Not mentioned / not found	44 items across 4 domains. Generation procedure not mentioned / not found.	No	Within section for professional judgment, the tool instructs to include protective factors.	Indicate if present by "ticking" the box.
The Northamptonshire Tackling Child Sexual Exploitation Toolkit *(Northamptonshire Safeguarding Children Board)*	Not mentioned / not found	No evidence	Children and young people for whom there are concerns they may be vulnerable to, being targeted for or involved in child sexual exploitation. Currently being used in the United Kingdom.	Not mentioned / not found	10 items and one professional judgment score. Not mentioned / not found	No	Not considered	Scores for items 1 to 10: 1 point 2 points 3 points 4 points 5 points Score of professional judgment: 1-10

NY State OFCS Child Sex Trafficking Indicators Tool *(New York State Office of Children and Family Services, (OFCS) (2016)*	Not mentioned / not found	No evidence	Children and youth at risk of sex trafficking. Currently being used in the United States.	Not mentioned / not found	38 items across 3 domains. Generation procedure not mentioned / not found	No	Not considered	Yes No
NY State OFCS Rapid Indicator Tool *(New York State Office of Children and Family Services (OFCS) (2016)*	Not mentioned / not found	No evidence	Children and youth at risk of sex trafficking. Currently being used in the United States.	Not mentioned / not found	11 items. Generation procedure not mentioned / not found.	No	Not considered	Yes No
LSCPB CSE Risk Assessment Tool *Leicester Safeguarding Children Partnership Board*	Not mentioned / not found	No evidence	Children and youth under the age of 18. Currently being used in the United Kingdom.	Not mentioned / not found	71 items across 4 domains. Generation procedure not mentioned / not found	No	Within section for professional judgment, the tool instructs to include protective factors (specified in	Yes No At risk indicators, medium risk indicators, and high risk indicators differentiate

(LSCPB)							the guidelines).	between current and historic (non-recent).
Barnardo's Research Toolkit of Indicators of CSE *(Barnardo's: Believe in Children - Children's Charity: See Baird et al., 2020)*	Yes	Limited evidence	Children and youth under the age of 18. Currently being used in the United Kingdom.	Not mentioned / not found	35 items across 3 domains. Based on Scotland's definition of CSE and National Framework (Scottish Government, 2016; Scottish Government, 2017); All Wales Protocol and Risk Assessment Framework developed by Barnardo's (2018); the template developed by Glasgow City Council (GCC) and Stirling University for GCC's Child Protection Committee (Rigby & Murie, 2013); and review of Serious Case Reviews in England (Baird et al., 2020).	No	Not considered	Vulnerability factors: Yes (1) No (0) Moderate risk

Table S2

Demographics of Participants (N = 91)

Demographic	Characteristic	N	%
Gender	Female	78	85.7
	Male	10	11
	Missing	3	3.3
Primary profession	Forensic psychologist	31	34.1
	Psychological therapist	10	11
	Social worker	9	9.9
	Counselling psychologist	4	4.4
	Academic	2	2.2
	Police officer	2	2.2
	Clinical psychologist	1	1.1
	Teacher	1	1.1
	Other	31	34.1
Years of work experience	<1 year	4	4.4
	1-5 years	28	30.8
	6-10 years	17	18.7
	11-15 years	14	15.4
	16-20 years	17	18.7
	21-25 years	6	6.6
	>25 years	5	5.5
Number of risk assessment reports completed per year	None	10	11
	1-10	32	35.2
	11-20	26	28.6

	21-30	6	6.6
	>30	13	14.3
	Missing	4	4.4
Number of reports regarding CSE concerns completed per year	None	28	30.8
	1-5	35	38.5
	6-10	10	11
	11-20	8	8.8
	>20	7	7.7
	Missing	3	3.3

Table S3. Set criteria for assessing the level of agreement.

Very low agreement	A theme that appears in the same section in both the research team and participants formulations, with a prevalence of 1% to 25% (meaning the theme is present in 1, 2 or 3 out of the 12 participants formulations).
Low agreement	A theme that appears in the same section in both the research team and participants formulations, with a prevalence of 26% to 42% (meaning the theme is present in 4 or 5 out of the 12 participants formulations).
Moderate agreement	A theme that appears in the same section in both the research team and participants formulations, with a prevalence of 43% to 59% (meaning the theme is present in 6 or 7 out of the 12 participants formulations).
High agreement	A theme that appears in the same section in both the research team and participants formulations, with a prevalence of 60% to 75% (meaning the theme is present in 8 or 9 out of the 12 participants formulations).
Very high agreement	A theme that appears in the same section in both the research team and participants formulations, with a prevalence of 76% to 99% (meaning the theme is present in 10 or 11 out of the 12 participants formulations).
Perfect agreement	A theme that appears in the same section in both the research team and participants formulations, with a prevalence of 100% (meaning the theme is present in 12 out of the 12 professionals' formulations).