

# **Toolbox of Measures Used in Experimental Research on Institutional Child Maltreatment**

**Technical Report**

**Gabrielle R. Hunt**

**December 2023**

## Acknowledgements and funding

This project was funded and supported by Porticus. Members of the Porticus team, as well as Professor Daryl Higgins from Australian Catholic University, provided input and guidance on the project.

## Author

**Gabrielle Hunt** is a Doctor of Philosophy student and Research Assistant at Australian Catholic University. Her research is focused on the prevalence and prevention of child sexual abuse and sexual harassment, with a particular focus on faith-based settings. Gabrielle's research and ongoing work are informed by her experience as a psychologist working with survivors of child sexual abuse as well as young people engaging in harmful sexual behaviours. She is supervised by the Director of the Institute of Child Protection Studies at Australian Catholic University, Professor Daryl Higgins.

## Institute of Child Protection Studies

Australian Catholic University

Postal: PO Box 256, Dickson, ACT, 2602

Email: [icps@acu.edu.au](mailto:icps@acu.edu.au)

Web: [www.acu.edu.au/icps](http://www.acu.edu.au/icps)

Twitter: @ACU\_ICPS

LinkedIn: <https://www.linkedin.com/company/institute-of-child-protection-studies/>

Phone: (02) 6209 1228

## Porticus

Porticus is a philanthropic organisation that aims to create a just and sustainable future where human dignity flourishes. Our work aims to strengthen the resilience of communities so that all people have ownership over their future, and natural resources are used in a sustainable way. We realize our objectives through strong networks of partners including local and global NGOs, communities, people with lived experience, policy makers and co-funders.

## To cite the Technical Report

Hunt, G. (2023). *Toolbox of Measures Used in Experimental Research on Institutional Child Maltreatment: Technical Report*. Australian Catholic University.

With the exception of the ACU and Porticus logos, content provided by third parties, and any material protected by a trademark, all textual material presented in this publication is provided under a Creative Commons Attribution 4.0 International licence (CC BY 4.0) <[creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/)>.

You may copy, distribute, and build upon this work for commercial and non-commercial purposes; however, you must attribute ACU as the copyright holder of the work. Content that is copyrighted by a third party is subject to the licensing arrangements of the original owner. Any copyright questions please contact ACU at <[copyright.officer@acu.edu.au](mailto:copyright.officer@acu.edu.au)>.

## Contents

1. Introduction.....	5
1.1. Preamble .....	5
1.2. Key Terminology .....	6
Psychometric Properties .....	6
Reliability .....	6
Validity .....	8
Likert Scale.....	9
Ceiling Effect .....	9
Floor Effect.....	10
Practice Effect .....	10
1.3. Rating Scale .....	10
2. Overview of Measures Used in the EGM .....	11
2.1. Institutional Safeguarding Practice .....	11
Institutional Culture .....	11
Institutional Environment.....	14
2.2. Adult Institutional Care Provider .....	17
Adult Institutional Care Provider’s Knowledge and Skills.....	17
Adult Institutional Care Providers’ Attitudes .....	25
Adult Institutional Care Provider’s Wellbeing .....	32
2.3. Child Maltreatment Disclosure .....	38
2.4. Child Maltreatment Occurrence/Recurrence .....	40
Administrative Data .....	49
2.5. Children’s Outcomes.....	50
Children’s Knowledge .....	50
Children’s Attitudes .....	75
Children’s Behaviours .....	84
Child Wellbeing .....	94

Administrative Data .....	94
2.6. Adult Perpetrators .....	94
2.7. Child/Youth Offender Maltreatment Behaviours .....	97
2.8. Parent/Caregiver .....	100
Parent Knowledge/Awareness and Skills .....	100
Parent Wellbeing .....	107
2.9. Intervention Feasibility and Implementation .....	110
3. Summary .....	110
4. References .....	111
5. Appendix .....	138
5.1. Child Wellbeing .....	138
Cognitive Functioning .....	138
Child Physical Health and Development .....	139
Emotional Functioning .....	140
Mental Health .....	141
5.2. Intervention Feasibility and Implementation Measures .....	143

# 1. Introduction

## 1.1. Preamble

Giving Evidence and partners produced an Evidence and Gap Map (EGM) of the rigorous experimental evidence available on interventions that address institutional child maltreatment (Finch et al., 2021; Thota et al., 2023). Their project collated available evidence from randomised control trials, quasi-experimental design studies, and systematic reviews. There has been a growing body of research in recent years, but there are still considerable gaps in the evidence base. Most studies focussed on the prevention of sexual abuse, with much less research focusing on the prevention of other types of child maltreatment. There were many studies focused on outcomes such as increasing children’s knowledge about abuse, but few focused on concepts such as organisational culture, adult perpetrators, or child/youth offending.

This project was aimed at understanding how the studies in this EGM have measured the intended outcomes. It is important to understand how outcomes were measured to guide further efforts to address factors associated with institutional child maltreatment and enhance protective factors that decrease its incidence. Using sound, reliable, and valid measures allows us to be more confident in our efforts to protect children and young people. This Technical Report provides further insights into the key findings of the EGM by providing an overview of the measures used to assess the achievement of safeguarding, along with their relative strengths and limitations.

There were nearly 300 ‘measures’ used across the 136 studies included in the EGM. There were many examples of questionnaires, scales, surveys, or measures with substantial inconsistencies in how they were applied to assess improvements in ‘safety’ in institutions. Many studies focused on ‘proxy’ measures, such as children’s knowledge, and it was not always clear if this directly led to a reduced incidence of institutional child maltreatment.

It was also common for researchers to create or develop their own measure of a particular concept rather than to use, modify, or adapt an existing measure. This was sometimes because there was not a suitable measure available at the time of the study. Unfortunately, researchers also often used measures with little or even no available support for their reliability or validity.

For this project, each paper published in English was reviewed and details about the measures used in the research were extracted and collated for review. Details of any analysis on the reliability and validity of the measures reported by the authors of the study, along with key findings from other research, were reviewed (such as the original development and validation research paper) to provide a comprehensive overview of the available evidence.

This resource can be used by funders, policymakers, practitioners, researchers, and others to help make evidence-informed decisions. Each measure has been placed under a broader construct that maps onto the constructs used by the EGM. The contents page has been hyperlinked to allow readers to quickly navigate throughout the Technical Report to review constructs and measures of interest. An executive summary is also available which provides a high level summary of the findings from this project.

## 1.2. Key Terminology

While statistical knowledge is not needed to read this Technical Report, we have provided an overview of key terms used below. This is by no means an exhaustive list, but it provides a basic overview of key processes in evaluating the psychometric properties of a measure.

### Psychometric Properties

- Refers to the qualities or characteristics of a measure. It is established through statistical and other techniques that fit broadly under the concepts of ‘reliability’ and ‘validity’.
- Measures that have good psychometric properties are measures we can be confident are consistently going to measure what we want to measure.

### Reliability

#### *Internal consistency*

- The degree to which items measure the same general construct.
- A higher level of internal consistency indicates that the items in the measure are closely related and provide a reliable measure of the target concept.
- It is most often assessed using Cronbach’s alpha. This is a statistical technique that provides an average correlation of each item with every other item. Values range from 0-1, with higher values indicating better internal consistency. The below table provides commonly used ‘cut-offs’ for Cronbach’s alpha values.

Cronbach’s alpha ( $\alpha$ )	Internal Consistency
> 0.9	Excellent
0.8-0.9	Good
0.7-0.8	Acceptable
0.6-0.7	Questionable
0.5-0.6	Poor
< 0.5	Unacceptable

#### *Test-retest reliability*

- The degree to which a measure is consistent or stable over time.
- Sometimes referred to as temporal stability.
- For example, we might test a measure to see if a participant will get the same or similar scores on two or more occasions under similar conditions.
- There should be a sufficient time gap between testing to minimise the influence of memory or practice effects.

- It is often tested using a correlation coefficient such as Pearson’s  $r$ .
  - $r$  closer to 1 indicates a stronger linear relationship
  - $r$  closer to 0 indicates a weaker or no linear relationship

*Inter-rater reliability*

- The degree to which two individuals agree on scoring a measure.
- A higher level of inter-rater reliability suggests that we can be confident that multiple raters of a measure are interpreting and applying the criteria or scoring in a consistent way.
- It is most commonly assessed using the Intraclass Correlation Coefficient (ICC), which measures the proportion of total variability due to differences between rates compared to the total variability. The below table provides an example of ‘cut-offs’ for ICC values, but others are sometimes used.

ICC	Inter-rater reliability
0.9-1.0	Excellent agreement
0.75 - <0.9	Good agreement
0.5 - <0.75	Moderate agreement
< 0.5	Poor agreement

- Cohen’s Kappa may also be used, especially when dealing with categorical data. The below table provides an example of ‘cut-offs’ for Cohen’s Kappa values, but others are sometimes used.

Cohen’s Kappa	Inter-rater reliability
0.81-1.00	Almost perfect agreement
0.61-0.80	Substantial agreement
0.41-0.60	Moderate agreement
0.20-0.40	Fair agreement
0.00-0.20	Slight agreement
<0.00	Poor agreement

- Higher inter-rater reliability (i.e., scores closest to 1) indicates strong agreement among raters.

- Inter-rater reliability on a measure can be improved by having clear guidelines on the criteria or scoring processes, as well as training and evaluations for raters.

### *Intra-rater reliability*

- The degree to which a person consistently scores something within the same measure.
- Similar statistical techniques can be used to assess this as described for inter-rater reliability.

## **Validity**

### *Face Validity*

- The degree to which a measure appears, at face value, to measure what it intends to measure
- This is a subjective assessment and cannot be used in isolation to determine the validity of a measure.

### *Content Validity*

- The degree to which a measure represents the entire content or facets of the concept being measured.
- It is typically established through expert judgement. A numerical indicator may be used to indicate the extent to which items are relevant and represent the concept being measured (e.g., Content Validity Index; CVI). A higher CVI (closer to 1) indicates stronger content validity.

### *Cross-Cultural Validity*

- The degree to which a measure is relevant and applicable for different cultural groups.
- This is often relevant to assess whether a translated or culturally adapted measure is an adequate reflection of the performance of the items of the original version of the measure.
- The cross-cultural validity of a measure can be improved by collaborating with researchers from different cultural backgrounds, piloting and refining measures in different cultural groups, and adapting measures to be culturally sensitive (e.g., changing terminology).

### *Criterion Validity*

- The degree to which the scores of a measure correlate with an established measure of comparison.
- For example, if a new anxiety measure is being developed, you would compare scores between the new measure and scores on a measure that has already been developed. You would expect that participants would get similar scores on each measure.
- It is often assessed using Pearson's  $r$ , with values ranging from 0 to 1.
  - $r$  closer to 1 indicates a stronger linear relationship
  - $r$  closer to 0 indicates a weaker or no linear relationship



## Construct Validity

- The extent to which something measures the theoretical concept it is designed to measure.
  - For example, if a personality scale claims to measure extroversion, construct validity would involve showing that the test scores correlate positively with other extroversion measures (convergent validity) and negatively with its opposite, such as an introversion measure (discriminate validity).
- It can be established through:
  - Convergent and discriminate validity (generally assessed using correlation coefficients such as Pearson's  $r$ ).
  - Factor analysis
    - A statistical technique to uncover the underlying structure (latent factors) that may be influencing the observed relationships among a set of variables.
    - Researchers would use Exploratory Factor Analysis (EFA) to explore the underlying structure without preconceived notions about the number or nature of the factors.
    - Researchers would use Confirmatory Factor Analysis (CFA) to assess specific hypotheses about the factor structure of a measure, where they want to test this predefined model.
  - Experimental manipulation
    - In an experiment, you can manipulate the construct and observe if the test responds as expected.
    - For example, you might implement an intervention designed to improve children's knowledge of abuse and see if their scores improve on a children's knowledge questionnaire after being exposed to the intervention.

## Likert Scale

This is a commonly used tool in surveys, questionnaires, or measures to assess a participant's attitudes, opinions, or perceptions about a question. It can be presented in many formats, but common categories include a 5-point scale:

- 5 = Strongly agree
- 4 = Agree
- 3 = Neither agree nor disagree
- 2 = Disagree
- 1 = Strongly disagree

## Ceiling Effect

- A ceiling effect occurs when a considerable number of participants score at, or very close to, the highest possible score on a measure.
- This might happen because the items are too easy, or the measure does not have a sufficient range to capture the full variability of the construct being measured.
- A ceiling effect creates issues in measuring a construct as it can make it challenging to identify differences in scores, detect improvements in scores, or identify high performers.

- For example, if there is a knowledge test where participants do very well before being exposed to a program, then it would be difficult to test whether participants' knowledge improves because of the program.

### Floor Effect

- A floor effect occurs when a considerable number of participants score at, or very close to, the lowest possible score on a measure.
- This might happen because the items are too difficult, or the measure does not have a sufficient range to capture the full variability of the construct being measured.
- A floor effect creates very similar issues to those described for a ceiling effect.

### Practice Effect

- A practice effect occurs when participants improve on a measure simply because they have completed the measure before.
- Participants may become familiar with the wording, content, or format of the measure or might remember their previous responses and adjust their responses.

## 1.3. Rating Scale

We have adopted a basic rating scale to indicate the available evidence on the psychometric properties of each measure. This was done to provide a quick snapshot of the available evidence on the relative strength of a measure. Our analysis is not exhaustive but provides a good overview to help guide decision-making processes.

Measures rated as **GREEN** have strong evidence for both the reliability and validity of the measure in a certain population. The best measures will have multiple values (e.g., internal consistency, inter-rater reliability, construct validity, and content validity) to support their psychometric properties. This means we can be confident in relying on these outcome measures in assessing a particular construct. It is important to note that even if a measure has a **GREEN** rating, it does not mean that it can be used in any cultural setting or with any participant. Researchers and others are responsible for ensuring that chosen measures are reliable and valid for their chosen population. This may require additional research and processes to assess the psychometric properties of an established measure, particularly when making adaptations, modifications, or translations.

Measures rated as **ORANGE** may have some preliminary evidence to support the reliability and validity of the measure but may have inconsistent findings or not sufficient evidence yet. For example, there may be findings to support the internal consistency of a measure and experimental evidence for its construct validity, but not its construct validity or test-retest reliability. Measures with an **ORANGE** rating would benefit from additional research to establish their psychometric properties, and caution is advised when interpreting findings based on these outcomes.

Measures rated as **RED** currently have little or no support for their use, or may have only reported one value to support its psychometric properties (e.g., Cronbach's alpha). These measures may be useful, but there is no, very little, or poor evidence on their psychometric properties currently. For example, preliminary experimental evidence for construct validity can be helpful but is not sufficient in isolation to determine a measures' psychometric properties. Some measures have also received

this rating due to key findings from the research in the EGM. We cannot be confident that measures with a **RED** rating have provided a reliable or valid measure of that construct.

A **GREY** rating has been given to measures that appear in protocols, and therefore, their relative psychometric properties have not yet been assessed or reported. Measures published in a language other than English have also been given this rating as we were not able to fully review the available evidence.

## 2. Overview of Measures Used in the EGM

All measures used in studies published in English on the EGM have been represented in some way below. They have been divided based on the construct that most aligns with what they intend to measure, with constructs matching the overarching categories used in the EGM. Under each construct, there is a table that provides a quick snapshot of the measures, including their rating. Measures that were developed by the researchers for the purposes of the study they were used in have been marked with an asterisk (\*). Further information about each measure can be found under the table for that construct. The name of the measure has been hyperlinked in the table to allow readers to quickly navigate to view details about that measure.

### 2.1. Institutional Safeguarding Practice

Institutional Safeguarding Practice is broadly aimed at understanding institutional operations, environments, and cultures that contribute to conditions of safety for children and young people. This might include having good policies in place, having an environment that supports reporting of concerns, and workers that are knowledgeable and committed to reducing the risk of harm.

#### Institutional Culture

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#">Delaware School Climate Survey</a>	Bear et al., (2011)		U.S.	Students, Teachers, and Parents	I feel safe at school.	Potentially limited to U.S. or Western contexts.
<a href="#">Classroom Assessment Scoring System (CLASS)-K3</a>	Pianta et al., (2008)		U.S.	Professional	<i>Observer rating of a teacher's overall emotional tone and connection to students.</i>	Requires a trained professional to observe a classroom.
<a href="#">*School Operational Culture</a>	Merrill et al., (2018)		Uganda	Students and School Staff	Would you say that students feel comfortable talking with you/want to confide in you if	No analysis was reported on the psychometrics. Some preliminary experimental

					they are unhappy about something at home or at school?	evidence on its construct validity.
<b>*Safety School Audit Checklist</b>	Wangamati et al., (2022)		Kenya and Tanzania	Teachers	Not reported.	Protocol. No analysis reported yet on the psychometrics.

### *Delaware School Climate Survey (Bear et al., 2011)*

The Delaware School Climate Survey is a self-report survey designed to assess a school’s overall climate from the perspectives of students, parents (referred to as Home), and teachers. Student-completed surveys are suitable from Grade 3. It is available in English, Spanish, and Haitian Creole. Resources for this scale are available on the University of Delaware website.

The Delaware School Climate Survey has been found to have good psychometric properties in a sample of Delaware schools (Bear et al., 2011). Confirmatory Factor Analyses showed that the survey was best represented with a bifactor model with five specific factors (Teacher-Student Relations, Student-Student Relations, Fairness of Rules, School Safety, and Liking of School) and one general factor. The internal consistency across the grade level, racial-ethnic, and gender groups ranged between 0.63 and 0.89 (Cronbach’s alpha) for the subscales and was excellent for the total scale (Cronbach’s alpha = 0.91-0.94; Bear et al., 2011). Scores on the Delaware School Climate Survey also correlated positively with academic achievement and correlated negatively with suspension and expulsion rates, providing support for the construct validity of the survey (Bear et al., 2011). The original scale did include a Student Conduct Problems scale which assessed students’ reports on harming others, bullying, fighting, drug use, stealing, and cheating. However, this scale was removed as there was no empirical support for its validity (Bear et al., 2011).

Subscales of the Delaware School Climate Survey have been used twice in U.S. based empirical research on institutional child maltreatment (Kim et al., 2019; Nickerson et al., 2019). One study measured teachers’ perceptions of the school staffs’ responsiveness, level of care, and ability to provide support (Kim et al., 2019). While Nickerson and colleagues (2019) assessed students’ perception of the quality of their interactions with teachers. As these studies only used subscales of the full survey, the internal consistency of the measure did decrease to an acceptable range (Cronbach’s alpha = .77-.78; Kim et al., 2019; Nickerson et al., 2019). Full-scale administration of this measure would also include items related to school safety and fairness, respect for diversity, classroom management techniques, and teacher-home communications. There is strong psychometric support for this measure in U.S. schools.

### *Classroom Assessment Scoring System (CLASS)-K3 (Pianta et al., 2008)*

This measure requires a trained professional to observe a Prep to Grade 3 teacher for 30 minutes, repeated up to six times over three hours. This measure can also be used for observations of classrooms up to the 12<sup>th</sup> grade. The observer rates the teacher based on three-domains of

teacher-student interaction; emotional support, classroom organisation, and instructional support (Pianta et al., 2008). These domains were proposed based on developmental and educational theories which suggest that these concepts are important for student learning and social development (Li et al., 2020). Each of the domains are rated on a 7-point scale from low to high. There are manuals available that explain the indicators, with examples, for each domain (Pianta et al., 2008).

The CLASS instrument is widely used to measure the quality of early childhood classrooms and has been shown to predict children's social and academic skills (Sabol et al., 2013). Li and colleagues' (2020) meta-analysis confirmed that the proposed three-factor structure had an excellent model-data fit and was based on a solid theoretical framework. The internal consistency of the domains was also good to excellent (Cronbach's alpha ranging between 0.8 and 0.9; Lei et al., 2020). This provided support for the reliability and construct validity of the CLASS. This measure has also been used in many countries, including Chile, Ecuador, Germany, Portugal, Thailand, and Tanzania (Araujo et al., 2016; Cadima et al., 2014; Leyva et al., 2015; Shavega et al., 2014; Udommana, 2011; von Suchodoletz et al., 2014).

The emotional support domain subscale was used once in U.S. empirical research (Baker-Henningham et al., 2019). This research found that training Grade 1 primary school teachers in Jamaica led to benefits to teacher practices and child outcomes. These benefits included reductions in the frequency of teachers' use of violence against children, increases in the emotional quality of the classroom environment, and benefits to children's learning, specifically related to children's oral language, self-regulation and reasoning skills. In addition, fewer teachers who participated in the intervention used violence against children (including physical violence, verbal abuse and other violence (e.g., intimidation)) over two full days of observation than teachers in the control group (11.1% vs. 0% for total violence respectively; Baker-Henningham et al., 2019).

The CLASS-K3 was also used in a Jamaican cluster-randomised controlled trial (Baker-Henningham et al., 2021a). The emotional (Cronbach's alpha = 0.74) and classroom organisation (Cronbach's alpha = 0.80) scales had acceptable to good internal consistency, but it was questionable for the instructional support scale on this sample (Cronbach's alpha = 0.61; Baker-Henningham et al., 2021).

#### *\*School Operational Culture (Merrill et al., 2018)*

In a Ugandan randomised-control trial, Merrill and colleagues (2018) cited a lack of available measures for school operational culture and normative beliefs on physical discipline practices. They created a questionnaire designed to be completed by school staff and students on relational, psychological, and structural domains. All questions were rated on a 4-point Likert scale. Questions included content about students' feelings of emotional support from teachers, staffs' perceived relationships with students, acceptance of physical discipline practices in school, and perceived level of involvement in school operations (Merrill et al., 2018). There was no reported analysis on the psychometrics of this measure. There was some experimental evidence for its construct validity. For example, students in intervention schools, reported greater feelings of emotional support from teachers than students from control schools (Merrill et al., 2018). However, with the lack of available information on the psychometrics of this survey, it is not clear yet if it is a reliable or valid measure of a school's operational culture.

3-items from the Student acceptance of teacher violence subscale were also included in Knight and colleagues' (2018) protocol for a nonrandomised quasi-experimental study in Uganda.

*\*Safety School Audit Checklist (Wangamati et al., 2022)*

In a protocol for an experimental study to be conducted in Kenya and Tanzania, a checklist was created to assess how the school's code of conduct addresses physical, emotional, and sexual violence in and around schools. It also plans to assess teachers' knowledge of criminal and violent behaviour, teachers' knowledge about how to respond in the case of an emergency, referral systems and pathways, and how to record incidents of violence or misbehaviour at school (Wangamati et al., 2022). There was no information included in the protocol on the design, validation, or psychometric properties of this checklist. Therefore, we cannot be sure currently on the reliability or validity of this measure.

### Institutional Environment

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#"><u>Emotional Climate in Organizations Scale (ECOS)</u></a>	Albuquerque et al., (2020)		Portugal	Adult	Not reported.	Unpublished manuscript. Details not publicly available.
<a href="#"><u>School-Level Environment Questionnaire – South Africa (SLEQ-SA)</u></a>	Aldridge et al., (2006)		South Africa	Teacher	Teachers frequently discuss teaching methods and strategies with each other.	More research may be needed to assess its psychometric properties in other cultural settings.
<a href="#"><u>Beyondblue School Climate Questionnaire (BBSCQ)</u></a>	Sawyer et al., (2010)		Australia	Students and Teachers	In this school, students' ideas are listened to and valued.	There is currently no known research on the psychometric properties of this questionnaire.
<a href="#"><u>*Classwide Prosocial Behavior</u></a>	Baker-Henningham et al., (2019)		Jamaica	Professional	<i>Observations of the frequency, intensity, and number of children involved in prosocial acts (e.g., sharing, helping, cooperating).</i>	There is currently no known research on the psychometric properties of this questionnaire.



<u>*Perception of School Environment</u>	Jewkes et al., (2019)		South Africa	Teacher	Our school has a sense of vision, and a mission that is recognised by all staff and learners.	There is currently little support for the psychometric properties of this measure.
--	-----------------------	--	--------------	---------	---	--

*Emotional Climate in Organizations Scale (ECOS; Albuquerque et al., 2020)*

The ECOS (Albuquerque et al., 2020) is 30-item self-report scale, with items rated on a 5-point scale (0 = never to 4 = always). The ECOS was referenced in Santos and colleagues' (2022) study on professional caregivers in a residential care facility in Portugal. However, this citation is for an unpublished manuscript, and we were unable to access the information.

According to Santos and colleagues (2022), the ECOS has two subscales, each with 15 items, aimed at measuring emotions felt in the workplace and motivations that lead workers to behave in certain ways. The scale was reported to have adequate internal consistency (Cronbach's alpha = 0.75 for the threat system, 0.83 for the soothing system, and 0.86 for the drive-system related emotions; Albuquerque et al., 2020). Details on other psychometric properties of the ECOS are not publicly available.

Professional caregivers were asked to complete the emotional subscales from the ECOS in Santos and colleagues (2022) study. The authors reported Cronbach's alphas of 0.75 for the threat and drive systems and 0.72 for the soothing system-related emotions. Univariate tests indicated significant and positive effects in compassion and fears of compassion (low and medium effect sizes), as well as in soothing related emotions (emotional climate) and social safeness (both with medium effect sizes), in favour of the treatment group. There seems to be some support for this measure, but further information and research is needed.

*School-Level Environment Questionnaire – South Africa (SLEQ-SA; Aldridge et al., 2006)*

The SLEQ was originally designed as a 56-item self-report tool designed to measure teachers' perceptions of their school environment. It measured several factors including student support, staff freedom, professional interest, participatory decision making, affiliation, innovation, resource adequacy, and work pressure (Rentoul & Fraser, 1983). Aldridge and colleagues (2006) adapted the measure to be used in South Africa. This version (SLEQ-SA) was found to have good to excellent internal consistency when using the school mean (Cronbach's alpha = 0.76-0.94), with the authors concluding that the overall internal consistency was acceptable. There was some overlap (measured with correlation coefficients) between the dimensions that each scale assessed (discriminant validity). However, factor analysis provided support for the independence of factor scores in the SLEQ-SA (Aldridge et al., 2006).

McElearney and colleagues (2021) used the SLEQ-SA which has 35 items across six subscales, in their cluster randomised controlled trial of Keeping Safe programs in Northern Ireland. There was no significant difference observed on scores on the SLEQ-SA between waitlist control and intervention schools after two years. This may suggest that either the intervention did not affect teachers' perceptions of school culture, or that the measure was not sensitive to any changes

(McElearney et al., 2021). The authors conducted a Confirmatory Factor Analysis (CFA), finding that the factor loadings for the SLEQ-SA were statistically significant. They reported a second-order factor model with seven first-order factors (participatory decision-making, staff freedom, professional interest, affiliation, parental involvement, work pressure, and resource adequacy). There seems to be support for the psychometric properties of the SLEQ-SA. It has not been used in any other study in the EGM.

### *Beyondblue School Climate Questionnaire (BBSCQ; Sawyer et al., 2010)*

The BBSCQ was created as a measure of student and teacher perceptions of the school environment. It was developed in an Australian study on school-based prevention for depression (Sawyer et al., 2010). It covers four aspects of the school climate: supportive teacher-pupil relations, sense of belonging, participation, and commitment. Students (specifically adolescents) are asked about their perceived relationships with teachers, their sense of belonging, and their level of participation in school activities. This measure contains 20 items, with scores ranging between 20 and 81, with higher scores indicating a more positive school environment (Sawyer et al., 2010). Teachers are asked 18 questions about their perception of the school environment, safety, participation in the school community, and the quality of relationships. Scores on the teacher section range between 18 and 90, with higher scores indicating a more positive school environment (Sawyer et al., 2010). To our knowledge, there has been no analysis on the psychometric properties of this questionnaire. Interestingly, there seems to be a relatively low correlation between student- and teacher-reported school climate (supportive relationship  $r = 0.29$ ; safety  $r = 0.44$ ). Currently, we cannot be certain that the BBSCQ is a reliable or valid measure of a school's environment.

It was only used in one study in the EGM. Wangamati and colleagues (2022) included this scale in their protocol for a mixed-methods pilot study of whole-school approaches to interventions to reduce violence in schools in Kenya and Tanzania.

### *School Questionnaire (Vorst, 1990)*

The School Questionnaire was developed in the Netherlands and was designed to measure changes in the relationship between students and teachers (Vorst, 1990). This scale is in Dutch so we were unable to review the full psychometric properties of this measure.

It was only used once in the EGM, where two subscales (Relationship with Teacher and Relationship with Classmates) were used in an experimental study in the Netherlands (Taal & Edelaar, 1997). This measure has good internal consistency (Cronbach's alpha = 0.83-0.9) and adequate test-retest reliability (0.66) in this sample.

### *\*Classwide Prosocial Behavior (Baker-Henningham et al., 2019)*

In a cluster-randomised controlled trial in Jamaica, a similar process used for the CLASS-K3 was used to assess children's engagement in prosocial behaviours. Mean scores over five observations of child prosocial behaviour were scored (Baker-Henningham et al., 2019). The score was designed to reflect the frequency, intensity, and number of children involved in prosocial acts, such as cooperating with one another, helping other students, or sharing. There was no reported analysis on the psychometrics of these observations, or the scoring used. There was some



preliminary experimental support for the construct validity, with improvements in teacher-reported social skills following the intervention (Baker-Henningham et al., 2019).

*\*Perception of school environment (Jewkes et al., 2019)*

This 18-item measure was used to assess perceptions of the school environment in a randomised control trial aimed at reducing gender-based violence in South African Grade 8s (Jewkes et al., 2019). Items were rated on a 4-point Likert from strongly agree to strongly disagree. The internal consistency of this measure was good (Cronbach’s alpha = 0.89), but no other detail about this scale was reported and there was no significant change in scores following the intervention (Jewkes et al., 2019). Therefore, we cannot currently be confident that this is a valid or reliable measure of school environment.

## 2.2. Adult Institutional Care Provider

Measures under this construct were broadly aimed at understanding the knowledge and skills, attitudes, and well-being of adult institutional care providers. This might include workers’ knowledge of abuse and reporting requirements, attitudes towards corporal punishment or other harsh discipline practices, or feelings of compassion.

### Adult Institutional Care Provider’s Knowledge and Skills

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#"><u>Knowledge Questionnaire</u></a>	Dilsiz & Mağden (2015)		Turkey	Teacher	English Translation in Peker et al., 2020: It may be high in students who cannot defend their rights, have poor problem-solving and coping skills.	Questionnaire in Turkish and was unable to ascertain the full psychometric properties. More research may be needed to determine its cross-cultural validity in other settings.
<a href="#"><u>Knowledge and Attitudes Scale</u></a>	Mathews et al., (2017)		U.S.	Early childcare and education providers	I would not report child abuse/neglect if I thought the child would be removed from his/her family.	Knowledge items are specific to U.S. law (Pennsylvania and Maine).
<a href="#"><u>Observational Record of the Caregiving Environment (ORCE)</u></a>	NICHD Early Child Care Research		U.S.	Professional	<i>Observation of caregivers.</i>	Requires a trained professional to observe and code.

	Network (1996)					
<b><u>Knowledge, Attitudes and Practices Survey (KAP)</u></b>	UNICEF (2017)		Macedonia	Professional	Most of our colleagues do not know how to proceed and whom to refer the case of violence against a child.	There is currently no known research on the psychometric properties of this survey.
<b><u>*Teacher Adoption of Positive Disciplinary Practices</u></b>	Fabbri et al., (2021)		Tanzania	Professional	<i>Observation of teachers.</i>	There is currently no known research on the psychometric properties of this observation strategy.
<b><u>*Knowledge of Content</u></b>	Gushwa et al., (2019)		U.S.	Teacher	Not reported.	No analysis was reported on the psychometrics and details about the scale were not available. Some preliminary experimental evidence on its construct validity.
<b><u>*Childcare Knowledge</u></b>	Hecker et al., (2021)		Tanzania	Institutional Caregivers	Which consequences does corporal punishment have?  [ a ] The child fears the person who uses corporal punishment.  [ b ] The child changes their behavior and behaves well.  [ c ] The child learns to solve	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.

					conflicts with violence.  [ d ] The child learns to behave well in future.	
<b><u>*Teachers' experience, professional role, perception, and intervention duration</u></b>	Hoefnagels et al., (2021)		Netherlands	Teacher	I regard it as part of my professional duties to detect child maltreatment.	There is some support, but more research is needed to determine the psychometric properties.
<b><u>*Parent and teacher self-report</u></b>	Kolko et al., (1989)		U.S.	Parents and teachers	How aware are you of the problem of children being physically/sexually touched in ways that make them uncomfortable ?	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Perceived Knowledge and Capabilities</u></b>	König et al., (2020)		Germany	Medical Staff	Not reported.	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Teacher Knowledge/Attitude Questionnaire</u></b>	Kraizer (1991)		U.S.	Teacher	Child sexual abuse is less widespread than the media makes it out to be.	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Parents' and Teachers' Knowledge and</u></b>	MacIntyre & Carr (1999)		Ireland	Parents and Teachers	If a girl is sexually assaulted she has always done	Further research would be needed to assess its psychometric

<b><u>Attitudes Questionnaire</u></b>					something to provoke it.	properties in other cultural settings.
<b><u>*Knowledge Questionnaire</u></b>	Martin et al., (2020)		Iran	Teacher	Children under the age of 6 need to have sex education.	Further research would be needed to assess its psychometric properties in other cultural settings.
<b><u>*CSA Knowledge Questionnaire</u></b>	Rheingold et al., (2015)		U.S.	Teacher	Not reported.	No analysis was reported on the psychometrics and details about the scale were not available. Some preliminary experimental evidence on its construct validity.
<b><u>*CSA Prevention Behaviours</u></b>	Rheingold et al., (2015)		U.S.	Teacher	Made changes in organizational policies regarding child sexual abuse.	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.

*Knowledge Questionnaire (Dilsiz & Mağden, 2015)*

This 53-item Turkish questionnaire was designed to measure teacher’s knowledge of five sub-dimensions, neglect, physical abuse, sexual abuse, risk recognition in abuse and neglect, and behavioural symptoms of children. Unfortunately, we were unable to obtain an English translation of this measure and were therefore unable to review the paper (Dilsiz & Mağden, 2015). The sexual abuse dimension was used by Peker and colleagues (2020) in their mixed-method study of counselling teachers’ knowledge of sexual abuse. The internal consistency of this dimension for this sample was reported as excellent (Cronbach’s alpha = 0.94). There was some support for the construct validity of this subscale where counselling teachers in the intervention group scored significantly higher on the knowledge questions, compared to the control group (Peker et al., 2020).

### Knowledge and Attitudes Scale (Mathews et al., 2017)

This measure was developed to assess childcare and education providers' knowledge and attitudes about child maltreatment, including risk factors and legal requirements for reporting (Mathews et al., 2017). The 21 knowledge items are specific to Pennsylvania laws on reporting child abuse and neglect. The content was modelled after previous scales (Crenshaw et al., 1995; Hawkins & McCallum, 2001; Fraser et al., 2010; McGrath et al., 1987), and a multi-stage process was conducted, including focus groups, cognitive interviews, field testing, and expert panel review (Mathews et al., 2017). The 13 items on attitudes towards child abuse and neglect were adapted from a previously validated scale (Walsh et al., 2012b). These attitudes items were also pilot-tested through expert panel review and field testing. Further experimental support for the construct validity of the scale was found through significant improvements in scores for educators subjected to the Look Out for Child Abuse (iLookOut) educational intervention (Mathews et al., 2017).

This questionnaire was adapted by Humphreys and colleagues (2021) to be relevant to Maine law. An expert panel from Maine's Office of Child and Family Services reviewed the changes. The internal consistency was acceptable for the knowledge questionnaire (Cronbach's alpha = 0.72) and good for the attitudes scale (Cronbach's alpha = 0.80; Humphreys et al., 2021).

There is strong support for this scale, but it is currently limited to US-based legal systems. Further research would be needed to adapt this scale for use in other jurisdictions.

### *Observational Record of the Caregiving Environment (ORCE; NICHD Early Child Care Research Network, 1996)*

The ORCE was developed to observe and record caregivers' behaviours towards infants, including responding to social signals and distress, stimulating cognitive development, and expressing positive feelings. Caregivers' experiences and beliefs about child rearing and safety are also assessed by the OCRE. These factors have been identified in the literature as contributing to social and cognitive development, and were informed by experts in the field (NICHD Early Child Care Research Network, 1996).

This measure was developed specifically for observing caregivers in different environments, including institutional caregivers. Previous equivalent measures were specific to home settings.

The ORCE has both behaviour scales (frequency of positive behaviours) and qualitative ratings of behaviours, which are scored over several observations. Coders must be trained and, to maintain certification must demonstrate exact agreement with master codes on example tapes. This helps to maintain the reliability of scores on this measure. The OCRE was found to have good internal consistency (Cronbach's alpha = 0.81) and its factor structure was confirmed through confirmatory factor analysis (NICHD Early Child Care Research Network, 1996). Scores on the OCRE were also found to be significantly correlated with scores on the Child Care HOME, providing support for its construct validity.

An adapted version of the measure was used in Romania to assess a child's caregiving experience in either institutional or family settings. Here, children were videotaped in their home (institution, foster, or family home) with their favourite caregiver for 1.5 hours (Johnson et al., 2010; Smyke et al., 2007). Coders were trained to establish interrater reliability (Zeanah et al., 2005). Caregiving quality was established by averaging the scores across the domains on the measure (sensitivity,

stimulation of development, positive regard for the child, detachment (reversed), and flat affect (reversed)). The internal consistency of this adapted version for this population was found to be good (Cronbach's alpha = 0.82; Smyke et al., 2010). There seems to be strong support for this measure when coded by trained professionals.

### *Knowledge, Attitudes and Practices Survey (KAP; UNICEF, 2017)*

This UNICEF survey was developed to measure professionals' knowledge, attitudes, and practices in relation to child disciplining practices and child abuse. It contains seven sections designed to assess: (1) knowledge on violent and alternative positive child discipline methods, (2) attitudes about child discipline practices, (3) beliefs about the effects and consequences of violence and positive child discipline practices, (4) practices in identifying, reporting, responding to and referring cases of violence against children, (5) empirical and normative expectations about violence against children, (6) conditional preferences to conform to empirical and normative expectations, and (7) networks and sources of information for advice and support in identifying, reporting, referral, and responding to violence against children. The survey is available in English and Macedonian. To our knowledge, there has been no analysis yet on the psychometric properties of this survey, therefore we cannot be confident that it is a reliable and valid measure.

The original survey has not been used yet in experimental research on institutional child maltreatment. Wangamati and colleagues (2022) included an adapted version of the survey in their protocol for a mixed methods pilot study aimed at reducing violence in schools in Kenya and Tanzania. There was mention of adaptations planned for use with both teaching and non-teaching staff, as well as parents. The parent version will also include items adapted from the [Alabama Parenting Questionnaire \(APQ\)](#). As this was a protocol, there was no analysis reported.

### *\*Teacher adoption of positive disciplinary strategies (Fabbri et al., 2021)*

In Fabbri and colleagues' (2021) cluster-randomised controlled trial in Tanzania, observers simply counted the number of 'positive practices' observed in a classroom setting. Positive practices may have included teachers' using de-escalation strategies or rewarding positive behaviour in children, although these details were not clearly defined in the paper. No analysis on inter-rater reliability was reported. There was mention of cognitive testing, but the details were not reported in the article. While this observation may give some sense of teachers' adoption of positive practices, it is important for measures, even simple ones, to be subjected to testing. Clear operational definitions of 'positive practices', clear details on pilot testing, and checks for inter-rater reliability would improve our confidence in the psychometric properties of this measure.

### *\*Knowledge of content (Gushwa et al., 2019)*

Gushwa and colleagues (2019) developed a 13-item measure designed to assess teachers' knowledge of sexual abuse, specifically, the content covered in the Enough! Training program. Items included content about the prevalence, impact, signs, and symptoms of sexual abuse, types of sexual abuse behaviours, the impact on children, reporting responsibilities, the veracity of children's reports, and responses to suspected abuse. Gushwa and colleagues (2019) also included 15 items on teachers' perceived increases in knowledge, reporting of future suspected or disclosed cases, the impact of identification of boundary-violating behaviours, and general attitudes towards the training. While these measures include important topics to assess teachers' knowledge of sexual abuse, and the authors made mention of 'reviewing' the questions, there is



little support yet for their psychometric properties. There was preliminary support for the construct validity of the measures as teachers in the intervention group showed significant improvements in their knowledge of sexual abuse, compared to the control group (Gushwa et al., 2019). However, we cannot yet be certain that they are reliable or valid measures of teachers' knowledge or perceived knowledge about sexual abuse.

#### *\*Childcare knowledge (Hecker et al., 2021)*

Hecker and colleagues (2021) developed an 11-item measure of childcare knowledge. It contained items designed to assess basic knowledge and competencies about working in childcare or with children. Each item had 3-4 possible answers, with one possible correct answer. There was no analysis reported on the reliability or validity of this scale. There was no reported analysis on the psychometric properties of this measure. There was some preliminary experimental support for the construct validity of the scale, with caregivers in the intervention displaying significant improvements in childcare knowledge. However, further research would be needed to determine the reliability and validity of this measure.

#### *\*Teachers' experience, professional role, perception, and intervention duration (Hoefnagels et al., 2021)*

In their non-randomised cluster-controlled trial, Hoefnagels and colleagues (2021) developed a questionnaire designed to collect information about teachers' sex, class composition, years of experience, their perceived ability to prevent maltreatment, and whether they had suspected child maltreatment in the last two months. A proxy for communication skills was also used by asking if teachers had talked to a parent about suspected child maltreatment, been in contact with medical services about suspected child maltreatment, or had a child confided in them. Teachers in the intervention group were also asked about the duration of the intervention sessions as well as their perceived role in relation to child maltreatment. The internal consistency of the questionnaire was acceptable (Cronbach's alpha = 0.71; Hoefnagels et al., 2021), although this is not necessarily an indication of the quality of the measure as it is designed to measure several not necessarily related constructs. There was some preliminary support for the construct validity of the scale, however, further research would be needed to determine the reliability and validity of this measure.

#### *\*Parent and teacher self-report (Kolko et al., 1989)*

In Kolko and colleagues' (1989) evaluation of a sexual victimisation awareness and prevention skills program, parents and teachers completed a 10-item questionnaire designed to assess their knowledge, awareness, and preparedness to act in response to sexual abuse. Each item was rated on a 5-point Likert scale. The parent and teacher versions were identical, but teachers were asked to report on children in their classroom, instead of on their own children. Factor analysis provided support for the three-factor structure, and there was preliminary experimental support for its construct validity. The questionnaire had questionable to acceptable internal consistency (Cronbach's alpha = 0.66-0.72). There is promising support, but further research would be needed.

#### *\*Perceived knowledge and capabilities (König et al., 2020)*

In a German randomised controlled trial, medical staff were asked about their perceived knowledge of child protection (item 1) and their perceived practical capabilities regarding child protection (item 2). Both items were rated on a 6-point Likert scale from 1 (very little) to 6 (very extensive). No

further detail was provided, and there was no analysis reported on the psychometrics of these items (König et al., 2020). There was some preliminary support for the construct validity of the scale, however, further research would be needed to determine the reliability and validity of this measure.

#### *\*Teacher Knowledge Attitude Questionnaire (Kraizer, 1991)*

Kraizer (1991) included two measures designed to assess differences among individual teachers. Teachers first completed a demographic questionnaire and then completed a 19-item Knowledge/Attitude Questionnaire. The questions were designed to assess knowledge, understanding, and attitudes about all forms of child abuse. There was some preliminary support for the construct validity of the questionnaire, but more research is needed to determine its psychometric properties.

#### *\*Parents' and Teachers' Knowledge and Attitudes Questionnaire (MacIntyre & Carr, 1999)*

This 38-item questionnaire was developed in an Irish study to assess parents' and teachers' knowledge and attitudes (MacIntyre & Carr, 1999). Parents and teachers were asked about the prevalence, indicators, and causes of child sexual abuse, characteristics of victims and offenders, and service-related information. Each item was rated on a 5-point Likert scale from "strongly agree" to "strongly disagree". The internal consistency of the measure was good (Cronbach's alpha = 0.80; MacIntyre & Carr, 1999). This study found that following the intervention, teachers' and parents' knowledge and attitudes had improved, providing experimental support for its construct validity. Factor analysis also revealed that the questionnaire was unidimensional (MacIntyre & Carr, 1999). It has not been used in any subsequent experimental research on institutional child maltreatment.

#### *\*Knowledge questionnaire (Martin et al., 2020)*

Martin and colleagues (2020) included a 31-item questionnaire designed to assess teachers' knowledge in their quasi-experimental study in Iran. This measure covered six domains, including sex education, sexual identity, stages of sexual development, approaches to sex education, children's sexual questions, masturbation, and child sexual abuse. Participants were asked to indicate if each statement is "true", "false", or "don't know". Correct answers were scored as 1 for each item. Higher scores on this measure indicated greater knowledge.

This measure was given to 10 midwives and reproductive health experts at Shahid Beheshti University of Medical Sciences to assess the content validity. The content validity index (CVI) and content validity ratio (CVR) were calculated and were found to be excellent (0.97 and 0.93, respectively; Martin et al., 2020). The internal consistency of the knowledge measure was adequate (Cronbach's alpha = 0.79). There was good support for this measure in assessing teachers' knowledge of sexual abuse and the importance of sex education in the prevention of child sexual abuse (Martin et al., 2020).

#### *\*CSA Knowledge Questionnaire (Rheingold et al., 2015)*

This questionnaire was created by Rheingold and colleagues (2015) for their U.S. based randomised controlled trial. The measure contained 12 true/false questions about child sexual abuse and covered concepts such as prevalence, risk factors, impacts, and prevention. No



analysis of the psychometric properties of this measure was reported, and there was no further detail provided in the article about the measure. There was some preliminary support for the construct validity of the questionnaire, but more research is needed.

*\*CSA Prevention Behaviours (Rheingold et al., 2015)*

Rheingold and colleagues (2015) also included a 21-item measure of teachers' child sexual abuse prevention behaviours in the past 3 months. This included behaviours such as preventing adults from having unsupervised time with children, talking to a child or other adult about sexual abuse, making organisational policy changes, or reporting sexual abuse to authorities. No analysis was reported on the reliability or validity of this measure. There was some preliminary experimental support for the construct validity of the questionnaire, but more research is needed.

### Adult Institutional Care Providers' Attitudes

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#"><u>Teacher Willingness to Teach Sexual Health Education Questionnaire</u></a>	Cohen et al., (2014)		Canada	Teacher	At this point in time how knowledgeable do you feel to teach children about healthy relationships?	There is good support, particularly for the teacher comfort scale. More research is needed.
<a href="#"><u>Childhood Sexual Abuse Myth Scale (CSAM)</u></a>	Collings (1997)		South Africa	Adults	Children who do not report ongoing abuse must want the sexual contact to continue.	There is a possible floor effect with some professionals (e.g., childcare workers). Endorsement of CSA myths may be affected by different cultural settings or types of professionals.
<a href="#"><u>Teachers' Attitudes Towards Reporting Child Sexual Abuse Scale (TRAS-CSA)</u></a>	Ekşi et al., (2018)		Turkey	Teacher	<i>Turkish version of TRAS-CSA.</i>	There is good support for the Turkish version.
<a href="#"><u>Violence Against Women Attitude Scale</u></a>	Kanbay et al., (2017)		Turkey	Adult	-	Questionnaire in Turkish and was unable to ascertain the full psychometric

						properties. More research may be needed to determine its cross-cultural validity in other settings.
<b><u>Educators and Child Abuse Questionnaire (ECAQ)</u></b>	Kenny (2000)		U.S.	Teacher	I am aware of the signs of child sexual abuse.	More research would be needed to assess its validity in other cultural settings.
<b><u>The Survey of Perceived Confidence in, and Attitudes towards, Approaches to Teaching and Learning about Sensitive Issues in Health and Personal Development</u></b>	Lynagh et al., (2010)		Australia	Teacher	At this point in time I feel confident and competent in being able to respond to difficult questions related to abuse from children.	There is some preliminary support for the construct validity of the original scale. There is more support for the adapted version used in McElearney et al., (2021).
<b><u>Healthcare Provider Attitudes Toward Child Maltreatment Reporting Scale</u></b>	Singh et al., 2017. Turkish translation: Turan & Erdoğan (2019)		U.S.	Professional (Healthcare version of TRAS-CSA)	I plan to report child maltreatment when I see it.	More research would be needed to assess its validity in other cultural settings.
<b><u>Teacher Reporting Attitude Scale – Child Sexual Abuse (TRAS-CSA)</u></b>	Walsh et al., (2012b)		Australia	Teacher	It is important for teachers to be involved in reporting child sexual abuse to prevent long-term consequences for children.	The scale is specific to child sexual abuse only.
<b><u>Gender Roles Attitude Scale (GRAS)</u></b>	Zeyneloğlu & Terzioğlu (2011)		Turkey	Adult	Woman should prefer to remain silent instead of arguing in case of a	More research would be needed to assess its

					conflict with their husbands.	validity in other cultural settings.
<b><u>*Attitudes on Child Beating</u></b>	Fabbri et al., (2021)  (based on Ruis-Caseres, 2011)		Tanzania	Teacher	It is acceptable to hit a student if the student is interrupting the class.	There is some support, but more research is needed to determine the psychometric properties.
<b><u>*Embarrassed or offended about talking about sex with a child</u></b>	Jewkes et al., (2019)		South Africa	Teacher	If your grade 8 child wanted to talk about sex with you, how embarrassed would you be?	There is currently no known research on the psychometric properties of these items.
<b><u>*Perspectives on one's own role in an institution</u></b>	König et al., (2020)		Germany	Medical manager	I feel better prepared to meet my responsibilities as manager in regard to institutional child protection.	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Attitude questionnaire</u></b>	Martin et al., (2020)		Iran	Teacher	In my opinion, sex education for children puts them at risk and abusing.	More research would be needed to assess its psychometric properties in other cultural settings.

### *Teacher Willingness to Teach Sexual Health Education Questionnaire (Cohen et al., 2004)*

This questionnaire contains 16 items designed to measure teacher willingness to teach sexual health education. In Part A, teachers are asked about their general opinions on including sexual health education in school, from (1) strongly disagree to (5) strongly agree (5-point Likert scale). Part B includes items assessing how important teachers feel it is cover 10 broad sexual health topics. These items are rated on a 5-point Likert scale from (1) not at all important to (5) extremely important. Part C asks at what grade level sexual health education should start. Part D asks teachers to rate how knowledgeable they feel to teach 26 topics on a 5-point Likert scale from (1) not at all knowledgeable to (5) extremely knowledgeable. Finally, teachers indicate if they have

received any training on providing sexual health education. There was support for the construct validity of this questionnaire, but no analysis was reported on its other psychometric properties.

The teacher comfort scale was used in a cluster-randomised trial in Northern Ireland (McElearney et al., 2021). This subscale assesses teachers' comfort in teaching safe messages and was adapted to include topics covered in the prevention programme used in the study (Keeping Safe programme). Confirmatory Factor Analysis in this study revealed a three-factor model for this subscale: knowledge, confidence, and comfort. There was also support for the construct validity of the adapted subscale, with teachers in the intervention group reporting significantly greater improvements in knowledge, confidence, and comfort in teaching sexual health education, compared to the control group (McElearney et al., 2021).

### *Childhood Sexual Abuse Myth Scale (CSAM; Collings, 1997)*

The CSAM has 15 statements on childhood sexual abuse myths. Participants rate the statements on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Scores can range from 15-74, with lower scores indicating lower endorsement of myths. The scale assesses three factors: blame diffusion, denial of abusiveness, and retrospective stereotypes (Collings, 1997). It was designed to assess adult endorsement of child sexual abuse myths and was found to have acceptable internal consistency (Cronbach's alpha = 0.76), good test-retest reliability ( $r = 0.87$ ), and acceptable levels of convergent and discriminate validity (Collings, 1997).

Rheingold and colleagues (2015) used this scale in their study of primary and secondary prevention efforts in childcare settings in the U.S. A key finding was that childcare professionals, on average, scored in the low 20s on the CSAM. This suggests a possible floor effect to the measure, meaning that it is difficult to differentiate between participant scores and between groups.

In other research, the CSAM has been found to be able to differentiate between endorsement of CSA myths in psychologists in South Africa (Collings, 2003). It seems that the CSAM may not be suitable to assess the endorsement of CSA myths with trained childcare professionals in some settings or cultures. It may be more useful with a general adult population or with other types of workers that do not necessarily have training on child abuse. Rheingold and colleagues (2015) concluded that decreasing CSA myths in childcare professionals may be a less relevant aim for future research. Further research would be needed to assess the validity of the CSAM in different cultural settings and with different workers.

### *Teachers' Attitudes Towards Reporting Child Sexual Abuse Scale (Ekşi et al., 2018)*

This Turkish scale was based on the [TRAS-CSA](#) and has three sub-dimensions aimed at assessing teachers' understanding of their reporting role, their confidence that the system will respond effectively to reports of child sexual abuse, and their concern about the potential consequences of reporting (Ekşi et al., 2018). Higher scores on this scale indicate more positive attitudes towards reporting sexual abuse. There is evidence to support its reliability and validity for measuring Turkish teachers' attitudes towards reporting CSA (Koç et al., 2020).

The Turkish version of this scale was used once in an experimental study on the EGM. In this study, the effect of a psychoeducation program on Turkish counselling teachers' risk recognition attitudes and reporting of sexual abuse (Peker et al., 2020) was explored. The overall scale had good internal consistency (Cronbach's alpha = 0.84), but each sub-dimension had questionable to

acceptable internal consistency (0.60-0.77; Peker et al., 2020). There was some support for the construct validity of the scale, with counselling teachers in the intervention group showing increased positive attitudes toward reporting sexual abuse.

### *Violence Against Women Attitude Scale (Kanbay et al., 2017)*

The Violence Against Women Attitude Scale was developed in Turkey with 30 items designed to measure attitudes towards violence against women. Scores range from 30-150, with lower scores indicating acceptance of violence attitudes. This scale has been found to have excellent internal consistency (Cronbach's alpha = 0.86) and test-retest reliability (0.81; Kanbay et al., 2017). As this scale was developed in Turkey, we were unfortunately unable to review the full details of the scale.

In a Turkish quasi-experimental study, this scale was used to assess nursing students' attitudes towards violence against women (Turan, 2022). Participants in the education group showed improvements in scores on this measure compared to the control group, providing experimental support for the construct validity (Turan, 2022).

### *Educators and Child Abuse Questionnaire (ECAQ; Kenny, 2000)*

The ECAQ is a self-report measure of educators' attitudes, competence, and experiences in the identification and reporting of child abuse and neglect. The first section collects information about the educators' personal data, including demographic information, definitions for child abuse and neglect, and incidences of reporting of child abuse (Kenny, 2000). The second section is aimed at the educators' educational background in child abuse reporting. Section Three then assesses the attitudes and personal beliefs of educators, including concepts such as perceptions of child discipline, perceived personal competence, and reporting procedures. The final section includes two hypothetical case studies that were adapted from James and DeVaney (1994). The vignettes were reviewed by a panel of experts and were confirmed as legally reportable incidents of child abuse in the U.S.

In the development of this scale, most teachers were found to have never made a child abuse report and reported inadequate training in child abuse signs, symptoms, and reporting procedures. Kenny concluded that there is a need for educating teachers on reporting, including interventions that focus on teachers' perceived deterrents to reporting (such as feeling as though child protection services do not help families, and fears about making an inaccurate report). The ECAQ has been found to have good content validity (Kenny, 2004), but has questionable (Cronbach's alpha = 0.64) to good (Cronbach's alpha = 0.85) internal consistency. The ECAQ may be a useful tool in assessing teachers' attitudes and competence in reporting child abuse.

Kim and colleagues (2019) used the third section in their study of teacher outcomes in the U.S. They did not report any specific validity checks but used this subscale of the ECAQ to assess teachers' awareness and knowledge of child abuse. Teachers were found to significantly increase their awareness and attitudes related to child abuse using this subscale, following the intervention program (Second Step Child Protection Unit), providing some support for the construct validity. This effect was particularly relevant for teachers who had lower knowledge and attitudes scores prior to the intervention (Kim et al., 2019).

*The Survey of Perceived Confidence in, and Attitudes towards, Approaches to Teaching and Learning about Sensitive Issues in Health and Personal Development (Lynagh et al., 2010)*

The questionnaire was developed in an evaluation of a 13-unit program for pre-service teachers in Australia. It contains three sections, with each item rated on a 5-point Likert scale from (1) strongly disagree to (5) strongly agree. The first section has 8 items that assess teachers' perceived confidence in teaching sensitive issues such as interpersonal relationships, sexual orientation, and reproduction. Section Two has 10 items that assess teachers' confidence in dealing with sensitive issues, such as helping a student who discloses that are being abused. The final section has 6 items that examine attitudes toward addressing sensitive issues and their beliefs about the importance of each issue (Lynagh et al., 2010). There was no analysis reported on the psychometric properties of this questionnaire, but there was some preliminary experimental support for the construct validity of the questionnaire. Pre-service teachers who attended the Units were found to increase in both confidence and competence in responding to sensitive issues (Lynagh et al., 2010). More research would be needed to assess the reliability and validity of the original scale.

McElearney and colleagues (2021) used items from the second part of the survey to assess teachers' confidence in managing sensitive issues. The authors adapted the subscale to be more targeted to components of the intervention (Keeping Safe programme) by having a focus on classroom management issues. There was support for the construct validity of the adapted subscale (McElearney et al., 2021).

*Healthcare Provider Attitudes Toward Child Maltreatment Reporting Scale (Singh et al., 2017)*

This scale is a revised version of the [TRAS-CSA](#), aimed at assessing healthcare workers' attitudes toward reporting child maltreatment (Singh et al., 2017). The wording of the TRAS-CSA was modified for use with medical students by removing terminology relevant to the education field. They also changed the terminology from 'child sexual abuse' to 'child maltreatment' to assess attitudes towards reporting a broad range of child abuse and neglect. This revised version was found to have good internal consistency and seemed to be useful in assessing attitudes for reporting child maltreatment in healthcare workers (Singh et al., 2017).

The original version of this scale has not been used in a study on the EGM. However, a Turkish-translated version of this scale (Turan & Erdoğan, 2019) was used in a quasi-experimental evaluation of a structured online educational program for nursing students in Turkey (Turan, 2022). A Cronbach's alpha of 0.911 was reported for this version in the study, indicating excellent internal consistency. No other analysis was reported on the translated version.

*Teacher Reporting Attitude Scale – Child Sexual Abuse (TRAS-CSA; Walsh et al., 2012b)*

The original TRAS-CSA had 21 items designed to assess teachers' attitudes towards reporting child sexual abuse. The original scale was developed through a five-stage process that included item development, panel review, focus groups, cognitive interviewing, and field testing (Walsh et al., 2010). Items are rated on a 5-point Likert scale from (1) strongly agree to (5) strongly disagree. Higher scores on this scale indicate more positive attitudes towards reporting.



In a validation study using a sample of Australian teachers, there were some item redundancies in the full 21-item version. A 14-item version of the scale was validated for use (Walsh et al., 2012b). Exploratory factor analysis demonstrated support for three distinct factors in the 14-item version: commitment to the reporting role, confidence in the system's effective response to reports, and concerns about the consequences of reporting. These findings demonstrated that Australian teachers' attitudes toward reporting were closely aligned with characteristics of the statutory child protection system (Walsh et al., 2012b).

Kim and colleagues (2019) used the 14-item version in their study to assess teachers' attitudes toward reporting. It was found to have good internal consistency (Cronbach's alpha = 0.82-0.84). There was experimental support for the construct validity of the TRAS-CSA in this study as teachers in the intervention group reported more positive attitudes towards reporting sexual abuse, compared to the control group (Kim et al., 2019).

### *Gender Roles Attitude Scale (GRAS; Zeyneloğlu & Terzioğlu, 2011)*

The GRAS was developed in Turkey and is a 38-item self-report measure of adult views on gender roles. This scale was normed on a population of university students. Scores can range from 38-190, with lower scores indicating support for more traditional gender roles. The GRAS measures five dimension: egalitarian gender roles, female gender roles, marriage gender roles, traditional gender roles, and male gender roles).

The GRAS was subjected to a review by an expert panel and factor analysis provided support for the validity of the scale. Scores on the GRAS significantly correlated with scores on the Attitude Toward Women Scale (AWS; Spence et al. 1997), providing support for the construct validity of the scale. It also had excellent internal consistency (Cronbach's alpha = 0.92; Zeyneloğlu & Terzioğlu, 2011), providing support for the reliability of the GRAS.

Turan (2022) used the GRAS to measure nursing students' attitudes on gender roles in their quasi-experimental study. It had excellent internal consistency in their sample (Cronbach's alpha = 0.93), and there was experimental support for its construct validity as well (Turan, 2022). Overall, there is good support for the GRAS for use in Turkish samples.

### *\*Attitudes on Child Beating (Fabbri et al., 2021)*

A teachers' self-report scale was used in Fabbri and colleagues' (2021) cluster-randomised controlled trial and was based on a parent attitude scale (Ruis-Caseres, 2011). This scale was designed to assess teachers' attitudes towards corporal punishment and includes items about a teacher's views on the acceptability of hitting a student if a student talks back, is absent from school, insults a teacher, does not take notes during a lesson, or if a student fights with other students. Items are rated on a 4-point Likert scale with a mean score calculated based on an average rating. This scale had good internal consistency in this sample of teachers in Tanzania (Cronbach's alpha = 0.89; Fabbri et al., 2021). There was no meaningful effect of the intervention on teacher outcomes, including attitudes towards child beating. This may have been because either the intervention had no significant effect on attitudes, or because the measure was not a valid measure of teacher attitudes (Fabbri et al., 2021). More research would be needed to assess the reliability and validity of this scale.

*\*Embarrassed or offended about talking about sex with a child (Jewkes et al., 2019)*

In Jewkes and colleagues (2019) two simple items were included that asked teachers if they felt embarrassed or offended about talking about sex with a child. The two items were rated on a 3-point Likert scale from very, somewhat, to not at all. There was no analysis reported on the reliability or validity of these items.

*\*Perspectives on one's own role in an institution (König et al., 2020)*

König and colleagues (2020) included 11 items designed to measure whether participation in the course, changed the medical managers' perspectives of their role in the context of institutional child protection. Items were scored on a 6-point Likert scale from 1 (don't agree at all) to 6 (agree completely). No analysis was reported on the psychometrics of this measure. There was some support for the construct validity based on experimental results, but more research would be needed.

*\*Attitude questionnaire (Martin et al., 2020)*

Similar to the above (see Martin et al., (2020) [Knowledge Questionnaire](#)), teachers were also asked about their attitudes towards sex education, sexual identity, stages of sexual development, correct approaches to sex education, children's sexual questions, masturbation, and child sexual abuse. Each item was rated on a 5-point Likert scale from 1 (completely agree) to 5 (completely disagree). Scores could range from 26 to 130, with higher scores indicating a better attitude.

Content validity ratio (CVR) and content validity index (CVI) were also excellent for the attitude questionnaire (0.9 and 0.99 respectively; Martin et al., 2020). Internal consistency was acceptable (Cronbach's alpha = 0.77). These two measures have good support for use in this context, but they have not been used in any subsequent experimental research on the prevention of institutional child maltreatment. Further research would be needed to determine the reliability and validity of these measures in other cultural settings.

**Adult Institutional Care Provider's Wellbeing**

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#">Teacher Self-Efficacy Scale</a>	Bandura (n.d)		U.S.	Teacher	How much can you do to prevent problem behavior on the school grounds?	More research may be needed to determine its cross-cultural validity in other countries.
<a href="#">Social Safeness and Pleasure Scale (SSPS)</a>	Gilbert et al., (2009)		U.K.	Adult	I feel a sense of belonging.	More research may be needed to determine its cross-cultural validity in other countries.



<b><u>Fears of Compassion Scale (FCS)</u></b>	Gilbert et al., (2009)		U.K.	Adult	People will take advantage of me if they see me as too compassionate.	More research may be needed to determine its cross-cultural validity in other countries.
<b><u>Compassion Scale (CS)</u></b>	Pommier et al., (2020)		U.S.	Adult	I pay careful attention when other people talk to me about their troubles.	More research may be needed to determine its cross-cultural validity in other countries.
<b><u>Self-Compassion Scale (SCS)</u></b>	Neff (2003)		U.S.	Adult	I'm tolerant of my own flaws and inadequacies.	More research may be needed to determine its cross-cultural validity in other countries.
<b><u>Centre for Epidemiological Studies Depression Scale (CES-D)</u></b>	Radloff (1977)		U.S.	Caregivers	I felt that I could not shake off the blues even with help from my family or friends.	More research may be needed to determine its cross-cultural validity in other countries.
<b><u>Teacher Burnout Scale</u></b>	Seidman et al., (1970)		U.S.	Teacher	The stressors in this job are more than I can bear.	More research may be needed to determine its cross-cultural validity in other countries.
<b><u>Brief Self-Control Scale</u></b>	Tangney et al., (2004)		U.S.	Adult	I am good at resisting temptation.	This seems to be a reliable and valid measure in the U.S., but there is little evidence for its psychometric properties in Africa.
<b><u>*Teacher Perception Questionnaire (TPQ)</u></b>	Wurtele et al., (1989)		U.S.	Teacher	Seems self-confident	There is limited support for the psychometric properties of this scale currently.

* <u>Teacher wellbeing in school</u>	Devries et al., (2015)		Uganda	Teacher	Do you feel that your employers care about your well-being?	There is limited support for the psychometric properties of this scale currently.
--------------------------------------	------------------------	--	--------	---------	---	---

*Teacher Self-Efficacy Scale (Bandura, n.d.)*

This 30-item self-report scale was designed to assess teachers’ understanding of the things that make activities in their school more difficult. Efficacy to influence decision-making, school resources, and instructional self-efficacy are assessed in this scale. There is strong evidence for the cross-cultural validity of this scale, with evidence for its reliability and validity in North America, East Asia, Europe, and Canada. It has also been found to be strongly correlated with other teacher outcomes, such as job satisfaction, providing support for its construct validity (Klassen et al., 2009).

This scale was originally developed in the U.S. but has only been used in experimental research on institutional child maltreatment in Jamaica (Baker-Henningham et al., 2019; Baker-Henningham et al., 2021a). These studies used subscales from the full scale. Baker-Henningham and colleagues (2019) used four subscales (instructional, disciplinary self-efficacy, efficacy to enlist parental involvement, and efficacy to create a positive school climate), whereas Baker-Henningham and colleagues (2021a) used three subscales (disciplinary self-efficacy, efficacy to enlist parental involvement, and efficacy to create a positive school climate). These studies did not report the efficacy of the use of the subscales in Jamaica and did not make mention of any adaptations made. There is some preliminary support for its cross-cultural validity in Africa but more research may be needed.

*Social Safeness and Pleasure Scale (SSPS; Gilbert et al., 2009)*

SSPS is an 11-item self-report scale for adults to report their individual experiences of positive feelings in their relationships, as well as how they experience the world as a safe place. Each item is rated on a 5-point Likert scale from (0) almost never to (4) almost all the time. The scale has been found to have strong construct and discriminate validity and demonstrates a high degree of reliability (Gilbert et al., 2009; Kelly et al., 2012). The scale also has excellent internal consistency (Cronbach’s alpha = 0.96).

The original English version has not been used in rigorous experimental research on the prevention of institutional child maltreatment. Santos and colleagues (2022) used a Portuguese version in their cluster randomised trial of compassionate mind training for caregivers in a residential youth facility. In this study, the scale had good internal consistency (Cronbach’s alpha = 0.88). No other analysis was reported on the translated version but there was some experimental support for its construct validity (Santos et al., 2022).

*Fears of Compassion Scale (FCS; Gilbert et al., 2011)*

Gilbert and colleagues (2011) developed three scales designed to assess fear of compassion for self, fear of compassion from others, and fear of compassion for others. Originally, 20 items were

created for each and subjected to review by a research team for face validity. The final version contained 45 items with each rated on a 5-point Likert scale from (0) don't agree at all to (4) completely agree. The FCS has acceptable to good internal consistency (Cronbach's alpha = 0.72-0.83; Gilbert et al., 2011). Exploratory factor analysis showed support for the construct validity of the FCS. Scores on the FCS were also related to other symptoms such as anxiety and depression, providing support for the construct validity of the scale (Gilbert et al., 2011).

The original version of the FCS has not been used in experimental research on institutional child maltreatment, but a Portuguese-translated version was used by Santos and colleagues (2022) in their cluster randomised trial. In their sample of residential caregivers, the internal consistency of the FCS was good to excellent (Cronbach's alpha = 0.91-0.96) and there was experimental support for its construct validity. More research may be needed to assess its psychometric properties in other cultural contexts.

### *Compassion Scale (CS; Pommier et al., 2020)*

The CS is a 24-item self-report scale designed to measure compassion for others. Each item is rated on a 5-point Likert scale from (1) almost never to (5) almost always. The CS contains six subscales: kindness, common humanity, mindfulness, indifference, separation, and disengagement. There is strong evidence for its psychometric properties, including its internal consistency (Cronbach's alpha = 0.90), test-retest reliability, construct validity, and content validity (Pommier et al., 2020).

The original version has not been used in experimental research on institutional child maltreatment, but a Portuguese-translated version was used by Santos and colleagues (2022). Same as the original version, the Portuguese version was found to have a two-factor model: compassion (kindness, common humanity, mindfulness) and disconnectedness (indifference, separation, and disengagement). This version had adequate to good internal consistency (Cronbach's alpha = 0.76-0.84) and there was experimental support for its construct validity (Santos et al., 2022). More research may be needed to assess its psychometric properties in other cultural contexts.

### *Self-Compassion Scale (SCS; Neff, 2003)*

The SCS was designed to assess adults' compassion towards themselves (Neff, 2003). It contains 26-items with each rated on a 5-point Likert scale from (1) almost never to (5) almost always. It has similar subscales as the [CS](#) but is aimed at self-compassion (self-kindness, self-judgement, common humanity, isolation, mindfulness, and over-identification). It was normed on a sample of college undergraduates and there was strong support for its reliability and validity (Neff, 2003). Exploratory and Confirmatory Factor Analysis provided support for its construct validity, internal consistency was adequate to good (Cronbach's alpha = 0.75-0.81), and test-retest reliability was excellent (Cronbach's alpha = 0.93; Neff, 2003).

The original version has not been used in experimental research on institutional child maltreatment, but a Portuguese-translated version was used by Santos and colleagues (2022). This version had good internal consistency (Cronbach's alpha = 0.87-0.88) and there was experimental support for its construct validity (Santos et al., 2022). More research may be needed to assess its psychometric properties in other cultural contexts.

### *Centre for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977)*

The CES-D is a 20-item scale designed to assess caregivers' symptoms of depression, including feelings of loneliness, restless sleep, and poor appetite. Participants rate each item from 0 to 3 (0 = rarely or none of the time, 1 = some or little of the time, 2 = moderately or much of the time, and 3 = most or almost most of the time). Higher scores indicate higher depressive symptoms (Radloff, 1977). The CES-D was found to have strong psychometric properties. It has good to excellent internal consistency (Cronbach's alpha = 0.84-0.90) and adequate test-retest coefficient given the scale being designed to be very sensitive to changes in depressive symptoms (Radloff, 1977). It was not designed to be a diagnostic tool for depression and was therefore sensitive to general and psychiatric in-patient populations. There was strong support for its validity, including content and construct validity as well (Radloff, 1977).

It was used once in an EGM study to assess depression symptoms in Jamaican teachers as part of an evaluation of a cluster randomised control trial aimed at preventing violence in schools (Baker-Henningham et al., 2019; Baker-Henningham et al., 2021). The authors did not report any analysis to test for the validity of this measure for use with Jamaican teachers. The intervention had no effect on depression scores for teachers, so it is not clear if this was due to the validity of the measure on this sample or because the intervention did not affect teachers' depression symptoms (Baker-Henningham et al., 2019).

### *Teacher Burnout Scale (Seidman et al., 1970)*

The Teacher Burnout Scale contains 21 items designed to measure burnout among public school teachers in the U.S. Factor analysis on a sample of 365 teachers, revealed four factors that were consistent with the theoretical constructs for burnout. These factors included career satisfaction, perceived administrative support, coping with job-related stress, and attitudes towards students (Seidman et al., 1970). Follow-up analysis also confirmed that the scale has adequate to good internal consistency (Cronbach's alpha = 0.72-0.89), stable test-retest reliability coefficients, and construct and predictive validity (Seidman et al., 1970). There is good support for the use of this measure to assess burnout in teachers in the U.S.

The Teacher Burnout Scale was used to assess burnout in Jamaican teachers (Baker-Henningham et al., 2019; Baker-Henningham et al., 2021). The authors did not report any analysis to test for the validity of this measure for use with this population. In the study, the intervention had no effect on burnout scores for teachers (Baker-Henningham et al., 2019). This may have been because either the intervention had no effect on burnout scores for teachers, or the measure was not a valid measure of burnout in Jamaican teachers (Baker-Henningham et al., 2019).

### *Brief Self-Control Scale (Tangney et al., 2004)*

The Brief Self-Control Scale is a 12-item measure designed to assess adults perceived skills in resisting temptation, having a hard time breaking bad habits, saying inappropriate things, and wishing they had more self-discipline. Items are rated on a 5-point Likert scale, with an overall score on self-control calculated based on a mean score on these items (Tangney et al., 2004). It was found to have good internal consistency (Cronbach's alpha = 0.83-0.85) and test-retest reliability (0.87-0.89; Tangney et al., 2004). Convergent validity was confirmed through correlations with higher scores on self-esteem, secure attachment, and interpersonal skills. Discriminant validity

was confirmed with negative correlations with binge eating and alcohol abuse scores. There is strong support for the reliability and validity of this scale for use with adults in the U.S.

Fabbri and colleagues' (2021) used the Brief Self-Control Scale in their cluster-randomised control trial in Tanzania. This scale had poor internal consistency on this sample (Cronbach's alpha = 0.59; Fabbri et al., 2021). There was also no meaningful effect of the intervention on teacher outcomes, including on the Brief Self-Control Scale. This may have been because either the intervention had no effect on teacher outcomes, or because the measure was not a valid measure of self-control in Tanzanian teachers (Fabbri et al., 2021). Further research is needed to assess the validity and reliability of this scale in other cultural contexts.

### *Teacher Perception Questionnaire (TPQ; Wurtele et al., 1989)*

The TPQ was developed in the U.S. and was designed to measure teachers' perceptions of positive or negative behaviour changes in children as a result of participating in a child sexual abuse prevention program (Wurtele et al., 1989). There is also a parent version (PPQ). These two scales are identical in format and content, aside from the PPQ also including a question on "difficulty separating from parent". Teachers are asked to rate how often different positive and negative behaviours occur, ranging from (1) never to (7) often. It has questionable internal consistency (Cronbach's alpha = 0.68; Wurtele et al., 1989).

The TPQ was first used in experimental research on the prevention of child maltreatment in institutions in 1992 by a team that included the developers of the scale (Wurtele et al., 1992a; Wurtele et al., 1992b). Grendel (1991) also included the TPQ in their dissertation on a child sexual abuse prevention program for Grade 1 students in the U.S. In this dissertation, the TPQ was adapted by removing some of the behavioural items as it was felt that they might be difficult for a teacher to observe. Overall, these studies have provided evidence that teachers do not perceive negative effects on student behaviour as a result of sexual abuse prevention education (Grendel, 1991; Wurtele et al., 1992a; Wurtele et al., 1992b). However, there is limited evidence currently for its psychometric properties.

### *\*Teacher wellbeing in school (Fabbri et al., 2021)*

Fabbri and colleagues (2021) used a short 5-item questionnaire designed to assess teacher wellbeing in school. This scale included items about teachers' feelings of enjoyment, feeling valued, taking pride in their work, financial compensation, and whether employers care about their well-being. Items were scored on a 4-point Likert scale, and it had questionable internal consistency (Cronbach's alpha = 0.68; Fabbri et al., 2021). There was also no meaningful effect of the intervention on teacher outcomes. This may have been because either the intervention had no effect on teacher outcomes, or because the measure was not a valid measure of teacher wellbeing in Tanzanian teachers (Fabbri et al., 2021).

Fabbri and colleagues (2021), in their supplementary materials, stated that they had used these questions from Devries and colleagues' (2015) study. However, we could not find a reference to these questions in the published details of the 2015 study. Currently, there is limited support for the psychometric properties of this scale.

## 2.3. Child Maltreatment Disclosure

Measuring child maltreatment disclosure is aimed at assessing children’s intentions to disclose through role-plays and scenarios, or collecting data (often from [administrative sources](#)) on children’s actual disclosures of maltreatment.

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#">Sexual Assault Disclosure Scale</a>	Chamberland (2003)		Canada	Child	I would guide her to someone who is competent to help.	Questionnaire in French and was unable to ascertain the full psychometric properties. More research may be needed to determine its cross-cultural validity in other countries.
<a href="#">Application of Protective Behaviors Test-Revised (APBT-R)</a>	White et al., (2018). Original: Dale et al., (2016)		Australia	Professional and child	This is Julie. This is you. Julie is telling you a secret. She is telling you that  someone has been hurting her. How would you feel?	There is promising support, but more research is needed to determine the psychometric properties.
<a href="#">*Child Report</a>	Kolko et al., (1989)		U.S.	Child	Has anyone ever touched you in a way that made you feel uncomfortable ?	No further details about these questions were reported.
<a href="#">*Maltreatment Disclosure Report Form (MDRF)</a>	Oldfield et al., (1996)		U.S.	Professional	<i>Administrative data</i>	No further details about this form were reported.

### *Sexual Assault Disclosure Scale (Chamberland, 2003)*

Daigneault and colleagues (2015) included Chamberland’s (2003) Sexual Assault Disclosure Scale in their Canadian cluster randomised control trial. Vignettes displayed a sexual assault in a dating



relationship and participants were asked to indicate the degree to which they considered the situation to constitute a sexual assault. Responses were rated on a 6-point Likert scale from (1) completely disagree to (6) completely agree, with higher scores indicating a greater ability to identify sexual assault. Participants are then asked 10 follow-up questions that were designed to assess how they would react to disclosure. This scale had questionable internal consistency (Cronbach's alpha = 0.60-0.69; Daigneault et al., 2015). There was some experimental evidence for the construct validity of the scale as participants in the intervention group improved in their ability to recognise sexual assault in a dating relationship. The original scale and publication are in French (Chamberland, 2003) so we were unfortunately unable to fully review the psychometrics of this scale.

### *Application of Protective Behaviors Test-Revised (APBT-R; White et al., 2018)*

The APBT-R is a revised version of the Application of Behaviors Test (Dale et al., 2016). This scale includes four unsafe and two safe scenarios that are verbally presented to primary school aged children to assess their intentions to disclose and help-seek in unsafe scenarios, disclosure confidence, and accuracy in correctly identifying unsafe situations. The unsafe scenarios include examples of a secret that another child had been hurt, violations of personal space, social exclusion by peers, and physical bullying. The two safe scenarios involve descriptions of keeping a secret for a birthday surprise and a child who had accidentally broken a crayon and then apologised. These safe scenarios were added to allow the calculation of safety identification skills. To assess disclosure intentions, responses from children are thematically analysed on a 4-point scale from (1) unhelpful, (2) neutral, (3) helpful, and (4) disclosure/help-seeking. If a child provides multiple responses, the highest-rated behaviour is used to calculate overall disclosure intentions. Higher scores on the APBT-R indicate higher safety behaviours. To assess disclosure confidence and safety identification skills, they are asked how confident they feel to perform those intentions (from (1) not sure to (4) completely sure), and if each scenario was safe (yes, no, or don't know). Higher scores indicate greater confidence and greater safety identification skills.

In the development of the APBT, the scenarios were reviewed by a panel of developmental psychologists and child abuse experts, who advised on content clarity and age appropriateness. The APBT has been found to have acceptable internal consistency (Cronbach's alpha = 0.74; Dale et al., 2016). More research is needed to ascertain the psychometric properties of the revised version.

The APBT-R has been used in randomised controlled trials in Australia (White et al., 2018, 2019). The interrater reliability was found to be excellent (0.997). However, in White and colleagues' (2018) randomised controlled trial, children did not significantly improve in their disclosure intentions or ability to discriminate between safe and unsafe behaviours, compared to the control (White et al., 2018). This was despite significant improvements in children's knowledge on the [ProBeQ](#). However, children who received in situ skills training in White and colleagues' (2019) study did show increased disclosure intentions but not increased safety identification skills. This suggests that increases in knowledge or exposure to interventions do not always translate to improved intentions to disclose or improved safety identification skills in children.

### *\*Child report (Kolko et al., 1989)*

All children in Kolko and colleagues' (1989) study were asked four questions about their experiences and actions following any incident of inappropriate touching. Parents and teachers

were also asked how many times the children had talked to them about being touched in an inappropriate physical or sexual nature, and how many of these had been reported to authorities. It would be difficult to ascertain the reliability of these reports, so we have chosen not to rate this approach.

### *Administrative Data*

It is difficult to make determinations about the reliability or validity of administrative data, but it can be an effective way to collect data about children's disclosures of maltreatment.

#### **\*Maltreatment Disclosure Report Form (MDRF; Oldfield et al., 1996)**

Oldfield and colleagues (1996) developed the MDRF in their study of the effectiveness of Project Trust as an elementary school-based victimisation prevention program in the U.S. The form recorded students' disclosures of maltreatment, including the date of the disclosure, the type of disclosure (initial or repeated), and the nature of the maltreatment (physical or sexual abuse). The MDRF also collected demographic details such as gender, age, socioeconomic status, and race (Oldfield et al., 1996). No other details about the MDRF were reported. There was no analysis on the reliability or validity of the measure reported, although this is to be expected as it seemed to have been aimed at collecting administrative data. The MDRF has not been used in any other experimental research on the prevention of child maltreatment in institutions.

### **Other**

There were two other methods used to collect administrative data on children's disclosures of abuse in studies in the EGM. For example, **Child Forensic Interviews** were used in a U.S. study of child sexual abuse disclosure rates in child advocacy centres (Elfreich et al., 2020). These interviews are conducted by trained professionals and are designed to have forensic value, be legally defensible in court, and be approached in a neutral and unbiased manner. For an interview to receive a code for 'disclosure', the child must have "disclosed any meaningful utterance or full and detailed account of the abuse experience in such a way that could be substantiated as an abuse disclosure in the state of Indiana" (Elfreich et al., 2020).

**Feedback Forms** were also used in U.S. study, where teachers and school social workers could make note of questions concerning disclosures of ongoing or past abuse, as well as delayed disclosures after children's exposure to the intervention program (Hazzard et al., 1991).

It is not always possible to ascertain the reliability or validity of administrative methods for capturing children's disclosures. Although administrative data can be an important way for researchers to assess outcomes in interventions. It is important to note that it is common for there to be an increase in disclosures after an intervention. This does not mean that there has been an increase in abuse, but rather may reflect that children felt more comfortable to disclose or were able to identify safe people to disclose to because of the intervention.

## **2.4. Child Maltreatment Occurrence/Recurrence**

Measures in this category were aimed at capturing specific data on children's self-reported experiences of maltreatment by asking if they had experienced certain events (such as sexual



abuse, sexual harassment, neglect, physical abuse, corporal punishment, and adversity). Similar to child maltreatment disclosure, [administrative sources](#) were also used.

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#">American Association of University Women Sexual Harassment Survey</a>	AAUW (1993)		U.S.	Child	During your whole school life, how often, if at all, has anyone made sexual comments, jokes, gestures, or looks when you did not want them to?	Includes experiences outside what is conceptually defined as sexual harassment (such as unwanted sexual comments and child sexual abuse). It may not be suitable for younger participants. There is also some evidence that it may not be a valid measure of sexual harassment.
<a href="#">Pediatric ACEs and Related Life Events Screener (PEARLS)</a>	ACES Aware, (n.d.)		U.S.	Parent/Caregiver	Has your child ever experienced sexual abuse?	Relies on parent/caregiver report for exposure to adversity.
<a href="#">Coddington's Child Life Events Scale (CLES)</a>	Coddington (1972)		U.S.	Child	Jail sentence of a parent for 1 year or more.	This scale does not differentiate between positive and negative events when making an assessment of the level of stress experienced by children.
<a href="#">WHO Multi-Country Study (WHO MCS) on Women's Health and Domestic</a>	Garcia-Moreno et al., (2005)		Multiple countries	Adults	Was kicked, dragged, or beaten up by an intimate partner.	The original version was designed for use with adult women and has strong

<b><u>Violence against Women</u></b>						psychometric properties. Adapted versions have been created for children and caregivers, but there is limited evidence currently for the reliability or validity of these versions.
<b><u>International Society for the Prevention of Child Abuse and Neglect Screening Tool (ICAST)</u></b>	IPSCAN (2006)		U.S.	Child	Have you seen adults in your home hurt each other physically (e.g., hitting, slapping, and kicking)?	The reliability and validity of adaptations of the scale need to be tested.
<b><u>Multidimensional Neglectful Behaviour Scale (MNBS)</u></b>	Kantor et al., (2004)		U.S.	Child and adult report available	This child's parent leaves him or her places where he or she does not feel safe.	This scale may not be as reliable or valid when used with younger children (6-9 years).
<b><u>Children's Feelings of Safety</u></b>	Schwab-Stone et al., (1995)		U.S.	Child	I feel safe in my class at school.	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>Child-Teen Witness to Woman Abuse Questionnaire</u></b>	Sudermann et al., (2000)		U.K.	Child	True or false: some hitting between a dad and mum is ok?	Full details about this questionnaire are not publicly available.
<b><u>Olweus Bullying Questionnaire (OPQ)</u></b>	Olweus (1996)		U.S.	Child	How many times have you been bullied in the past couple of months?	More research may be needed to determine its cross-cultural validity in other countries.

<b><u>Peer Victimisation</u></b>	Orpinas & Kelder (1995)		U.S.	Child	A student pushed or shoved me.	Unpublished manuscript. Full details not publicly available.
<b><u>*Sexual victimisation and perpetration</u></b>	Daigneault et al., (2015)		Canada	Adolescent	Have you had a sexual relationship, sexual behaviours, attempted to have a sexual relationship or made someone behave sexually while the other person did not want to?	No analysis was reported on the psychometrics.
<b><u>*Bystander Opportunity</u></b>	Edwards et al., (2022)		U.S.	Child	Saw or heard about a student sending a naked photo of another student without that person's permission	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Corporal punishment at school</u></b>	Karmaliani et al., (2020)		Pakistan	Child	How often within in the past 4 weeks... Did a teacher twist your ear?	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Violence and neglect at home</u></b>	Karmaliani et al., (2020)		Pakistan	Child	In the past 4 weeks have you been beaten so hard at home that your were injured?	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.

### *American Association of University Women Sexual Harassment Survey (AAUW, 1993)*

This 14-item questionnaire assesses students' self-report of experiences of sexual harassment (AAUW, 1993). The survey includes experiences of unwanted sexual comments as well as non-contact ("spied on you as you dressed or showered at school") and contact ("forced you to do something sexual") sexual abuse. The survey has been widely used and has good internal consistency (Cronbach's alpha = 0.87; AAUW, 1993, 2001).

Some findings suggest that it may not be suitable for younger children (11-13 years) as understanding abstract wording (such as "in a sexual way") may be more difficult for younger students (McMaster et al., 2002). Similarly, there is limited evidence on the content or construct validity of the original scale as this research is commonly conducted on modified versions of the original survey. Exploratory factor analysis on a version modified to capture both victimisation and perpetration of sexual harassment, found two factors – sexual harassment and forced sexual contact – but that these had low explained variance, indicating that they did not fully capture the conceptualisation of peer sexual harassment (Espelage et al., 2012; Valik et al., 2022). This may be because the survey captures a wide range of experiences including child sexual abuse and unwanted sexual comments, which are conceptually different from sexual harassment (Mathews & Collin-Vézina, 2017). Internationally recognised definitions define sexual harassment as unwanted or unwelcome sexual behaviour that is experienced as offensive, humiliating, or intimidating (Australian Human Rights Commission, n.d.), not just as unwanted.

Versions of this survey have been used twice in experimental research on institutional child maltreatment. In Edwards and colleagues' (2022) study of the effectiveness of a sexual assault self-defence program, participants were asked two questions from the full version. Participants were asked about their self-reported experiences of sexual comments and rumours. Taylor and colleagues (2010) used an adapted version to measure the prevalence and incidence of being a victim and/or perpetrator of sexual harassment (including verbal, written, physical, and voyeuristic behaviours) among U.S. high school students. Neither study reported on the psychometric properties of the versions used in their sample.

### *Pediatric ACEs and Related Life Events Screener (PEARLS; ACES Aware, n.d.)*

The PEARLS is a parent/caregiver tool designed to identify their child's exposure to childhood adversity and stress factors. It is based on the well-established Adverse Childhood Experiences Questionnaire (ACEs), which has been found to be a valid and reliable measure of childhood adversity (Koita et al., 2018). The PEARLS has also been found to be a valid measure in that it is effective at identifying children at high risk for clinically significant outcomes (Thakur et al., 2020). The PEARLS contains the original ten ACEs as well as exposure to discrimination, community violence, death of a caregiver, forced separation from a caregiver, housing instability, food insecurity, and physical illness/disability of a caregiver.

Each item is answered as "yes" or "no". Scores of 4 or more indicate a high risk for toxic stress and potential symptomatology (ACES Aware, n.d.). The PEARLS is available in English and Spanish. The validity of the PEARLS was determined through cognitive interviews and reviews by parents/caregivers and clinic staff.

It has been included in one protocol for experimental research on institutional child maltreatment (Perez Jolles et al., 2021).

### *Coddington's Child Life Events Scale (CLES; Coddington, 1972)*

The CLES is a child self-report scale designed to assess how specific events have affected a child's personal growth and adjustment. There are pre-school (5 years and under, CLES-P), child (6-12 years, CLES-P), and adolescent (13-19 years, CLES-A) versions available. It includes both 'positive' events (e.g., outstanding personal achievement) 'negative' events (e.g., death of a parent), and events that may be experienced as either (e.g., beginning or moving school). It conceptualises change as contributing to stress in children.

It has been found to have good test-retest reliability, internal consistency, and inter-parent agreement (Berden et al., 1990), providing support for the reliability of the questionnaire. Support for the validity of the scale has come from studies demonstrating that scores on this measure correlate with several relevant child outcomes. A limitation of the CLES is that it does not differentiate between events experienced as positive or negative. It only provides an overall assessment of the amount of 'events' or change that has occurred, with more events suggesting potential greater stress. However, not all of the events may be experienced as stressful.

The CLES was adapted for use in a Bucharest Early Intervention Project (BEIP) study to assess children's self-reported events in the past 12 months (Wade et al., 2019). In this study only 8.5% of the sample reported between 8 and 12 events, so the authors re-scaled the measure by combining children reporting 7 or more events to correct for this in their analysis. In this study, institutionalised children with more stressful life events at age 12 then had higher levels of externalising problems at age 16 (Wade et al., 2019). The same was not true for children who had lived with foster carers or who had never been institutionalised. These results suggested that family care may buffer the effect of stressful life events on later psychopathology (Wade et al., 2019). It was noted that in this study they made efforts to separate out events that children had little control over ("independent life events" e.g., "you changed schools") from those they did have some degree of control ("dependent event" e.g., "you began dating"). This was done to potentially address some of the limitations of the CLES discussed above.

### *WHO Multi-Country Study (WHO MCS) on Women's Health and Domestic Violence against Women (Garcia-Moreno et al., 2005)*

The WHO MCS was a tool used to measure exposure to physical and sexual violence and controlling behaviours by an intimate partner in a multi-country study (Garcia-Moreno et al., 2005). The WHO MCS has been used widely internationally, having been validated for use with adult males and females in a range of settings (Fulu et al., 2013; Nybergh et al., 2013).

Merrill and colleagues (2018) adapted behaviourally specific acts of violence from the WHO MCS and the [ICAST](#) in their study. The adapted version included two child self-report and two caregiver self-report items, but they did not report any analysis on the validity or reliability of their adaptation.

Jewkes and colleagues (2019) included adaptations of the WHO MCS (Garcia-Moreno et al., 2005; Jewkes et al., 2011) to measure the incidence of intimate partner violence (IPV) in dating relationships among adolescents as well as emotional, sexual, and physical IPV in female parents. No analysis was reported on the psychometric properties of the adapted version.

Knight and colleagues (2020) also included an adapted version of the WHO MCS and the [Conflict in Adolescent Dating Relationships Inventory](#) (Antônio & Hokoda, 2009; Wolfe et al., 2001) in

their protocol for a Ugandan quasi-experimental study. Adolescents will be asked about their experiences of both victimisation and perpetration of IPV in dating relationships.

More research is needed to ascertain the psychometric properties of this scale when used with children, adolescents, and caregivers.

### *International Society for the Prevention of Child Abuse and Neglect Screening Tool (ICAST; IPSCAN, 2006)*

The ICAST is a children's self-report questionnaire designed to assess experiences of abuse and neglect. It includes consideration of maltreatment that has occurred in the previous weeks, months, year, and lifetime. The reliability and construct validity of the original ICAST and the child institutional version (ICAST-CI) was established in four countries (IPSCAN, 2006). The scale is commonly used globally and has been translated into 20 languages. It has been found to be a valid and reliable measure of exposure to child abuse and neglect (IPSCAN, 2006; Zolotor et al., 2009). Internal consistency has been found to be acceptable to good (Cronbach's alpha = 0.72-0.86). The only exception is the violence exposure scale (Cronbach's alpha = 0.69; Zolotor et al., 2009). However, test-retest reliability for the ICAST is not well established, which is common for violence measures.

In experimental research on the prevention of institutional child maltreatment, the ICAST or ICAST-CI has been used multiple times, but only in Africa (Kenya, Tanzania, and Uganda; Devries et al., 2015; Devries et al., 2017; Devries et al., 2018; Fabbri et al., 2021; Merrill et al., 2018; Wangamati et al., 2022). Merrill and colleagues (2018) adapted the scale by including behaviourally specific acts of violence from both the ICAST and [WHO MCS](#). They included two child self-report and two caregiver self-report items in their version but did not report any analysis on the validity or reliability of their adaptation.

In a cluster-randomised control trial of the EmpaTeach intervention in Nyarugusu Refugee Camp in Tanzania, an adapted and translated version of the ICAST-CI was used (Fabbri et al., 2021). The ICAST-CI was translated to Kiswahili and Kirundi, which was cognitively tested with small sample of five students and five teachers. The details of this process were not reported. The authors noted in their conclusion that they did not test for reliability and validity, and therefore cannot be sure that they had an accurate report of the prevalence of violence (Fabbri et al., 2021).

In a cluster-randomised controlled trial of the Good Schools Toolkit in Ugandan primary schools, the ICAST-CI was translated and adapted to include additional items and past week time frames (Devries et al., 2015, 2017, 2018). Additional items included experiences of physical, emotional, and sexual violence, if they had experienced any injury because of violence from school staff, as well as experiences of peer violence (physical, sexual, and emotional). Items on the ICAST were also adapted to include teachers' self-report of their use of violence towards students in the past week. This adapted version was also used by Knight and colleagues (2018) in their Ugandan study. The changes were pretested for understanding and relevance to the Ugandan context in a pilot study (Devries et al., 2013). Further details about this process were not reported. There was experimental support for the construct validity in this sample as students who reported experiencing violence in the past week also reported higher levels of mental health difficulties (Devries et al., 2015).



Wangamati and colleagues (2022) also included a plan to make a slight adaptation to the ICAST-CI for their study by only capturing reports of violence from the past weeks or months. As this paper was a protocol for a planned mixed-methods study in schools in Kenya and Tanzania, no analysis was included.

### *Multidimensional Neglectful Behaviour Scale (MNBS; Kantor et al., 2004)*

This scale was developed in the U.S. to measure the extent to which a child's environment meets four basic developmental needs including: physical (e.g., food, shelter, clothing, and medical care), supervisory (e.g., attending to misbehaviour, setting limits, and knowing child's activities), cognitive (e.g., reading to child and explaining things), and emotional (e.g., support, affection, and companionship). The scale was originally developed for adult report, but has been adapted for retrospective recall with older youth (12-14), as well as a child report version for 6-15 years (MNBS-CR).

The full version (MNBS-CR) was found to have excellent internal consistency (Cronbach's alpha = 0.94) for older children (10-15 years). However, the internal consistency for younger children (6-9 years) is only questionable (Cronbach's alpha = 0.66), suggesting that may not be as reliable with younger children. There is strong evidence for its construct validity when used with older children as it has been shown to correlate significantly with children's reports of exposure to violence, parental alcoholism and mental health, clinician reports of behavioural disorders, and scores on the Child Behaviour Checklist (CBCL; Kantor et al., 2004).

McElearney and colleagues (2021) adapted the MNBS by using it as a knowledge questionnaire. Children were asked to identify if certain examples represented neglect, instead of being asked if they had experienced it themselves. There was some preliminary experimental evidence for the construct validity of this adapted version, as children showed improvements in knowledge about what constitutes neglect because of the intervention (McElearney et al., 2021), although this difference was not significant. Currently, there is not sufficient support for this adapted version, and more research would be needed to ascertain the psychometric properties of the MNBS as a knowledge of neglect scale.

### *Children's Feelings of Safety (Schwab-Stone et al., 1995)*

Schwab-Stone and colleagues (1995) created a measure designed to assess children's feelings of safety in the home, neighbourhood, school, and community. This measure was then adapted by Henry (2000) and has 11-items. Participants indicate the degree to which they feel safe in the different scenarios on a 3-point Likert scale (0 = never, 1 = sometimes, 2 = always). Higher scores indicate a greater sense of safety. It was found to have good internal consistency (Cronbach's alpha = 0.89), and has preliminary experimental support for its construct validity (Dahlberg et al., 2005).

A French translation of this measure was used in Daigneault and colleagues (2012) evaluation of a sexual abuse prevention workshop in Canada. This version had good internal consistency (Cronbach's alpha = 0.87), and there was some experimental evidence for its construct validity. More research would be needed to determine the psychometric properties of the original and French translations of this measure.

### *Child-Teen Witness to Woman Abuse Questionnaire (Sudermann et al., 2000)*

The Child-Teen Witness to Woman Abuse Questionnaire was designed as part of an evaluation study on a group treatment programme for children who had witnessed woman abuse (Sudermann et al., 2000). There was no analysis of the psychometric properties of this questionnaire in this study.

The Knowledge/Attitudes to Woman Abuse subscale of this questionnaire was used in McElearney and colleagues' (2021) cluster randomised control trial in Northern Ireland. This subscale focuses on children's knowledge of domestic abuse. This measure is culturally appropriate and has been used in previous needs assessment research in Northern Ireland schools (McElearney et al., 2011a). However, this research is no longer public, and we were unable to review this information.

McElearney and colleagues (2021) found that there was a greater increase in knowledge relating to domestic abuse in children exposed to the intervention than children in the control group. These findings provide preliminary experimental support for its construct validity. There seems to be some promising findings, but the detail is not publicly available.

### *Olweus Bullying Questionnaire (OPQ; Olweus, 1996)*

The OPQ is a 42-item measure designed to measure bullying problems in schools. It is a slightly revised version of the Olweus Bully/Victim Questionnaire (Olweus, 1996), and is typically used for students from grades 3 to 12. It has good internal consistency (Cronbach's alpha = 0.80) on a sample of students in grades 4 and 5, as well as on middle school and junior high samples (Soldberg & Olweus, 2007). There is strong construct validity as higher scores on the "being bullied" question were associated with the frequency of other types of victimisation (Solberg & Olweus, 2007). Criterion validity of the OPQ was determined with peer ratings of being bullied or bullying other students. There is strong psychometric support for the OPQ.

It has only been used in one study on the EGM on the prevention of institutional child maltreatment (McElearney et al., 2021). This study used two global items from the OPQ to measure children's self-report of the frequency of bullying victimisation and perpetration. Evidence for the cultural validity of the OPQ in Northern Ireland was found in a Department of Education survey of bullying in schools. The authors included Olweus' definition of bullying to promote the content and construct validity of the items (Evans et al., 2014).

### *Peer Victimization (Orpinas & Kelder, 1995)*

Orpinas and Kelder (1995) developed a 10-item peer victimisation scale. However, this comes from an unpublished manuscript so we were unable to review the full psychometric properties of this measure.

Daigneault and colleagues (2012) used a French translation of this scale in their Canadian evaluation of a sexual abuse prevention workshop. The translation was done by two independent translators and had excellent internal consistency (Cronbach's alpha = 0.91). There was preliminary experimental support for its construct validity, with students reporting reduced rates of peer victimisation. Further research would be needed to assess the psychometric properties of the French translation of the measure.

*\*Sexual victimisation and perpetration (Daigneault et al., 2021)*

Two items were included that assessed sexual victimisation and perpetration in Daigneault and colleagues' (2021) cluster randomisation study. Participants were asked, "Have you had a sexual relationship, were subjected to sexual behaviours or were forced to behave sexually with one of these people while you did not want to?" and "Have you had a sexual relationship, sexual behaviours, attempted to have a sexual relationship or made someone behave sexually while the other person did not want to?". No analysis was reported on the psychometric properties of these items.

*\*Bystander Opportunity (Edwards et al., 2022)*

4-items were included in Edwards and colleagues' (2022) U.S. self-defense program study that were designed to assess bystander opportunity. These items were based on the work of Banyard (2015) and Coker and colleagues (2011). They included scenarios of seeing or hearing about another student experiencing sexual violence, and each item was rated as (0) 0 times, (1) 1-2 times, (3) 3-5 times, (6) 6-9 times, and (10) 10 or more times. There was preliminary experimental evidence for its construct validity, but no analysis was reported on the reliability or validity of this scale. Therefore, we cannot currently be confident that it is a reliable or valid measure of girls' self-defence knowledge.

*\*Corporal punishment at school (Karmaliani et al., 2020)*

A 6-item measure of corporal punishment at school was included in a Pakistani cluster-randomised controlled trial aimed at reducing peer violence among children (Karmaliani et al., 2020). Students reported the number of times (never, once, 2-3 times, or 4+ times) they had experienced each type of corporal punishment at school. The scale had acceptable internal consistency (Cronbach's alpha = 0.758). There was also some preliminary experimental support for the construct validity of the scale, with students reporting a reduction in the frequency of corporal punishment after the intervention (Right to Play; Karmaliani et al., 2020). While there is some preliminary evidence for the psychometrics of this scale, more research is needed.

*\*Violence and neglect at home (Karmaliani et al., 2020)*

In Karmaliani and colleagues' (2020) study, children completed items on their experiences in the last four weeks of violence and neglect at home. There was one item on fathers fighting, two items on witnessing violence against their mother, and two items on the frequency and severity of physical punishment at home. Children were also asked two items on if they had gone without food in the last four weeks. There was no analysis reported on the reliability or validity of these items.

## **Administrative Data**

It is difficult to make determinations about the reliability or validity of administrative data, but it can be an effective way to collect data about the occurrence of child maltreatment.

*\*CPS records (Cerezo & Pons-Salvador, 2004)*

The number of children registered in Child Protection Services (CPS) or reported to the Local Coordination Team (LCT) with signs of maltreatment were reviewed in Cerezo and Pons-Salvador's (2004) study of child maltreatment detection systems in Spain.

**\*Abuse Substantiation (Elfreich et al., 2020)**

Following Child Forensic Interviews in Elfreich and colleagues’ (2020) study, administrative data was collected from the Indiana Department of Child Services on abuse substantiations. After reviewing available evidence, the Department makes a decision about whether the disclosure is substantiated or unsubstantiated.

**\*State Agency Child Welfare Administrative Data (Green et al., 2020)**

In Green and colleagues’ (2020) they collected child welfare administrative data from 16 Early Head Start (EHS) sites. This data included information about child welfare involvement, substantiated reports, and out-of-home placements. While difficult to obtain, the authors noted the value in including administrative data in their study.

**\*Police and Child Protection Data (Herbert & Bromfield, 2021)**

Police and Child Protection Data were obtained for Herbert and Bromfield’s (2021) Australian quasi-experimental study. Alleged offender characteristics, characteristics of the reported abuse, involvement of agencies in responses and types of support, case outcomes, and child demographic details were collected.

**\*Administrative data of reported incidents of maltreatment (Johnson-Motoyama et al., 2023)**

The National Child Abuse and Neglect Data System (NCANDS) was used in Johnson-Motoyama and colleagues’ (2023) study. The NCANDS includes data about reported incidents of child maltreatment accepted for investigation or assessment, the total number of substantiated reports, foster care services resulting from child protection responses, and children with substantiated reports.

**\*Guidance counsellor incident reports (Kolko et al., 1989)**

Kolko and colleagues (1989) included guidance counsellor incident reports that included reports from children or adults of inappropriate physical or sexual touching. Details about the type of incident and the schools' response were collected for analysis.

## 2.5. Children’s Outcomes

This was the construct with the greatest number of measures and focus in experimental research. Measures under this construct were broadly aimed at understanding the knowledge, attitudes, behaviours, and well-being of children. This included measures of children’s knowledge of abuse, attitudes towards gendered violence, and observations of children’s behaviours in role-plays and scenarios.

### Children’s Knowledge

Measure	Author	Rating	Location	Who	Example Item	Considerations
---------	--------	--------	----------	-----	--------------	----------------

<b><u>Protective Behaviours Questionnaire (ProBeQ)</u></b>	Dale et al., (2016)		Australia	Child	When you are feeling scared, what is your body trying to tell your brain?	More research may be needed to determine its cross-cultural validity in other countries.
<b><u>Knowledge about CSA Protection Strategies</u></b>	del Campo Sánchez & Sánchez López (2006)		South America	Child	English translation published by Bustamante et al., (2019): Is it ok to say “no” to a grown up when they ask you to do something bad?	There is promising support published by Bustamante and colleagues (2019). This questionnaire is in Spanish, and we were unable to ascertain the full psychometric properties and details. More research may be needed to determine its cross-cultural validity in other countries.
<b><u>Good Touch Bad Touch Curriculum Test (GTBTCT)</u></b>	Harvey et al., (1988)		U.S.	Child	If a person forces or tricks you into a bad touch, should you tell?	There is some support for the original version. There is more support available for the Turkish version.
<b><u>Good/Bad Touch Questionnaire (GBT)</u></b>	Kenny & Wurtele, (2010)		U.S.	Child	Would it be okay for [person in vignette] to touch your private parts?	There is limited support for the psychometrics of this measure.
<b><u>Children’s Safety Knowledge and Skills Questionnaire (CSKS-Q)</u></b>	Kraizer (1986)		U.S.	Child	If someone touches you in a way you don’t like, what do you do?	There is a potential ceiling effect when used with older children (adolescents).

<b><u>Children's Knowledge</u></b>	Nelson (1981)		U.S.	Child	True or false: Children are often to blame for causing adult to have hit them.	Potential ceiling effect for some items and no analysis reported on the psychometric properties.
<b><u>What if Situation Test (WIST)</u></b>	Nemerofsky (1986)		U.S.	Child	What would you do if someone touched your private parts?	There may be a ceiling effect when used with older children.
<b><u>Personal Safety Questionnaire (PSQ)</u></b>	Saslowsky & Wurtele (1986)		U.S.	Child	Would you tell an adult if someone touched your private parts?	There may be a ceiling effect when used with older children.
<b><u>Children's Knowledge of Abuse Questionnaire (CKAQ)</u></b>	Tutty (1992, 1997)		Canada	Child	If someone touches you in a way you don't like, you should tell a grown-up you trust.	Some items would benefit from being updated to reflect key characteristics of abuse.
<b><u>Genital Body Parts (BODY)</u></b>	Wurtele et al., (1992)		U.S.	Child and researcher	"Now I'm going to point to different parts of the their bodies, and I want you to tell me what the body part is called."	There is some promising support.
<b><u>*Child Report Survey of Knowledge of Potentially Risky Situations</u></b>	Bright et al., (2022)		U.S.	Child	Is it safe if someone looks at your private parts but doesn't touch them.	There is limited support for the psychometrics of this measure and it has a potential ceiling effect.
<b><u>*Children's Awareness of Scary Secrets Questionnaire (CASSQ)</u></b>	Chen et al., (2012)		Taiwan	Child	Not reported	Poor internal consistency and a ceiling effect for older children.



<p><b><u>*Children's Sexual Knowledge Questionnaire (CSKQ)</u></b></p>	<p>Chen et al., (2012)</p>		<p>Taiwan</p>	<p>Child</p>	<p>Not reported</p>	<p>There is limited support for the psychometrics of this measure.</p>
<p><b><u>*Child knowledge questionnaire</u></b></p>	<p>Conte et al., (1985)</p>		<p>U.S.</p>	<p>Child</p>	<p>Someone you care about and trust tries to touch you in places you do not want to be touched and tells you to keep it a secret. Would you: (1) Let him touch you because you do not want to hurt his feelings: (2) Say no, run and tell: or (3) Leave but not tell anyone</p>	<p>There is support for inter-rater reliability, but we cannot be confident that this is a reliable or valid measure of children's knowledge.</p>
<p><b><u>*Sexual Abuse Knowledge Inventory (SAKI)</u></b></p>	<p>Crowley (1989)</p>		<p>U.S.</p>	<p>Child</p>	<p>If someone touches you in a way that feels uncomfortable you should...</p> <p>(A) Say no only if they are younger than you</p> <p>(B) Say ok but only this time</p> <p>(C) Say no very loudly</p> <p>(D) Say nothing</p>	<p>There may be a slight ceiling effect when used with older children.</p>
<p><b><u>*Case Vignettes</u></b></p>	<p>Czerwinski et al., (2018)</p>		<p>Germany</p>	<p>Child</p>	<p>Joan should tell the teacher that they will tell other people if he does it again</p>	<p>There is some support, but more research is needed to determine the</p>

						psychometric properties.
<b><u>*General Knowledge about Violence</u></b>	Daigneault et al., (2012)		Canada	Child	True or false: If you laugh at somebody's appearance, it is not really violence.	There is limited support for the psychometrics of this measure.
<b><u>*Sexual Assault Knowledge Questionnaire</u></b>	Daigneault et al., 2015		Canada	Child	True or false: Voyeurism is a form of sexual assault	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Children's knowledge and efficacy</u></b>	Dake et al., (2003)		U.S.	Child	If an adult tells you to do something that you think is wrong, should you always do it?	There is limited support for the psychometrics of this measure and a potential ceiling effect.
<b><u>*Children's Knowledge</u></b>	Dhooper & Schnieder, (1995)		U.S.	Child	You tell someone you trust that your mom hurts you a lot. He or she does nothelp you with the problem. Is it best to just quit talking about it?	There is some support, but more research is needed to determine the psychometric properties.
<b><u>*Maltreatment Knowledge</u></b>	Diaz et al., (2021)		U.S.	Child	Abusers are very good at hiding their bad intentions from the people whom they abuse.	More research may be needed to determine its cross-cultural validity in other countries.
<b><u>*Knowledge Questionnaire</u></b>	Edwards et al., (2019)		U.S.	Child	According to the FBI, _____ of rapes that are reported to	There is promising support, but more research

					the police are false reports. 2%, 10%, 30%, or 60%	is needed to determine the psychometric properties.
<b><u>*Girls' Self-Defence Knowledge</u></b>	Edwards et al., (2022)		U.S.	Child	Which are the best ways to defend yourself if you are attacked?	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Differentiation, Knowledge, and Application</u></b>	Harvey et al., (1988)		U.S.	Child	Do you think that children should decide with whom they want to share their bodies?	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*What I know about touching scale</u></b>	Hazzard et al., (1991)		U.S.	Child	Not reported	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Children's Knowledge</u></b>	Hébert et al., (2001)		Canada	Child	Not reported	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*CAN recognition and talkability</u></b>	Hoefnagels, (2021)		The Netherlands	Child	Do you recognise this as child maltreatment?	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its

						construct validity.
<b><u>*Knowledge and Skills of Sexual Abuse Prevention</u></b>	Jin et al., (2017)		China	Child	Can child refuse when not wanting to be touched?	Questionable internal consistency and no other analysis was reported on the psychometrics.
<b><u>*Knowledge of CSA</u></b>	Kang et al., (2022)		South Korea	Child	Not reported	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Car, Traffic, Fire, Gun Safety Questionnaire (SAFETY)</u></b>	Kenny et al., (2012)		U.S.	Child	What should you do when you are going to cross the street?	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Self-protection skill form</u></b>	Kızıtepe et al., (2022)		Turkey	Child	What can you do if a stranger person wants to touch you?	There is some support, but more research is needed to determine the psychometric properties.
<b><u>*Knowledge, experiences, actions</u></b>	Kolko et al., (1987)		U.S.	Children, parents/teachers	How often are children your age touched in the private parts of their bodies in a confusing way- a way that you don't understand or that makes you feel uncomfortable ?	There is some support, but more research is needed to determine the psychometric properties.

<b><u>*Child Self-Report</u></b>	Kolko et al., (1989)		U.S.	Child	What should you do if you tell someone about  a bad touch, but he/she doesn't listen?	There is inconsistent support for the psychometric properties of this measure.
<b><u>*Knowledge/Attitude Test</u></b>	Kraizer (1991)		U.S.	Child	If someone touches you in a way you don't like, what do you do?	There is preliminary support, but further research is needed.
<b><u>*Measurement of Knowledge and Assertiveness</u></b>	Neherta et al., (2017)		India	Child	Not reported	Details about the questionnaire were not reported.
<b><u>*Good Secrets, Bad Secrets Quiz</u></b>	Synder (1986)		U.S.	Child	True or false: Sometimes it is okay to say "no" to a grown-up	There is inconsistent support for the psychometric properties of this measure.
<b><u>*Knowledge Related to Gender Violence and Harassment</u></b>	Taylor et al., (2010)		U.S.	Child	Not reported	No analysis was reported on the psychometrics or details of the questionnaire. Some preliminary experimental evidence on its construct validity.
<b><u>*Knowledge Scale</u></b>	Telljohann et al., (1997)		U.S.	Child	Are strangers the only people who try to touch children's private parts in a way they don't like?	There is limited support for the psychometrics of this measure and it has a potential practice effect.
<b><u>*Drawn vignettes child knowledge</u></b>	Warden et al., (1997)		Scotland	Child	Here is a girl walking past the cinema with her	Possible practice effect and ceiling effect. We

					mother. This man is standing outside, and he starts talking to the girl. The man is very friendly; he tells the girl his name and says, "Have you seen this film? I hear it's really good. Would you like to come and see it with me?"	cannot be confident that this is a reliable or valid measure of children's knowledge.
<u>*Student assessment</u>	Weeks et al., (2021)		U.S.	Child	Your aunt wants to give you a hug. Is it okay to say "no thank you"?	Further research would be needed to assess its psychometric properties. Poor inter-item correlations and potential ceiling effect for some items.
<u>Child Sexual Abuse Knowledge Form</u>	Yilmaz Irmak et al., (2018)		Turkey	Child	Is it OK for an older person/adult to touch you when you don't want this?	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.

### *Protective Behaviours Questionnaire (ProBeQ; Dale et al., 2016)*

The ProBeQ contains 12 items designed to assess children's knowledge of six core interpersonal safety concepts taught in the Learn to be Safe with Emmy and Friends™ program. Items include concepts such as personal space, private and public body parts, safe and unsafe secrets, emotion recognition, and early warning signs. High scores indicate greater knowledge of protective behaviours (Dale et al., 2016).



The ProBeQ has poor internal consistency (Cronbach's alpha = 0.55), which may be explained by its focus on six different concepts. It has adequate test-retest reliability (ICC = 0.772), and there is support for its construct and content validity (Dale et al., 2016).

This questionnaire has been used once in an Australian randomised controlled trial (White et al., 2018). The authors did not report any psychometric properties of the ProBeQ on their sample. The children in this study did improve their knowledge of interpersonal safety as measured by the ProBeQ after the intervention, providing experimental support for its construct validity (White et al., 2018).

### *Knowledge about CSA Protection Strategies (del Campo Sánchez & Sánchez López, 2006)*

This questionnaire was developed and validated in a Spanish-speaking population. It contains 35 items designed to assess children's knowledge about child sexual abuse prevention strategies. Unfortunately, we were unable to review the full details of this questionnaire as it was published in Spanish.

An adapted version was used by Bustamante and colleagues (2019) in their study of a school-based child sexual abuse prevention program in Ecuador. Two questions from the original questionnaire were replaced with items more specific to the content covered in their prevention program. Bustamante and colleagues (2019) reported strong support for the reliability and validity of this adapted questionnaire. It reportedly had good internal consistency (Kuder Richardson  $r = 0.83$ ) and test-retest reliability ( $r = 0.92$ ). Scores on this questionnaire also correlated strongly with the [Children's Knowledge of Abuse Questionnaire-Revised](#), providing support for its construct validity.

There seems to be strong support for the psychometric properties in Spanish-speaking populations. More research may be needed to ascertain its reliability and validity in other cultural contexts.

### *Good Touch Bad Touch Curriculum Test (GTBTCT; Harvey et al., 1988)*

The GTBTCT is a 10-item questionnaire designed to assess children's sexual abuse knowledge. Items assess differentiation between good and sexually abusive touches, application of knowledge to specific situations, and basic knowledge about coping with sexual abuse. Each item is scored as 0 (incorrect answer or doesn't know) or 1 (correct answer), with higher scores indicating greater knowledge about sexual abuse (Harvey et al., 1988). However, there is limited information on the reliability and validity of the original version reported.

The original version has not been used in experimental research on the prevention of institutional child maltreatment. A Turkish translation of the GTBTCT was used in an evaluation of a psychoeducation school-based child sexual abuse prevention program for Turkish elementary school students (Cecen-Erogul & Kaf Hasirci, 2013). In this study, the authors conducted Confirmatory Factor Analysis and the results indicated an acceptable fit to the data with the hypothesised one-factor model, providing support for its construct validity. The internal consistency for this sample of Turkish elementary school students was acceptable (Cronbach's alpha = 0.78). It was also found to have high temporal stability, providing support for the test-retest reliability of the questionnaire. It was concluded that the Turkish translation of the GTBTCT had satisfactory reliability and validity (Cecen-Erogul & Kaf Hasirci, 2013).

### *Good/Bad Touch Questionnaire (GBT; Kenny & Wurtele, 2010)*

The GBT contains four scenarios designed to assess if children are able to recognise a “good” person as a potential perpetrator of an inappropriate touch (Kenny & Wurtele, 2010). Two of the scenarios include a “good” person requesting to touch a child’s private parts, and the other two include a “bad” person. Following each scenario, children are asked, “Would it be okay for [person in vignette] to touch your private parts?”. Children’s responses are scored as either 0 (incorrect “yes” or “don’t know”) or 1 (correct “no”). The internal consistency for both the “good” (Cronbach’s alpha = 0.90) and “bad” (Cronbach’s alpha = 0.92) scenarios were excellent (Kenny & Wurtele, 2010).

This measure was used once in an experimental study on the prevention of institutional child maltreatment (Kenny et al., 2012). In this study, Latino preschoolers exposed to the intervention showed significantly increased GBT scores compared to the control group from pretest to post-test. This provides preliminary experimental support for the construct validity of the GBT, but the authors did not report any analysis on the psychometrics of the GBT for use on their sample (Kenny et al., 2012).

The limited analysis available on the psychometric properties of the questionnaire, and the reliance on “good” and “bad” people means that we cannot be confident that it represents a reliable or valid measure of children’s knowledge of abuse.

### *Children’s Safety Knowledge and Skills Questionnaire (CSKS-Q; Kraizer, 1986)*

The CSKS-Q is a questionnaire designed to assess children’s knowledge of safety-related concepts and specific safety skills. Items cover concepts such as the right to say “no” to adult, body ownership, strangers, dealing with inappropriate touches, secrecy, and being disbelieved following disclosure.

There is support for the criterion validity of the CSKS-Q as scores on this questionnaire have been found to predict resistance to stranger abduction (Kraizer et al., 1989; Fryer et al., 1987a, 1987b). It has adequate internal consistency, providing some support for the reliability of the measure (Barron & Topping, 2013; MacIntyre & Carr, 1999).

MacIntyre and Carr (1999) adapted the CSKS-Q for their Irish study on child sexual abuse prevention. Minor changes were made to the wording to be culturally sensitive, e.g., word “mother” was replaced with “mammy”. One item was also removed as the authors deemed it to be unrelated to the content taught in their prevention program. The internal consistency of the adapted version on this sample was adequate (Cronbach’s alpha = 0.79). There was experimental support for the construct validity with children (7 and 10 years) showing improvements in knowledge and skills scores as a result of the intervention (MacIntyre & Carr, 1999).

Barron and Topping (2013) included an adapted version of the CSKS-Q (Fryer et al., 1987) in their evaluation of a school-based child sexual abuse prevention program for Grade 6, 7 and 8 students. The internal consistency of the questionnaire on this sample was adequate (Cronbach’s alpha = 0.78; Barron & Topping, 2013). However, the adolescents in this sample made small knowledge and skills gains. This indicates that either the intervention was not effective or that the measure was not sensitive to changes in this sample.

### *Children's Knowledge (Nelson, 1981)*

Wolfe and colleagues (1986) included simulated abuse situations in their evaluation of a school-based prevention of physical and sexual abuse program. The original details of the simulations are not publicly available as they are in an unpublished manuscript (Nelson, 1981).

Two five-minute plays, followed by one-hour classroom discussion periods, were conducted. Both plays involved a child at school who was upset about an event that happened at home the previous evening. The discussion focused on five themes: description of misdeeds, the child's feelings, possible actions the child could take, who was responsible for the act, and the importance of seeking help (Wolfe et al., 1986). Children were then presented with a shorter version of a questionnaire presented in an unpublished manuscript (Nelson, 1981). Three items were dropped due to "their inability to contribute to the validity of the measure", although the details were not reported. Questions were presented in a true-false format based on findings from a pilot study (Wolfe et al., 1986). Details of this pilot study and analysis of the psychometrics of this approach were not reported by the authors.

The results of the study suggested that there may be a ceiling effect with some items, as 82-86% of children answered three of the items correctly before being exposed to the intervention (Wolfe et al., 1986). There was also no change in knowledge scores after the intervention for some items, suggesting either that the intervention was not effective at improving knowledge in these areas, or that the questionnaire was not a valid or reliable measure of children's knowledge. Based on these findings, and the lack of details available about the measure, we cannot be confident that this questionnaire is a valid or reliable measure of children's knowledge.

### *What if Situation Test (WIST; Nemerofsky, 1986)*

The WIST is a widely used assessment of children's knowledge about child sexual abuse and was originally constructed from the learning objectives in the Children's Primary Prevention Training Program (Nemerofsky et al., 1986). In the original design, children are asked by their teacher how they would respond in a variety of situations. Responses are written down verbatim and scored using a key. Scores can range from 0-64, with higher scores indicating better understanding of child sexual abuse prevention skills and concepts (Nemerofsky, 1986).

The WIST includes 29 items that address concepts of both appropriate (e.g., doctor's visit) and inappropriate (e.g., requests for touching by others, requests to keep secrets) requests to touch or look at a child's genitals, as well as other general knowledge (e.g., name and location of private parts, actions to be taken if a child was afraid or uncomfortable). 17 items address a child's actions in response to abuse and receive 1 point for an assertive or motoric response, 2 points for disclosure, and 3 points for a disclosure and an assertive response.

The WIST has been found to have good internal consistency (Cronbach's alpha = 0.83) on a sample of 1,044 6-year-old children (Nemerofsky, 1991). Children exposed to a sexual abuse prevention training program score significantly higher on the WIST compared to a control group, providing support for the construct validity of the measure (Nemerofsky et al., 1994).

There is evidence to suggest that there is a potential ceiling effect when used with older children. Older children consistently score better on the WIST, perhaps due to their higher conceptualisation of sexual abuse, but also because they have better planning ability, verbal skills, and potentially have had more exposure to the topic through media, peers, or family (Saslowsky & Wurtele, 1986).

The WIST was the most commonly used children's knowledge measure used in experimental research on the prevention of child maltreatment in institutional settings. U.S based research commonly used the original version of the WIST (Grendel, 1991; Kenny et al., 2012; Wurtele et al., 1986; Wurtele et al., 1992a; Wurtele et al., 1992b), with the creators of the WIST also being included in the EGM (Saslowsky and Wurtele, 1986). However, it was more common for researchers to adapt, modify, and/or translate the WIST for their study.

Nickerson and colleagues (2019) used the revised version (WIST-R), which has adequate internal consistency (Cronbach's alpha = 0.72). A translated version of the WIST-III (Tunc et al., 2018) was used in Turkey (Citak et al., 2018). The Turkish version had questionable to excellent internal consistency and test-retest reliability (Cronbach's alpha = 0.68-0.90;  $r = 0.48-0.92$ ; Citak et al., 2018).

Urbann and colleagues (2020) translated the WIST to German and German sign language and developed illustrations for each item. This translated version was pilot tested with eight children to test the length of the interview and whether they understood the instructions. Some items were later excluded from the final data analysis due to insufficient reliability. For example, the "Touches" (Cronbach's alpha = 0.223) and "Secrets and Getting Help" (Cronbach's alpha = 0.423) were removed (Urbann et al., 2020).

Zhang and colleagues (2014) translated the WIST to Chinese. A team of psychologists and teachers reviewed the translation and suggested changes to improve the content validity. This updated version was then piloted with five children to ensure they were able to understand the instructions. The authors reported a one-month test-retest reliability of 0.61-0.70, but no other analysis was reported to check for the reliability and validity of the Chinese translation of the WIST (Zhang et al., 2014).

Ratto and Bogat (1990) made changes to the wording of the WIST to make it appropriate for younger children and omitted one bad touch vignette. The modified version had good internal consistency (Cronbach's alpha = 0.84) and they reported excellent inter-rater reliability (Cronbach's alpha = 0.99). Results from this study suggest that either this modified version of the WIST was not a valid measure of children's knowledge, or the intervention was not effective, as participation in the program did not affect children's scores on the WIST. In particular, children in the experimental group did not improve in their ability to disclose abuse or accurately verbalise an abusive action (Ratto & Bogat, 1990).

McElearney and colleagues (2021) adapted the WIST for their Northern Ireland study by including five additional 'inappropriate request' scenarios that included online facilitated and harmful sexual abuse. Local stakeholders, safeguarding experts, and children were included in the process of developing these new scenarios. Confirmatory Factor Analysis (CFA) showed that this adapted WIST had adequate to close fit, with acceptable to good internal consistency (Cronbach's alpha = 0.75-0.90), and good test-retest reliability ( $r = 0.71-0.84$ ; McElearney et al., 2021; Wurtele et al., 1998).

Huang and Cui (2020) only used one vignette and two questions from the WIST, where inter-rater reliability was reported as excellent (0.98). No other analysis was reported on the psychometrics of this short version.

Chen and colleagues (2012) adapted the WIST by using culturally appropriate names and terminology to be relevant to the Taiwanese culture. This translation was reviewed by a team of

psychologists and speech and language therapists. Six simulated situations were presented as cartoons, with brief scenarios on the back to assess the children's self-protection skills and their ability to discriminate between appropriate and inappropriate scenarios (Chen et al., 2012). The inappropriate scenarios were not reliable, with unacceptable one-month test retest scores for late elementary school children ( $r = 0.27$ ) and for the overall sample ( $r = 0.55$ ). Children correctly identified two out of the three simulated scenarios as inappropriate, thus suggesting a ceiling effect with this measure, particularly for late elementary school children. This meant that it was not possible to determine if the child sexual abuse prevention program had a positive impact on children's knowledge and skills relating to protective behaviours (Chen et al., 2012).

Daigneault and colleagues (2021) made substantial changes to the WIST. In this study one subscale was derived that matched the scenarios presented in their prevention program (ESPACE workshop). The measure was translated to French. Scoring was also changed as participants were provided with 4 possible responses in the form of cartoons and asked to pick which response they would recommend. This adapted version had questionable internal consistency (Cronbach's alpha = 0.69; Daigneault et al., 2021).

Of note, teachers rarely administered this measure, despite this being the original design. However, the WIST seemed to be a valid and reliable measure in several cultural contexts.

### *Personal Safety Questionnaire (PSQ; Saslawsky & Wurtele, 1986)*

The PSQ is a 13-item questionnaire designed to assess children's knowledge about sexual abuse. Each item covers concepts commonly taught in sexual abuse prevention programs such as definitions, descriptions of victims and perpetrators, and responses to potential abusers. Items are rated as "yes", "no", or "I'm not sure", with higher scores indicating greater knowledge. The PSQ has good internal consistency (Cronbach's alpha = 0.78), and it has a one-week test-retest reliability of 0.64 (Saslawsky & Wurtele, 1986). Construct validity of the PSQ has been supported by appropriate moderate correlation ( $r = 0.54$ ,  $p < .01$ ) with the ["What If" Situations Test](#) (Wurtele et al., 1989).

The PSQ is a common tool used to assess children's knowledge, and has been used in studies in the U.S., Canada, and China (Crowley, 1986; Grendel, 1991; Hébert et al., 2001; Kenny et al., 2012; Nickerson et al., 2019; Ratto & Bogat, 1990; Saslawsky & Wurtele, 1986; Wurtele et al., 1992a; Wurtele et al., 1992b; Wurtele et al., 1986; Zhang et al., 2014). Most used the original version of the measure.

Grendel (1991) used a modified version in their dissertation. Minor changes were made to the wording to align with the presentation in the study, but the meaning of the items was not changed. One issue discovered in the pilot of this modified version was that children had difficulty understanding the question ("Do children have rights and control over their own bodies?", original version: "are you the boss of your body?"). In this case, the interviewer would then explain to the child, "that means, are you the boss of your body?". No other analyses were reported on the validity of this modification. There was some experimental evidence for the construct validity, with improvements in knowledge observed (Grendel, 1991).

Hébert and colleagues (2001) derived an 11-item knowledge questionnaire based on the PSQ and [CKAQ](#), as well as other items designed by the research team. The initial derived questionnaire contained 82 items, and after a pretest was reduced to a brief version for use with both 1<sup>st</sup> and 3<sup>rd</sup>



grade students. This version was administered collectively with the class. The internal consistency of this derived version was acceptable (Cronbach's alpha = 0.70) and had a satisfactory test-retest score ( $r = 0.80$ ) in a sample of 36 children (Hébert et al., 2001). No other analyses were reported on the validity of this derived measure, therefore we cannot be confident in the validity of this version.

Both Kenny and colleagues (2012) and Zhang and colleagues (2014) used four out of the 11 questions, with Zhang and colleagues (2014) also translating the measure into Chinese. One-month test-retest reliability of 0.51 (Pearson's  $r$ ) was reported for the English-shortened version (Kenny et al., 2012). A pilot study for the Chinese version was conducted, where developmental and educational psychologists were consulted on the content and five preschoolers were administered the questionnaire to ensure they were able to comprehend the instructions. No changes were made to the translated version as a result of the pilot study. The one-month test-retest reliability of these 4-item versions was poor (0.51 for Kenny et al., 2012; 0.59 for Zhang et al., 2014). No other analyses were reported on the reliability or validity of the translated and shortened versions of the PSQ (Kenny et al., 2012; Zhang et al., 2014). There was experimental evidence for the construct validity in both studies, with children demonstrating improvements in knowledge as a result of the intervention (Kenny et al., 2012; Zhang et al., 2014), but more research would be needed.

Across all studies that used the PSQ, they reported that students gained significant knowledge as measured by the PSQ, following the intervention used, and these gains were often maintained after a few months (Crowley, 1986; Grendel, 1991; Hébert et al., 2001; Kenny et al., 2012; Nickerson et al., 2019; Ratto & Bogat, 1990; Saslawsky & Wurtele, 1986; Wurtele et al., 1992a; Wurtele et al., 1992b; Wurtele et al., 1986; Zhang et al., 2014). These results provide experimental support for the construct validity of the PSQ.

This measure may be best suited for use with younger children. Some studies report a ceiling effect with older children (Saslawsky & Wurtele, 1986).

### *Children's Knowledge of Abuse Questionnaire (CKAQ; Tutty, 1992, 1997)*

The CKAQ was designed to measure elementary school-aged children's (6-12 years) knowledge of abuse concepts that are commonly taught in prevention programs. The original 40-item version had good psychometric properties (Tutty, 1990). A new factor analysis suggested that a 24-item version would be adequate. The 24-item version has good internal consistency (Cronbach's alpha = 0.87) and one-month test-retest reliability ( $r = 0.88$ ; Tutty, 1995). The CKAQ has been found to be sensitive to change, which provides informal support for the construct validity of the measure. The measure does not introduce a testing effect by inadvertently teaching concepts when used as a pre-test measure, meaning that you can be confident that the scores measured are an accurate measure of knowledge in children. It also correlates with scores on the [PSQ](#), providing evidence for convergent validity (Tutty, 1995). Younger children do have consistently scores lower than older children, but there are no significant gender differences (Tutty, 1997).

The CKAQ has been used nine times in experimental research on the prevention of institutional child maltreatment (Baker et al., 2013; Czerwinski et al., 2018; Daigneault et al., 2012; Jones et al., 2020; McElearney et al., 2021; Nickerson et al., 2019; Oldfield et al., 1996; Pulido et al., 2015; Tutty, 1997). These studies were conducted in Australia, Canada, Germany, the United Kingdom, and the United States of America.

Oldfield and colleagues (1996) used the original version in their evaluation of Project Trust in U.S. elementary schools. In this sample, the CKAQ had good internal consistency (Cronbach's alpha = 0.84). There was experimental evidence for the construct validity of the CKAQ, with students in the intervention group showing greater improvements in knowledge and prevention skills (Oldfield et al., 1996).

The Inappropriate Touch Subscale was used in experimental research on the prevention of child sexual abuse in institutions in the U.S. (Nickerson et al., 2019; Pulido et al., 2015), as well as a French-translation in Canada (Daigneault et al., 2012). The French-translation did reduce the internal consistency of the measure (from 0.84 to 0.78; Daigneault et al., 2012). The authors did not report any further analysis of the psychometric properties of the subscale in their sample (Daigneault et al., 2012; Nickerson et al., 2019; Pulido et al., 2015).

The CKAQ-R-III was then used in three studies (Baker et al., 2013; Czerwinski et al., 2018; Jones et al., 2020). Baker and colleagues (2013) modified the measure by making changes to the wording of the questions to suit the Hawaiian cultural context. An additional item was also added to evaluate the particular curriculum included in their study. This modified version had questionable internal consistency (Cronbach's alpha = 0.68) and there was no other analysis reported in the psychometrics of the modification (Baker et al., 2013).

Czerwinski and colleagues (2018) used a German version of the CKAQ-R-III which had been developed specifically for German primary school students (Andresen et al., 2015). In this version, the items are scored on a 4-point scale instead of the original 3-point scale. During a pilot study, which included evaluation of the psychometrics of the items on their sample, the authors excluded six items that were found to have low item difficulty (i.e., many students scored correctly on these items prior to the intervention). A further five items were excluded because the authors deemed that they were unrelated to the subject matter of the intervention. Six new items were then added that were specific to their intervention. Czerwinski and colleagues (2018) included a remaining 20 items for their study, which had acceptable internal consistency (Cronbach's alpha = 0.78). No other analysis of the final version of this knowledge measure was included.

Jones and colleagues (2020) also modified the CKAQ-R-III by removing two items. The item "you can trust your feelings about whether a touch is good or bad" was removed as research suggests that this item may be problematic as abusive touch can sometimes be experienced as pleasurable, but is still abusive (Scholes et al., 2014; Steiler-Hunt et al., 2014). The item "If you get separated from your parents in a shopping mall, it's OK to ask a salesclerk or a security guard for help, even if they are strangers" was removed as research has indicated that people use positions of authority as a way of building trust with children to perpetrate abuse.

The CKAQ has been found to be a valid and reliable measure of children's knowledge of abuse. Some items may benefit from refinement based on key research on the characteristics of abuse.

### *Genital Body Parts (BODY; Wurtele et al., 1992)*

This measure is used to determine children's knowledge of genital terminology. Children are asked to provide the names of various body parts on a drawing of a nude male and female body. Responses are coded as 0 (incorrect, slang, or doesn't know) or 1 (correct). Raters of this scale have been found to be consistent (inter-rater reliability = 91-100%; Kenny & Wurtele, 2008). To our knowledge no other analysis has been reported on the psychometric properties of the BODY.



Kenny and colleagues (2012) used the BODY in their evaluation of a personal safety program for Latino preschoolers. Spanish terms were included for use with this population and were provided by a bilingual pediatrician who specialises in the assessment and treatment of child maltreatment. This measure was previously used in a bilingual population for the authors (Kenny, 2010; Kenny & Wurtele, 2008). No analysis of the psychometrics of this measure for their sample was reported in this sample, but participants in the personal safety program made greater genital terminology knowledge gains than the control group, providing some evidence for the construct validity of the measure (Kenny et al., 2012).

#### *\*Child Report Survey of Knowledge of Potentially Risky Situations (Bright et al., 2022)*

A 14-item measure of children's knowledge of safe and risky situations was developed by Bright and colleagues' (2022) for their randomised control trial. This survey was modelled on Tutty's (1997) [CKAQ](#) as well as specific concepts covered in the curriculum (Child Safety Matters). A 16-item version of the survey was pilot-tested with 25 students. This pilot test served to check for student's comprehension of the items as well as how long it took to complete. Two items were removed and they reported slight adjustments in wording to improve comprehension of the remaining items. No other testing on the psychometrics of this measure was reported.

There is little support currently for use of this measure. In this study, a potential ceiling effect was found for some items, where students (over 75%) scored correctly at the pretest (Bright et al., 2022). Similarly, there were some items in which students did poorly at both the pretest and the posttest, suggesting either that the intervention did not improve children's knowledge on these particular concepts or that the items were not a valid or reliable measure of children's knowledge.

#### *\*Children's Awareness of Scary Secrets Questionnaire (CASSQ; Chen et al., 2012)*

The CASSQ was developed by Chen and colleagues (2012) to assess children's ability to distinguish between okay and not okay secrets. It has 6 items, with correct responses receiving 1 point and incorrect receiving 0 points.

One-month test-retest reliability was reported as 0.45, and the internal consistency was poor for early elementary school children (Cronbach's alpha = 0.57) and unacceptable for late elementary school children (Cronbach's alpha = 0.42; Chen et al., 2012). The CASSQ was found to be too easy for the late elementary school children, with most scoring correctly on most items before the intervention. This suggests a ceiling effect for this version of the CASSQ, particularly for older children. Currently, there is limited support for the psychometric properties of the CASSQ.

#### *\*Children's Sexual Knowledge Questionnaire (CSKQ; Chen et al., 2012)*

The CSKQ was developed by Chen and colleagues (2012) in their evaluation of a sexual abuse prevention program for Taiwanese children. It was designed to assess knowledge of body parts, gender differences, puberty, and conception. The items were based on texts used to assist in program development, with each answered as true, false, or don't know.

The questionnaire had adequate test-retest reliability (0.71), but poor internal consistency for both early (Cronbach's alpha = 0.35) and late (Cronbach's alpha = 0.49) elementary school children (Chen et al., 2012). There was a potential ceiling effect for this measure as children did well on this measure before being exposed to the intervention. Subsequently, there was no significant

improvement in knowledge as a result of the intervention (Chen et al., 2012). Currently, there is limited support for the psychometric properties of the CSKQ.

#### *\*Child knowledge questionnaire (Conte et al., 1985)*

Conte and colleagues (1985) developed a questionnaire of children's knowledge in their evaluation of a sexual abuse prevention program. Items covered concepts such as body part names, secrets, support systems, disclosure, and the differences between affectionate and non-affectionate touch. Inter-rater reliability was excellent (98% agreement between raters), but no other analysis was reported on the psychometric properties of the questionnaire.

While there was an increase in knowledge overall, children did not learn more than half the concepts taught in the prevention program (Conte et al., 1985). We currently cannot be certain whether this was because the prevention program was ineffective at teaching those concepts, or because the questionnaire was not psychometrically sound.

#### *\*Sexual Abuse Knowledge Inventory (SAKI; Crowley, 1989)*

The SAKI was created by Crowley (1989) as part of their dissertation. This measure contains 27 items designed to measure concepts taught in the Good Touch/Bad Touch (GTBT) program. It was designed to be administered in a group format. Scores range from 0-27, with higher scores indicating greater knowledge of sexual abuse concepts (Crowley, 1989).

Seven mental health professionals involved in child sexual abuse prevention education were asked to contribute to a pool of questions designed to assess concepts taught in the GTBT program. 113 potential questions were then rewritten to a consistent format and a lower reading difficulty, and were then reviewed by a panel of five experts. Experts independently rated the items on a 7-point Likert scale and the best three questions for each of the nine concepts in the GTBT program then became the 27-item SAKI.

The internal consistency (Cronbach's alpha = 0.7383) and test-retest reliability (Pearson's  $r = 0.8922$ ) were acceptable. Scores on the SAKI also correlated with scores on the [PSQ](#), providing support for the criterion validity of the SAKI. Participants in the treatment group also showed significant increases in knowledge on the SAKI compared to the control group. There was evidence that there might be a marked ceiling effect with older children (Crowley, 1989). The SAKI has not been used in any subsequent experimental research on the prevention of institutional maltreatment.

#### *\*Case Vignettes (Czerwinski et al., 2018)*

In a German study on the effectiveness of a school-based prevention program to prevent child sexual abuse, Czerwinski and colleagues (2018) developed a case vignette to assess children's gains in assessing actions in response to potentially hazardous situations. The vignette described a child's guitar teacher asking a 9-year-old child (Joan) to stay longer and physically soliciting them after a lesson. 15 items were then listed and were designed to assess multiple ways to deal with the situation (e.g., Joan should tell the teacher that they will tell other people if he does it again or Joan should better not fight back because the guitar teacher is stronger). Some items were adopted from Andresen and colleagues (2015), whereas others were developed based on content taught in the prevention program. The scale was found to have good internal consistency (Cronbach's alpha = 0.81). There was some preliminary evidence for the construct validity of the

vignettes as there was a medium effect on scores after the intervention (Czerwinski et al., 2018). More research would be needed to ascertain the psychometric properties of the vignette and items, as well as its validity in other cultural contexts.

#### *\*General Knowledge about Violence (Daigneault et al., 2012)*

Daigneault and colleagues (2012) developed a ten-item scale designed to measure children's general knowledge about violence. Nine true or false questions were included that assessed the content covered in the workshop in this study. The last item included three definitions of aggression, where participants were asked to select the correct definition. Higher scores on this measure indicated greater knowledge about violence. This measure had questionable internal consistency (Cronbach's alpha = 0.68), and no other analysis was reported on the reliability or validity of this scale (Daigneault et al., 2012).

#### *\*Sexual Assault Knowledge Questionnaire (Daigneault et al., 2015)*

Daigneault and colleagues (2015) included 10 items designed to assess children's sexual assault knowledge in their Canadian cluster randomised control trial. Items were rated as "true" or "false", with higher scores indicating greater knowledge. Participants were also asked who they would contact to get help if they, or someone they know, experienced sexual assault. Overall, the intervention showed improvements in knowledge for eight of the ten items, providing preliminary experimental evidence for the construct validity of the questionnaire. Participants, however, demonstrated weaker knowledge about the identity of the perpetrator and legal issues relating to alcohol or drug intoxication and consent (Daigneault et al., 2015). No analysis was reported on the psychometric properties of the questionnaire.

#### *\*Children's knowledge and efficacy (Dake et al., 2003)*

Dake and colleagues (2003) developed a 26-item scale designed to assess children's knowledge and efficacy. 16 items covered different aspects of child abuse, 5 items covered how confident children were in their ability to respond effectively in abuse situations, 4 items covered demographic information, and the last item asked if children knew an adult outside of the family who they could talk to about an abusive situation (Dake et al., 2003).

The reliability of this scale was tested using control group data. The internal consistency of the knowledge and efficacy subscales were questionable (Cronbach's alpha = 0.64) and poor (Cronbach's alpha = 0.53), respectively. The stability of the knowledge and efficacy subscales was tested using pre-test and post-test data from the control group. This showed that the subscales had questionable test-retest reliability (Pearson product moment correlation coefficients = 0.67 and 0.57, respectively; Dake et al., 2003).

While the intervention group did significantly increase in their knowledge as measured by this scale, there were no significant increases in efficacy from pre-test to post-test. There was also a potential ceiling effect for two knowledge items, as more than 90% of participants answered these items correctly at the pre-test (Dake et al., 2003). Currently, there is limited support for the psychometric properties of this scale.

### *\*Children's Knowledge (Dhooper & Schnieder, 1995)*

In an evaluation of a school-based child abuse prevention program, Dhooper and Schnieder (1995) developed a 12-item questionnaire designed to assess children's knowledge. It assessed four dimensions of child abuse (general understanding, the ability to discriminate between discipline and physical abuse, the difference between appropriate and inappropriate touch, and how to respond to situations of physical and sexual abuse), each with three items. This measure was readable at a third-grade level, according to the Fry Readability Estimate Scale (Dhooper & Schneider, 1995). Each item was read aloud by a teacher, without the use of voice inflection, while the children read along. Children were then asked to circle "yes", "no", or "don't know" in response.

This measure had questionable internal consistency (Cronbach's alpha = 0.68). The face and content validity were determined as 'high' by a panel of child abuse experts. However, details about this process were not reported (Dhooper & Schneider, 1995). There was preliminary experimental evidence for its construct validity, with children showing improved scores following the prevention program. More research would be needed to ascertain the reliability and validity of the questionnaire.

### *\*Maltreatment Knowledge (Diaz et al., 2021)*

Diaz and colleagues (2021) developed a 16-item survey designed to assess children's knowledge about maltreatment. Items were rated on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). The survey also included six questions about the specific strategy taught (RESIST strategy: Run, Escape, Scream, Ignore, Stay Away, and Tell). Participants were asked to identify the correct strategy for each letter in RESIST.

A principal component analysis was conducted on the 16 child maltreatment knowledge items and the six resistance strategy items, providing support for the validity of the questionnaire. Internal consistency was good for the 16-item scale (Cronbach's alpha = 0.82) and acceptable for the six resistance strategy items (Cronbach's alpha = 0.73; Diaz et al., 2021). There was also experimental evidence for the construct validity of the survey, with students having improved scores following exposure to the intervention. These analyses provide good support for this knowledge scale on this population.

### *\*Knowledge Questionnaire (Edwards et al., 2019)*

Edwards and colleagues (2019) created six items to measure children's knowledge about interpersonal violence. They were designed to be similar to previous outcome evaluation studies (Banyard et al., 2007; Foshee & Langwick, 2010). Nine items from the Knowledge Questionnaire (KQ) were also included and were based on the information presented in the intervention program of the study. Higher scores on this measure indicated greater knowledge about interpersonal violence. Internal consistency was good for this measure (Cronbach's alpha = 0.84-0.87 across time points; Edwards et al., 2019). Those in the intervention condition did significantly increase in their knowledge of interpersonal violence compared to the control group. However, there was no other analysis on the reliability or validity reported on this measure. There is promising support but more research is needed.

### *\*Girls' Self-Defence Knowledge (Edwards et al., 2022)*

Edwards and colleagues (2022) included 4-items on girls' self-defence knowledge. These items were also modelled after items used in a Malawi cluster-randomised controlled trial (Decker et al., 2018). There was preliminary experimental evidence for its construct validity, but no analysis was reported on the reliability or validity of this scale. Therefore, we cannot currently be confident that it is a reliable or valid measure of girls' self-defence knowledge.

### *\*Differentiation, Knowledge, and Application (Harvey et al., 1988)*

In Harvey and colleagues' (1988) evaluation of a sexual abuse prevention program, kindergarten-age children completed a three-stage questionnaire. First, children were shown 10 pictures designed to assess their ability to differentiate between good and sexually abusive touches. Five pictures presented "good touches", such as a hug, and five pictures presented sexually abusive touches, such as a man touching a child between their legs. Children were asked to simply identify if the scenario represented a good or abusive touch. Five items designed to assess basic knowledge about abuse were then presented to the children. Children were then assessed on their ability to apply knowledge by exposure to two scenes. A researcher read each scene to the child and then asked the child six questions about whether the abuse was wrong and what the child should do (Harvey et al., 1988).

There was no analysis reported on the psychometric properties of this measure. There was some preliminary experimental support for its construct validity, but without further evaluation, we cannot be certain that it is a reliable or valid measure of children's knowledge.

### *\*What I know about touching scale (Hazzard et al., 1991)*

A 25-item questionnaire designed to assess children's knowledge about sexual abuse was used in Hazzard and colleagues' (1991) evaluation of child sexual abuse prevention program in six U.S. schools. It was developed following a pilot test of previous scales (Plummer, 1984; Wurtele et al., 1986). Details about this pilot test were not reported in this study. Items covered concepts such as definitions of abuse, characteristics of abusers, saying no, disclosing, and sexual abuse not being a child's fault. Items were answered as yes, no, or I don't know, with higher scores indicating greater knowledge (Hazzard et al., 1991).

The questionnaire was found to have acceptable internal consistency (Cronbach's alpha = 0.75) and test-retest reliability (0.77; Hazzard et al., 1991). There was some preliminary experimental support for its construct validity, with scores on this questionnaire increasing following the intervention. More research would be needed to ascertain its psychometric properties.

### *\*Children's Knowledge (Hébert et al., 2001)*

Hébert and colleagues' (2001) derived an 11-item measure based on their own created items as well as items from the [CKAQ](#) and [PSQ](#). Participants answered each item with "yes", "no", or "I don't know", with higher scores indicating greater knowledge about abuse.

An initial pool of 82 items were pretested and reduced to the 11-item version used in this study. Details about this pretest were not reported in this study. The internal consistency of the questionnaire was acceptable (Cronbach's alpha = 0.70) and it had good test-retest reliability ( $r =$



0.80; Hébert et al., 2001). There was preliminary experimental evidence for its construct validity, but more research would be needed to analyse its validity.

#### *\*CAN recognition and talkability (Hoefnagels, 2021)*

Adapted vignettes depicting emotional and physical abuse (Hoefnagels, 2001) were used in Hoefnagels and colleagues' (2021) non-randomised cluster-controlled trial in the Netherlands. Vignettes depicting neglect were also added for the purposes of this study. Recognition of child maltreatment was assessed with the question: "Do you recognise this as child maltreatment?" and talkability was assessed with the question: "Did you talk about child maltreatment after school in the last month?".

The stories were reviewed by an expert panel and a pilot was also conducted to assess the face value of the vignettes (Hoefnagels et al., 2021). The details of this review and pilot were not reported, nor was there any other analysis reported on the psychometric properties. There was preliminary experimental evidence for its construct validity, but more research would be needed to analyse its psychometric properties.

#### *\*Knowledge and Skills of Sexual Abuse Prevention (Jin et al., 2017)*

Jin and colleagues (2017) developed a two-part survey in their Chinese study of sexual abuse prevention in schools. This survey was designed to measure children's knowledge and skills in sexual abuse prevention. 10-items were designed to measure knowledge of sexual abuse prevention including children's judgement of body safe rules, appropriate and inappropriate touches, and child sexual abuse risk. The skills part included three scenarios that were designed to assess three key skills: saying no, getting away, and telling a trusted adult. The total score for each part ranged from 0-10 with higher scores indicating greater knowledge and skills.

This measure had questionable internal consistency for both the knowledge (Cronbach's alpha = 0.668) and total skill (Cronbach's alpha = 0.600) parts. No other analyses were reported on the psychometrics of this measure. There was some preliminary experimental support for the validity of the measure, with participants showing improved knowledge and self-protection skills after the program. More research would be needed to assess the reliability and validity of this measure.

#### *\*Knowledge of CSA (Kang et al., 2022)*

Kang and colleagues (2022) revised a measure of participants' knowledge of child sexual abuse used in a previous evaluation of a child sexual abuse prevention program (Kim & Kang, 2017). The original measure was subjected to a review by an expert panel, but the details were not reported. The measure included three subcategories of knowledge (concept of CSA, causes and occurrence of CSA, and difficulties after CSA), each with 10 items. The content validity index was above 0.8, and there was experimental support for its construct validity, with improvements in scores observed (Kang et al., 2022). No analysis was reported on the reliability of this measure, therefore further research would be needed.

#### *\*Car, Traffic, Fire, Gun Safety Questionnaire (SAFETY; Kenny et al., 2012)*

The SAFETY tool was developed by Kenny and colleagues (2012) in their evaluation of a personal safety program for Latino preschoolers in the U.S. It was designed to assess children's knowledge of general safety knowledge of concepts relating to cars, traffic, fire, and guns. There was some

preliminary experimental support for its construct validity, with improvements in personal safety knowledge after the intervention (Kenny et al., 2012). This measure has not been assessed for its psychometric properties.

*\*Self-protection skill form (Kızıltepe et al., 2022)*

Children were asked, “What can you do if a stranger person wants to touch you?”, in Kızıltepe and colleagues (2022) study in Turkey. This question was developed by Yılmaz Irmak and colleagues (2018) to assess children’s knowledge of skills to protect themselves from sexual abuse. Children were prompted to provide multiple responses, with each being coded by two trained postgraduate students. ‘Effective’ responses were summed and included responses such as telling someone, escaping or getting away, saying no, and hitting or shouting. Inter-rater reliability was excellent, ranging between 0.94 and 0.98 at pre-test, post-test, and follow-up (Kızıltepe et al., 2022). There was some preliminary experimental support for its construct validity, but no other analysis was reported on the reliability or validity of this question. Importantly, this item was limited to stranger situations and does not capture some of the complexity in other types of child sexual abuse, such as grooming.

*\*Knowledge, experiences, actions (Kolko et al., 1987)*

In Kolko and colleagues’ (1987) evaluation of a school-based sexual abuse awareness and prevention program, a pre-post questionnaire was developed to assess changes in children, teachers’, and parents’ knowledge and opinions, experiences, and actions.

The questionnaire had poor to adequate test-retest reliability ( $r = 0.49-0.68$ ) and poor overall internal consistency (Cronbach’s alpha = 0.34). There was some preliminary experimental support for the construct validity of the scale, with children and parents exposed to the intervention showing improvements in knowledge. There was no impact on knowledge for teachers (Kolko et al., 1987). This may be because the intervention was not effective for teachers, or because the measure was not a valid or reliable measure of teachers’ knowledge.

*\*Child Self-report (Kolko et al., 1989)*

In Kolko and colleagues’ (1989) study, children completed a 25-item questionnaire designed to assess children’s awareness of abuse, subjective disturbance to abuse, and their likelihood of talking. Children completed questions designed to assess their ability to implement the concepts and skills taught in the intervention program. For example, children were asked “When should children keep secrets about confusing touches?”. After the intervention, children were also asked about three preventative strategies taught in the intervention and whether they would be (1) no more likely to (5) a lot more likely to disclose if an adult touched them in an uncomfortable way (Kolko et al., 1989).

Factor analysis provided some support for the construct validity, but it had poor (Cronbach’s alpha = 0.32 for program concepts/skills) to questionable (Cronbach’s alpha = 0.60 for awareness and 0.69 for likelihood of talking) internal consistency. The questions also had poor to questionable test-rest reliability ( $r = 0.43-0.60$ ; Kolko et al., 1989). There was some experimental evidence for the construct validity of the questionnaire, where children showed improvements in awareness and knowledge of the program concepts and skills. However, there was no significant difference in children’s level of distress about sexually intimidating encounters or in their likelihood of disclosure.



Kolko and colleagues (1989) concluded that there is a “need for continued refinement of measures of knowledge”. This measure was not used in any subsequent studies in the EGM, and the evidence for the psychometric properties of the questionnaire is inconsistent.

*\*Knowledge/Attitude Test (Kraizer, 1991)*

This 20-item self-report instrument was developed by Kraizer in their dissertation and was designed to assess a child’s cognitive awareness, attitudes, and understanding of issues associated with risk. Scores on this measure were related to scores on self-esteem measures and protective behaviour measures, providing support for its construct validity (Kraizer, 1991). The author made mention of checks for test-retest reliability but these details were not reported. There was experimental support for its construct validity, with scores increasing following exposure to a prevention program (Kraizer, 1991).

It was used twice by the authors who developed it in experimental research on the prevention of institutional child maltreatment (Fryer et al., 1987a; Kraizer, 1991). Neither study reported any analysis on the psychometrics of this measure. There is preliminary support, but further analysis would be needed.

*\*Measurement of Knowledge and Assertiveness (Neherta et al., 2017)*

Neherta and colleagues (2017) made reference to a measure of knowledge and assertiveness in their evaluation of sexual abuse prevention interventions in Padang, India. However, the paper did not include any details about this measure or its psychometric properties.

*\*Good Secrets, Bad Secrets Quiz (Synder, 1986)*

This quiz was designed to specifically assess concepts taught in the “Good Secrets, Bad Secrets” prevention program in Synder’s (1986) dissertation. A 20-item version was piloted but a potential ceiling effect was found as all children scored 80% or better before seeing the prevention program. The quiz was then expanded to 35 items with correct answers being made “less obvious”. Questions covered concepts such as distinguishing between appropriate and sexual touches, recognising how to get help, recognising that sexual touches may come from known adults and not just strangers, and general safety habits (Synder, 1986). The final version was reviewed by a panel of qualified professionals in the field of child sexual abuse. The professionals agreed that it was a valid instrument and test-retest reliability in a pilot was good ( $r = 0.91$ ; Synder, 1986). While children did improve in their knowledge in this study, mastery (defined as 80% or better at post-test) was only possible for about 41% of children. Synder (1986) made some recommendations in their paper about improvements to item discrimination and processes in delivering prevention programs, but this measure has not been used again in experimental research on the prevention of institutional child maltreatment.

*\*Knowledge Related to Gender Violence and Harassment (Taylor et al., 2010)*

Taylor and colleagues (2010) included questions relating to state rape laws, sexual harassment myths, resources for support, and definitions of abuse and sexual harassment, in their evaluation of a gender violence/harassment prevention program in middle schools. The authors reported that the items were subjected to a pilot test, but the details were not included in their paper. There was no analysis reported on the reliability or validity of this scale. There was preliminary experimental

support for its construct validity, with those in the treatment group having significantly higher scores than the control group. Currently, we cannot be confident that this is a reliable or valid measure.

#### *\*Knowledge Scale (Telljohann et al., 1997)*

Telljohann and colleagues (1997) developed a measure of children's knowledge of the content covered in a sexual abuse prevention program. Content analysis was conducted on the program as well as the third grade curriculum and a review of the literature. Curriculum designers then reviewed the measure. No other details about this review were reported in this study. A pilot study of a 30-item version was conducted on a convenience sample of 61 third grade students. Item analysis indicated a ceiling effect on six items, which were removed from the final version. The final 24-item measure included 18 knowledge items, 2 behavioural intention items, and one item asking about a safe adult they could disclose to in the case of victimisation (Telljohann et al., 1997).

For the knowledge scores, the internal consistency (Cronbach's alpha = 0.59) and the test-retest reliability ( $r = 0.56$ ) were poor. For the behavioural items, test-retest reliability ( $r = 0.44$ ) and internal consistency were unacceptable (Cronbach's alpha = 0.49). The poor internal consistency may be explained by this measure only having two items. Modest gains in knowledge were found for the intervention group and the behavioural items had a potential ceiling effect (Telljohann et al., 1997). Similarly, the control group improved in their ability to identify a safe adult to talk to from pre-test to post-test, despite not being exposed to the intervention. This suggests that this item may sensitise participants, and therefore it may not be an accurate measure of true intention to disclose.

#### *\*Drawn vignettes child knowledge (Warden et al., 1997)*

In Warden and colleagues' (1997) evaluation of a children's safety training in Scotland, drawn vignettes were used to assess children's knowledge of the four key safety situations taught in the prevention program (Kidscape). The drawings were included in a pilot study to ensure children were able to comprehend the task and to test the temporal stability of the vignettes. The test-retest reliability of the vignettes was adequate to good (Kendall's tau = 0.48-0.70; Warden et al., 1997). There was some preliminary support for the construct validity of the vignettes, as awareness of the children who participated in the programme had improved. However, the authors noted that students in the control group also improved in their awareness, suggesting that there was a possible practice effect to the vignettes (e.g., the act of students being exposed to the vignettes twice was educational and could explain the improvements in awareness). Similarly, improvements in knowledge were only notable for younger children. Older participants had performed well on the vignettes before the intervention (Warden et al., 1997), indicating a possible ceiling effect. Therefore, we cannot be confident that these vignettes are a reliable or valid measure of children's knowledge.

#### *\*Student assessment (Weeks et al., 2021)*

Weeks and colleagues (2021) revised a student assessment scale (Finkelhor et al., 2018) aimed at assessing student knowledge for their evaluation of a victimisation prevention curriculum for elementary school students in the U.S. Items were revised to reduce bias. Children completed between 20 and 24 items depending on their grade level. There was generally low correlation between items (0.08-0.39) and the internal consistency was questionable (Cronbach's alpha =

0.62). While there were improvements in knowledge following the intervention, the improvements were modest. Currently, there is limited support for the psychometric properties of this measure.

*\*Child Sexual Abuse Knowledge Form (Yilmaz Irmak et al., 2018)*

The Child Sexual Abuse Knowledge Form is a Turkish version of the [PSQ](#) (Saslowsky & Wurtele, 1986). It was developed for use by Yilmaz Irmak and colleagues (2018) for use in an evaluation of a teacher-based child sexual abuse prevention program in Turkey (Kiziltepe et al., 2022). It includes 17 questions designed to assess children’s knowledge of bodily autonomy and awareness of sexual abuse concepts (Yilmaz Irmak et al., 2018).

In this study, only 10 questions were used as they better differentiated between groups in the analysis (Kiziltepe et al., 2022). There was experimental evidence for the construct validity of the questionnaire, with children exposed to the intervention gaining more knowledge than children on the wait-list comparison group (Kiziltepe et al., 2022). To our knowledge, there has been no other analysis of the psychometric properties of this translated version.

### Children’s Attitudes

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#">Readiness to Help</a>	Banyard et al., (2014)		U.S.	Child	There is not much need for me to think about relationship abuse and/or sexual assault among high school students.	There is strong support for this scale on college samples. There is promising support for high school samples for one subscale. More research would be needed to ascertain psychometrics of the full scale on younger participants.
<a href="#">Victim Empathy Scale (VES)</a>	Beckett & Fisher (1994)		U.S.	Adult  (Used in EGM with children)	I can empathise with the emotions of a victim of relationship abuse and/or sexual assault.	There is strong support for use with adult offender populations. More research would be needed to determine its psychometric properties for use with children.

<b><u>Interpersonal Reactivity Index (IRI)</u></b>	Davis (1980)		U.S.	Adolescent	I sometimes find it difficult to see things from the “other guy’s” point of view.	More research may be needed to assess its psychometric properties in younger cohorts and in other cultural settings.
<b><u>The Pros and Cons of Bystander Action Scale (PCBAS)</u></b>	Edwards et al., (2018)		U.S.	Adolescents	Even if I don’t know the person, I can still help.	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>Relationship Media Literacy Scale (RMLS)</u></b>	Edwards et al., (2018)		U.K.	Adolescents	I am bothered by the media’s portrayal of relationship abuse and sexual assault.	There is some support, but more research is needed to determine the psychometric properties.
<b><u>Discomfort with Sexist Situations- Revised (DSS-R)</u></b>	Kilmartin et al., (1999)		U.S.	Adolescents and adults	While talking with a friend about a girl he likes, he says angrily, I took her out to dinner and a movie and she didn’t even kiss me.	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>Illinois Rape Myth Acceptance Short Form (IRMA-SF)</u></b>	Payne et al., (1999)		U.S.	Adolescents and adults	When women are raped, it is often because the way they said ‘no’ was unclear	More research may be needed to assess its psychometric properties in younger cohorts and in other cultural settings.
<b><u>Gender Equitable Men Scale (GEM)</u></b>	Pulerwitz & Baker (2008)		Brazil	Adolescents and adults (male)	There are times when a woman deserves to be beaten.	More research may be needed to analyse its psychometrics in other cultural contexts.

<b><u>Student Attitudes Toward Gender Violence</u></b>	Ward (2002)		U.S.	Adolescents and adults	If a guy forces his girlfriend to have sex with him when she doesn't want to, it is rape.	There is promising support, but more research is needed.
<b><u>*Sexual Violence Attitude Scale</u></b>	Daigneault et al., 2015		Canada	Child	Most of the time, girls falsely report sexual assault to get attention	There is some support, but more research is needed to determine the psychometric properties.
<b><u>*Confidence to Resist</u></b>	Edwards et al., (2022)		U.S.	Children	If I am attacked by a strong man, I feel confident that I can defend myself.	No analysis was reported on the psychometrics.
<b><u>*Behavioural intentions, social norms, attributions</u></b>	Hoefnagels et al., (2021)		The Netherlands	Child	If you were Bas, would you seek for help?	There is some support, but more research is needed to determine the psychometric properties.
<b><u>*Gender attitudes, social norms, and communication with girlfriend/boyfriend</u></b>	Jewkes et al., (2019)		South Africa	Child	I think a man should have the final say in family matters.	Some support for its internal consistency but no other analysis was reported on the psychometrics.
<b><u>*Child gender attitudes</u></b>	Karmaliani et al., (2020)		Pakistan	Child	I think the wives in the family should have a say in how money in their family is spent.	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Control in Sexual Conflicts</u></b>	Taal & Edelaar (1997)		The Netherlands	Child	Not reported	There is some support, but more research is needed. There is also a

						potential ceiling effect.
<u>*Empathy, Adverse Sexual Beliefs, and Attitudes Toward Dating Violence</u>	van Lieshout et al., (2019)		The Netherlands	Children	I think a boy is allowed to force a girl into having intercourse when she has allowed him to take off her pants.	Limited support for the psychometric properties.

### *Readiness to Help (Banyard et al., 2014)*

The Readiness to Help Scale is based on a 36-item Readiness for Change Scale (Banyard et al., 2010). This new version contains 33-items designed to measure attitudes related to sense of responsibility, problem awareness, and taking action regarding sexual and relationship abuse and stalking. Items are rated on a 5-point scale. This measure was normed on a sample of college students in the U.S., where internal consistency was good to excellent (Cronbach’s alpha = 0.87-0.93). Content validity was first established through collaboration with practitioner and research experts in the field of bystander intervention (Banyard et al., 2014). Scores also correlated with attitudes more supportive of taking responsibility and taking bystander action, providing support for its construct validity. Factor analysis revealed a three-factor solution: action, taking responsibility, and no awareness. Overall, there is strong support for use of this measure with a U.S. college sample.

The Denial subscale of this measure was used to assess attitudes in high school students in the U.S. following a bystander-focused interpersonal violence prevention program (Edwards et al., 2019). Previous research identified this concept as important in understanding bystander prevention in college samples (Moynihan et al., 2015). A 4-item scale rating was used (1 = disagree strongly to 4 = agree strongly). In this study, internal consistency ranged from acceptable to good (Cronbach’s alpha = 0.69-0.80; Edwards et al., 2019). There was experimental support for the construct validity of the subscale, with students showing a significant decrease in denial compared to students in the control conditions (Edwards et al., 2019). There seems to be promising support for this subscale in high school samples, but more research would be needed, including for the full Readiness to Help scale.

### *Victim Empathy Scale (VES; Beckett & Fisher, 1994)*

The VES is a 28-item self-report tool developed to measure emotional empathy for victims of violent crime and was originally designed to be used with adult sex offenders (Beckett & Fisher, 1994). However, Edwards and colleagues (2019) used this measure to assess adolescent Native American girls’ victim empathy. The authors reported good internal consistency (Cronbach’s alphas = 0.80-0.86) across time-points, but there was no mention from the authors of any adjustments made to the original measure for use with their sample. The VES has strong psychometric properties when used on adult populations (Mizwa, 1993), but it is not clear if this measure is reliable or valid when used with non-offending populations or with children.



### *Interpersonal Reactivity Index (IRI; Davis, 1980)*

The IRI is a 28-item scale designed to measure empathy in adolescents and adults. It has four subscales, each with seven items, including: perspective taking, fantasy, empathic concern, personal distress. Items are rated on a 5-point Likert scale from “does not describe me well” to “describes me very well” (Davis, 1980). It has strong psychometric properties. The internal consistency is adequate (Cronbach’s alpha = 0.70-0.78) and has adequate test-retest reliability (0.61-0.81; Davis, 1980). There is also support for its content, construct, and predictive validity. For example, IRI has been found to correlate with the Empathy Quotient (Melchers et al., 2015) and the Basic Empathy Scale (Jolliffe & Farrington, 2006), and higher scores on the IRI are associated with higher self-esteem and better social functioning. Most of the research has focussed on college samples. More research may be needed to assess its psychometric properties when used with children.

It has only been used once in experimental research on the prevention of institutional child maltreatment. A French translation (Pelletier et al., 1998) of the empathic concern subscale was used in an evaluation of a sexual abuse prevention workshop in multicultural, impoverished urban area in Canada (Daigneault et al., 2012). The French translation was found to have questionable to acceptable internal consistency (Cronbach’s alpha = 0.68-0.73), but for use in Daigneault and colleagues’ (2012) sample, the internal consistency was poor (Cronbach’s alpha = 0.52). There was also no change in empathy scores in this study. It is not clear if this was because the intervention did not affect empathy in this sample or if the measure was not suitable for this population.

### *The Pros and Cons of Bystander Action Scale (PCBAS; Edwards et al., 2018)*

This 8-item measure was designed to assess students’ perceptions of the pros and cons of bystander action, with a particular focus on situations relating to relationship abuse and sexual assault (Edwards et al., 2018). Items are rated on a 4-point Likert scale from (1) disagree strongly to (4) agree strongly.

The scale originally contained 10 items, but two were removed to improve the validity of the scale. Factor analysis revealed a two-factor model provided the best fit. The internal consistency was questionable to adequate (Cronbach’s alpha = 0.63-0.76; Edwards et al., 2018).

It was later used once by the developers of the measure in an evaluation of a bystander-focused interpersonal violence prevention program for high school students in the U.S. (Edwards et al., 2019). In this sample, the PCBAS had questionable to acceptable internal consistency across time points (Cronbach’s alpha = 0.61-0.74). There was a reduction in some forms of violence and short-term changes in bystander barriers/facilitators in students exposed to the intervention. However, there was very little long-term impact on actual bystander behaviour (Edwards et al., 2019). More research is needed to analyse the psychometric properties of the PCBAS.

### *Relationship Media Literacy Scale (RMLS; Edwards et al., 2018)*

This 6-item scale was designed to assess media literacy with items rated on a 4-point Likert scale from (1) disagree strongly to (4) agree strongly. It was specifically designed to assess perceptions of how the media portrays relationships. However, there was limited support for its validity and reliability (Edwards et al., 2018).



Edwards and colleagues (2019) used the Bothered by Media subscale in their bystander prevention program evaluation in the U.S. This subscale includes three items and had questionable to acceptable internal consistency across time points in this sample (Cronbach's alpha = 0.65-0.74). Long term changes in media literacy were observed in students that received the bystander prevention program, compared to the control group, providing preliminary experimental support for its construct validity (Edwards et al., 2019).

Further research would be needed to analyse its psychometric properties.

### *Discomfort with Sexist Situations-Revised (DSS-R; Kilmartin et al., 1999)*

The DSS-R involves rating how comfortable participants feel in seven situations that are sexist or rape supportive. Items are rated from 1 (very uncomfortable) to 5 (very comfortable), with higher scores suggesting greater discomfort with sexist or rape supportive situations. The DSS-R was originally designed for use with a college population.

Hillenbrand-Gunn and colleagues (2010) modified the DSS-R for use with male high school-aged children in an evaluation of a rape prevention intervention in the U.S. The internal consistency of this modified version on this sample was acceptable (Cronbach's alpha = 0.72). Both female and male students then completed the DSS-R based on a "typical guy at your school" to gather information about social norms. This typical guy version had acceptable internal consistency (Cronbach's alpha for male students = 0.79; Cronbach's alpha for female students = 0.71; Hillenbrand-Gunn et al., 2010). No other analysis was reported on the reliability or validity of this modified version of the DSS-R. There was experimental support for the construct validity, with scores improving following the rape prevention program.

There is promising support, but further research would be needed to analyse the psychometric properties of the DSS-R, including on samples of adolescents.

### *Illinois Rape Myth Acceptance Short Form (IRMA-SF; Payne et al., 1999)*

The IRMA-SF is a 20-item self-report scale that contains seven subscales aimed at measuring attitudes and beliefs about sexual assault. Each item was designed to assess attitudes and beliefs that are generally false but are culturally persistent and serve to "deny and justify male sexual aggression against women" (Lonsway & Fitzgerald, 1994, p.134). Items are rated on a 5-point Likert scale (1 =strongly disagree to 5 = strongly agree). Higher scores on this measure indicate higher endorsement of rape myths such as the victim asked for it or lied about it happening.

The IRMA-SF was validated on a population of 780 U.S. university students. There is also a 45-item version. The internal consistency of the full scale was reported to be excellent (Cronbach's alpha = 0.93), with item-total correlations ranging between 0.31 and 0.67 (Payne et al., 1999). There is also strong support for the validity of the measure. Exploratory and confirmatory factor analysis revealed a general myth component and seven subcomponents. Scores on the IRMA-SF also correlate strongly with a number of related measures such as the Sex-Role Stereotyping Scale (Burt, 1980), Adversarial Sexual Beliefs Scale (Burt, 1980), and Hostility Toward Women Scale (Lonsway & Fitzgerald, 1995), providing evidence for its construct validity (Payne et al., 1999).

The IRMA-SF was used twice to assess children's rape myth acceptance in U.S. based experimental research on prevention of institutional child maltreatment (Edwards et al., 2019;

Hillenbrand-Gunn et al., 2010). Edwards and colleagues (2019) used an even shorter version containing six items, that had been used previously in a study on bystander prevention in college campuses (Coker et al., 2011). Both studies reported adequate to good internal consistency (Cronbach's alpha = 0.72-0.88). There was experimental support for the construct validity, with both studies reporting significant decreases in rape-supportive attitudes (Edwards et al., 2019; Hillenbrand-Gunn et al., 2010). More research may be needed to assess the psychometrics of the 6-item version.

### *Gender-Equitable Men Scale (GEM; Pulerwitz & Barker, 2008)*

The GEM is a 24-item measure of attitudes toward gender norms among young men. Items assess concepts such as violence, sexual and reproductive health, sexual relations, domestic work, and homophobia (Pulerwitz & Baker, 2008). There is strong support for its reliability and validity in a sample of 15-24 year-old men in Brazil. It has good internal consistency (Cronbach's alpha = 0.81) and factor analyses provided support for two subscales. Higher scores on the GEM were also significantly correlated with less self-reported partner violence, higher education level, and more contraceptive use, providing support for its construct validity (Pulerwitz & Barker, 2008). The GEM scale was validated for use on a younger sample of 10-14 and 15-24 year-olds in Uganda (Vu et al., 2017).

It has not yet been used in experimental research on institutional child maltreatment but was included in Knight and colleagues' (2020) protocol for a nonrandomised cluster-controlled trial in Uganda.

There is strong support for the psychometrics of the GEM scale. More research may be needed to analyse its validity and reliability in other cultural contexts.

### *Student Attitudes Toward Gender Violence (Ward, 2002)*

This scale was developed by Ward (2002) in their dissertation and was designed to measure student attitudes towards male violence against women. It contains 16 items rated on a 5-point Likert scale from strongly disagree to strongly agree. Higher scores on this scale indicate sexist attitudes toward gender violence. The survey was subjected to pilot testing on a sample of college students. The internal consistency was found to be good (Cronbach's alpha = 0.82-0.84). Factor analysis confirmed the unidimensional structure of the survey, providing preliminary evidence for its validity. Further research would be needed to ascertain the psychometric properties of this survey.

Taylor and colleagues (2010) adapted the items for their randomised experimental evaluation in U.S. middle schools. A series of questions were included to assess students' acceptability of violence, abusive, and harassing behaviours, perceived social norms, and the students' motivation to comply with these norms. A factor analysis was conducted which revealed six underlying dimensions, each providing a good overall summary of the attitudinal measures across six content areas (e.g., inappropriate attribution of girls' fault in sexual harassment, belief that gender violence and sexual harassment is not a problem, attitudes that reduce sexual harassment, intention to confront sexual harassment, attitude toward prevention of sexual harassment, and personal space). There was experimental support for its construct validity, with improvements in scores following the intervention. No other analysis was reported on the psychometrics of this adapted version. More research would be needed to assess the psychometric properties of this version.

### *\*Sexual Violence Attitude Scale (Daigneault et al., 2015)*

Daigneault and colleagues (2015) included 25 items designed to assess children's attitudes towards sexual violence in their Canadian cluster randomised control trial. Items were derived from several previous studies including the Rape Attitude Scale (Hall et al., 1986), Dating Violence Attitude and Knowledge Scale (Lavoie et al., 1997), Attitudes Towards Rape-Revised Scale (Harrison et al., 1991), Child Sexual Abuse Myth Scale (Collings, 1997), Sexual Assault Disclosure Scale (Chamberland, 2003) and two additional researcher-developed scale. Items were scored on a 6-point Likert scale from (1) completely agree to (6) completely disagree, with higher scores indicating a less favourable attitude to sexual assault. The internal consistency of this derived scale was good (Cronbach's alpha = 0.82) and there was preliminary experimental evidence for its construct validity (Daigneault et al., 2015). More analysis would be needed to ascertain the reliability and validity of this derived scale.

### *\*Confidence to Resist (Edwards et al., 2022)*

One item was included in Edwards and colleagues' (2022) study on the effectiveness of a sexual assault self-defence program, to assess girls' confidence to resist a potential attacker. This item read: "If I am attacked by a strong man, I feel confident that I can defend myself", and was modelled after previous research (Decker et al., 2018). Participants indicated this item with a 'yes' or 'no' response. It is difficult to ascertain the reliability or validity of single-item measures. Therefore, it is not clear what the utility of this item is in assessing girls' confidence to resist an attack.

### *\*Behavioural intentions, social norms, attributions (Hoefnagels et al., 2021)*

Hoefnagels and colleagues (2021) included a vignette designed to ascertain students' behavioural intention to disclose possible abuse or neglect in a non-randomised cluster-controlled trial in the Netherlands (Hoefnagels et al., 2021). Students were asked, "If you were Bas, would you seek for help?" and "(...) would you tell to your teacher?" (yes, no, maybe). Students were then asked, "(...) do you think your teacher believes him?" (yes, no, maybe) to assess children's perception of the teacher's potential to believe a child. Students were also asked about their perceptions of social norms about talking about child abuse and neglect ("(...) do you think your peers agree that this is something you can talk about?"), their attribution of blame ("Is Samira the one who's to blame for her father being so angry?"), and their opinion about the admissibility of abuse and neglect (e.g., "Is Bas' father allowed to insult him?"). Internal consistency was acceptable (Cronbach's alpha = 0.74), but no other analysis on the psychometrics of these items was reported (Hoefnagels et al., 2021). There was some mixed experimental evidence for its construct validity. Currently, we cannot be confident that this is a valid or reliable measure of children's intentions to disclose.

### *\*Gender attitudes, social norms, and communication with girlfriend/boyfriend (Jewkes et al., 2019)*

Jewkes and colleagues (2019) adapted 5-items on gender attitudes, 9-items on individual gender attitudes, and 9-items on social norms that had been used in a previous study (Jewkes, 2002). These items had poor (Cronbach's alpha for girls = 0.552, for boys = 0.595), adequate (Cronbach's alpha for girls = 0.76, for boys = 0.79), and good (Cronbach's alpha for girls = 0.83, for boys = 0.88) internal consistency on this sample, respectively (Jewkes et al., 2019). Three items designed to assess students' communication with their girlfriend or boyfriend were also included. These

items had adequate internal consistency (Cronbach's alpha = 0.713-0.763; Jewkes et al., 2019). There was no significant effect on outcomes as a result of the intervention, so we cannot be certain currently that these were reliable or valid measures.

*\*Child gender attitudes (Karmaliani et al., 2020)*

In Karmaliani and colleagues' (2020) Pakistani cluster-randomised controlled trial, 13 items were included to measure children's gender attitudes, roles, and norms. For example, items assessed attitudes such as a husband's right to punish their wife, wives obeying their husbands, and limits placed on women's participation in social events and employment. The internal consistency for these items was acceptable (Cronbach's alpha = 0.74), and there was preliminary experimental support for its construct validity (Karmaliani et al., 2020). There was no other analysis was reported on the reliability or validity of these items.

*\*Control in Sexual Conflicts (Taal & Edelaar, 1997)*

This questionnaire was designed to measure the feeling of control over unwanted sexual encounters in children. Four stories were presented about unwanted sexual encounters in Taal and Edelaar's (1997) evaluation of a child sexual abuse prevention program in the Netherlands. Two stories describe an initially innocent encounter, and the other two stories describe a firm refusal of the sexual approach. Both internal (children's own abilities or competencies) and external (circumstances beyond the child's control) control are assessed. Children rate how much each of the statements applies to them from 1 (very untrue) to 5 (very true). Scores range from 24 to 120 with higher scores indicating a higher feeling of control.

It has been found to have good internal consistency (Cronbach's alpha = 0.84), but poor test-retest reliability ( $r = 0.20$ ; Taal & Edelaar, 1997). Immediately after participation in the program, the youngest and oldest children felt less in control of an abusive interaction, and the youngest pupils thought refusal was less feasible. In the long term, children in this study thought refusing abusive interactions was more feasible, providing preliminary support for its construct validity (Taal & Edelaar, 1997). However, the questionnaire has a potential ceiling effect, with children reporting high levels of perceived control before the intervention. While there is some support, more research would be needed.

*\*Empathy, Adverse Sexual Beliefs, Attitudes Toward Dating Violence, Rape Attitudes, Social Norms, and Outcome Expectancies (van Lieshout et al., 2019)*

van Lieshout and colleagues (2019) included a number of research-developed measures for their study on a sexual harassment prevention program for Dutch residential care youth.

First, a short scenario in which a boy assaulted a girl on the street in their study on Dutch residential care youth. The participants were asked to indicate if they felt "empathy/compassion for the girl" on a 7-point Likert scale from (1) totally disagree to (7) totally agree. This approach was based on previous research (Bos et al., 2007; Van Alphen et al., 2011).

Eight items designed to assess rape attitudes were rated on a 7-point Likert scale from (1) totally disagree to (7) totally agree. The items were based on previous research (Maxwell et al., 2014) and had excellent internal consistency (Cronbach's alpha = 0.91).

Three items designed to assess adverse sexual beliefs were also included. Items were rated on a 7-point Likert scale from (1) totally disagree to (7) totally agree. These items were adapted from Burt (1980) and had questionable internal consistency (Cronbach’s alpha = 0.68; van Lieshout et al., 2019).

Nine items designed to assess attitudes toward dating violence were rated on a 7-point Likert scale from (1) totally disagree to (7) totally agree. These items were based on previous research (Price et al., 1999) and had good internal consistency (Cronbach’s alpha = 0.81; van Lieshout et al., 2019).

Four items designed to assess social norms toward sexual harassment were rated on a 7-point Likert scale from (1) totally disagree to (7) totally agree. The items had questionable internal consistency (Cronbach’s alpha = 0.61; van Lieshout et al., 2019).

Three items designed to assess the extent to which participants expect to become angry, start persuading, or start whining when they want to say sex but their girlfriend does not, were rated on a 7-point Likert scale from (1) totally disagree to (7) totally agree. The items had acceptable internal consistency (Cronbach’s alpha = 0.78; van Lieshout et al., 2019).

The scale Outcome Expectancies was based on O’Donohue et al. (2003). The scale measured to what extent the participant would expect seven outcomes that could occur as a consequence of persuading a girlfriend who is not in the mood to have sex. This scale had good internal consistency (Cronbach’s alpha = 0.89; van Lieshout et al., 2019).

Some items from the Reasoned Action Approach (based on Fishbein and Ajzen, 2010) scale were selected. This scale was designed to assess antecedents to sexually offensive behaviour. No analysis was reported on this modified version.

Participants were also asked about their age, ethnicity, religion, education level, sexual history/relationship experience, and sexual harassment behaviours. No analysis was reported on the psychometrics of these items.

There was no other analysis reported on the psychometric properties of these items, and there was no treatment effect on these outcomes. Similarly, the trainers for the intervention program reported that many of the boys had difficulty completing the questionnaire due to the length, and some also had difficulty interpreting the questions. Therefore, we cannot be certain that these were valid or reliable measures.

### Children’s Behaviours

Measure	Author	Rating	Location	Who	Example Item	Considerations
<u>Strange Situation Procedure</u>	Ainsworth et al., (1978)		U.S.	Observer	n/a	Requires a trained professional to observe and code. The original design is specific to a Western



						conceptualisation of attachment.
<b><u>What Would You Do? (WWYD)</u></b>	Hillenbrand-Gunn (2003)		U.S.	Adolescents	How likely are you to stop the first time a girl says <i>no</i> to your sexual advances?	The full details of the scale are not published. There is promising support, but more details and research are needed.
<b><u>Play it Safe! Scale</u></b>	Thompson et al., (2020)		U.S.	Child	Vignette 3. May played tag in the pool with her Uncle Troy. He rubbed May on the bottom of her bathing suit three times.	There is good preliminary support for this scale. Factor analysis would provide additional support for its psychometrics.
<b><u>Parent and Teacher Evaluation Forms</u></b>	Yilmaz Irmak et al., (2018)		Turkey	Parents and teachers	Having a nightmare, fear of strangers, not wanting to go to school.	There is currently limited support for its psychometric properties.
<b><u>*Touch Discrimination Task</u></b>	Blumberg et al., 1991		U.S.	Child	Mary went to the doctor because she fell off her bike and hurt her private parts. The doctor took her panties down and he touched Mary's private parts.	Limited support for its psychometric properties and has a potential ceiling effect.
<b><u>*Touch aversion</u></b>	Czerwinski et al., (2018)		Germany	Child	Not reported	There was some support for its reliability, but details about the scale were not reported and there was no reported

						analysis on the validity of the scale.
<b><u>*What Would You Do?</u></b>	Hazzard et al., (1991)		U.S.	Child	Do you think this situation is safe or unsafe?	Potential ceiling effect for some items and no analysis reported on the psychometric properties.
<b><u>*Vignettes of Skills in Abusive Situations</u></b>	Hébert et al., (2001)		Canada	Child	Not reported	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Self-Protective Behaviors Measure</u></b>	Hillenbrand-Gunn et al., (2010)		U.S.	Female adolescents	If your boyfriend forced you to have sex with him after you had too much to drink at a party, how likely would you be to...?	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Self-protective behaviours against CSA</u></b>	Kang et al., (2022)		South Korea	Child	Not reported	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Self-protective skills</u></b>	Krahé & Knappert (2009)		Germany	Child	Every Saturday, Anna goes swimming with her granddad. Anna really likes going swimming with him. To change into their swimsuits,	There is promising support, but more research is needed to determine the psychometric properties.



					they always share a changing booth, and her granddad says that this is their secret; nobody should know that they do this. Anna would prefer to have a changing booth for herself. She feels really bad about this secret.	
<b><u>*Simulation of a real-life scenario</u></b>	Kraizer et al., (1988)		U.S.	Observer	n/a	There is currently limited support for its psychometric properties.  Ethical issues with simulations.
<b><u>*Prevention of sexual abuse role play</u></b>	Kraizer (1991)		U.S.	Observer	What if your grandmother came over and was pinching your cheek (...). If you don't like it, what would you say and do?	There was some support, but there are important ethical issues with role-plays.
<b><u>*The Choice of Safety Strategy Questionnaire</u></b>	Taal & Edelaar (1997)		The Netherlands	Child	Not reported	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*The Touch Questionnaire</u></b>	Taal & Edelaar (1997)		The Netherlands	Children	Not reported	There is promising support, but more research is needed to

						determine the psychometric properties.
<b>*Observed Protective Behaviours Test (OPBT)</b>	White et al., (2018)		Australia	Observer	<i>Observations of children</i>	There is currently limited support for its psychometric properties.

### *Strange Situation Procedure (Ainsworth et al., 1978)*

The Strange Situation Procedure assesses children’s attachment behaviour by observing their response to caregivers and strangers entering and leaving the room. The child is observed playing for 21 minutes, and the stressfulness of the situation varies throughout the observation time. It can be used for children aged 9 to 30 months. Four patterns of attachment are assessed: secure, anxious-avoidant, anxious-ambivalent/resistant, and disorganised/disoriented (Ainsworth et al., 1978).

It has been used twice in Romanian experimental studies on institutional child maltreatment to examine the attachment of children from the BEIP studies (Smyke et al., 2010; Tang et al., 2021). In both instances, the MacArthur Preschool system was used for coding (Crittenden, 1992).

Tang and colleagues (2021) examined foster care children with their foster mothers and institutionalised children with their “favourite” caregiver, but did not report any analysis to check for the psychometrics for this measure on their sample. Smyke and colleagues (2010) used a home-based version, and 75% of the procedures were double-coded. They reported 87.7% agreement on attachment classification between coders, where disagreements were then resolved in conference. There was experimental support for the construct validity but there was no other analysis of the psychometrics reported.

A common critique is that the underlying conceptualisation of attachment for this approach is based on Western ideas about attachment (e.g., mothers as the primary caregiver). Additional research may be needed to ascertain its validity in other cultural settings (Ziv & Hotam, 2015).

### *What Would You Do? (WWYD; Hillenbrand-Gunn, 2003)*

The WWYD is a 9-item measure with two subscales designed to assess male participants’ behaviour in rape-supportive situations. The first subscale includes three items that were taken from a larger inventory ([IRMA](#)) that assesses the participant’s own coercive behaviour. The second subscale includes six researcher-constructed items that assess the participants’ willingness to intervene in rape-supportive situations. Items are rated on a 5-point Likert scale from (1) not at all likely to (5) very likely. The WWYD has acceptable internal consistency (Cronbach’s alpha = 0.74; Hillenbrand-Gunn, 2003).

There is also a WWYD-Typical Guy (WWYD-T) version, which is identical to the WWYD but can be completed by both male and female students. In this version, students are asked to rate the items based on “what you think the typical guy at your school would answer”, to assess their perceptions

of social norms. Higher scores on the WWYD-T indicate the participant views their peers as less rape supportive. It has acceptable internal consistency for male (Cronbach's alpha = 0.73) students and good internal consistency for female (Cronbach's alpha = 0.84) students.

Both the WWYD and the WWYD-T were used by the same author in a U.S. based study on the efficacy of a high school rape prevention intervention (Hillenbrand-Gunn et al., 2010), but have not been used in any other study in the EGM. In this study, the willingness to engage in rape-preventative behaviours among male students was associated with the level of discomfort with sexist situations (Hillenbrand-Gunn et al., 2010). Unwillingness to engage in rape-preventative behaviours was associated with the acceptance of rape myths, providing experimental support for the construct validity (Hillenbrand-Gunn et al., 2010).

The full details of the WWYD are not published. While there is promising support, more details and research are needed.

### *Parent and Teacher Evaluation Forms (Yılmaz Irmak et al., 2018)*

These evaluation forms were designed to assess short-term positive and negative side effects of a prevention program on children. It includes 17 behaviours for the parent form and 16 behaviours for the teacher form. Parents and teachers are asked to rate how many times the child has displayed these behaviours in the past two weeks (between 0-4). Both forms were used by Kızıltepe and colleagues (2022) in their evaluation of a child sexual abuse prevention program in Turkey. To our knowledge, there has been no analysis on the psychometric properties of these forms.

### *Play it Safe! Scale (Thompson et al., 2020)*

The Play it Safe! Scale has 14 items and 5 vignettes depicting different physical and sexual abuse. The 5 scenarios describe physical abuse from a parent, a babysitter taking photos of a child while undressing, confusing touches from an adult family member in a swimming pool, confusing touch from an adult, and meeting strangers online. Each vignette had four response options (Thompson et al., 2020). It was pilot-tested with a group of 13 primary school children. Questions were reworded to reduce confusion.

The Play it Safe! Scale has adequate internal consistency (Cronbach's alpha = 0.77). It was found to correlate with other similar measures, providing support for its reliability and construct validity (Child Safety Matters and Kidpower; Thompson et al., 2020).

The scale was used once in a study on the EGM by the same authors in an evaluation of a school-based child physical and sexual abuse prevention program (Thompson et al., 2022). There was experimental evidence for its construct validity, with significant improvements in knowledge observed, particularly for younger children (Thompson et al., 2022).

### *\*Touch Discrimination Task (TDT; Blumberg et al., 1991)*

The TDT was designed to assess a child's understanding of the difference between okay and not-okay touches. It also assesses if children become over-sensitised to touch as a result of a sexual abuse prevention program. The TDT includes seven vignettes where an adult touches a child in some way. Three vignettes include inappropriate touches of private parts, two involve appropriate touches of private parts, and two involve other appropriate touches (Blumberg et al., 1991).

It was developed by the researchers of an experimental study on the prevention of institutional child maltreatment but has not been used again (Blumberg et al., 1991). In this study, children who received a role-play based sexual abuse prevention program, significantly improved on the TDT compared to the control group. However, this improvement in scores for children in the role-play group was almost exclusively on the appropriate touch questions and not on the sexually abusive touch items. There was also no significant difference in TDT scores between the multimedia and control group. This finding may be explained by a ceiling effect, in that children demonstrated high levels of knowledge about inappropriate touch questions, even before the intervention. Blumberg and colleagues (1991) concluded that prevention programs “spend too much time teaching children to do something at which they already excel”. They called for a focus on more complex items to assess the true effects of knowledge gains in children.

#### *\*Touch aversion (Czerwinski et al., 2018)*

Czerwinski and colleagues (2018) developed a scale designed to assess children’s aversion to touch. It was intended to ascertain if there were any possible side effects to the intervention. A list of items was pilot-tested and revised, although the details of the pilot or the revisions were not reported. A final 8-item scale was used for children and a 16-item version for parents. The response scales were adapted to those used for the [CKAQ](#) (child version) and SCARED (parent version). Higher scores on this scale indicated a greater aversion to touch in children. Both had acceptable internal consistency (Cronbach’s alpha = 0.71 and 0.70, respectively) and there was no effect on scores as a result of the intervention. No other analysis on the psychometrics or the details of the scale were not reported. Therefore, we cannot be confident that this is a reliable or valid measure of touch aversion.

#### *\*What would you do? (Hazzard et al., 1991)*

In Hazzard and colleagues’ (1991) study, they used six videotaped vignettes to assess children’s ability to distinguish between safe and potentially abusive situations and their ability to use primary prevention skills in abusive situations. Each vignette was 30 seconds, with four presenting potentially abusive scenarios. In the development of the vignettes, ten psychologists were asked to rate each scene as either safe or unsafe, but no details on the psychometric properties of the approach were reported. Children were asked: (1) do you think this situation is unsafe, (2) Why?, (3) If this was happening to you, what would you say?, and (4) what would you do? (Hazzard et al., 1991).

Results of the study indicated that children exposed to the intervention did demonstrate a better ability to differentiate between safe and unsafe situations compared to children in the control group. (Hazzard et al., 1991). However, there was no difference in preventative skill responses (e.g., say no, tell an adult, leave). This was largely because both intervention and control children gave approximately two of the three correct responses (Hazzard et al., 1991), suggesting a potential ceiling effect. Currently, we cannot be confident that this was a reliable or valid measure of children’s knowledge.

#### *\*Vignettes of Skills in Abusive Situations (Hébert et al., 2001)*

Hébert and colleagues (2001) developed five vignettes to assess children’s skills in preventing abuse in their evaluation of a child sexual abuse prevention program for elementary school children in Canada. Four abusive situations and one non-abusive situation were filmed. The

situations included bullying, a potentially abusive situation with a stranger, a potentially abusive situation with a known adult, a disclosure by a peer of sexual abuse by a known aggressor, and a non-abusive situation with a known adult (Hébert et al., 2001). Children were presented with the films in groups of four to five and then asked individually in a face-to-face interview to describe how they would react in the situation. A list of questions was created to standardise the process as much as possible. Children's ability to recognise the inappropriateness of the situation and display preventative behaviours was assessed. A coding form distinguished between categories of protective behaviours: assertiveness, disclosure, getting peer support, and self-defence (Hébert et al., 2001). These categories were confirmed through content analysis, although the details of this process were not reported. Inter-rater reliability (0.86-0.97) and test-retest reliability (0.74) were good, but the internal consistency was questionable (Cronbach's alpha = 0.69). There was preliminary experimental support for its construct validity, with improvements in protective behaviours observed for children exposed to the intervention (Hébert et al., 2001). There is promising support but further research would be needed to assess the validity of the measure.

#### *\*Self-Protective Behaviors Measure (Hillenbrand-Gunn et al., 2010)*

This 21-item questionnaire was designed by Hillenbrand-Gunn and colleagues (2010) for their U.S. study on a high school rape prevention intervention. It was designed to assess adolescents' willingness to engage in rape-preventative and pre-cautionary behaviours. The items were taken from a larger inventory created for female adolescent students for the ICASA project. The first 17 items asked participants are presented with three scenarios (e.g., "If your boyfriend forced you to have sex with him after you had too much to drink at a party, how likely would you be to...?"), and then asked to rate behavioural items on a scale from 1 (very likely) to 5 (not at all likely). For items 18 to 21, participants are asked to rate precautionary behaviours on a scale of 1 (never) to 5 (always). It was found to have good internal consistency (Cronbach's alpha = 0.83) in their sample, and there was preliminary experimental support for its construct validity. Students exposed to the intervention showed improvements in their willingness to engage in rape-preventative and self-protective behaviours (Hillenbrand-Gunn et al., 2010). Further research would be needed to ascertain the psychometric properties of this measure.

#### *\*Self-protective behaviours against CSA (Kang et al., 2022)*

Kang and colleagues (2022) revised a simulated abuse situation that had been used in a previous evaluation of a child sexual abuse prevention program (Kim & Kang, 2017). Three subcategories were included: "behaviour in safe/unsafe situations" (identify and avoid potentially abused situations), "coping strategies" (What to do), and "self-assertiveness" (take ownership of their own body). Multiple-choice questions were used to assess students' responses, with scores ranging from 0 to 10. A high score indicated higher competence in self-protective behaviour against CSA. There was preliminary experimental support for its construct validity, with students displaying significant differences in self-protective behaviours (Kang et al., 2022). The content validity index was reported as above 0.8, but no analysis was reported on the reliability of this measure. Therefore, further research would be needed.

#### *\*Self-protective skills (Krahé & Knappert, 2009)*

In Krahé & Knappert's (2009) evaluation of a theatre-based sexual abuse prevention program in Germany, eight scenarios were developed to assess gains in self-protective skills. The scenarios depicted children being uneasy or uncertain about an adult's behaviour and were accompanied



with a simple cartoon. Each scenario was followed by four questions designed to assess a child's ability to distinguish between 'good' and 'bad' secrets, being able to talk to someone, feeling confident to say no, distinguishing between 'good' touch and 'bad' touch, and not leaving with strangers (Krahé & Knappert, 2009). The internal consistency was questionable to adequate (Cronbach's alpha = 0.61-0.72), and there was preliminary experimental evidence for its construct validity as scores improved following the intervention. There is promising support, but further research would be needed to ascertain the psychometric properties of this measure.

#### *\*Simulation of a real-life scenario (Kraizer et al., 1988)*

In Kraizer and colleagues (1988) study of prevention of sexual abuse and abduction, they included a simulated situation in a school environment, where children had the opportunity to leave the school building with a stranger. A researcher who was unknown to the child would approach a child and ask for their assistance in bringing in some materials for a puppet show. If the child agreed, they were told that the stranger would come for them later. If the child refused, the stranger stated "Thanks anyway". Children were then debriefed by letting them know that the stranger had gone to the office for help as they should have. Kraizer and colleagues (1988) reported that this simulation "did not create anxiety or upset" for the children and that the simulations were "tightly controlled". This simulation was repeated in a later study (Kraizer, 1991) as well. However, there are several ethical considerations for a simulation of this nature. Similarly, it only assesses a child's response to a request for help from an unknown adult and may not necessarily reflect their actual level of protective skills or risk for abduction or abuse.

However, an important finding in this study was that improvements in knowledge, did not significantly predict success in the simulated situation. Therefore, children knowing the right answer on a paper-and-pencil based knowledge questionnaire did not necessarily indicate their ability to implement those strategies in a simulation (Kraizer et al., 1988). Past performance was also not necessarily indicative of future performance. Three children who passed the first simulation in the control group then failed the second simulation (i.e. agreed to follow the stranger). These results suggest that simulations may not be a reliable outcome measure of protective behaviour skills.

#### *\*Prevention of sexual abuse role play (Kraizer, 1991)*

A role play developed by Kraizer (1988) was used subsequently by the same author in their dissertation (1991). It was designed to assess a child's ability and willingness to terminate unwanted touch. For example, children are asked "what if your mom had a friend over and she was playing with your hair. Do you like people to play with your hair?". The researcher then plays with the child's hair and asks the child: "if you don't like it, what would you say and do?". The scenarios then escalate in level of intrusiveness (Kraizer, 1991). Scores were allocated by a trained observers and were based on a child's verbal response and body language. Points were also provided when children indicate that they will tell someone. There was some experimental support for the construct validity of the scores on this role play, with children exposed to the intervention having improved scores (Kraizer, 1991). However, overall scores were still relatively low after the intervention. The inter-rater reliability was 1.0 in previous publications (Fryer et al., 1987a, 1987b; Kraizer et al., 1988), providing support for the reliability of scores on this measure. More research would be needed to ascertain its psychometric properties. However, there are several ethical considerations needed for role-plays of this nature.



*\*The Choice of Safety Strategy Questionnaire (Taal & Edelaar, 1997)*

This questionnaire was designed by Taal and Edelaar (1997) in their evaluation of a child sexual abuse prevention program in the Netherlands. It contains 20 situations designed to assess how children might respond in a sexually abusive situation. Four strategies are included in this questionnaire, including problem-solving, contending, avoiding conflict, or yielding. It had acceptable internal consistency (Cronbach's alpha = 0.77). There were short-term effects on the choice of safety strategies for students exposed to the prevention program, providing preliminary experimental support for its construct validity (Taal & Edelaar, 1997). There is some support, but more research would be needed.

*\*The Touch Questionnaire (Taal & Edelaar, 1997)*

The Touch Questionnaire was developed by Taal and Edelaar (1997) and was designed to measure feelings of discomfort toward nonsexual touch in children. It was included in their evaluation of a child sexual abuse prevention program in the Netherlands to test for unintended impacts on children's comfort with nonsexual touch as a result of the intervention. It has 10 items, rated on a 5-point Likert scale from (1) not nice at all to (5) very nice. Scores are added to calculate a total "Touch" score ranging from 10-50. Higher scores indicate greater comfort toward nonsexual touch (Taal & Edelaar, 1997). The questionnaire had acceptable internal consistency (Cronbach's alpha = 0.77). There were some grade-level interaction effects whereby sixth-grade children had higher comfort levels with nonsexual touch, but the eighth-grade children indicated lower comfort with nonsexual touch immediately after the treatment ended. The Touch Questionnaire had a test-retest reliability of 0.50 in this sample, but no other analysis was reported on the psychometrics of this questionnaire (Taal & Edelaar, 1997). There is some preliminary support, but more research would be needed.

*\*Observed Protective Behaviors Test (OPBT; White et al., 2018)*

The OPBT was developed as part of an Australian randomised controlled trial (White et al., 2018). It was a two-part in-vivo situation where children were assessed on their ability to implement personal safety skills. It was based on past approaches to behavioural skills training (Gunby et al., 2010; Johnson et al., 2005; Johnson et al., 2006) and a secret-keeping scale (Dunkerley & Dalenberg, 2000). The first part of the simulation examines children's ability to recognise and respond to an unsafe situation (stranger simulation). Children are rated by a confederate on their verbal (agreed to leave, neutral, refused) and motor responses (approached, neutral, withdrew). The second part of the OPBT assesses whether children disclose the simulated unsafe situation. Children are given six opportunities or prompts to disclose, and the timing of the disclosure (or lack thereof) is rated. At the end of the interview, an anxiety measure is administered and the child is made aware of the simulation in a debrief, and the confederate (stranger) acknowledges that he was wrong to ask the child to keep a secret and apologises.

While the intervention did improve interpersonal safety knowledge and skills, it did not significantly impact children's disclosure intentions, safety identification skills, or interpersonal safety skills as measured by the OPBT (White et al., 2018). Therefore, we cannot be confident that the OPBT is a reliable or valid measure of protective behaviours in children.

## Child Wellbeing

There are a range of measures that have been used in experimental research on the prevention of institutional child maltreatment to assess child outcomes, including their physical and mental health, trauma symptoms, cognitive functioning, and emotional well-being. These measures have been used for multiple purposes, including screening, assessing impacts of trauma, physical health assessments, assessing potential negative impacts of the intervention (e.g., increased anxiety), and examining potential protective factors (e.g., self-esteem or self-efficacy). Many of these measures have been well established in their psychometric properties and these measures (or other similar ones) can be useful when providing a well-rounded assessment of child outcomes in experimental research on the prevention of child maltreatment. They are not necessarily specific to child maltreatment, as they may be used for many other purposes. For this reason, we have not provided an in-depth overview as we have for the other measures used in the EGM. We have included a list of these measures for fullness in [Appendix 5.1](#).

## Administrative Data

It is difficult to make determinations about the reliability or validity of administrative data, but it can be an effective way to collect data about children's outcomes.

### School Attendance (Baker-Henningham et al., 2021; Fabbri et al., 2021)

The frequency of school attendance was collected from school records and student-self-report.

### Headstart Family Information Systems (HSFIS) Enrolment Forms (Green et al., 2020)

These enrolment forms collect demographic information about children and families, including race/ethnicity, age of mother, mother's occupation and education, family status (housing, single-parent household, etc.), housing instability, family income, family substance abuse, and the total number of children in the home.

### Child Demographics (Webb et al., 1991)

Information about children's age, sex, ethnic background, previous education, and achievement level were collected.

## 2.6. Adult Perpetrators

Measures of adult perpetrators were aimed at understanding adult's self-reports of the perpetration of child maltreatment.

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#">Conflict Tactics Scale</a>	Straus (1979)		U.S.	Parents  (Child self-report version also available)	Threatened to hit or throw something at the other one.	Further research may be needed to assess its psychometric properties in

						other cultural settings.
<b><u>*Teachers Use of Violence</u></b>	Baker-Henningham et al., (2019)		Jamaica	Observer	<i>Observation of teachers</i>	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Parent report of spanking</u></b>	Green et al., (2020)		U.S.	Parent	Have you spanked your child in the past week?	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Perpetration of corporal punishment</u></b>	Jewkes et al., (2019)		South Africa	Teacher	In the past 6 months I beat a learner or sent one to another teacher for beating.	No analysis was reported on the psychometrics.
<b><u>*Teachers negative behaviour</u></b>	Jewkes et al., (2019)		South Africa	Teachers	Teachers often flirt with the learners at our school.	Good internal consistency but no other analysis was reported on the psychometrics.

### *Conflict Tactics Scale (CTS; Straus, 1979)*

The CTS contains 80 items designed to assess parent's use of violence, verbal aggression, and reasoning within the family. 20 items are administered to the parent about their relationship with their child. Then 40 items are aimed at understanding parent-child interactions, 20 about the parents' interactions with their child, and 20 about their perceptions of their partner's interactions. The last 40 items contain questions about the interactions between the parent and their partner (Straus, 1979).

The CTS was normed on a national probability sample of 2,143 couples in the U.S. The three factors were chosen based on the "catharsis theory" of violence control. It has adequate internal consistency and there is support for its construct validity (Straus, 1974, 1998).

Despite having been developed in the U.S., it has not been used in any experimental research on the prevention of institutional child maltreatment in the U.S. It has been used several times in experimental research in Africa, including Haiti, Ghana, Jamaica, Tanzania, and Uganda (Francis & Baker-Henningham, 2021; Hecker et al., 2021; Hecker et al., 2021; López García et al., 2022; Masath et al., 2020; Nkuba et al., 2018; Scharpf et al., 2021; Ssenyonga et al., 2022). In all instances, the CTS was modified or adapted for the purposes of the study.

López García and colleagues (2022) included the original CTS in their protocol for a randomised controlled trial of violence prevention in Haitian primary schools. The paper made mention of adapting the measure, but further detail was not included.

A modified and adapted version of the CTS was included in a randomised control trial of violence prevention programs in schools in Ghana, Tanzania, and Kenya (Masath et al., 2020; Scharpf et al., 2021; Ssenyonga et al., 2022). The items were adapted to include a child's self-report of their own experiences of violence (including physical assault, psychological aggression, non-violent discipline, and neglect). A teacher's self-report of their use of both physical and emotional violence, as well as their attitudes towards the use of violent discipline were also included. This teacher's modified version of the CTS had been used in previous studies in Tanzania and Uganda. There is support for its feasibility as a teachers' self-report measure of their use of violence in the classroom (Ssenyonga et al., 2022).

The CTS Parent-Child (Straus & Mattingly, 2007) was adapted for use in three studies (Hecker et al., 2021; Francis & Baker-Henningham, 2021; Nkuba et al., 2018). The original CTS Parent-Child has poor (physical violence subscale, Cronbach's alpha = 0.55) to questionable (emotional violence subscale, Cronbach's alpha = 0.60) internal consistency. This may be explained by the items being designed to measure rare events, reducing the correlation between the items (Nkuba et al., 2018).

Francis and Baker-Henningham (2021) made a slight modification for their randomised control trial in Jamaica by only asking parents about the last two weeks. The internal consistency was reported as questionable (Cronbach's alpha = 0.69) with good test-retest reliability (Intraclass correlation coefficient = 0.88). Whereas Hecker and colleagues (2021) adapted the CTS Parent-Child (Straus & Mattingly, 2007) for their cluster-randomised control trial in Tanzania by changing the time to the past month and the language to be suitable for caregivers working in orphanages. The internal consistency for this adaptation was reported as acceptable for both the caregiver's (Cronbach's alpha = 0.73) and children's (Cronbach's alpha = 0.77) report. Nkuba and colleagues (2018) used the physical and emotional violence subscales from the CTS Parent-Child in their cluster randomised control trial in Tanzania.

#### *\*Teachers Use of Violence (Baker-Henningham et al., 2019, 2021)*

Teachers were observed and a total sum of teachers' use of violent behaviour was calculated in a Jamaican cluster randomised controlled trial (Baker-Henningham et al., 2019, 2021). Teachers were observed over one school day and over five 20-minute observation periods on another school day. Incidents of physical punishment included counts of teachers hitting a student's hand, hitting students with an object, pushing, pulling, poking, or throwing an object, or making a child stand or kneel in uncomfortable positions. Verbal abuse included teachers calling a student a derogatory name, threatening a student, or frightening a student. Other types of abuse included intimidating a student or non-verbal threats (e.g., using a ruler to threaten a student with physical punishment).

There was no analysis reported on the reliability or validity of these observations. There was some preliminary experimental evidence for the construct validity of the observations, with reductions in the frequency of teachers’ use of violence (Baker-Henningham et al., 2019, 2021). However, without analysis on the psychometrics, we cannot be confident that this is a valid and reliable measure of teachers’ use of violence. Typically, psychometrically sound observations would include training for observers and checks for inter-rater reliability as a minimum.

*\*Parent report of spanking (Green et al., 2020)*

In Green and colleagues’ (2020) study on the effectiveness of Early Head Start on long-term reductions of child maltreatment, parent interviews were conducted at 24- and 36- months. In these interviews, parents were asked one question about whether they had spanked their child within the past week. While there was a reduction in parent reports of spanking, providing some preliminary experimental support for the construct validity (Green et al., 2020), there was no analysis reported on the reliability or validity of this approach.

*\*Perpetration of corporal punishment (Jewkes et al., 2019)*

Jewkes and colleagues (2019) included a yes/no question asking teachers if they had “beat a learner or sent one to another teacher for beating” in the past 6 months. No analysis was reported on the reliability or validity of this approach. There was also a non-significant difference in teachers’ use of corporal punishment (Jewkes et al., 2019). It is not clear if this was because the intervention did not significantly affect this outcome, or because this was not a valid or reliable measure of teachers’ use of corporal punishment.

*\*Teachers negative behaviour (Jewkes et al., 2019)*

Jewkes and colleagues (2019) included nine items designed to assess teachers’ negative behaviours at school. Items were rated on a 4-point Likert scale (strongly agree, agree, disagree, strongly disagree). The internal consistency of the items was good (Cronbach’s alpha = 0.8). There was a non-significant reduction in teachers’ reports of negative behaviour, so we cannot be certain that these results provide evidence for its construct validity. More research would be needed to ascertain this scale’s psychometric properties.

## 2.7. Child/Youth Offender Maltreatment Behaviours

Measures of child/youth maltreatment behaviours were aimed at understanding children’s self-reports of engaging in violent, abusive, or bullying behaviours.

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#"><u>Sexual Coercion Measure</u></a>	Cook-Craig et al., (2014)		U.S.	Child	Another student had sexual activities with you although you did not really want to because either the student	There is promising support, but more research is needed to determine the psychometric properties.

					threatened to end your friendship or romantic relationship if you didn't or you felt pressured by the student's constant arguments or begging.	
<b><u>*Sexual victimisation and perpetration</u></b>	Daigneault et al., (2015)		Canada	Adolescent	Have you had a sexual relationship, sexual behaviours, attempted to have a sexual relationship or made someone behave sexually while the other person did not want to?	No analysis was reported on the psychometrics.
<b><u>*Peer Perpetration Scale</u></b>	Karmaliani et al., (2020)		Pakistan	Child	How often within the past 4 weeks have you... called another child bad names?	Good internal consistency but no other analysis was reported on the psychometrics.
<b><u>*Bullying at school</u></b>	Jewkes et al., (2019)		South Africa	Child	I like to make fun of others at school, especially the weak ones.	Questionable internal consistency and no other analysis was reported on the psychometrics.
<b><u>*Teacher Report of Bullying at school</u></b>	Jewkes et al., (2019)		South Africa	Teachers	Learners may be bullied at school if they are thought to be gay or lesbian.	Good internal consistency but no other analysis was reported on the psychometrics.



### *Sexual Coercion Measure (Cook-Craig et al., 2014)*

Cook-Craig and colleagues (2014) developed a measure of victimisation and perpetration of sexual coercion. It included items from previously developed questionnaires, such as the Illinois Rape Myth Scale (Payne et al., 1999), the General Dating Violence Acceptance Scale (Forshee et al., 1998), and the Peer Support for Violence (DeKeseredy, 1990). There were also several researcher-developed items aimed at assessing bystander behaviours as well as victimisation and perpetration of stalking, dating violence, sexual harassment, sexual violence, and contraceptive interference. Each section on the measure had acceptable to good internal consistency (Cronbach's alpha = 0.70-0.86), and there was support for the construct validity of the scale with factor loadings ranging between 0.52 and 0.84 (Cook-Craig et al., 2014).

Variations of this measure have been used twice in experimental research on institutional child maltreatment. In Edwards and colleagues' (2022) evaluation of a self-defense program in U.S. high schools, three items were selected from the Sexual Coercion Measure. One item from the Youth Risk Behavior Survey (YRBS; Centers for Disease Control and Prevention, 2014) was also included in this research. Participants answered "yes" or "no" to indicate if each behaviour had occurred in the past six months. No analysis was reported on the psychometrics of this adapted version, but there was some preliminary experimental support for its construct validity. Following the intervention (IMpower), participants reported lower rates of sexual assault (Edwards et al., 2022).

The Sexual Coercion Measure was also used by Edwards and colleagues (2019) in their evaluation of a bystander prevention program (BITB-HSC) for U.S. high school students. Following exploratory factor analysis of the measure, some changes were made to the final measure used in the study. Participants reported on their experiences of sexual harassment and stalking victimisation and perpetration, and sexual assault and dating violence victimisation and perpetration, in the past 2 months and in the past year (Edwards et al., 2019). The details of the factor analysis were not reported, and it was not clear if there was any other analysis on the psychometric properties of the adapted version. There was some preliminary experimental support for its construct validity, with rates of sexual harassment and stalking perpetration in past 2 months significantly reduced between time collection one (pre-intervention) and three (an average of 98 days after time one). However, there were no other significant changes in reported victimisation or perpetration of sexual coercion (Edwards et al., 2019). It is not clear if this is because the intervention was not effective at addressing these issues or if the measure was not a valid or reliable measure of sexual coercion.

There is promising support for this measure, but more research would be needed.

### *\*Peer Perpetration Scale (Khuwaja et al., 2018)*

Khuwaja and colleagues (2018) developed a Peer Perpetration Scale that was modelled on a Peer Victimization Scale. The wording of the items was changed to ask students about the perpetration of the items (physical and verbal victimisation, social manipulation, property attacks). Participants are asked to indicate how many times they have done each of the 16 behaviours in the past four weeks (e.g., never, once, a few times (2-3), or many more times (4+)). It was reported as having good internal consistency (Cronbach's alpha = 0.86-0.89; Khuwaja et al., 2018).

Karmaliani et al., 2020 used this in their Pakistan cluster-randomised controlled trial. It was reported as having good internal consistency overall in their sample (Cronbach’s alpha = 0.89) and there was preliminary experimental support for its construct validity. Further research would be needed to assess the psychometric properties of this measure.

*\*Bullying at School (Jewkes et al., 2019)*

Jewkes and colleagues (2019) included three items designed to assess children’s self-report of bullying at school. Items were rated on a 4-point Likert scale (strongly agree, agree, disagree, strongly disagree). The internal consistency of the items was questionable (Cronbach’s alpha = 0.675-0.688). There was a non-significant reduction in self-reported bullying, so we cannot be certain that these results provide evidence for its construct validity. More research would be needed to ascertain this scale's psychometric properties.

*\*Teacher Report of Bullying at School (Jewkes et al., 2019)*

Jewkes and colleagues (2019) included eleven items designed to assess teachers’ reports of bullying at school. Items were rated on a 4-point Likert scale (strongly agree, agree, disagree, strongly disagree). The internal consistency of the items was good (Cronbach’s alpha = 0.88). There was a non-significant reduction in teachers’ reports of bullying, so we cannot be certain that these results provide evidence for its construct validity. More research would be needed to ascertain this scale's psychometric properties.

## 2.8. Parent/Caregiver

Measures under this construct were broadly aimed at understanding the knowledge and skills, attitudes, and well-being of parents. This might include parents’ knowledge of abuse, caregiving quality and attachment, and parenting stress.

### Parent Knowledge/Awareness and Skills

Measure	Author	Rating	Location	Who	Example Item	Considerations
<a href="#"><u>Parent Protective Behaviors Checklist (PPBC)</u></a>	Dale et al., (2016)		Australia	Parent	Not reported	Further research would be needed to assess its psychometric properties in other cultural settings.
<a href="#"><u>Alabama Parenting Questionnaire (APQ)</u></a>	Frick (1991)		U.S.	Parent	You get so busy that you forget where you child is and what he/she is doing.	Further research would be needed to assess its psychometric properties in other cultural settings.

<b><u>Knowledge of Infant Development Inventory (KIDI)</u></b>	MacPhee (1981)		U.S.	Parent	Do you agree or disagree: Two-year-olds are able to reason logically, much like an adult would.	Strong support when used with mothers. More research would be needed when used with professionals or fathers.
<b><u>Parent Trauma Response Questionnaire (PTRQ)</u></b>	Williamson et al., (2018)		U.K.	Parent	Our family will not get back to the way we were before the event happened.	Further research would be needed to assess its psychometric properties in other cultural settings.
<b><u>*Current caregiving quality</u></b>	Colich et al., (2021)		Romania	Observer	n/a	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Parent Involvement</u></b>	Francis & Baker-Henningham, (2021)		Jamaica	Parent	Spend 10–15 min or more doing something fun with the child	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Parent Supportiveness and Assistance</u></b>	Green et al., (2020)		U.S.	Observer	Not reported	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its construct validity.
<b><u>*Parent questionnaire</u></b>	Hazzard et al., (1991)		U.S.	Parents	Been more open about discussing feelings	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its

						construct validity.
<b><u>*Caregiver Communication, Knowledge, and Parenting Practices</u></b>	Jewkes et al., (2019)		South Africa	Child and caregiver	How often do you ask your Grade 8 child how was his or her day or how was school?	There is some support, but further research would be needed.
<b><u>*Parent literacy</u></b>	Karmaliani et al., (2020)		Pakistan	Child	Can your mother/father read and write?	No analysis reported on the psychometric properties.
<b><u>*Parent and teacher self-report</u></b>	Kolko et al., (1989)		U.S.	Parents and teachers	How aware are you of the problem of children being physically/sexually touched in ways that make them uncomfortable ?	There is promising support, but more research is needed to determine the psychometric properties.
<b><u>*Parents' and Teachers' Knowledge and Attitudes Questionnaire</u></b>	MacIntyre & Carr (1999)		Ireland	Parents and Teachers	If a girl is sexually assaulted she has always done something to provoke it.	Further research would be needed to assess its psychometric properties in other cultural settings.
<b><u>*Parenting and Child Sexuality Questionnaire</u></b>	McElearney et al., (2021)		Ireland	Parent	I feel that I have adequate knowledge to provide 'Keeping Safe' education to my children.	Further research would be needed to assess its psychometric properties in other cultural settings.
<b><u>*Caregiver attitudes toward physical discipline</u></b>	Merrill et al., (2018)		Uganda	Caregivers	Not reported	No analysis was reported on the psychometrics. Some preliminary experimental evidence on its

						construct validity.
<u>*Parent Perception Questionnaire (PPQ)</u>	Wurtele et al., (1992b)		U.S.	Parent	Difficulty separating from a parent.	No analysis reported on the psychometric properties.

### *Parent Protective Behaviors Checklist (PPBC; Dale et al., 2016)*

The PPBC was developed in an Australian study evaluating a school-based protection program for young children. It was designed to measure parent or caregiver perspectives about their child’s understanding and application of protective behaviour concepts (Dale et al., 2016). It includes 10 items where parents or caregivers record their perceptions about each statement on a scale from (1) not at all/unsure to (4) very/always. Scores are summed with higher scores indicating higher perceived knowledge and skills in their child.

The PPBC was reviewed by an expert panel of five developmental psychologists. It was found to have good internal consistency (Cronbach’s alpha = 0.84) and adequate test-retest reliability (ICC = .68). There was support for its construct validity through factor analysis and experimental evidence (Dale et al., 2016).

The PPBC was used by White and colleagues (2018) in their Australian-based randomised controlled trial. They reported good internal consistency for use on their sample of parents (Cronbach’s alpha = 0.823). No other analysis was reported on the psychometrics of the PPBC for their sample.

### *Alabama Parenting Questionnaire (APQ; Frick, 1991)*

The APQ is a 42-item questionnaire designed to measure five domains of etiology and treatment relating to parenting child externalising problems. These include; positive involvement with children, supervision and monitoring, use of positive discipline techniques, consistency in the use of positive discipline, and use of corporal punishment.

There is strong support for the psychometric properties of the APQ. It has good internal consistency and test-retest reliability, providing evidence for the reliability of the questionnaire (Shelton et al., 1996). The APQ has also been found to be significantly associated with other, well-established measures of parenting, providing support for its construct validity. Scores on the APQ also distinguish families of children with disruptive behaviour diagnoses from control families, providing support for the predictive validity of the APQ (Shelton et al., 1996).

The APQ has not yet been used in experimental research on the prevention of institutional child maltreatment. Wangamati and colleagues (2022) included an adapted version of the APQ in their protocol for a mixed methods pilot study on a whole-school approach to reduce violence in Kenyan and Tanzanian schools. In this protocol, it was not clear what adaptations were being made to the APQ.

### *Knowledge of Infant Development Inventory (KIDI; MacPhee, 1981)*

The KIDI was designed to measure parents' and caregivers' experiences and knowledge of parenting practices, norms and milestones for children from birth to 6 years of age, developmental processes, and health and safety guidelines. It measures a professional-based criterion of knowledge as the items were based on pediatricians' views on what all mothers should know (MacPhee, 1981). The KIDI contains 75 items and can be used as a screening tool to identify parents who might benefit from further support and intervention. It can also be used as an evaluation tool for parenting programs. Parents indicate if they agree, disagree or are not sure to each item (MacPhee, 1981).

The KIDI has good internal consistency when used with mothers (Cronbach's alpha = 0.82). However, it has poor internal consistency when used with child-care professionals (Cronbach's alpha = 0.50) and questionable internal consistency when used with college students (Cronbach's alpha = 0.67; MacPhee, 1981). Test-retest reliability is excellent when used with mothers ( $r = 0.91$ ). Construct validity of the KIDI has been supported by its positive relationship with a wide variety of indicators of knowledge of infants (Hamilton & Orme, 1990), such as the Knowledge of Child Development Inventory (Larsen & Juhasz, 1986) and the Parent Opinion Questionnaire (Azar et al., 1984).

It has only been used once in experimental research on the prevention of child maltreatment in institutional settings. In this study, families involved in the Early Head Start program had lower family conflict and more positive parent-child interactions, which later resulted in a reduced likelihood of children being involved in the child welfare system (Green et al., 2020).

### *Parent Trauma Response Questionnaire (PTRQ; Williamson et al., 2018)*

The PTRQ was designed to measure parental responses to childhood trauma. It has two subscales: parental appraisal and parental support. Appraisal items are rated on a 4-point Likert scale from (0) don't agree at all to (3) completely agree. Parent support items are rated on a 4-point Likert scale from (0) not at all to (3) a lot (Williamson et al., 2018).

An initial list of items was generated based on a review of the literature, cognitive and behavioural models of PTSD, measures of adult post-trauma cognitions, and existing measures of general parenting. This list of items was then reviewed by an expert panel and then subjected to statistical analysis. There was strong support for both its validity and reliability (Williamson et al., 2018). The internal consistency was adequate to good (Cronbach's alpha = 0.68-0.90). It had good test-retest reliability. Scores on the PTRQ were significantly associated with scores on the Parental Overprotection Scale (POS; Edwards et al., 2010), providing support for its construct validity.

The support style subscale was included in a protocol for a randomised controlled trial of a group intervention for young people in care with elevated posttraumatic stress symptoms (Hiller et al., 2021). For this study, the subscale will be completed by carers, and the wording will be changed to 'carers' to reflect this. This subscale includes 10 items that measure whether a carer has an avoidant (e.g., discouraging conversation) or approach (e.g., allowing the young person to talk openly) style.



### *\*Current caregiving quality (Colich et al., 2021)*

In Colich and colleagues (2021) evaluation of caregiving environments for Romanian children in the Bucharest Early Intervention Project (BEIP), caregivers were asked to rate the quality of the caregiving environment at 8, 12, and 16 years. Caregiving environments were rated as dangerous, unacceptable, marginal, mixed, or acceptable. The scoring had excellent inter-rater reliability (ICC = 0.93). There was some preliminary experimental support for its construct validity, with higher-quality caregiving environments being associated with lower internalising and externalising symptoms in adolescents (Colich et al., 2021).

### *\*Parent Involvement (Francis & Baker-Henningham, 2021)*

A 12-item questionnaire designed to assess parents' level of involvement was included in Francis and Baker-Henningham's (2021) cluster randomised controlled trial in Jamaica. Parents are asked to report how often they engage in each of the behaviours on a 6-point scale from (0) none to (6) more than once a day. It was based on a shorter version used by the authors in a previous study (Baker-Henningham & Francis, 2018). It was reported as having adequate internal consistency (Cronbach's alpha = 0.71) and good test-retest reliability (ICC = 0.96). There was preliminary experimental support for its construct validity, with parental involvement improving after the intervention. There is promising support but further research is needed.

### *\*Parent Supportiveness and Assistance (Green et al., 2020)*

In Green and colleagues' (2020) evaluation of the Early Head Start program, parents were videotaped doing semi-structured play and puzzle tasks with their children. Parents were then rated by an observer on the level of supportiveness (sensitivity to child, cognitive stimulation, and positive emotional regard) and the quality of assistance provided. Strict protocols for coding were mentioned, but details were not reported. There was preliminary experimental evidence for the construct validity, but no analysis was reported on the psychometric properties of these observations or scoring practices used.

### *\*Parent questionnaire (Hazzard et al., 1991)*

In Hazzard and colleagues' (1991) evaluation of a child sexual abuse prevention program, parents completed a 16-item checklist on their child's behaviours and were asked about their perceptions of the overall impact of the program. The items were adapted from previous research (Miller-Perrin & Wurtele, 1986), but no other details on the survey, the adaptations made, or analysis of the psychometrics were reported. There was some preliminary experimental evidence for its construct validity, with some changes in behaviour observed. For example, parents of children who attended the program reported that their children talked more about fires and fire safety, were more open to discussing feelings, and asked more questions about sexual abuse and sex (Hazzard et al., 1991).

### *\*Caregiver Communication, Knowledge, and Parenting Practices (Jewkes et al., 2019)*

In Jewkes and colleagues' (2019) randomised controlled trial in South Africa, five items were included that were designed to assess caregiver communication with their child. Both children and caregivers were asked about their level of communication. The items had questionable internal consistency for the children's report (Cronbach's alpha = 0.65-0.68) and adequate to good internal consistency for the caregiver's report (Cronbach's alpha = 0.77-0.83).

Thirteen items were also included that were designed to assess a caregivers' knowledge of the child. These items had good internal consistency (Cronbach's alpha = 0.85-0.87).

Caregivers also completed 29 items on positive parenting practices, such as involvement and monitoring at home, and seven items on negative parenting practices. These items were based on previous research (Rickel & Biasseti, 1982). The positive items had excellent internal consistency (Cronbach's alpha = 0.91-0.92) and the negative items had acceptable internal consistency (Cronbach's alpha = 0.71-0.73); Jewkes et al., 2019).

There was some preliminary experimental support for the construct validity of these items, with communication and knowledge scores improving after the intervention, and a reduction in negative parenting practices (Jewkes et al., 2019). There was no change in positive parenting practices and there was no other analysis reported on the psychometric properties of these measures.

#### *\*Parent literacy (Karmaliani et al., 2020)*

In Karmaliani and colleagues' (2020) Pakistani study, two items were included designed to assess children's reports of their parent's literacy skills. No analysis was reported on the psychometric properties of these items.

#### *\*Parent and teacher self-report (Kolko et al., 1989)*

In Kolko and colleagues' (1989) evaluation of a sexual victimisation awareness and prevention skills program, parents and teachers completed a 10-item questionnaire designed to assess their knowledge, awareness, and preparedness to act in response to sexual abuse. Each item was rated on a 5-point Likert scale. The parent and teacher versions were identical, but teachers were asked to report on children in their classroom, instead of on their own children. Factor analysis provided support for the three-factor structure, and there was preliminary experimental support for its construct validity. The questionnaire had questionable to acceptable internal consistency (Cronbach's alpha = 0.66-0.72). There is promising support, but further research would be needed.

#### *\*Parents' and Teachers' Knowledge and Attitudes Questionnaire (MacIntyre & Carr, 1999)*

This 38-item questionnaire was developed in an Irish study to assess parents' and teachers' knowledge and attitudes (MacIntyre & Carr, 1999). Parents and teachers were asked about the prevalence, indicators, and causekks of child sexual abuse, characteristics of victims and offenders, and service-related information. Each item was rated on a 5-point Likert scale from "strongly agree" to "strongly disagree". The internal consistency of the measure was good (Cronbach's alpha = 0.80; MacIntyre & Carr, 1999). This study found that following the intervention, teachers' and parents' knowledge and attitudes had improved, providing experimental support for its construct validity. Factor analysis also revealed that the questionnaire was unidimensional (MacIntyre & Carr, 1999). It has not been used in any subsequent experimental research on institutional child maltreatment.

#### *\*Parenting and Child Sexuality Questionnaire (McElearney et al., 2021)*

McElearney and colleagues (2021) adapted two scales from Morawska and colleagues' (2015) Parenting and Child Sexuality Questionnaire, for their Irish experimental study. It was designed to assess parents' self-report of their knowledge and confidence in communicating about keeping safe measures with their children. The original version was designed in Australia.

Six items specific to parent knowledge, and eleven items for parent confidence were used. Four items from Walsh and colleagues’ (2012a) Current and Future Confidence scale were added as well. This scale was found to have adequate internal consistency for the knowledge items (Cronbach’s alpha = 0.77) and excellent internal consistency for confidence items (Cronbach’s alpha = 0.95; Morawska et al., 2015).

Confirmatory Factor Analysis provided support for the factor structure of the adapted questionnaire, and there was experimental support for its construct validity. Parents displayed improvements in their knowledge and confidence in communicating with their children about keeping safe measures (McElearney et al., 2021).

*\*Caregiver attitudes toward physical discipline (Merrill et al., 2018)*

In Merrill and colleagues (2018) cluster randomised-controlled trial in Uganda, caregiver attitudes toward violence were assessed. There were four items on beliefs about physical discipline practices at school and four items on beliefs about physical discipline practices at home. No analysis was reported on the psychometric properties of these items. There was preliminary experimental evidence for its construct validity, but further research would be needed.

*\*Parent Perception Questionnaire (PPQ; Wurtele et al., 1992b)*

The PPQ was developed by Wurtele and colleagues (1992b) in their study on the effects of child sexual abuse prevention program for preschool students. It was designed to assess for parents perceptions of both negative and positive effects of an intervention program. It includes items relating to positive behaviours (e.g., physically affectionate), negative behaviours (e.g., refuses to obey), if they have engaged in any other prevention programs, how fearful their children are of nine people and situations, and their overall assessment of the effect of the program on their child. It contains 17 behaviours and parents are asked to indicate if any of the behaviours increased after the intervention, and if the increase was a problem. There is also a teacher version (TPQ). The two scales are identical in format and content, aside from the PPQ also including a question on “difficulty separating from parent”. There has been no known analysis on the psychometric properties of the PPQ.

The PPQ was used by Grendel (1991) in their dissertation. An adapted version was then later used by Hébert and colleagues (2001). Full details of the adapted version were not reported in their paper, but parents were asked to comment on positive and negative behaviour changes. No analysis was reported on the reliability or validity of the adaptation.

**Parent Wellbeing**

Measure	Author	Rating	Location	Who	Example Item	Considerations
<b><u>Parenting Stress Index-Short Form (PSI-SF)</u></b>	Abidin (1995)		U.S.	Parent	Never able to do things that I like to do.	Further research would be needed to assess its psychometric properties in

						other cultural settings.
<b><u>Home Observation for Measurement of the Environment (HOME)</u></b>	Bradley & Caldwell (1988)		U.S.	Observer	Child's play environment appears safe and free of hazards.	Requires a trained professional to observe. The presence of an observer may distort behaviour in families. Further research would be needed to assess its psychometric properties in other cultural settings.
<b><u>Family Environment Scale (FES)</u></b>	Moos et al., (1974)		U.S.	Parent and child	We fight a lot in our family	Further research would be needed to assess its psychometric properties in other cultural settings.
<b><u>*Stress score</u></b>	Jewkes et al., (2019)		South Africa	Caregivers	I have felt a sense of hopelessness about my child or children	Good internal consistency but no other analysis was reported on the psychometrics.

### *Parenting Stress Index-Short Form (PSI-SF; Abidin, 1995)*

The PSI-SF is a 36-item self-report measure of three subscales relating to parenting stress: parental distress (PD), parent-child dysfunctional interaction (PCDI), and difficult child (DC). It is designed for parents with children from 0-12 years of age and is an abbreviated version of the full-scale version (PSI-4).

There is strong support for the PSI-SF. It has adequate to excellent internal consistency (Cronbach's alpha = 0.78-0.90; Roggman et al., 1994). There are inconsistent findings on the underlying factor structure, but there seems to be some support for the underlying three-factor structure (Díaz-Herrero et al., 2011; Reitman et al., 2002). Scores on the PSI-SF are associated with scores on the Goldberg General Health Questionnaire, providing support for its construct validity (Aracena et al., 2016).

The Parent Distress (PD) and Parent-Child Dysfunctional Interaction (PCDI) subscales were used in Green and colleagues' (2020) evaluation of the Early Head Start program in the U.S. on reducing child maltreatment. No analysis of the reliability of these subscales on this sample was reported in this study. Families involved in the intervention in this study had lower parenting distress scores and more positive child interactions, which led to later reductions in their children being involved in the U.S. child welfare system (Green et al., 2020). These results provide support experimental for its construct validity.

### *Home Observation for Measurement of the Environment (HOME; Bradley & Caldwell, 1988)*

The HOME includes an interview, multi-response maternal report, and observer ratings aimed at assessing a child's home environment. This includes the emotional support provided by the child's family and the quality of cognitive stimulation (Bradley & Caldwell, 1988). There are versions available depending on the age of the child (under 3, 3-5 years, 6-9 years, and 10-15 years).

The HOME has excellent internal consistency and inter-rater reliability, providing support for its reliability (Elardo & Bradley, 1981). Test-retest reliability is adequate ( $r = 0.62$ ; Caldwell & Bradley, 1979), indicating the the scale is sensitive to changes in family dynamics, providing support for its psychometric properties. Scores on the HOME correlate with maternal and paternal education, paternal presence and occupation, socioeconomic status, as well as child outcomes (such as IQ, success in school, language development), providing support for the construct and criterion validity of the HOME (Elardo & Bradley, 1981).

Despite being a widely used measure of parent support for learning and development, the HOME was only used in one study in the EGM (Green et al., 2020). In this evaluation of Early Head Start in the U.S., the HOME was collected at 15, 24, and 36 months of age. No analysis was reported on the reliability or validity of the HOME in this study, but there was experimental support for its construct validity (Green et al., 2020).

### *Family Environment Scale (FES; Moos et al., 1974)*

The FES was designed to assess three conceptual domains of family interactions: relationship, personal growth or goal orientation, and system maintenance. It contains 90 true or false items, and there are parent and child reports available (Moos et al., 1974).

The FES has questionable to acceptable internal consistency (Cronbach's alpha = 0.61-0.78) and adequate test-retest reliability (0.68-0.86; Moos & Moos, 2002). Factor analysis on the FES has revealed three-factor solutions, consistent with the design of the scale (Lanz & Maino, 2014). The construct validity of the FES was established through significant associations with scores on the Family Adaptability and Cohesion Evaluation Scales (FACES-II; Dickerson & Coyne, 1987), Structural Family Interaction Scale – Revised (Perosa & Perosa, 1990), and Parental Bonding Instrument (Parker et al., 1979).

The Family Conflict subscale was used in one U.S. experimental study on the prevention of institutional child maltreatment (Green et al., 2020). The authors did not report any analysis on the psychometrics of this subscale on their population.

### *\*Stress Score (Jewkes et al., 2019)*

An adapted version of the Compassion Fatigue Scale-Revised (Gentry et al., 2002) was used once in a randomised controlled trial in South Africa (Jewkes et al., 2019). Changes were made for it to be a child-raising fatigue scale. Overall internal consistency was good for male caregivers (Cronbach's alpha = 0.87) and excellent for female caregivers (Cronbach's alpha = 0.90). More research would be needed to assess its psychometric properties when used as a child-raising fatigue scale.

## **2.9. Intervention Feasibility and Implementation**

In some included studies in the EGM, assessments were used to test for the feasibility and/or implementation of the intervention. These included assessments such as the Abbreviated Acceptability Rating Profile (AARP) or the Feasibility Questionnaire (used in Kim et al., 2019 and Taal & Edelaar, 1997, respectively), as well as numerous researcher-developed questionnaires created for the purposes of the study. As these assessments are more focused on the fidelity of the intervention in experimental research, they have not been reviewed here. Some researchers may find using measures like these helpful to ensure the reliability of the implementation of the intervention in different settings. This might be particularly relevant in cases where the implementation of the intervention is carried out by a third party or participants (such as teachers) and cannot be closely monitored by the research team. Some studies also asked participants to comment on their satisfaction with the intervention. See [Appendix 5.2](#) for a list of studies that used intervention feasibility or implementation measures.

## **3. Summary**

If we want to take seriously the prevention of child maltreatment in institutional settings, then we need to do our due diligence in choosing sound outcomes in our research designs. It is generally not best practice for researchers to develop their own measure without doing (or reporting) the required processes to assess the psychometric properties of the measure. Often, time and money are barriers to effectively accessing, using, adapting, modifying, or translating established measures with strong support. Measures are also often created with Western, English-speaking populations in mind, so generally, more resources are needed to adapt and translate measures for use in other cultural contexts. We need to prioritise the efforts and resources needed to use measures that are valid and reliable. When we do this, we can then be more confident that our interventions and prevention programs are effective at protecting children and young people.



## 4. References

- AAUW. (1993). Hostile hallways: The AAUW survey on sexual harassment in America's schools. *Journal of School Health, 63*(8), 355-358.
- AAUW. (2001). *Hostile hallways: Bullying, teasing, and sexual harassment in school*.  
<https://www.aauw.org/files/2013/02/hostile-hallways-bullying-teasing-and-sexual-harassment-in-school.pdf>
- Abidin, R. R. (1995). *Parenting Stress Index: Professional Manual [PSI]*. PAR, Psychological Assessment Resources.
- ACEs Aware. ACEs screening clinical workflows. n.d. Available from:  
<https://www.acesaware.org/wp-content/uploads/2019/12/ACE-Clinical-Workflows-Algorithms-and-ACE-Associated-Health-Conditions.pdf>.
- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, MI: Erlbaum.
- Aldridge J.M., Laugksch R.C. & Fraser B.J. (2006) School-level environment and outcomes-based education in South Africa. *Learning Environments Research, 9*, 123–47.
- Anderson-Varney, T. (1987). *The evaluation of a personal safety curriculum*. Unpublished master's thesis, Michigan State University, East Lansing, MI.
- Andresen, S., Gade, J. D., & Grunewalt, K. (2015). Prävention sexueller Gewalt in der Grundschule: Erfahrungen, Überzeugungen und Wirkungen aus Sicht von Kindern, Eltern, Lehr- und Fachkräften. Beltz Juventa.
- Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 5*(4), 237-249.
- Antônio T, Hokoda A. (2009). Gender variations in dating violence and positive conflict resolution among Mexican adolescents. *Violence Vict, 24*(4), 533-545.
- Aracena, M., Gómez, E., Undurraga, C., Leiva, L., Marinkovic, K., & Molina, Y. (2016). Validity and reliability of the Parenting Stress Index Short Form (PSI-SF) applied to a Chilean sample. *Journal of Child and Family Studies, 25*, 3554-3564.
- Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics, 131*(3), 1415-1453.
- Australian Human Rights Commission. (2022). *Time for Respect: Fifth National Survey on Sexual Harassment in Australian Workplaces*. Australian Human Rights Commission.
- Ayers, T. S., & Sandler, I. N. (1999). *Manual for the children's coping strategies checklist & the how I coped under pressure scale*. Arizona State University; Universiteit van Amsterdam.

- Azar, S. T., Robinson, D. R., Hekimian, E., & Twentyman, C. T. (1984). Unrealistic expectations and problem-solving ability in maltreating and comparison mothers. *Journal of Consulting and Clinical Psychology, 52*(4), 687.
- Baker, C. K., Gleason, K., Naai, R., Mitchell, J., & Trecker, C. (2013). Increasing knowledge of sexual abuse: A study with elementary school children in Hawai'i. *Research on Social Work Practice, 23*(2), 167-178.
- Baker-Henningham, H., & Francis, T. (2018). Parents' use of harsh punishment and young children's behaviour and achievement: a longitudinal study of Jamaican children with conduct problems. *Global Mental Health, 5*, e32.
- Baker-Henningham, H., Scott, Y., Bowers, M., & Francis, T. (2019). Evaluation of a violence-prevention programme with Jamaican primary school teachers: A cluster randomised trial. *International Journal of Environmental Research and Public Health, 16*(15), 2797.
- Baker-Henningham, H., Bowers, M., Francis, T., Vera-Hernández, M., & Walker, S. P. (2021a). The Irie Classroom Toolbox, a universal violence-prevention teacher-training programme, in Jamaican preschools: A single-blind, cluster-randomised controlled trial. *Lancet Global Health, 9*(4), e456-e468.
- Baker-Henningham, H., Scott, Y., Francis, T., & Walker, S. P. (2021b). Effects of a teacher-training violence prevention program in Jamaican preschools on child behaviours, academic achievement, and school attendance in grade one of primary school: Follow up of a cluster randomised trial. *Frontiers in Psychology, 12*, 652050
- Bandura, A. (n.d.). *Teacher Self-Efficacy Scale*. Available online: <http://u.osu.edu/hoy.17/files/2014/09/BanduraInstr-1sdm5sg.pdf>
- Banyard V. L. (2015). *Toward the next generation of bystander prevention of sexual and relationship violence: Action coils to engage communities* (1st ed.). Springer International Publishing.
- Banyard, V. L., Eckstein, R. P., & Moynihan, M. M. (2010). Sexual violence prevention: The role of stages of change. *Journal of Interpersonal Violence, 25*(1), 111-135.
- Banyard, V. L., Moynihan, M. M., & Plante, E. G. (2007). Sexual violence prevention through bystander education: An experimental evaluation. *Journal of Community Psychology, 35*, 463–481.
- Banyard, V. L., Moynihan, M. M., Cares, A., & Warner, R. (2014). How do we know if it works? Measuring outcomes in bystander-focused violence prevention on campuses. *Psychology of Violence, 4*, 101– 115.
- Barron, I. G., & Topping, K. J. (2013). Prevention programs and legal considerations for child sexual abuse cases: Exploratory Evaluation of a School-Based Child Sexual Abuse Prevention Program. *Journal of Child Sexual Abuse, 22*(8), 931-948.
- Battle, J. (1981). *Manual for the Culture-Free Self-Esteem Inventory for Children*. Seattle, WA: Special Child Publications.

- Bayley, N. (1993). *Bayley Scales of Infant Development* (2<sup>nd</sup> ed.). New York: Psychological Corporation.
- Bear, G. G., Gaskins, C., Blank, J., & Chen, F. F. (2011). Delaware School Climate Survey-Student: its factor structure, concurrent validity, and reliability. *Journal of School Psychology, 49*(2), 157–174.
- Bear, G. G., Yang, C., Pell, M., & Gaskins, C. (2014). Validation of a brief measure of teachers' perceptions of school climate: Relations to student achievement and suspensions. *Learning Environment Research, 17*, 339–354.
- Beckett, R. C., & Fisher, D. (1994). *Assessing victim empathy: A new measure. Paper presented at the 13th Annual Conference of The Association for the Treatment of Sexual Abusers*. San Francisco, CA.
- Berden, G. F., Althaus, M., & Verhulst, F. C. (1990). Major life events and changes in the behavioural functioning of children. *Journal of Child Psychology and Psychiatry, 31*(6), 949-959.
- Bick, J., Zhu, T., Stamoulis, C., Fox, N. A., Zeanah, C., & Nelson, C. A. (2015). Effect of early institutionalization and foster care on long-term white matter development: a randomized clinical trial. *JAMA pediatrics, 169*(3), 211-219.
- Blumberg, E. J., Chadwick, M. W., Fogarty, L. A., Speth, T. W., & Chadwick, D. L. (1991). The touch discrimination component of sexual abuse prevention training: Unanticipated positive consequences. *Journal of Interpersonal Violence, 6*(1), 12-28.
- Bos A. E. R., Dijker A. J. M., Koomen W. (2007). Sex differences in emotional and behavioral responses to HIV + individuals' expression of distress. *Psychology & Health, 22*, 493-511.
- Bosworth, K., & Espelage, D. (1995). *Teen conflict survey*. Bloomington, IN: Center for Adolescent Studies, Indiana University.
- Bradley, R. H., & Caldwell, B. M. (1988). Using the home inventory to assess the family environment. *Pediatric nursing, 14*(2), 97-102.
- Bradley, B.P., Mogg, K., White, J., Groom, C., & De Bono, J. (1999). Attentional bias for emotional faces in generalized anxiety disorder. *British Journal of Clinical Psychology, 38*(3), 267–278.
- Bright, M. A., Huq, M. S., Patel, S., Miller, M. D., & Finkelhor, D. (2022). Child safety matters: Randomised control trial of a school-based, child victimization prevention curriculum. *Journal of Interpersonal Violence, 37*(1-2), 538-556.
- Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of personality and social psychology, 38*(2), 217.
- Bustamante, G., Andrade, M. S., Mikesell, C., Cullen, C., Endara, P., Burneo, V., ... & Grunauer, M. (2019). "I have the right to feel safe": Evaluation of a school-based child sexual abuse prevention program in Ecuador. *Child Abuse & Neglect, 91*, 31-40.

- Cadima, J., Peixoto, C., & Leal, T. (2014). Observed classroom quality in first grade: Associations with teacher, classroom, and school characteristics. *European Journal of Psychology of Education, 29*, 139-158.
- Cairns, R. B., Leung, M. C., Buchanan, L., & Cairns, B. D. (1995). Friendships and social networks in childhood and adolescence: Fluidity, reliability, and interrelations. *Child Development, 66*(5), 1330-1345.
- Caldwell, B. M., & Bradley, R. H. (1979). *Home observation for measurement of the environment*. Little Rock, AR: University of Arkansas at Little Rock.
- Cecen-Erogul, A. R., & Kaf Hasirci, O. (2013). The Effectiveness of Psycho-educational School-based Child Sexual Abuse Prevention Training Program on Turkish Elementary Students. *Educational Sciences: Theory and Practice, 13*(2), 725-729.
- Centers for Disease Control and Prevention. (2014). *Youth risk behavior surveillance: United States, 2013*. (Surveillance Summaries, Issue).
- Cerezo, M. A., & Pons-Salvador, G. (2004). Improving child maltreatment detection systems: a large-scale case study involving health, social services, and school professionals. *Child Abuse & Neglect, 28*(11), 1153-1169.
- Chamberland, A. (2003). Évaluation des effets du volet information et sensibilisation du programme jeunes en action contre la violence sexuelle (J'AVISE). Québec:(M.S.C.), Université Laval
- Chen, Y. C., Fortson, B. L., & Tseng, K. W. (2012). Pilot evaluation of a sexual abuse prevention program for Taiwanese children. *Journal of Child Sexual Abuse, 21*(6), 621-645.
- Citak Tunc, G., Gorak, G., Ozyazicioglu, N., Ak, B., Isil, O., & Vural, P. (2018). Preventing child sexual abuse: Body safety training for young children in Turkey. *Journal of Child Sexual Abuse, 27*(4), 347-364.
- Coddington, R. D. (1972). The significance of life events as etiologic factors in the diseases of children—II a study of a normal population. *J. Psychosom. Res, 16*, 205–213.
- Cohen, J. N., Byers, E. S., Sears, H. A. & Weaver, A. D. (2004) Sexual Health Education: Attitudes, Knowledge, and Comfort of Teachers in New Brunswick Schools. *The Canadian Journal of Human Sexuality, 13*(1): 65–74.
- Cohen-Kettenis, P. T., & Dekking, Y. M. (1980). *Cognitive aspects of children's social fear. A psychometric study*. Lisse, The Netherlands: Swets & Zeitlinger.
- Coker A. L., Cook-Craig P. G., Williams C. M., Fisher B. S., Clear E. R., Garcia L. S., & Hegge L. M. (2011). Evaluation of Green Dot: An active bystander intervention to reduce sexual violence on college campuses. *Violence Against Women, 17*(6), 777–796.
- Colich, N. L., Sheridan, M. A., Humphreys, K. L., Wade, M., Tibu, F., Nelson, C. A., Zeanah, C. H., Fox, N. A., & McLaughlin, K. A. (2021). Heightened sensitivity to the caregiving environment during adolescence: Implications for recovery following early-life adversity. *Journal of Child Psychology and Psychiatry, 62*(8).

- Collings, S. J. (1997). Development, reliability, and validity of the child sexual abuse myth scale. *Journal of Interpersonal Violence, 12*(5), 665-674.
- Collings, S. J. (2003). Child sexual abuse myth acceptance among aspirant, trainee, and registered psychologists in Durban, South Africa. *Social Behavior and Personality: An International Journal, 31*(8), 835-842.
- Conte, J. R., Rosen, C., Saperstein, L., & Shermack, R. (1985). An evaluation of a program to prevent the sexual victimization of young children. *Child Abuse & Neglect, 9*(3), 319-328.
- Cook-Craig P. G., Coker A. L., Clear E. R., Garcia L. S., Bush H. M., Brancato C. J., Williams C. M., & Fisher B. S. (2014). Challenge and opportunity in evaluating a diffusion-based active bystander prevention program green dot in high schools. *Violence Against Women, 20*(10), 1179–1202.
- Crenshaw W. B., Crenshaw L. M., Lichtenberg J. W. (1995). When educators confront child abuse: An analysis of the decision to report. *Child Abuse & Neglect, 19*(9), 1095–113.
- Crittenden, P. M. (1992). Quality of attachment in the preschool years. *Development and Psychopathology, 4*, 209–241.
- Crowley, K. J. (1989). *Evaluation of good touches/bad touches: A program to prevent child sexual abuse in school-age children*. Pace University.
- Czerwinski, F., Finne, E., Alfes, J., & Kolip, P. (2018). Effectiveness of a school-based intervention to prevent child sexual abuse—Evaluation of the German IGEL program. *Child Abuse & Neglect, 86*, 109-122.
- Dahlberg, L. L., Toal, S. B., Swahn, M., Behrens, C. B. (2005). *Measuring Violence-Related Attitudes, Behaviors, and Influences Among Youths: A Compendium of Assessment Tools*, 2nd ed., Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.
- Daigneault, I., Hébert, M., McDuff, P., & Frappier, J. Y. (2012). Evaluation of a sexual abuse prevention workshop in a multicultural, impoverished urban area. *Journal of Child Sexual Abuse, 21*(5), 521-542.
- Daigneault, I., Hébert, M., McDuff, P., Michaud, F., Vezina-Gagnon, P., Henry, A., & Porter-Vignola, E. (2015). Effectiveness of a sexual assault awareness and prevention workshop for youth: A 3-month follow-up pragmatic cluster randomization study. *The Canadian Journal of Human Sexuality, 24*(1), 19-30.
- Dake, J. A., Price, J. H., & Murnan, J. (2003). Evaluation of a child abuse prevention curriculum for third-grade students: assessment of knowledge and efficacy expectations. *Journal of School Health, 73*(2), 76-82.
- Dale R., Shanley D. C., Zimmer-Gembeck M. J., Lines K., Pickering K., White C. (2016). Empowering and protecting children by enhancing knowledge, skills and well-being: A randomized trial of Learn to BE SAFE with Emmy™. *Child Abuse & Neglect, 51*, 368–378.



- Danzer, V. A., Gerber, M. F., Lyons, T. M., & Voress, J. K. (1991). *DABERON-2: Screening for school readiness*. Pro-Ed.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.
- Decker M. R., Wood S. N., Ndinda E., Yenokyan G., Sinclair J., Maksud N., Ross B., Omondi B., & Ndirangu M. (2018). Sexual violence among adolescent girls and young women in Malawi: A cluster-randomized controlled implementation trial of empowerment self-defense training. *BMC Public Health, 18*(1), 1–12.
- De Boo G. M. (2005). *Coping Strategieën Lijst voor Kinderen* (Children's Coping Strategies Checklist). Universiteit van Amsterdam.
- DeKeseredy W. S. (1990). Male peer support and women abuse: The current state of knowledge. *Sociological Focus, 23*, 129-139.
- Dekking, Y. M. (1977). Voorlopige handleiding bij de sociale angst-schaal voor kinderen: S.A.S.–K [Provisional manual for the social anxiety scale for children: S.A.S.–K]. Swets & Zeitlinger.
- del Campo Sánchez, A., & Sánchez, F. L. (2006). Evaluación de un programa de prevención de abusos sexuales a menores en Educación Primaria. *Psicothema, 18*(1), 1-8.
- Devries, K. M., Knight, L., Allen, E., Parkes, J., Kyegombe, N., & Naker, D. (2017). Does the Good Schools Toolkit reduce physical, sexual and emotional violence, and injuries, in girls and boys equally? A cluster-randomised controlled trial. *Prevention Science, 18*, 839-853.
- Devries, K. M., Knight, L., Child, J. C., Mirembe, A., Nakuti, J., Jones, R., ... & Naker, D. (2015). The Good School Toolkit for reducing physical violence from school staff to primary school students: a cluster-randomised controlled trial in Uganda. *The Lancet Global Health, 3*(7), e378-e386.
- Devries, K., Kuper, H., Knight, L., Allen, E., Kyegombe, N., Banks, L. M., ... & Naker, D. (2018). Reducing physical violence toward primary school students with disabilities. *Journal of Adolescent Health, 62*(3), 303-310.
- DeWalt, D. A., Thissen, D., Stucky, B. D., Langer, M. M., Morgan DeWitt, E., Irwin, D. E., ... & Varni, J. W. (2013). PROMIS Pediatric Peer Relationships Scale: development of a peer relationships item bank as part of social health measurement. *Health Psychology, 32*(10), 1093.
- Dhooper, S. S., & Schneider, P. L. (1995). Evaluation of a school-based child abuse prevention program. *Research on Social Work Practice, 5*(1), 36-46.
- Díaz-Herrero, Á., López-Pina, J. A., Pérez-López, J., de la Nuez, A. G. B., & Martínez-Fuentes, M. T. (2011). Validity of the Parenting Stress Index-Short Form in a sample of Spanish fathers. *The Spanish Journal of Psychology, 14*(2), 990-997.
- Diaz, M. J., Moreland, D., & Wolfersteig, W. (2021). Assessing the effects of Childhelp's Speak Up be Safe child abuse prevention curriculum for high school students. *Journal of Child & Adolescent Trauma, 14*(3), 425-432.



- Dickerson, V. C., & Coyne, J. C. (1987). Family cohesion and control: A multitrait-multimethod study. *Journal of Marital and Family Therapy, 13*(3), 275-285.
- Dilsiz, H. & Mağden, D. (2015). Determining teachers' knowledge and risk recognition levels on child abuse and neglect. *Hacettepe University Faculty of Health Sciences Journal, 1*(2), 678-693.
- Döpfner, M., Görtz-Dorten, A., & Lehmkuhl, G. (2008). DISYPS-II Diagnostik-System für psychische Störungen nach ICD-10 und DMS-IV für Kinder und Jugendliche – II. Bern: Huber.
- Dryden, E. M., Desmarais, J., & Arsenault, L. (2014). Effectiveness of the IMPACT: Ability program to improve safety and self-advocacy skills in high school students with disabilities. *Journal of School Health, 84*(12), 793-801.
- Dunkerley, G. K., & Dalenberg, C. J. (2000). Secret-keeping behaviors in black and white children as a function of interviewer race, racial identity, and risk for abuse. *Journal of Aggression Maltreatment Trauma, 2*(2), 13–35.
- Dunn, L. M., & Dunn, L. M. (1965). *Peabody Picture Vocabulary Test*. Springer.
- Edwards, K. M., Banyard, V. L., Sessarego, S. N., Stanely, L. R., Mitchell, K. J., Echstein, R. P., et al. (2018). Measurement tools to assess relationship abuse and sexual assault prevention program effectiveness among youth. *Psychology of Violence, 8*, 537–545.
- Edwards, K. M., Banyard, V. L., Sessarego, S. N., Waterman, E. A., Mitchell, K. J., & Chang, H. (2019). Evaluation of a bystander-focused interpersonal violence prevention program with high school students. *Prevention Science, 20*, 488-498.
- Edwards, K. M., Siller, L., Wheeler, L. A., Charge, L. L., Charge, D. P. L., Bordeaux, S., ... & Simon, B. (2022). Effectiveness of a sexual assault self-defense program for American Indian girls. *Journal of Interpersonal Violence, 37*(15-16), NP13245-NP13267.
- Edwards, S. L., Rapee, R. M., & Kennedy, S. (2010). Prediction of anxiety symptoms in preschool-aged children: examination of maternal and paternal perspectives. *Journal of Child Psychology and Psychiatry, 51*(3), 313-321.
- Ekşi, H., Koç, S. & Türk, T. (2018). *Turkish psychometric properties of the scale of teachers' attitudes towards reporting child sexual abuse*. International Education and Science Congress, Afyon.
- Elardo, R., & Bradley, R. H. (1981). The Home Observation for Measurement of the Environment (HOME) Scale: A review of research. *Developmental Review, 1*(2), 113-145.
- Elfreich, M. R., Stevenson, M. C., Sisson, C., Winstead, A. P., & Parmenter, K. M. (2020). Sexual abuse disclosure mediates the effect of an abuse prevention program on substantiation. *Child Maltreatment, 25*(2), 215-223.

- Espelage, D. L., Basile, K. C., & Hamburger, M. E. (2012). Bullying perpetration and subsequent sexual violence perpetration among middle school students. *Journal of Adolescent Health, 50*(1), 60-65.
- Essau, C. A., Muris, P., & Ederer, E. M. (2002). Reliability and validity of the Spence Children's Anxiety Scale and the Screen for Child Anxiety Related Emotional Disorders in German children. *Journal of Behavior Therapy and Experimental Psychiatry, 33*(1), 1–18.
- Essex, M. J., Boyce, W. T., Goldstein, L. H., Armstrong, J. M., Kraemer, H. C., & Kupfer, D. J. (2002). The confluence of mental, physical, social, and academic difficulties in middle childhood. II: Developing the MacArthur health and behavior questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 588–603.
- Evans, C., Fraser, M.W. & Cotter, K.L. (2014) The effectiveness of school-based bullying prevention programs: A systematic review. *Aggression and Violent Behavior, 19*(5): 532–544.
- Eyberg, S., and Ross, A. (1978). Assessment of child behavior problems: the validation of a new inventory. *J. Clin. Child Psychol. 7*, 113–116.
- Fabbri, C., Rodrigues, K., Leurent, B., Allen, E., Qiu, M., Zuakulu, M., ... & Devries, K. M. (2021). The EmpaTeach intervention for reducing physical violence from teachers to students in Nyarugusu Refugee Camp: A cluster-randomised controlled trial. *PLoS medicine, 18*(10), e1003808.
- Feldmann, J., Storck, C., & Pfeffer, S. (2018). ReSi: Evaluation eines Programms zur Kompetenzförderung und Prävention sexuellen Missbrauchs im Kindergarten. *Praxis der Kinderpsychologie und Kinderpsychiatrie, 67*(8), 720-735.
- Finch, M., Featherson, R., Chakraborty, S., Bjørndal, L., Mildon, R., Albers, B., Fiennes, C., Taylor, D. J. A., Schachtman, R., Yang, T., Shlonsky, A. (2021). Interventions that address institutional child maltreatment: An evidence and gap map. *Campbell Systematic Reviews, 17*(1), e1139.
- Fishbein M., & Ajzen I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.
- Foa, E. B., Asnaani, A., Zang, Y., Capaldi, S., & Yeh, R. (2018). Psychometrics of the Child PTSD Symptom Scale for DSM-5 for trauma-exposed children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 47*(1), 38-46.
- Foshee V. A., Bauman K. E., Arraiga X. B., Helms R. W., Koch G. G., Linder G. F. (1998). An evaluation of Safe Dates, and adolescent dating violence prevention program. *American Journal of Public Health, 88*, 45-50.
- Foshee, V. A., & Langwick, V. L. (2010). *Safe Dates: An adolescent dating abuse prevention curriculum* (2nd ed.). Center City, MI: Hazeldon.

- Francis, T., & Baker-Henningham, H. (2021). The Irie Homes Toolbox: A cluster randomized controlled trial of an early childhood parenting program to prevent violence against children in Jamaica. *Children and Youth Services Review*, 126, 106060.
- Fraser J, Mathews B, Walsh K, Chen L, Dunne M. (2010). Factors influencing child abuse and neglect recognition and reporting by nurses: A multivariate analysis. *Int J Nurs Stud*, 47(2): 146–53.
- Frick, P. J. (1991). *The Alabama Parenting Questionnaire*. University of Alabama.
- Fryer Jr, G. E., Kraizer, S. K., & Mlyoshi, T. (1987a). Measuring actual reduction of risk to child abuse: A new approach. *Child Abuse & Neglect*, 11(2), 173-179.
- Fryer Jr, G. E., Kraizer, S. K., & Mlyoshi, T. (1987b). Measuring children's retention of skills to resist stranger abduction: Use of the simulation technique. *Child Abuse & Neglect*, 11(2), 181-185.
- Fulu, E., Warner, X., Miedema, S., Jewkes, R., Roselli, T., & Lang, J. (2013). Why do some men use violence against women and how can we prevent it? Quantitative findings from the United Nations multi-country study on men and violence in Asia and the Pacific.
- Furman, W., & Buhrmester, D. (2009). Methods and measures: The network of relationships inventory: Behavioral systems version. *International Journal of Behavioral Development*, 33, 470–478.
- García-Moreno, C., Jansen, H. A. F. M., Ellsberg, M., Heise, L., & Watts, C. (2005). *WHO Multi-country study on women's health and domestic violence against women: Initial results on prevalence, health outcomes and women's responses*. World Health Organisation.
- Gentry J. E., Baranowsky A. B., Dunning K. (2002). ARP: The accelerated recovery program (ARP) for compassion fatigue. In Figley C. R. (Eds.), *Treating compassion fatigue* (pp. 123–137). New York: Brunner-Routledge.
- Gilbert P., McEwan K., Ranjana M., Richter A., Leigh F., Mills A., et al. (2009). An exploration of different types of positive affect in students and patients with a bipolar disorder. *Clin Neuropsychiatry*, 6, 135–43.
- Gilbert, P., McEwan, K., Matos, M., & Ravis, A. (2011). Fears of compassion: Development of three self-report measures. *Psychology and Psychotherapy: Theory, Research and Practice*, 84(3), 239-255.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345.
- Green, B. L., Ayoub, C., Bartlett, J. D., Furrer, C., Chazan-Cohen, R., Buttitta, K., ... & Regalbuto, E. (2020). Pathways to prevention: Early Head Start outcomes in the first three years lead to long-term reductions in child maltreatment. *Children and Youth Services Review*, 118, 105403.
- Grendel, M. A. (1991). *Cognitive and emotional effects of a brief child sexual abuse prevention program for first-graders*. University of Cincinnati.

- Gunby, K. V., Carr, J. E., & Leblanc, L. A. (2010). Teaching abduction prevention skills to children with autism. *Journal of Applied Behavior Analysis, 43*, 107–112.
- Gushwa, M., Bernier, J., & Robinson, D. (2019). Advancing child sexual abuse prevention in schools: An exploration of the effectiveness of the enough! Online training program for K-12 teachers. *Journal of Child Sexual Abuse, 28*(2), 144-159.
- Hall, E. R., Howard, J. A., Boezio, S. L. (1986). Tolerance of rape: A sexist or antisocial attitudes? *Psychology of Women Quarterly, 10*(2), 101-118.
- Hamilton, M. A., & Orme, J. G. (1990). Examining the construct validity of three parenting knowledge measures using LISREL. *Social Service Review, 64*(1), 121-143.
- Harrison, P. J., Downes, J., Williams, M. D. (1991). *Date and acquaintance rape: Perceptions and attitude change strategies*. US: ACPA Executive Office
- Harvey, P., Forehand, R., Brown, C., & Holmes, T. (1988). The prevention of sexual abuse: Examination of the effectiveness of a program with kindergarten-age children. *Behavior Therapy, 19*(3), 429-435.
- Hawkins R, McCallum C. (2001). Effects of mandatory notification training on the tendency to report hypothetical cases of child abuse and neglect. *Child Abuse Rev, 10*(5), 301–22.
- Hazzard, A., Webb, C., Kleemeier, C., Angert, L., & Pohl, J. (1991). Child sexual abuse prevention: Evaluation and one-year follow-up. *Child Abuse & Neglect, 15*(1-2), 123-138.
- Hébert, M., Lavoie, F., Piché, C., & Poitras, M. (2001). Proximate effects of a child sexual abuse prevention program in elementary school children. *Child Abuse & Neglect, 25*(4), 505-522.
- Hecker, T., Mkinga, G., Kirika, A., Nkuba, M., Preston, J., & Hermenau, K. (2021). Preventing maltreatment in institutional care: A cluster-randomized controlled trial in East Africa. *Preventive medicine reports, 24*, 101593.
- Helfinstein, S. M., Kirwan, M. L., Benson, B. E., Hardin, M. G., Pine, D. S., Ernst, M., & Fox, N. A. (2013). Validation of a child-friendly version of the monetary incentive delay task. *Social Cognitive and Affective Neuroscience, 8*(6), 720-726.
- Heidotting, T. (1996). *A meta-analytic evaluation of school-based sexual abuse and personal safety prevention programs for children*. University of Cincinnati.
- Henry, D. (2000). *Initial report of the pilot study for the evaluation of the SAFE TO LEARN Demonstration Project*. Chicago, IL: Child Health Data Lab. Children’s Memorial Hospital.
- Herbert, J. L., & Bromfield, L. M. (2021). A quasi-experimental study of the Multi-Agency Investigation & Support Team (MIST): A collaborative response to child sexual abuse. *Child Abuse & Neglect, 111*, 104827.
- Hermenau, K., Goessmann, K., Rygaard, N. P., Landolt, M. A., & Hecker, T. (2017). Fostering child development by improving care quality: A systematic review of the effectiveness of structural interventions and caregiver trainings in institutional care. *Trauma, Violence, & Abuse, 18*(5), 544-561.

- Hillenbrand-Gunn, T. L. (2003). *Discomfort With Sexist Situations-Revised and What Would You Do? scales*. Unpublished scales.
- Hillenbrand-Gunn, T. L., Heppner, M. J., Mauch, P. A., & Park, H. J. (2010). Men as allies: The efficacy of a high school rape prevention intervention. *Journal of Counseling & Development, 88*(1), 43-51.
- Hiller, R. M., Davis, R. S., Devaney, J., Halligan, S. L., Meiser-Stedman, R., Smith, P., ... & MacNeill, S. (2021). Protocol for the RELATE trial: a feasibility and pilot randomised controlled trial of a low-intensity group intervention for young people in care with elevated posttraumatic stress symptoms. *Pilot and feasibility studies, 7*(1), 1-11.
- Hoefnagels, C., Onrust, S., van Rooijen, M., Jonkman, H., van Spanje-Hennes, A., & Breeman, L. D. (2021). Changing the classroom climate to lower the threshold for child abuse and neglect self-disclosure: a non-randomized cluster controlled trial. *Children and Youth Services Review, 130*, 106196.
- Huang, S., & Cui, C. (2020). Preventing child sexual abuse using picture books: The effect of book character and message framing. *Journal of Child Sexual Abuse, 29*(4), 448-467.
- Humphreys, K. L., Gleason, M. M., Drury, S. S., Miron, D., Nelson, C. A., Fox, N. A., & Zeanah, C. H. (2015). Effects of institutional rearing and foster care on psychopathology at age 12 years in Romania: follow-up of an open, randomised controlled trial. *The Lancet Psychiatry, 2*(7), 625-634.
- Humphreys, K. L., Piersiak, H. A., Panlilio, C. C., Lehman, E. B., Verdiglione, N., Dore, S., & Levi, B. H. (2021). A randomized control trial of a child abuse mandated reporter training: Knowledge and attitudes. *Child Abuse & Neglect, 117*, 105033.
- Hymel, K. P., Armijo-Garcia, V., Foster, R., Frazier, T. N., Stoiko, M., Christie, L. M., ... & Wang, M. (2014). Validation of a clinical prediction rule for pediatric abusive head trauma. *Pediatrics, 134*(6), e1537-e1544.
- Hymel, K. P., Armijo-Garcia, V., Musick, M., Marinello, M., Herman, B. E., Weeks, K., ... & Pediatric Brain Injury Research Network (PediBIRN) Investigators. (2021). A cluster randomized trial to reduce missed abusive head trauma in pediatric intensive care settings. *The Journal of Pediatrics, 236*, 260-268.
- Hymel, K. P., Willson, D. F., Boos, S. C., Pullin, D. A., Homa, K., Lorenz, D. J., ... & Pediatric Brain Injury Research Network (PediBIRN) Investigators. (2013). Derivation of a clinical prediction rule for pediatric abusive head trauma. *Pediatric Critical Care Medicine, 14*(2), 210-220.
- International Society for the Prevention of Child Abuse and Neglect. (2006). *ICAST-C: the ISPCAN Child Abuse Screening Tool—Child Version. Manual and proposed guidelines for pilot administration*. Aurora: International Society for the Prevention of Child Abuse and Neglect.
- James, S. H., & DeVaney, S. B. (1994). Reporting suspected sexual abuse: A study of counselor and counselor trainee responses. *Elementary School Guidance & Counselling, 28*(4), 257-263.



- Jewkes, R., Gevers, A., Chirwa, E., Mahlangu, P., Shamu, S., Shai, N., & Lombard, C. (2019). RCT evaluation of Skhokho: A holistic school intervention to prevent gender-based violence among South African Grade 8s. *PLoS one*, *14*(10), e0223562.
- Jin, Y., Chen, J., Jiang, Y., & Yu, B. (2017). Evaluation of a sexual abuse prevention education program for school-age children in China: a comparison of teachers and parents as instructors. *Health Education Research*, *32*(4), 364-373.
- Johnson-Motoyama, M., Ginther, D. K., Phillips, R., Beer, O. W., Merkel-Holguin, L., & Fluke, J. (2023). Differential response and the reduction of child maltreatment and foster care services utilization in the US from 2004 to 2017. *Child Maltreatment*, *28*(1), 152-162.
- Johnson, B. M., Miltenberger, R. G., Egemo-Helm, K., Jostad, C. J., Flessner, C., & Gatheridge, B. (2005). Evaluation of behavioral skills training for teaching abduction-prevention skills to young children. *Journal of Applied Behavior Analysis*, *38*, 67–78.
- Johnson, B. M., Miltenberger, R. G., Knudson, P., Egemo-Helm, K., Kelso, P., Jostad, C., et al. (2006). A preliminary evaluation of two behavioral skills training procedures for teaching abduction-prevention skills to schoolchildren. *Journal of Applied Behavior Analysis*, *39*, 25–34.
- Johnson, D. E., Guthrie, D., Smyke, A. T., Koga, S. F., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2010). Growth and associations between auxology, caregiving environment, and cognition in socially deprived Romanian children randomized to foster vs ongoing institutional care. *Archives of Pediatrics & Adolescent Medicine*, *164*(6), 507-516.
- Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, *29*(4), 589-611.
- Jones, C., Scholes, L., Rolfe, B., & Stieler-Hunt, C. (2020). A serious-game for child sexual abuse prevention: An evaluation of orbit. *Child Abuse & Neglect*, *107*, 104569.
- Kanbay, Y., Aslan, Ö., Işık, E., & Tektaş, P. (2017). *Development study on ISKEBE violence against women attitude scale* (ISKEBE Attitude Scale).
- Kang, S. R., Kim, S. J., & Kang, K. A. (2022). Effects of child sexual abuse prevention education program using hybrid application (CSAPE-H) on fifth-grade students in South Korea. *The Journal of School Nursing*, *38*(4), 368-379.
- Kantor, G. K., Holt, M. K., Mebert, C. J., Straus, M. A., Drach, K. M., Ricci, L. R., ... & Brown, W. (2004). Development and preliminary psychometric properties of the multidimensional neglectful behavior scale-child report. *Child Maltreatment*, *9*(4), 409-428.
- Karmaliani, R., McFarlane, J., Khuwaja, H. M. A., Somani, Y., Bhamani, S. S., Saeed Ali, T., ... & Jewkes, R. (2020). Right To Play's intervention to reduce peer violence among children in public schools in Pakistan: a cluster-randomized controlled trial. *Global Health Action*, *13*(1), 1836604.



- Kelly, A. C., Zuroff, D. C., Leybman, M. J., & Gilbert, P. (2012). Social safeness, received social support, and maladjustment: Testing a tripartite model of affect regulation. *Cognitive Therapy and Research*, 36, 815-826.
- Kenny, M. and Wurtele, S. K. (2010). Children's abilities to recognize a "good" person as a potential perpetrator of childhood sexual abuse. *Child Abuse & Neglect*, 34, 490 – 495.
- Kenny, M. C. (2000). Educators and child abuse questionnaire. *Child Abuse & Neglect*.
- Kenny, M. C. (2004). Teachers' attitudes toward and knowledge of child maltreatment. *Child Abuse & Neglect*, 28(12), 1311-1319.
- Kenny, M. C. (2010). Child sexual abuse education with ethnically diverse families: A preliminary analysis. *Children and Youth Services Review*, 32(7), 981-989.
- Kenny, M. C., & Wurtele, S. K. (2008). Preschoolers' knowledge of genital terminology: A comparison of English and Spanish speakers. *American Journal of Sexuality Education*, 3(4), 345-354.
- Kenny, M. C., Wurtele, S. K., & Alonso, L. (2012). Evaluation of a personal safety program with Latino preschoolers. *Journal of Child Sexual Abuse*, 21(4), 368-385.
- Khuwaja, H. M. A., Karmaliani, R., McFarlane, J., et al. (2018). The intersection of school corporal punishment and associated factors: baseline results from a randomized controlled trial in Pakistan. *PLoS One*, 13, e0206032
- Kilmartin, C., Conway, A., Friedberg, A., McQuoid, T., & Tschan, T. (1999). *Social conformity and sexism in all-male peer groups*. Paper presented at the annual convention of the Virginia Psychological Association, Virginia Beach, VA.
- Kim, S. J., & Kang, K. A. (2017). Effects of the child sexual abuse prevention education (C-SAPE) program on South Korean fifth-grade students' competence in terms of knowledge and self-protective behaviors. *The Journal of School Nursing*, 33(2), 123-132.
- Kim, S., Nickerson, A., Livingston, J. A., Dudley, M., Manges, M., Tulledge, J., & Allen, K. (2019). Teacher outcomes from the Second Step Child Protection Unit: moderating roles of prior preparedness, and treatment acceptability. *Journal of Child Sexual Abuse*, 28(6), 726-744.
- Kızıltepe, R., Eslek, D., Irmak, T. Y., & Güngör, D. (2022). "I am Learning to Protect Myself with Mika:" A Teacher-based Child Sexual Abuse Prevention Program in Turkey. *Journal of Interpersonal Violence*, 37(11-12), NP10220-NP10244.
- Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong, I. Y., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology*, 34(1), 67-76.
- Knight, L., Allen, E., Mirembe, A., Nakuti, J., Namy, S., Child, J. C., ... & Devries, K. M. (2018). Implementation of the Good School Toolkit in Uganda: a quantitative process evaluation of a successful violence prevention program. *BMC Public Health*, 18, 1-13.

- Knight, L., Atuhaire, L., Allen, E., Namy, S., Anton-Erxleben, K., Nakuti, J., ... & Devries, K. (2020). Long-term outcomes of the Good School Toolkit primary school violence prevention intervention among adolescents: protocol for a nonrandomized quasi-experimental study. *JMIR research protocols*, *9*(12), e20940.
- Koç, S., Ekşi, H., & Türk, T. (2020). Child Sexual Abuse Scale: Turkish form. *Ilkogretim Online*, *19*(1), 173-182.
- Koita, K., Long, D., Hessler, D., Benson, M., Daley, K., Bucci, M., ... & Burke Harris, N. (2018). Development and implementation of a pediatric adverse childhood experiences (ACEs) and other determinants of health questionnaire in the pediatric medical home: A pilot study. *PloS one*, *13*(12), e0208088.
- Kolko, D. J., Moser, J. T., & Hughes, J. (1989). Classroom training in sexual victimization awareness and prevention skills: An extension of the Red Flag/Green Flag people program. *Journal of Family Violence*, *4*, 25-45.
- Kolko, D. J., Moser, J. T., Litz, J., & Hughes, J. (1987). Promoting awareness and prevention of child sexual victimization using the Red Flag/Green Flag program: An evaluation with follow-up. *Journal of Family Violence*, *2*, 11-35.
- König, E., Maier, A., Fegert, J. M., & Hoffmann, U. (2020). Development and randomized controlled trial evaluation of E-learning trainings for professionals. *Archives of Public Health*, *78*, 1-8.
- Konijnendijk, A. A., Boere-Boonekamp, M. M., Haasnoot, M. E., & Need, A. (2019). Effects of a computerised guideline support tool on child healthcare professionals' response to suspicions of child abuse and neglect: a community-based intervention trial. *BMC medical informatics and decision making*, *19*(1), 1-12.
- Kovacs, M. (1992). *Children's depression inventory*. Toronto Ontario.
- Krahé, B., & Knappert, L. (2009). A group-randomized evaluation of a theatre-based sexual abuse prevention programme for primary school children in Germany. *Journal of Community & Applied Social Psychology*, *19*(4), 321-329.
- Kraizer, S. (1986). *Children need to know personal safety program*. The Educator's Guide to Preventing Child Sexual Abuse.
- Kraizer, S. (1991). *The Safe Child Program for the prevention of child abuse: Development and evaluation of a school-based curriculum*. The Union Institute.
- Kraizer, S. K., Fryer, G. E., & Miller, M. (1988). Programming for preventing sexual abuse and abduction: what does it mean when it works?. *Child Welfare*, 69-78.
- Kraizer, S., Witte, S. S., & Freyer, G. E. (1989). *Child sexual abuse prevention programs: What makes them effective in protecting children?*. Columbia.
- Lanz, M., & Maino, E. (2014). Family environment scale. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 2170-2174). Springer.

- Larsen, J. J., & Juhasz, A. M. (1986). The knowledge of child development inventory. *Adolescence, 21*(81), 39.
- Lavoie, F., Dufort, F., Hébert, M., Vézina, L. (1997). Évaluation d'un programme de prévention de la violence lors des fréquentations: Une évaluation de VIRAJ selon une approche quasi expérimentale. Sainte Foy, Quebec: Département de psychologie. Université Laval Rapport de recherche soumis au Conseil québécois de la recherche sociale.
- Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher–child interactions in Chile and their associations with prekindergarten outcomes. *Child Development, 86*(3), 781-799.
- Li, H., Liu, J., & Hunter, C. V. (2020). A meta-analysis of the factor structure of the classroom assessment scoring system (CLASS). *The Journal of Experimental Education, 88*(2), 265-287.
- Lonsway, K. A., & Fitzgerald, L. F. (1994). Rape myths. In review. *Psychology of Women Quarterly, 18*(2), 133-164.
- Lonsway, K. A., & Fitzgerald, L. F. (1995). Attitudinal antecedents of rape myth acceptance: A theoretical and empirical reexamination. *Journal of Personality and Social Psychology, 68*(4), 704.
- López García, A. I., Scharpf, F., Hoeffler, A., & Hecker, T. (2022). Preventing violence by teachers in primary schools: study protocol for a cluster randomized controlled trial in Haiti. *Frontiers in Public Health, 9*, 797267.
- Lynagh, M., Gilligan, C., & Handley, T. (2010) Teaching about, and dealing with sensitive issues in schools: How confident are pre-service teachers? *Asia-Pacific Journal of Health, Sport & Physical Education, 1*(3–4), 5–11.
- MacIntyre, D., & Carr, A. (1999). Evaluation of the effectiveness of the stay safe primary prevention programme for child sexual abuse. *Child Abuse & Neglect, 23*(12), 1307-1325.
- MacPhee, D. (1981). *Manual: Knowledge of infant development inventory*. Unpublished manuscript, University of North Carolina.
- Martin, J., Riazi, H., Firoozi, A., & Nasiri, M. (2020). A sex education program for teachers of preschool children: a quasi-experimental study in Iran. *BMC Public Health, 20*, 1-9.
- Masath, F. B., Hermenau, K., Nkuba, M., & Hecker, T. (2020). Reducing violent discipline by teachers using Interaction Competencies with Children for Teachers (ICC-T): study protocol for a matched cluster randomized controlled trial in Tanzanian public primary schools. *Trials, 21*(1), 1-11.
- Mathews, B., & Collin-Vézina, D. (2019). Child sexual abuse: Toward a conceptual model and definition. *Trauma, Violence, & Abuse, 20*(2), 131-148.
- Mathews, B., Yang, C., Lehman, E. B., Mincemoyer, C., Verdiglione, N., & Levi, B. H. (2017). Educating early childhood care and education providers to improve knowledge and

attitudes about reporting child maltreatment: A randomized controlled trial. *PLoS one*, 12(5), e0177777.

- McElearney, A., Adamson, G., Murphy, C., Rogers, S., Radcliffe, D. (2021). Results and learning from a cluster randomised trial on the impact of Keeping Safe in Northern Ireland, an education programme that teaches children about abuse. NSPCC Learning.
- McElearney, A., Scott, J., Adamson, G., Turtle, K., McBride, O. & Stephenson, P. (2011a) *Establishing the need to teach “keeping safe” messages in primary schools in Northern Ireland: what do children currently know and understand?* Belfast: NSPCC.
- McGrath P, Cappelli M, Wiseman D, Khalil N, Allan B. (1987). Teacher awareness program on child abuse: A randomized controlled trial. *Child Abuse & Neglect*, 11(1): 125–32.
- McMaster, L. E., Connolly, J., Pepler, D., & Craig, W. M. (2002). Peer to peer sexual harassment in early adolescence: A developmental perspective. *Development and Psychopathology*, 14(1), 91-105.
- Melchers, M., Montag, C., Markett, S., & Reuter, M. (2015). Assessment of empathy via self-report and behavioural paradigms: data on convergent and discriminant validity. *Cognitive Neuropsychiatry*, 20(2), 157-171.
- Merrell, K. W. (2002). *School Social Behavior Scales, 2nd Edition (SSBS-2)*. Eugene, OR: Assessment Intervention Resources.
- Merrill, K. G., Knight, L., Namy, S., Allen, E., Naker, D., & Devries, K. M. (2018). Effects of a violence prevention intervention in schools and surrounding communities: Secondary analysis of a cluster randomised-controlled trial in Uganda. *Child Abuse & Neglect*, 84, 182-195.
- Miller-Perrin, C., & Wurtele, S. (1986). *Harmful effects of school-based sexual abuse prevention programs? Reassure the parents*. Paper presented at Annual Convention of American Psychological Association, Washington, D.C.
- Mizwa, M. (1993). *Validation of the Victim Empathy Scale: A Measure of Emotional Empathy for Victims of Violent Crime*. University of North Dakota.
- Mogg, K., Bradley, B.P., De Bono, J., & Painter, M. (1997). Time course of attentional bias for threat information in non-clinical anxiety. *Behaviour Research and Therapy*, 35(4), 297–303.
- Moos, R. H., & Moos, B. S. (2002). *Family environment scale manual: Development, applications, research*. Third ed Palo Alto, CA: Center for Health Care Evaluation.
- Moos, R., Insel, P., & Humphrey, B. (1974). *Combined preliminary manual: family, work, & group environment scales*. Palo Alto, Calif.: Consulting Psychologists Press.
- Morawska, A., Walsh, A., Grabski, M. & Fletcher, R. (2015). Parental confidence and preferences for communicating with their child about sexuality. *Sex Education*, 15(3): 235–248.

- Moynihan, M. M., Banyard, V. L., Cares, A. C., Potter, S. J., Williams, L. M., & Stapleton, J. G. (2015). Encouraging responses in sexual and relationship violence prevention: What program effects remain one year later? *Journal of Interpersonal Violence, 30*, 110–132.
- Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity, 2*(3), 223-250.
- Neherta, M., Machmud, R., & Damayanti, R. (2017). The difference in intervention of sexual abuse prevention by two variance professions on primary school children in Padang. *Indian Journal of Community Health, 29*(1), 118-122.
- Nelson, D. (1985). *An evaluation of the student outcomes and instructional characteristics of the "You're in charge" program*. Unpublished manuscript prepared for the Utah State Office of Education, Salt Lake City.
- Nemerofsky, A. G. (1986). *The What-If Situations Test*. Baltimore, MD.
- Nemerofsky, A. G. (1991). *Child sexual abuse prevention: Teacher and child variables affecting the learning of skills and concepts in a sexual abuse prevention program*. Unpublished doctoral dissertation, The Johns Hopkins University, Baltimore, MD.
- Nemerofsky, A. G., Carran, D. T., & Rosenberg, L. A. (1994). Age variation in performance among preschool children in a sexual abuse prevention program. *J. Child Sexual Abuse, 3*, 85.
- NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Child Res Q., 11*, 269–306.
- Nickerson, A. B., Tullidge, J., Manges, M., Kesselring, S., Parks, T., Livingston, J. A., & Dudley, M. (2019). Randomized controlled trial of the Child Protection Unit: Grade and gender as moderators of CSA prevention concepts in elementary students. *Child Abuse & Neglect, 96*, 104101.
- Nkuba, M., Hermenau, K., Goessmann, K., & Hecker, T. (2018). Reducing violence by teachers using the preventative intervention Interaction Competencies with Children for Teachers (ICC-T): a cluster randomized controlled trial at public secondary schools in Tanzania. *PLoS One, 13*(8), e0201362.
- Nowicki, S. (1973). Children's Nowicki-Strickland Internal-External Locus of Control Manual. *Journal of Consulting and Clinical Psychology, 40*(1), 148-54.
- Nybergh, L., Taft, C., & Krantz, G. (2013). Psychometric properties of the WHO Violence Against Women instrument in a female population-based sample in Sweden: a cross-sectional survey. *BMJ Open, 3*(5).
- O'Donohue W., Yeater E. A., Fanetti M. (2003). Rape prevention with college males: The roles of rape myth acceptance, victim empathy, and outcome expectancies. *Journal of Interpersonal Violence, 18*, 513-531.
- Oldfield, D., Hays, B. J., & Megel, M. E. (1996). Evaluation of the effectiveness of Project Trust: an elementary school-based victimization prevention strategy. *Child Abuse & Neglect, 20*(9), 821-832.



- Olweus, D. (1996). *The revised Olweus Bully/Victim Questionnaire for Students*. Bergen, Norway: University of Bergen.
- Orpinas, P., & Kelder, S. (1995). *Students for Peace Project. Second student evaluation*. Houston, TX: University of Texas Health Science Center at Houston, School of Public Health.
- Pagano, M. E., Cassidy, L. J., Little, M., Murphy, J. M., & Jellinek, A. M. S. (2000). Identifying psychosocial dysfunction in school-age children: The Pediatric Symptom Checklist as a self-report measure. *Psychology in the Schools, 37*(2), 91-106.
- Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology, 29*, 611–621.
- Parker G., Tupling H., Brown L. B. (1979). A parental bonding instrument. *Br J Med Psychol 52*, 1–10.
- Payne, D. L., Lonsway, K. A., & Fitzgerald, L. F. (1999). Rape myth acceptance: Exploration of its structure and its measurement using the Illinois rape myth acceptance scale. *Journal of Research in Personality, 33*, 27–68.
- Peker, A., Cengiz, S., & Çelik, A. K. (2020). The Effect of Psycho-Education Program Developed for Sexual Abuse on Counseling Teachers' Reporting Sexual Abuse and Information and Risk Recognition Attitudes. *International Journal of Education and Literacy Studies, 8*(4), 74-86.
- Pelletier, V., Tourigny, M., Clément, M.-E. & Lavoie, F. (1998). Incidence et facteurs associés à la violence dans les fréquentations amoureuses chez les jeunes [Incidence and factors associated with violence in dating relationships in youth], Hull, Quebec, Canada: Université du Québec à Hull, Département de psychoéducation: Rapport de recherche présenté au CALACS Laurentides .
- Perez Jolles, M., Mack, W. J., Reaves, C., Saldana, L., Stadnick, N. A., Fernandez, M. E., & Aarons, G. A. (2021). Using a participatory method to test a strategy supporting the implementation of a state policy on screening children for adverse childhood experiences (ACEs) in a Federally Qualified Health Center system: a stepped-wedge cluster randomized trial. *Implementation Science Communications, 2*, 1-14.
- Perosa, L. M., & Perosa, S. L. (1990). Convergent and discriminant validity for family self-report measures. *Educational and Psychological Measurement, 50*(4), 855-868.
- Perrin, S., Meiser-Stedman, R., & Smith, P. (2005). The Children's Revised Impact of Event Scale (CRIES): Validity as a screening instrument for PTSD. *Behavioural and Cognitive Psychotherapy, 33*(4), 487-498.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Paul H. Brookes Publishing Co.
- Pitts, C. (2015). *Child sexual abuse prevention programs for pre-schoolers: A synthesis of current evidence*. Sydney: Royal Commission into Institutional Responses to Child Sexual Abuse.



- Pommier, E., Neff, K. D., & Tóth-Király, I. (2020). The development and validation of the Compassion Scale. *Assessment, 27*(1), 21-39.
- Price E. L., Byers E. S. The Dating Violence Research Team. (1999). The attitudes towards dating violence scales: Development and initial validation. *Journal of Family Violence, 14*, 351-375.
- Pulerwitz, J., & Barker, G. (2008). Measuring attitudes toward gender norms among young men in Brazil: Development and psychometric evaluation of the GEM Scale. *Men and Masculinities, 10*(3), 322-338.
- Pulido, M. L., Dauber, S., Tully, B. A., Hamilton, P., Smith, M. J., & Freeman, K. (2015). Knowledge gains following a child sexual abuse prevention program among urban students: A cluster-randomized evaluation. *American Journal of Public Health, 105*(7), 1344-1350.
- Quadara, A., Nagy, V., Higgins, D., & Siegel, N. (2015). *Conceptualising the prevention of child sexual abuse*. Australian Institute of Family Studies.
- Radford, L., Richardson Foster, H., Barter, C. A., & Stanley, N. (2017). *Rapid Evidence Assessment: What can be learnt from other jurisdictions about preventing and responding to child sexual abuse*. University of Central Lancashire.
- Radloff L. (1977). The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Measurement, 1*, 385–401.
- Ratto, R., & Bogat, G. A. (1990). An evaluation of a preschool curriculum to educate children in the prevention of sexual abuse. *Journal of Community Psychology, 18*(3), 289-297.
- Ravens-Sieberer, U., Herdman, M., Devine, J., Otto, C., Bullinger, M., Rose, M., & Klasen, F. (2014). The European KIDSCREEN approach to measure quality of life and well-being in children: development, current application, and future advances. *Quality of Life Research, 23*, 791-803.
- Reitman, D., Currier, R. O., & Stickle, T. R. (2002). A critical evaluation of the Parenting Stress Index-Short Form (PSI-SF) in a head start population. *Journal of Clinical Child and Adolescent Psychology, 31*(3), 384-392.
- Rentoul, J. A., & Fraser, B. J. (1983). Development of a school—level environment questionnaire. *Journal of Educational Administration, 21*(1), 21-39.
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*, 119–125.
- Reynolds, C. R., & Richmond, B. O. (1985). *Revised children's manifest anxiety scale (RCMAS) manual*. Los Angeles, CA: Western Psychological Services.
- Reynolds, C.R. & Richmond, B.O. (2008). *Revised children's manifest anxiety scale, second edition (RCMAS-2): Manual*. Los Angeles: Western Psychological Services.
- Rheingold, A. A., Zajac, K., Chapman, J. E., Patton, M., de Arellano, M., Saunders, B., & Kilpatrick, D. (2015). Child sexual abuse prevention training for childcare professionals: An

independent multi-site randomized controlled trial of stewards of children. *Prevention Science*, 16, 374-385.

- Ricardo, C., Eads, M., & Barker, G. T. (2011). *Engaging boys and young men in the prevention of sexual violence: A systematic and global review of evaluated interventions*. Sexual Violence Research Initiative and Promundo. Pretoria, South Africa.
- Rickel, A. U., & Biasetti, L. L. (1982). Modification of the Block Child Rearing Practices Report. *Journal of Clinical Psychology*, 38(1), 129–134.
- Robbins, T. W., James, M., Owen, A. M., McKinnes, L., & Rabbitt, P. (1994). Cambridge Automated Neuropsychological Test Battery (CANTAB). A factor analytic study in a large sample of normal volunteers. *Dementia*, 5, 266-281.
- Roggman, L. A., Moe, S. T., Hart, A. D., & Forthun, L. F. (1994). Family leisure and social support: Relations with parenting stress and psychological well-being in Head Start parents. *Early Childhood Research Quarterly*, 9(3-4), 463-480.
- Ruis-Caseres, M. (2011). *Child Protection Knowledge, Attitudes and Practices in Central and Western Liberia*. Save the Children.
- Rutter, M. (1967). A children's behavior questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, 8, 1–11.
- Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Torrance, CA: Western Psychological Services.
- Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-K programs predict children's learning?. *Science*, 341(6148), 845-846.
- Sachser, C., Keller, F., & Goldbeck, L. (2017). Complex PTSD as proposed for ICD-11: Validation of a new disorder in children and adolescents and their response to Trauma-Focused Cognitive Behavioral Therapy. *Journal of Child Psychology and Psychiatry*, 58(2), 160-168.
- Santos, L., do Rosário Pinheiro, M., & Rijo, D. (2022). Compassionate mind training for caregivers of residential youth care: Early findings of a cluster randomized trial. *Child Abuse & Neglect*, 123, 105429.
- Saslowsky, D. A., & Wurtele, S. K. (1986). Educating children about sexual abuse: implications for pediatric intervention and possible prevention. *Journal of Pediatric Psychology*, 11(2), 235-245.
- Sawyer M. G, Harchak T. F., Spence S. H., et al. (2010). School-Based prevention of depression: a 2-year follow-up of a randomized controlled trial of the beyondblue schools research initiative. *Adolesc Health*, 47, 297–304.
- Scharpf, F., Kirika, A., Masath, F. B., Mkinga, G., Ssenyonga, J., Nyarko-Tetteh, E., ... & Hecker, T. (2021). Reducing physical and emotional violence by teachers using the intervention Interaction Competencies with Children—for Teachers (ICC-T): study protocol of a multi-country cluster randomized controlled trial in Ghana, Tanzania, and Uganda. *BMC Public Health*, 21, 1-15.

- Scholes, L. J., Jones, C. M., & Nagel, M. C. (2014). Boys and CSA prevention: Issues surrounding gender and approaches for prevention. *Australian Journal of Teacher Education (Online)*, 39(11), 1-15.
- Schwab-Stone, M. E., Ayers, T. S., Kaspro, W., Voyce, C., Barone, C., Shriver, T., & Weissberg, R. P. (1995). No safe haven: A study of violence exposure in an urban community. *Journal of the American Academy of Child & Adolescent Psychiatry*, 34(10), 1343-1352.
- Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. J. Weinman, S. Wright, & M. Johnston, *Measures in health psychology: A user's portfolio. Causal and control beliefs*, 35, 37.
- Seidman, S. & Zager, J. (1970). The Teacher Burnout Scale. *Educational Research Quarterly*, 11, 26-33.
- Shavega, T. J., Brugman, D., & van Tuijl, C. (2014). Children's behavioral adjustment in pre-primary schools in Tanzania: A multilevel approach. *Early Education and Development*, 25(3), 356-380.
- Shelton, K. K., Frick, P. J., & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, 25(3), 317-329.
- Sherr, L., Roberts, K. J., & Gandhi, N. (2017). Child violence experiences in institutionalised/orphanage care. *Psychology, Health & Medicine*, 22(sup1), 31-57.
- Singh, S., Knox, M., Pelletier, H. (2017). Exploratory factor analysis and psychometric evaluation of the healthcare provider attitudes toward child maltreatment reporting scale. *Children's Health Care*, 46(4), 356-365.
- Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, 22(2), 173-187.
- Smyke, A.T., Dumitrescu, A., & Zeanah, C.H. (2002). Attachment disturbances in young children: The continuum of caretaking casualty. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(8), 972-982.
- Smyke, A. T., Koga, S. F., Johnson, D. E., et al. (2007). The caregiving context in institution-reared and family-reared infants and toddlers in Romania. *J Child Psychol Psychiatry*, 48(2), 210-218.
- Smyke, A. T., Zeanah, C. H., Fox, N. A., Nelson, C. A., & Guthrie, D. (2010). Placement in foster care enhances quality of attachment among young institutionalized children. *Child Development*, 81(1), 212-223.
- Snyder, C. M. (1986). *An evaluation of the "good secrets, bad secrets" sexual assault prevention program (molestation, abuse, erie)*, Pennsylvania. Doctoral dissertation, United States International University.
- Solberg, M. E., Olweus, D., & Endresen, I. M. (2007). Bullies and victims at school: Are they the same pupils?. *British Journal of Educational Psychology*, 77(2), 441-464.

- South, S., Shlonsky, A., Mildon, R., Pourliakas, A., Falkiner, J., & Laughlin, A. (2014). *Scoping review: Evaluations of out-of-home care practice elements that aim to prevent child sexual abuse*. Parenting Research Centre.
- Spence, J. T., & Hahn, E. D. (1997). The attitudes toward women scale and attitude change in college students. *Psychology of Women Quarterly*, 21(1), 17-34.
- Spielberger, C. D. (1973). *Manual for the State-Trait Inventory for Children*. Palo Alto, CA: Consulting Psychological Press.
- Ssenyonga, J., Katharin, H., Mattonet, K., Nkuba, M., & Hecker, T. (2022). Reducing teachers' use of violence toward students: A cluster-randomized controlled trial in secondary schools in Southwestern Uganda. *Children and Youth Services Review*, 138, 106521.
- Stevens, K., & Ratcliffe, J. (2012). Measuring and valuing health benefits for economic evaluation in adolescence: an assessment of the practicality and validity of the child health utility 9D in the Australian adolescent population. *Value in Health*, 15(8), 1092-1099.
- Stieler-Hunt, C., Jones, C. M., Rolfe, B., & Pozzebon, K. (2014). Examining key design decisions involved in developing a serious game for child sexual abuse prevention. *Frontiers in Psychology*, 5, 73.
- Straus, M. A. (1974). Leveling, civility, and violence in the family. *Journal of Marriage and the Family*, 13-29.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics Scales. *Journal of Marriage and the Family*, 41, 75-88.
- Straus, M. A., & Mattingly, M. J. (2007). *A short form and severity level types for the Parent-Child Conflict Tactics Scales*. Durham, NH: Family Research Laboratory, University of New Hampshire.
- Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W., & Runyan, D. (1998). Identification of child maltreatment with the Parent-Child Conflict Tactics Scales: Development and psychometric data for a national sample of American parents. *Child Abuse & Neglect*, 22(4), 249-270.
- Sullivan, P. M., Scanlan, J. M., Brookhouser, P. E., Schulte, L. E., & Knutson, J. F. (1992). The effects of psychotherapy on behavior problems of sexually abused deaf children. *Child Abuse & Neglect*, 16(2), 297-307.
- Taal, M., & Edelaar, M. (1997). Positive and negative effects of a child sexual abuse prevention program. *Child Abuse & Neglect*, 21(4), 399-410.
- Tang, A., Almas, A., Zeytinoglu, S., Zeanah, C. H., Nelson, C. A., & Fox, N. A. (2021). Long-Term Effects of Institutional Care and Enhanced Attachment Relationships on Close Adolescent Friendships. *Child Development*, 92(6), 2431-2446.
- Tangney J. P., Baumeister R. F., & Boone A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *J Pers*, 72, 271–324.

- Taylor, B., Stein, N., & Burden, F. (2010). The effects of gender violence/harassment prevention programming in middle schools: A randomized experimental evaluation. *Violence and Victims, 25*(2), 202-223.
- Taylor, E. K., Dopp, A. R., Lounsbury, K., Thompson, Y., Miller, M., Jorgensen, A., & Silovsky, J. F. (2021). Enhancing Early Engagement (E3) in mental health services training for children's advocacy center's victim advocates: feasibility protocol for a randomized controlled trial. *Pilot and Feasibility Studies, 7*, 1-18.
- Telljohann, S. K., Everett, S. A., & Price, J. H. (1997). Evaluation of a third grade sexual abuse curriculum. *Journal of School Health, 67*(4), 149-153.
- Thakur, N., Hessler, D., Koita, K., Ye, M., Benson, M., Gilgoff, R., ... & Harris, N. B. (2020). Pediatrics adverse childhood experiences and related life events screener (PEARLS) and health in a safety-net practice. *Child Abuse & Neglect, 108*, 104685.
- Thompson E. L., Garg A., Collier Esser K., Caddy D., Matthes S., Rohr D., Spence E. (2020). *Validation of a child physical and sexual abuse prevention scale for a school-based prevention program*. American Public Health Association Annual Meeting, San Francisco, CA.
- Thompson, E. L., Zhou, Z., Garg, A., Rohr, D., Ajoku, B., & Spence, E. E. (2022). Evaluation of a school-based child physical and sexual abuse prevention program. *Health Education & Behavior, 49*(4), 584-592.
- Thota, A., Fiennes, C., White, H., & Jain, S. (2023). *Institutional Responses to Child Maltreatment Evidence and Gap Map: Report on First Update*. Giving Evidence.
- Topping, K. J., & Barron, I. G. (2009). School-based child sexual abuse prevention programs: A review of effectiveness. *Review of educational research, 79*(1), 431-463.
- Troller-Renfree, S., McDermott, J. M., Nelson, C. A., Zeanah, C. H., & Fox, N. A. (2015). The effects of early foster care intervention on attention biases in previously institutionalized children in Romania. *Developmental science, 18*(5), 713-722.
- Tunc, G. T., Gorak, G., Ozyazicioglu, N., Ak, B., Isil, B., & Vural, P. (2018). Determining the appropriateness of the "what if" situations test (WIST) with Turkish pre-schoolers. *Journal of Child Sexual Abuse, 27*(3), 292-304.
- Turan, F. D. (2022). Effects of a structured online educational program course on nursing students' attitudes toward gender roles and women and children's violence abuse reports: a quasi-experimental evaluation. *Nurse Education Today, 108*, 105191.
- Turan, T., Erdoğan, Ç., (2019). Validity and reliability of the healthcare provider attitudes toward child maltreatment reporting scale on nursing students. *JAREN 5*, 46–52.
- Tutty, L. (1990). Preventing child sexual abuse: A review of current research and theory. In M. Rothery & G. Cameron (Eds.), *Child maltreatment: Expanding our concepts of helping* (pp. 259-275). Hillsdale, NJ: Lawrence Erlbaum.



- Tutty, L. M. (1992). The ability of elementary school children to learn child sexual abuse prevention concepts. *Child Abuse & Neglect, 16*(3), 369-384.
- Tutty, L. M. (1995). The revised Children's Knowledge of Abuse Questionnaire: Development of a measure of children's understanding of sexual abuse prevention concepts. *Social Work Research, 19*(2), 112-120.
- Tutty, L. M. (1997). Child sexual abuse prevention programs: Evaluating who do you tell. *Child Abuse & Neglect, 21*(9), 869-881.
- Udommana, P. (2011). *Measuring classroom quality: Lessons learned from applying Western measures in Thai early childhood classrooms*. University of California, Los Angeles.
- UNICEF. (2017). *Survey on Professional groups' knowledge, attitudes, practices and social norms associated with violence against children*. Available: <https://www.unicef.org/northmacedonia/media/>
- Urbann, K., Bienstein, P., & Kaul, T. (2020). The evidence-based sexual abuse prevention program: strong with Sam. *The Journal of Deaf Studies and Deaf Education, 25*(4), 421-429.
- Valik, A., Holmqvist Gattario, K., Lunde, C., & Skoog, T. (2022). PSH-C: A measure of peer sexual harassment among children. *Journal of Social Issues*.
- van Alphen L. M., Dijker A. J. M., Bos A. E. R., van den Borne B. H. W., Curfs L. M. G. (2011). Explaining not-in-my-backyard responses to different social groups: The role of group characteristics and emotions. *Social Psychological and Personality Science, 2*, 245-252.
- van Lieshout, S., Mevissen, F. E., van Breukelen, G., Jonker, M., & Ruiters, R. A. (2019). Make a move: A comprehensive effect evaluation of a sexual harassment prevention program in Dutch residential youth care. *Journal of Interpersonal Violence, 34*(9), 1772-1800.
- Vaux, A. (1988). *Social support: Theory, research, and intervention*. New York, NY: Praeger Publishers.
- Von Suchodoletz, A., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher–child interactions in German preschools. *Early Childhood Research Quarterly, 29*(4), 509-519.
- Vorst, H. C. M. (1990). *Manual and elucidation of the school questionnaire*. Nijmegen, The Netherlands: Berkhout.
- Wade, M., Fox, N. A., Zeanah, C. H., & Nelson III, C. A. (2019). Long-term effects of institutional rearing, foster care, and brain activity on memory and executive functioning. *Proceedings of the National Academy of Sciences, 116*(5), 1808-1813.
- Wade, M., Zeanah, C. H., Fox, N. A., & Nelson, C. A. (2020a). Global deficits in executive functioning are transdiagnostic mediators between severe childhood neglect and psychopathology in adolescence. *Psychological Medicine, 50*(10), 1687-1694.



- Wade, M., Zeanah, C. H., Fox, N. A., & Nelson, C. A. (2020b). Social communication deficits following early-life deprivation and relation to psychopathology: a randomized clinical trial of foster care. *Journal of Child Psychology and Psychiatry*, 61(12), 1360-1369.
- Wade, M., Zeanah, C. H., Fox, N. A., Tibu, F., Ciolan, L. E., & Nelson, C. A. (2019). Stress sensitization among severely neglected children and protection by social enrichment. *Nature Communications*, 10(1), 5771.
- Walsh, K., Brandon, L. & Chirio, L. (2012a) Mother-Child Communication about Sexual Abuse Prevention. *Journal of Child Sexual Abuse*, 21(4), 399–421.
- Walsh, K., Rassafiani, M., Mathews, B., Farrell, A., Butler D. (2012b). Exploratory factor analysis and psychometric evaluation of the Teachers Reporting Attitude Scale for Child Sexual Abuse. *J Child Sex Abus*, 21(5), 489–506.
- Walsh, K., Zwi, K., Woolfenden, S., & Shlonsky, A. (2015). School-based education programmes for the prevention of child sexual abuse: A systematic review. *Campbell Systematic Reviews*, 11(1), 1-180.
- Wangamati, C. K., Mdala, I., Ogutu, B., Sokoine, K., Ochieng, M., Majikata, S., ... & Kelly, S. A. (2022). Assessment of whole school approach intervention to reduce violence affecting children in and around schools in Kenya and Tanzania: protocol for a before-and-after, mixed-methods pilot study. *BMJ Open*, 12(5), e055231.
- Ward, K. J. (2002). *Making the invisible visible: A feminist evaluation of an adolescent gender violence prevention program*. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.
- Warden, D., Moran, E., Gillies, J., Mayes, G., & Macleod, L. (1997). An evaluation of a children's safety training programme. *Educational Psychology*, 17(4), 433-448.
- Weatherley, R., Hajar, A. S., Noralina, O., John, M., Preusser, N., & Yong, M. (2012). Evaluation of a school-based sexual abuse prevention curriculum in Malaysia. *Children and Youth Services Review*, 34(1), 119-125.
- Wechsler, D. (2012). *Wechsler preschool and primary scale of intelligence—fourth edition*. The Psychological Corporation San Antonio, TX.
- Weeks, E. A., Whitaker, D. J., Pendarvis, S., Finkelhor, D., Neal-Rossi, C., & Rivers, D. (2021). Evaluation of the child safety matters curriculum for improving knowledge about victimization among elementary school children: a randomized trial. *Journal of Child Sexual Abuse*, 30(8), 977-993.
- Wehmeyer, M. L. (1995). *The Arc's Self-Determination Scale: Procedural Guidelines*. The Arc.
- White, C., Shanley, D. C., Zimmer-Gembeck, M. J., Walsh, K., Hawkins, R., Lines, K., & Webb, H. (2018). Promoting young children's interpersonal safety knowledge, intentions, confidence, and protective behavior skills: Outcomes of a randomized controlled trial. *Child Abuse & Neglect*, 82, 144-155.

- White, C., Shanley, D. C., Zimmer-Gembeck, M. J., Walsh, K., Hawkins, R., & Lines, K. (2019). Outcomes of In Situ Training for Disclosure as a Standalone and a Booster to a Child Protective Behaviors Education Program. *Child Maltreatment, 24*(2), 193-202.
- Williamson, V., Hiller, R. M., Meiser-Stedman, R., Creswell, C., Dalgleish, T., Fearon, P., et al. (2018). The parent trauma response questionnaire (PTRQ): development and preliminary validation. *Eur J Psychotraumatol. 9*(1), 1478583.
- Wolfe, D. A., MacPherson, T., Blount, R., & Wolfe, V. V. (1986). Evaluation of a brief intervention for educating school children in awareness of physical and sexual abuse. *Child Abuse & Neglect, 10*(1), 85-92.
- Wolfe, D. A., Scott, K., Wekerle, C., Pittman, A. L. (2001). Child maltreatment: risk of adjustment problems and dating violence in adolescence. *J Am Acad Child Adolesc Psychiatry, 40*(3), 282-289.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Riverside Publishing.
- Woodcock, R. W., Mather, N., & Schrank, F. A. (2004). *Woodcock-Johnson III diagnostic reading battery*. Riverside Publishing.
- World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th ed.).
- Wurtele, S. K., Gillispie, E. I., Currier, L. L., & Franklin, C. F. (1992). A comparison of teachers vs. parents as instructors of a personal safety program for preschoolers. *Child Abuse & Neglect, 16*(1), 127-137.
- Wurtele, S. K., Hughes, J., & Owens, J. S. (1998) An Examination of the Reliability of the “What If” Situations Test: A Brief Report. *Journal of Child Sexual Abuse, 7*(1), 41–52.
- Wurtele, S. K., Kast, L. C., & Melzer, A. M. (1992). Sexual abuse prevention education for young children: A comparison of teachers and parents as instructors. *Child Abuse & Neglect, 16*(6), 865-876.
- Wurtele, S. K., Kast, L. C., Miller-Perrin, C. L., Kondrick, P. A. (1989). Comparison of programs for teaching personal safety skills to preschoolers. *Journal of Consulting and Clinical Psychology, 57*(4), 505-511.
- Wurtele, S. K., Saslawsky, D. A., Miller, C. L., Marrs, S. R., & Britcher, J. C. (1986). Teaching personal safety skills for potential prevention of sexual abuse: a comparison of treatments. *Journal of Consulting and Clinical Psychology, 54*(5), 688.
- Yılmaz Irmak T., Kızıltepe R., Aksel Ş., Güngör D., & Eslek D. (2018). “I’m learning to protect myself with Mika.” Efficacy of sexual abuse prevention program. *Turkish Journal of Psychology, 33*(81), 41–61.
- Yılmaz Irmak, T., Kızıltepe, R., Eslek, D., & Aksel, Ş. (2020). *I am learning to protect myself with Mika: Sexual abuse prevention program for preschool children a manual for practitioners*. Ege University.

- Zeanah, C. H., Smyke, A. T., Koga, S. F., Carlson, E. (2005). Attachment in institutionalized and community children in Romania. *Child Dev*, 76(5), 1015–1028.
- Zeyneloğlu, S., & Terzioğlu, F. (2011). *Development and psychometric properties gender roles attitude scale*. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi.
- Zhang, W., Chen, J., Feng, Y., Li, J., Liu, C., & Zhao, X. (2014). Evaluation of a sexual abuse prevention education for Chinese preschoolers. *Research on Social Work Practice*, 24(4), 428-436.
- Ziv, Y., & Hotam, Y. (2015). Theory and measure in the psychological field: The case of attachment theory and the strange situation procedure. *Theory & Psychology*, 25(3), 274-291.
- Zolotor, A. J., Runyan, D. K., Dunne, M. P., Jain, D., Péters, H. R., Ramirez, C., ... & Isaeva, O. (2009). ISPCAN Child Abuse Screening Tool Children's Version (ICAST-C): Instrument development and multi-national pilot testing. *Child Abuse & Neglect*, 33(11), 833-841.

## 5. Appendix

### 5.1. Child Wellbeing

#### Cognitive Functioning

Daberon Screening for School Readiness (Daberon-2; Danzer et al., 1991)

Used in Francis and Baker-Henningham (2021).

Peabody Picture Vocabulary Test (Dunn & Dunn, 1981)

Used in Fryer and colleagues (1987) and MacIntyre and Carr (1999).

Monetary Incentive Delay (MID) task (Helfinstein et al., 2013)

Used in Colich and colleagues (2021).

Cambridge Automated Neuropsychological Test Battery (CANTAB; Robbins et al., 1994)

Used in Colich and colleagues (2021), Wade and colleagues (2019), and Wade and colleagues (2020).

Wechsler Preschool Primary Scale of Intelligence II (Wechsler, 2000)

Used in Johnson and colleagues (2010).

Wehmeyer's Self-Determination Scale (Wehmeyer, 1995)

Used in Dryden and colleagues (2014).

Woodcock-Johnson III (Woodcock et al., 2001, 2004)

Used in Baker-Henningham and colleagues (2019) and Baker-Henningham and colleagues (2021).

\*Educational tests (Devries et al., 2015)

\*Child engagement and sustained attention (Green et al., 2020)

\*Children's recall (Huang & Cui, 2020)

\*IQ (Humphreys et al., 2015)

\*Child performance at school and absence (Karmaliani et al., 2020)

\*Staying in school (Knight et al., 2020)

\*Numeracy and literacy test (López García et al., 2022)

\*Academic performance (Masath et al., 2020)

\*Numeracy and literacy test (Scharpf et al., 2021)

\*Academic performance (Scharpf et al., 2021)

\*Motivation to learn (Taylor et al., 2021)

\*Observations of interactions between children (Troller-Renfree et al., 2015)

## **Child Physical Health and Development**

ACE-Related Health Conditions Checklist (ACES Aware, n.d.)

Used in Perez Jolles and colleagues (2021).

Inventory of Parent and Peer Attachment (Armsden et al., 1987)

Used in Hiller and colleagues (2021).

Bayley Scales of Infant and Toddler Development II (BSID-II; Bayley, 1993)

Used in Johnson and colleagues (2010) and Smyke and colleagues (2010).

Dot probe task (Bradley et al., 1999; Mogg et al., 1997)

Used in Troller-Renfree and colleagues (2015).

Youth Risk Behaviour Survey (YRBS; Centres for Disease Control and Prevention, 2014)

Used in Edwards and colleagues (2022) and Taylor and colleagues (2010).

DISYPS-II (Döpfner et al., 2008)

Used in Urbann and colleagues (2020).

Health and Behavior Questionnaire (HBQ; Essex et al., 2002)

Used in Colich and colleagues (2021), Troller-Renfree and colleagues (2015), Wade and colleagues (2019), Wade and colleagues (2020a, 2020b).

PediBIRN-4 (Hymel et al., 2013, 2014)

Used in Hymel and colleagues (2021).

Pediatric Symptom Checklist (PSC; Pagano et al., 2000)

Used in López García and colleagues (2022), Perez Jolles and colleagues (2021), and Scharpf and colleagues (2021), and Taylor and colleagues (2021).

KIDSCREEN-10 (Ravens et al., 2014)

Used in López García and colleagues (2022) and Scharpf and colleagues (2021).

Disturbances of Attachment Interview (DAI; Smyke et al., 2002)

Used in Troller-Renfree and colleagues (2015).

Child Health Utility 9D (CHU-9D; Stevens et al., 2012)

Used in Hiller and colleagues (2021).

ICD-10 (World Health Organization, 2019)

Used in Urbann and colleagues (2020).

DTI Scan Protocol and Image Preprocessing

Used in Bick and colleagues (2015).

PsychLab 101 (App available)

Used in López García and colleagues (2022) and Scharpf and colleagues (2021).

Retinal Examination and Skeletal Survey

Used in Hymel and colleagues (2021).

\*Inhibitory control (Baker-Henningham et al., 2021)

\*Physical growth (Johnson et al., 2010)

## **Emotional Functioning**

Fear Scale (Anderson-Varney, 1987)

Used in Ratto and Bogat (1990).

Children's Coping Strategies Checklist (Ayers & Sandler, 1999; De Boo, 2005)

Used in Hoefnagels and colleagues (2021).

Culture Free Self-Esteem Inventory (Battle, 1981)

Used Kraizer (1991) and MacIntyre and Carr (1999).

Teen Conflict Survey (Bosworth & Espelage, 1995)

Used in Daigneault and colleagues (2012).

Social Cognitive Map (SCM) Technique (Cairns et al., 1995)

Used in López García and colleagues (2022) and Scharpf and colleagues (2021).

PROMIS Pediatric Peer Relationship Scale (DeWalt et al., 2013)

Used in López García and colleagues (2022) and Scharpf and colleagues (2021).

Network of Relationship Inventory (NRI; Furman & Buhrmester, 2009)

Used in Tang and colleagues (2021).



School Social Behaviour Scales 2 (SBSS-2; Merrell, 2008)

Used in Baker-Henningham and colleagues (2021).

Friendship Quality Questionnaire (FQQ; Parker & Asher, 1993)

Used in Tang and colleagues (2021).

Marlowe-Crowne Social Desirability Scale-Form C (MCSD-C; Reynolds, 1982)

Used in Hillenbrand-Gunn and colleagues (2010).

Rutter Teachers' Scale (Rutter, 1967)

Used in MacIntyre and Carr (1999).

Social Communication Questionnaire (SCQ; Rutter et al., 2003)

Used in Wade and colleagues (2020).

Generalized Self-Efficacy (GES; Schwarzer et al., 1995)

Used in Dryden and colleagues (2014).

Preschool Self-Regulation Assessment (PSRA; Smith-Donald et al., 2007)

Used in Baker-Henningham and colleagues (2019) and Baker-Henningham and colleagues (2021).

Vaux Social Support Record (Vaux, 1988)

Used in Daigneault and colleagues (2012).

\*Harter Perceived Competence Scale for Children (Fryer et al., 1987)

\*Perceived norm and self-efficacy (Huang & Cui, 2020)

\*Delinquency (Jewkes et al., 2019)

\*Agency, self-regulation, psychological assets, and peer connectedness (Knight et al., 2020)

\*Self-efficacy subscale (McElearney et al., 2021)

\*Observation of play with a friend (Tang et al., 2021)

\*Self-esteem and self-efficacy toward self-control and communication (van Lieshout et al., 2019)

## **Mental Health**

Child Behaviour Checklist (CBCL; Achenbach & Rescorla, 2000)

Used in Green and colleagues (2020) and Sullivan and colleagues (1992).

Mood and Feelings Questionnaire (MFQ; Angold et al., 1995)

Used in Fabbri and colleagues (2021) and Hiller and colleagues (2021).

Dutch Social Anxiety Scale for Children (Cohen-Kettenis & Dekking, 1980)

Used in Taal and Edelaar (1997).

Social Anxiety Scale for Children (Dekking, 1977)

Used in Hoefnagels and colleagues (2021).

Child Anxiety Related Disorders Questionnaire (SCARED; Essau et al., 2002)

Used in Czerwinski and colleagues (2018).

Eyberg Child Behaviour Inventory (ECBI; Eyberg & Ross, 1978)

Used in Baker-Henningham and colleagues (2021) and Francis and Baker-Henningham (2021).

Child PTSD Symptom Scale – Interview Schedule (CPSS-I; Foa et al., 2018)

Used in Hiller and colleagues (2021).

Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001)

Used in Baker-Henningham and colleagues (2019, 2021a, 2021b), Devries and colleagues (2015), Francis and Baker-Henningham (2021), Hiller and colleagues (2021), and Masath and colleagues (2020).

Children’s Depression Inventory 2 (CDI-2; Kovacs, 2010)

Used in Karmaliani and colleagues (2020).

Children’s Nowicki-Strickland Internal External Locus of Control Inventory (Nowicki & Strickland, 1973)

Used in Kraizer (1991).

CRIES-8 (Perrin et al., 2005)

Used in Hiller and colleagues (2021).

Revised Children’s Manifest Anxiety Scale (RCMAS and RCMAS-2; Reynolds & Richmond, 1985, 2008)

Used in McElearney and colleagues (2021), Oldfield and colleagues (1996), White and colleagues (2018), and White and colleagues (2019).

Child and Adolescent Trauma Screen (CATS; Sachser et al., 2017)

Used in Hiller and colleagues (2021).

State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973)

Used in Hazzard and colleagues (1991) and Oldfield and colleagues (1996).

\*Childhood trauma and depression scores (Jewkes et al., 2019)

\*DISC-IV symptoms (Humphreys et al., 2015)

## **5.2. Intervention Feasibility and Implementation Measures**

\*Student satisfaction questionnaire (Barron & Topping, 2013)

\*Program fidelity checklist (Barron & Topping, 2013; Nickerson et al., 2019; Scharpf et al., 2021; Ssenyonga et al., 2022)

\*Implementation Fidelity (Bright et al., 2022)

\*Exposure to intervention (Devries et al., 2018)

\*Children's reaction to prevention program (Grendel, 1991)

\*Children's satisfaction with the workshop (Hébert et al., 2001)

\*Motivation to participate and acceptability of integration (Hecker et al., 2021)

\*Outcome Measurement System (OMS; Herbert & Bromfield, 2021)

\*Level of satisfaction with education (Kang et al., 2022)

\*Program implementation and exposure survey (Knight et al., 2018)

\*User satisfaction (König et al., 2020)

\*Time spent on guides (Konijnendijk et al., 2021)

\*Professional adherence to guides (Konijnendijk et al., 2019)

\*Child Exit Interview (Kraizer, 1991)

\*Teacher questionnaire (Kraizer, 1991)

\*Program evaluation (MacIntyre & Carr, 1999)

\*Evaluation of intervention (Masath et al., 2020)

\*ICC-T Training Evaluation (Nkuba et al., 2018)

\*Semi-structured interviews and reflective diaries (Peker et al., 2020)

\*Parent-child workbook (Ratto & Bogat, 1990)

- \*Webinar evaluation (Taylor et al., 2021)
- \*Teacher questionnaire (Warden et al., 1997)
- \*Pre/posttest questionnaire (Weatherley et al., 2012)
- \*Child attendance (White et al., 2018)
- \*End of session checklist (White et al., 2018)
- \*Parent satisfaction (White et al., 2018)