

**Institutional Responses to Child Maltreatment:
Guidebook to the Evidence
(Summaries and Syntheses of Rigorous Studies ‘What
Works’ Studies)**

August 2023

Prepared with technical experts of the Campbell South Asia

Corresponding author:

Caroline Fiennes, Director, Giving Evidence

Caroline.fiennes@giving-evidence.com

Other products related to this Guidebook – such as the Evidence and Gap Map from which this is drawn – are published at www.giving-evidence.com/csa

Introduction to the Guidebook

In order that work in child protection can be as effective and evidence based as possible, Porticus, a philanthropic funder, wanted to find the rigorous causal evidence around ‘what works’ in child abuse in institutional contexts. We searched for such studies, collated them, coded them, and presented them in the Evidence and Gap Map¹ (EGM) (The summary report and the full report are at www.giving-evidence.com/csa). Each study on the EGM examines the effect of (at least) one intervention of interest on (at least) one outcome of interest. The EGM has a grid (see Appendix 1), in which the rows are interventions, and the columns are outcomes: a study examining the effect of Intervention X on Outcome Y appears in the cell XY, and if the study examines several interventions (and/or several outcomes), it will appear in the several relevant cells. The EGM is designed for non-specialists, to guide them to relevant evidence. The EGM shows where there is such evidence and where there isn’t; it does not show what the evidence says.

This ‘guidebook’ complements the EGM by summarising what the evidence on the Evidence and Gap Map says. Our aim is to help funders, policymakers, practitioners and others to make evidence-informed decisions, and also to find easily the evidence relevant to the decisions they need to make.

This Guidebook begins with a summary of main findings of the Evidence and Gap Map. It then has three types of content, presented in three sections. There is a glossary at the end.

<p>Section 1:</p> <ul style="list-style-type: none"> a. Introduction and comment about the whole evidence-base b. Some points on reading social science studies, and this Guidebook 	<p>1a: Overall findings (e.g., that most programmes which have been studied do work 😊). Also, issues commonly reported, commonly omitted, comments which apply to many or all studies, e.g., on the types of outcomes used, the sample sizes, the sizes of effects observed, and quality of studies.</p> <p>1b: This section includes guidance on reading social science studies and hence using the material, especially for applying results from one place to another. Also, we describe how the Guidebook treated cells of particular types, protocols, etc.</p>
<p>Section 2: Syntheses of the studies in cells which have three or more studies (which we call ‘heavy’ cells)</p>	<p>Twenty-four cells on the EGM have 3+ studies, and we have synthesized the studies in each one to provide a single view of the evidence in each such cell. They are presented in order in which you would encounter the cells if ‘reading’ the EGM in the normal way: top to bottom (so, starting with prevention), and left to right. These cells vary hugely in their number of studies: most have 3-8 studies; some 9-18; and one has >80.</p>
<p>Section 3: Summaries of individual studies which appear in cells on the EGM which have just one or two studies (which we call ‘light’ cells)</p>	<p>These are presented in alphabetical order. There are eight such studies.</p>

¹ This updated EGM was published in July 2023. It is based on a search conducted in early 2022, and therefore should contain all relevant studies published up to that point.

Contents

Introduction to the Guidebook.....	2
Contents.....	3
Reminder: main findings of the EGM: where there is evidence and where there isn't	6
Section 1a: Introduction and comment about the whole evidence-base	7
1. Overall findings in the Guidebook: what that evidence says.....	7
Effect sizes.....	9
2. Some comments on the studies on the EGM	10
Geography.....	10
Theories on which the interventions are based / theories of change	10
Age of the studies, and whole-school approaches	11
Outcomes which are measured	11
Length of trials, and follow-up periods.....	12
Cost information	12
Sample sizes in the studies	13
Who created and ran the interventions, and the theories on which they are based?	14
What else do we know about these topics?	14
Section 1b. What this Guidebook contains.....	15
Studies vs papers	15
Studies that appear in multiple cells and/or have multiple outcomes.....	15
Protocols	16
Treatment of systematic reviews in cells which contain primary studies	16
Treatment of cells with just systematic reviews and no primary studies	16
Section 2: Syntheses of studies in cells which contain 3+ studies.....	17
Prevention / Institutional Safeguarding Practice: Operations.....	17
Prevention / Institutional Safeguarding Practice: Environment.....	23
Prevention / Adult Institutional Caregiver.....	28
Prevention / Disclosure: Disclosure rates	32

Prevention / Child Safety: Maltreatment Occurrence/Recurrence	38
Prevention / Child wellbeing: cognitive functioning.....	43
Prevention / Child Wellbeing: Knowledge and Awareness	46
Prevention / Child Wellbeing: Mental Health.....	54
Prevention / Child Wellbeing: Social-Emotional Functioning	60
Prevention / Adult Perpetrator: Desistance	66
Prevention / Child or Youth Offender: Maltreatment Behaviour.....	70
Prevention / Parent or Caregiver: Knowledge and Awareness	73
Disclosure / Child Maltreatment Disclosure	79
Disclosure / Child Wellbeing: Child Knowledge and Awareness.....	82
Response / Institutional Safeguarding Practice: Operations	86
Response / Institutional Safeguarding Practice: Environment	93
Response / Adult institutional caregiver: behaviour, knowledge, attitudes	98
Response / Disclosure: Disclosure rates	102
Response / Child maltreatment occurrence/recurrence.....	105
Response / Child Wellbeing: Mental Health.....	108
Response / Child Wellbeing: Child Knowledge and Awareness.....	111
Treatment / Child Wellbeing: Cognitive Functioning	114
Treatment / Child Wellbeing: Social Emotional Functioning.....	119
Treatment / Child Wellbeing: Mental Health	123
Section 3: Summaries of individual studies which appear in cells with only one or two studies.....	130
Summary of three studies from one RCT (Bucharest Early Intervention Project): Johnson (2010), Smyke (2010), Bick (2015) (Treatment / child physical health, and child social-emotional functioning).....	130
Edwards (2019) (Prevention / adult maltreatment behaviour).....	138
Herbert, Bromfield (2021) (Disclosure; Operational Practice)	142
Hoefnagels (2021) (Disclosure; Adult Institutional Care Provider).....	146
Karmaliani (2020) (Prevention; Adult Perpetrator: Maltreatment Behaviours)	150
Merrill (2018) (Prevention / institutional culture).....	153

Nkuba (2018) (Prevention / institutional culture)159

White (2019) (Disclosure; Child Mental Health)164

Glossary167

Appendix 1: The Visual Evidence and Gap Map170

Appendix 2: Systematic Reviews on the EGM which are in cells empty apart from themselves1

Appendix 3: Guidance on reading social science studies, and using this Guidebook8

 Counterfactuals / comparison groups8

 Translating evidence to a difference context9

 Measures11

 When results are measured12

 Attenuation / effects which fade12

 Attrition13

 Programme participants13

 Quality of studies / risk of bias14

Reminder: main findings of the EGM: where there is evidence and where there isn't

The main findings of the EGM were as follows. There is more detail about this in the summary report and the full report about the EGM¹. The visual map is in Appendix 1.

- The map has a total of 136 studies. Of these, 108 are complete primary studies (including 79 RCTs and 29 QEDs), 8 ongoing primary studies (study protocols) and 20 systematic reviews.
- **Geographically, the studies don't seem to match where the world's population is.** Most studies are from high-income countries. The US dominates, with 44 of the 108 completed primary studies. The EGM has no primary studies from India, only one from Pakistan and three from China. Africa is represented better with 10 completed primary studies. Many interventions which have 'worked' in developed countries have not yet been tested in low- and middle-income countries.
- **The major concentration of studies is in education-based prevention programmes,** including both early education and school settings. 93 of the 108 primary studies (including 8 study protocols) included in the EGM assessed these programmes.
 - The most commonly studied interventions in primary studies (77 studies) were about preventing sexual abuse by teaching children, in schools, about how to recognise it (e.g., good touches vs bad touches), and how to deal with it / report it. These interventions of course put the onus for prevention on the child. There were many fewer studies of institutional responses to prevention (e.g., on training staff to recognize abuse and change their practices to help prevent it. Glass walls in classrooms would be another example, not that we found any studies of this).
- **Most studies reported intermediate outcomes,** such as children's acquisition and retention of knowledge, but not actual disclosure of incidence. Most studies are quite short duration (presumably for reasons of funding), which presents challenges for ascertaining the real effects of the programmes, especially on outcomes such as disclosure.
- **Most of the studies are about sexual abuse:** this is the focus for 77 of the 108 completed primary studies.
- **Most of the studies are about prevention:** 93 primary studies (completed and on-going studies), and 17 systematic reviews. We found three primary studies and three systematic reviews of interventions specifically aiming to facilitate disclosure².
- **No study is specifically in church settings.** One study (Rheingold) includes clergy, but amongst other childcare professionals and does not split out results for clergy.
- **There is a paucity of primary studies about treating survivors from the last nearly 20 years.** The study from Romanian orphanages started in 2001, and there is one non-randomised trial published in 1992.
- **Seven studies focused on children particularly at-risk.**
- A majority of studies were rated as having moderate level overall [risk of bias](#)³ (explained [here](#)). Most of the systematic reviews had high risk of bias. Reassuringly, a considerable number of primary students (20) were rated as 'low risk of bias' which indicates high confidence in their reported results.
- **Only one study looked at educational attainment as an outcome** (the *Good School Toolkit* study in Uganda).
- **Institutional safeguarding practice was studied in 17 primary studies.**
- **Very few studies came from practitioners and non-profits.**

² Disclosure is sometimes a consequence of some of the interventions studied, such as teaching children about abuse and how to prevent it. We are distinguishing here between studies that are primarily about prevention, vs primarily to encourage disclosure.

³ This Guidebook takes the assessments of risk of bias and study quality from the EGM.

- **Cost information was missing from almost all studies.**
- **Very few studies came from South Asia.** This is an evidence gap.
- **Growth in studies from Africa.** There is a good growth in studies published from African countries recently, most of them are published during 2020-22. The focus of studies is on preventive interventions, no study on interventions related to disclosure or response, a few systematic reviews related to treatment.
- **Growth in studies on interventions related to response and treatment, but disclosure is still an evidence gap.**
- **Most studies are based on only small sample.** Small sample size studies are able to detect when the effect is large, whereas large sample size can detect even smaller effects of an intervention. This is a possible reason why in many of the studies findings are inconclusive and it demands more research with bigger sample size.

Section 1a: Introduction and comment about the whole evidence-base

This section discusses the overall evidence in the EGM. We look at its quality, the overall findings, describe the types of content in the Guidebook. Section 1b talks about precisely what is in the Guidebook, e.g., how we treated cells with particular numbers of studies, and particular types of studies. Some points to remember when reading and using any social science studies – including those on the Guidebook – is in the appendix: readers unfamiliar with social science studies are advised to read that first.

1. Overall findings in the Guidebook: what that evidence says

Most of the interventions studied have some positive effect(s). Very few of the interventions which have been studied have no effect on any of their intended outcomes. However, to be clear, a positive effect means that the intervention produces *some positive effect*: they do not eliminate the problem, but rather reduce it somewhat. Most of the studied interventions have a modest effect.

Where programmes do work, the effects are usually fairly small. An effective programme may improve knowledge by 20-30 percent and reduce abuse by 10-20%. (The modest-ness of effects is true of most social interventions in any sector.) For example, the Good Schools Toolkit, which is studied in multiple papers on the EGM, reduced violence from being experienced by 80% of students in the previous term (clearly a giant amount) to being experienced by ‘only’ 60% of them.

And where there are effects, they tend to weaken over time. For example, the most studied interventions are to teach children about good touches vs bad touches – and people forget things over time. This attenuation supports the need for institutionalizing proven approaches into repeated procedures rather than one-off interventions. An example is in Rheingold, an RCT of training for childcare professionals: the knowledge of the trained group was improved relative to the control group immediately after the training, but three months later, the gap had shrunk.

On the other hand, none of the studied interventions seems to create harm :-). Remarkably, we found **no study reporting any adverse effects of the interventions*** - though many studies did look for them: such as whether children’s anxiety increased when they learned about ‘bad touches’.

*There are some studies that appeared to point to adverse effects, but none is very clear, and they may all be related to increased *reporting* rather than increased *incidence*. One is the *Bringing in the Bystander* programme (see Edwards 2019). This aimed to prevent gender-based violence and sexual harassment. Students who got the intervention reported committing significantly more violence against their dating partners than did students in the control group. This is probably because the programme sensitised them to what constitutes this type of violence, so they reported it more, rather than because the programme made them commit it more. Similarly, one of the studies of the *Red Flag / Green Flag People* programme, which trained children to avoid abuse, found that six months after the programme, the intervention group children reported more abusive encounters than did children in the control group (who reported none). Again, this may well be because the intervention encouraged reporting behaviour, rather than increases actual abuses. A third was in Taylor 2010, in which students who got the programme reported (themselves) committing more violence against their dating partners, though this might be because the programme taught them that behaviours, they had hitherto considered normal were in fact violent. For example, many students did not know that sex between minors is legally considered rape.

Obviously, this does not guarantee that other interventions - existing or new - will not create harm, so stay on your guard about them. It simply means that none of the studies we identified had found evidence of harm.

More evidence is needed in many cells on the EGM: some of the cells have too little evidence to enable a clear view of the effects of the programmes, and/or the evidence is too weak. E.g., cells with interventions related to “disclosure” and “treatment” of child maltreatment and cells with outcomes related to adult perpetrator and outcomes related to child/youth offender. Even where studies exist, there are often deficiencies in study design or implementation which reduce our confidence in what the evidence says. We have not written in every such cell that ‘more evidence is needed’ simply to avoid repetition, but it is nonetheless true.

Some interventions have no effect, or at least, no effect on some outcomes. For example, a programme run in the Netherlands with at-risk boys living in residential care (studied in van Lieshout) aimed to reduce sexual harassment by them but found no effect. This finding is consistent with a finding across social sectors (ie., outside child protection), that around 80 per cent of interventions in all sectors have small or no effect.

There are some mixed results. Some interventions found a positive effect on some outcomes but no effect on others. For example, Edwards (2019) examined a bystander programme in US high schools: it found no statistically significant effects on participants stopping harassment, speaking against blame or excuses, or talking to an upset person, but did find improvement in victim empathy and denying that rape is possible or had occurred. And some cells with multiple studies found that some interventions work and some achieve nothing. The van Lieshout study mentioned is in a cell with four other primaries: of these five, two found a positive effect, and three found no effect. This is also unsurprising because, even within one cell, interventions, populations, comparison groups and outcome measures can vary substantially.

We re-state here our earlier comment about the quality of the evidence on the EGM: studies which give only low confidence, or which have high risk of bias, are quite likely to report ‘answers’ that are not correct, and hence to be misleading. Whilst the direction of bias introduced by evaluation deficiencies is not always known, it is generally the case that weaker evaluation designs find larger effects than stronger

designs – the evaluator Rossi called this the ‘Stainless Steel Law of Evaluation’. The reality of these interventions’ true effectiveness may not be as rosy as this overall finding implies.

Effect sizes

The effect sizes of the interventions studied varied quite considerably but were generally pretty modest. This is normal for social science studies. There is no vaccine for child abuse.

The size of an intervention’s effect is the extent to which the intervention increases or decreases an outcome: for example, the intervention might increase literacy by five percentage points. An example is in *the Good School Toolkit* (GST) in Uganda: here, amongst students whose schools did not get the programme, 80% said that they experienced violence in the previous term, whereas amongst students whose schools did do the programme, ‘only’ around 60% reported that. This implies that the GST reduced violence by 20 percentage points, or a quarter.

Often, the results of the studies on the EGM are rather unclear. This is for a few reasons.

First, some studies report results in ways that are challenging to translate to a percent change. Literally all the outcomes assessed in the studies on the EGM were self-reported, e.g., by parents or teachers, or by children: sometimes children’s outcomes were based on parents’ views, e.g., of their child’s confidence. This means that various measurement scales were used by studies to assess things such as knowledge levels or behaviour change.

Second, though some of the scales were validated, i.e., they had been certified to be reliable for a given outcome, others were not. The researchers may have created a scale especially for the study, and not validated it. This makes it difficult to assess how well an intervention works, i.e., how much change in outcomes it brought about, across all studies. While there are statistical methods (meta-analysis) to try to convert the results from different scales into a common metric, that is not usual practice for an EGM, which is an overview of what evidence is available rather than what it says.

And third, some studies in the EGM reported results unclearly so we do not repeat them in our summaries. For example, Merrill (2018) uses different ranges of Likert scales⁴ for different outcomes: one outcome is measured on a scale 0-3, but for another outcome, the scale is 0-12. This prevents us comparing them or identifying the size of the impact.

However, we do sometimes have information on whether the results were statistically significantly better for the intervention group compared to the control group (that did not get the intervention or got a different intervention), which provides us with a good understanding of whether an intervention was likely effective or not.

Please note that when a summary talks about ‘a significant reduction’, it means that the reduction was *statistically* significant (i.e., not just caused by random noise), rather than meaning that the reduction was substantial (large). If a study has a very large sample, then even quite a small effect may be found to be statistically significant.

⁴ A Likert scale is a type of survey scale, which has a question with a series of answers to choose from, ranging from one extreme attitude to another, normally with a moderate or neutral option. For example: “How likely are you to recommend this product?: very likely, somewhat likely, no opinion, somewhat unlikely, very unlikely.”

Some interventions showed fairly dramatic results. For example:

- The *Good School Toolkit* (in Devries 2015) appeared to reduce violence by school staff against pupils in the last week by more than half: from 40% of staff using it, to 15%.
- School-based interventions can be very effective at increasing disclosure (see page 29 for a synthesis of ten papers about nine studies.) A programme *Red Flag, Green Flag People* had 20 children disclosing vs none in the control group, when studied in both 1987 and 1989; despite small sample sizes, the *Good School Toolkit* generated over 400 additional referrals due to disclosures; and a programme in Spain had eight disclosures vs two in the control group.

2. Some comments on the studies on the EGM

Geography

The studies are overwhelmingly from high-income countries. Western Europe, the US and Canada account for 63% of the completed primary studies. By contrast, the EGM has zero studies from India, one from Pakistan, two from Iran. A striking change from the original EGM (published in 2020 based on a search in 2019) to the updated map (published in 2023 based on a search in 2022) is that Africa now is much better represented in the map with 14 primary studies and 4 systematic reviews (compared we had only 3 last time). There are five studies from China. (As a reminder, our search specified no time period, so this represents all findable studies from any year till early 2022.)

This means that there is an 'easy' and important opportunity to test the effectiveness of these interventions in low- and middle-income countries. For example, the most tested interventions are programmes to prevent child sexual abuse by teaching children to distinguish good touches from bad touches. They have appeared to be effective in each of the 75 primary studies of them on the EGM. However, around half of these are from US alone and rest are from Canada, the UK, Europe, Australia, Central America, East Asia, a few from Africa. In other words, there is a strong need for these interventions to be evaluated in South Asia, Africa or South America a programme which appears to be able to save children from sexual abuse.

Theories on which the interventions are based / theories of change

Most studies in the EGM do not explicitly state the theory on which their intervention is based.

Most studies in the EGM focus on raising children's knowledge and skills to prevent abuse. The core idea of such prevention programmes is to establish a clear understanding of what constitutes unacceptable behaviour, so that children recognise and report such behaviour.

The logic of personal safety prevention programmes targeted at young children is that children are active agents in preventing child sexual abuse. If children are aware of what constitutes abuse (e.g., good touch vs bad touch) they are likely to recognise it, object to it, walk away from it and report it, all of which reduce the likelihood of abuse occurring. Programmes teach children skills for managing and reporting abusive situations.

This process is supported by parent and teacher training so they can reinforce the approach and listen to children when they discuss these issues, including reporting abuse.

Although not stated as part of the theory of change, the intervention and evaluation activities provide an opportunity for disclosure.

When the principal target of the intervention are caregivers (most frequently teachers), the programmes are often based on stages of change theories. One prominent example is the Transtheoretical Model of Behaviour Change, a six-step process that starts by making people aware of the problem of physical and

sexual abuse, and then supports planning and implementation of behaviours to deal with the problem. The final stage is when new behaviours have become the norm.

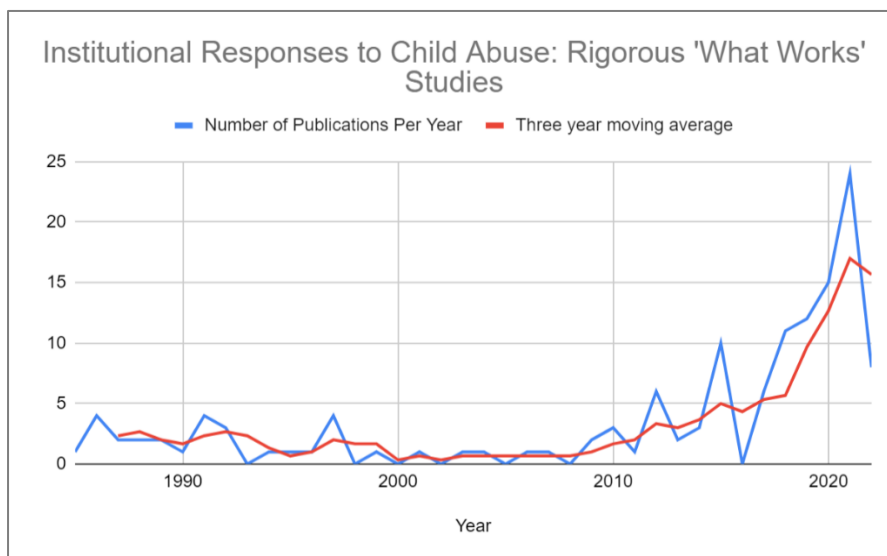
The Guidebook gives the theory of change where it is explained in the study, and sometimes we have been able to identify the implicit theory used if it wasn't stated explicitly.

Age of the studies, and whole-school approaches

The first study published in 1985 (now 38 years ago), and since then till year 2011 there were only 38 studies published i.e. 38 studies in 26 years. This also means on an average there were >1.5 studies published/year. But the publication rate grew substantially during the last decade from 2012 onwards. Around 72% of total papers included in the map are published during last one decade period (from year 2012-year 2022). 25 papers in the map are from the year 2021 alone.

Many of the EGM's school-based interventions happen just in the classroom. By contrast, the *Good Schools Toolkit* in Uganda takes a whole school approach. By analogy, in bullying, more recent work has taken a whole school approach rather than just classroom-based interventions, which can be atomised and disjointed. This may be because of the studies' age.

The rate of production of studies in this area has grown considerably:



Outcomes which are measured

Few studies record actual abuse, or disclosures of abuse. That is, few measure the effect of the intervention on the thing that we really care about. (The measures used are [discussed more here](#).) This may be for various understandable reasons: many survivors (e.g., of sexual abuse) take a long time to disclose, so trials with that outcome would need to be very long; if a form of abuse is pretty rare, researchers would need a giant sample size to detect a statistically significant number of cases; and so on. Nonetheless, do be aware that most of the studies measure some other outcome - some outcome which is *related* to our main outcome of interest but is not itself that outcome, and often an intermediate outcome.

Outcomes measured relate to the intervention's theory of change.

- Many studies measure acquisition and retention of knowledge. That may be knowledge gained by children (e.g., in the many programmes to teach children to distinguish good touches from bad touches), or by parents or staff (e.g., in the institutional training interventions).
- Some measure changes in attitudes, e.g., the proportion of participants who believe the ‘rape myths’, i.e., beliefs which blame victims and excuse perpetrators and so are less likely to recognize abuse as being abuse.
- Some studies do record incidences of abuse, e.g., the *Good School Toolkit* recorded (students’ recollections of) actual violence of various types by teachers.
- Some studies record the effects of abuse. *The Bucharest Early Intervention Project* papers examine the effect of severe neglect in early childhood on, for example, children’s height, weight, head circumference and brain matter development.

There is only one study on the EGM which reported on educational outcomes, which is the *Good School Toolkit*. This is remarkable because, though there is very little funding to address child abuse, there is much more for education / educational attainment. Quite possibly if somebody could demonstrate that a particular child abuse prevention programme increased educational attainment, it may become considerably easier to fund it.

Length of trials, and follow-up periods

The trials were generally quite short. Most measured outcomes at baseline (immediately before the intervention starts), end-line (immediately when the intervention ends) and in the months / years afterwards: in fact, many only measured up to about six months after the intervention ended. This is obviously pretty short, given that abuse can occur (or be perpetrated) at any point in a person’s life. We know that knowledge attenuates quite fast (people forget rapidly), so six months - or even 18 months - is rather unsatisfactory for showing the meaningful effect of a knowledge-gain intervention.

The reason that many trials are short is money: following participants for longer is obviously more expensive. It can also be that funding for the study is often linked to funding for the programme, so that the funding ceases when (or soon after) programme delivery finishes.

There is also the chance of greater attrition (explained below), as people lose contact, lose interest in being measured, move away, or even die from unrelated causes. It is a great credit to the funders involved in *BEIP* that they set it up with enough funding to last the course: three philanthropic foundations⁵ and a government research agency.

The exception is some of the *BEIP* papers. The trial began when the children were toddlers, and some of the papers describe measures when they were teenagers.

Cost information

Very few studies reported on the cost of the intervention(!): (For one intervention the *BEIP*, we found cost information from another source.) Consequently, we do not know the cost-effectiveness of the interventions. This is a clear evidence gap.

⁵ John D. and Catherine T. MacArthur Foundation, the Binder Family Foundation, the Sinneave Family Foundation.

None of the studies for which we wrote summaries (in Section 3) reported on the cost of the intervention. For the *GST*, we found cost data reported elsewhere. (Obviously those data pertain to all the papers written about *GST*, because they all concern the same intervention).

This is clearly a great shame, because cost is (obviously!) a major determinant of decisions by practitioners, funders and policy-makers about which programmes to run. Programme costing is not straightforward (e.g., there are multiple ways to allocate fixed costs across programme sites), so even where estimates are produced they may well not be comparable, and they are likely not transferable to other settings. This is not to say that studies should not collect cost data, but they should do so being aware of these issues.

However, it is sadly not unusual for social science studies to contain scant information about the programme design, and little or no information about programme costs⁶: reportedly, some economics journals will require cost information to be removed.

Consequently, we can say almost nothing about the cost-effectiveness of the various interventions studied.

We have made a few comments, to try to be helpful. Some interventions are pretty clearly cheaper than others, e.g., delivering training online is probably cheaper than delivering it in-person.

Sample sizes in the studies

Sample size is important in the design of studies. If effects are not likely to be large – and remember that effects sizes are generally small in real life – then a large sample is needed to detect it. If the programme is assigned across schools – that is, some schools implement the programme and a comparison group of schools do not – then it is the number of schools that matters most, whereas the number of children in each school matters far less. A sample of 500 children may sound large, but if they come from just two schools, one doing the programme and the other not, then the study is very likely ‘under-powered’ i.e., it is likely not to find a programme effect even though there is one. In many cases the number of schools is very low.

Some studies in the EGM have very small samples. This reduces the chance that any differences observed were really due to the programme(s) rather than to random chance or other factors.

For instance, several programmes ran in fewer than seven schools. This means that maybe only three schools got it and three did not. There are any number of other factors that could have influenced the result: maybe those three that got it happened to have very experienced headteachers, or happen to be in wealthier areas, or happen to have better mid-day meals than the other schools – and those factors could have affected the outcomes. By contrast, if the study had had 50 schools in the group which got the intervention and the group which didn’t, it is much less likely that all 50 would share some characteristic not shared by the other group which would influence the outcome, and hence any observed differences in outcomes is much more likely to be due to the programme.

⁶ A nice paper about why so few impact studies by economists include cost data, by a then-World Bank development economist, is here. In short, economists aren’t trained to look at costs, aren’t interested in costs(!) and cost analysis hard: <https://blogs.worldbank.org/impactevaluations/why-don-t-economists-do-cost-analysis-their-impact-evaluations> Researchers in other disciplines may differ.

Researchers should undertake ‘power calculations’, which show how they determined sample size used, and the sample size required to detect the anticipated effect. Research commissioners should require power calculations as part of the research design. The calculations should be externally reviewed. It is not so common that the calculations are wrong, though that may be the case, but that the assumptions made are too optimistic or restrictive.

Who created and ran the interventions, and the theories on which they are based?

The studies did not consistently report this: some did, some did not. In *BEIP*, the foster care programme was created by researchers and eventually supported by the local government; the *Stewards of Children* programme examined in Rheingold was run by Darkness to Light, an American NGO. But some studies do not state who ran the intervention/s. This is a shame because it might be possible to gain more information if we knew, e.g., from the implementing agency’s public materials.

Many programmes are ‘branded programmes’ meaning they are available on a commercial basis, often via non-profits working with, or set up by, research teams at US universities. *Stewards of Children*, examined in Rheingold (2015) is one such. Another is *Enough! Preventing Child Sexual Abuse in My School*, examined in Gushwa (2018). Sometimes branded programmes are evaluated by the programme designers, who sell the right to use the programme, which creates a clear conflict of interest: this is precisely what plagues pharmaceutical research, much of which is companies evaluating their own products. Unsurprisingly, in these branded social programmes (and pharmaceuticals) ‘own-evaluations’ find larger effects than do independent evaluations. There is thus a need for independent evaluations of programmes where the evidence comes from own-evaluations.

There is also an issue around branded programmes versus usual practice. The use of branded programmes is most pervasive in US education. The What Works Clearing Houseⁱⁱ lists 231 programmes to improve literacy. But surely there aren’t 231 different ways of teaching children to read. The alternative approach is to identify the elements – or components – which matter in successful programmes and to build those into standard practice. Intensity and duration normally correlate to effectiveness of social programmes (unsurprisingly). This identification requires using a strong base of primary studies combined with well-conducted systematic reviews.

Similarly, some study reports state the theories on which they were based, and some did not. Sometimes we have been able to infer or guess, based on knowledge of child development and other psycho-social theories.

What else do we know about these topics?

Where possible, the Guidebook talks about what else is known about the intervention or outcome, from other studies which are not on the EGM: many summaries and syntheses have boxed texts about this. That material from other studies outside of the EGM, e.g., research into child abuse outside of institutions.

Section 1b. What this Guidebook contains

As mentioned, the Guidebook has basically two types of content:

- For each cell which has three or more studies (which we call 'heavy' cells), the Guidebook has a synthesis of all the studies in that cell.
- Summaries of studies which appear in cells on the EGM which have just one or two studies. We call these 'light' cells.

Within that system, we had to make a number of determinations of precisely what to include ('policy decisions'), and we list those now.

Studies vs papers

We distinguish between studies and papers. A study is one set of participants, though multiple papers may be written about them.

For example, in Romania, the Ceausescu regime ran many awful orphanages. After the regime fell, researchers set up a foster programme to move children into families: there were more children than they could accommodate, so they chose the children randomly, and set up an RCT, the *Bucharest Early Intervention Project (BEIP)*. The children (both the group who remained in the orphanages and those who moved to foster care) have been studied at various points since. There was one 'study' - i.e., one experiment with one set of participants - but multiple 'papers' have been written about them: the EGM has eleven papers about that study.

Similarly, multiple papers have been written about one study (one experiment) of the *Good School Toolkit (GST)* in Uganda.

Studies that appear in multiple cells and/or have multiple outcomes

We have summaries of some studies (or papers) that appear in 'light cells'. Many studies look at multiple outcomes, so some studies appear in multiple cells. We have written just one summary of such studies - to avoid duplication and general confusion. For example, White 2019 (an RCT) relates to both prevention and disclosure. Thus, the study appears in several light cells, including cells under the intervention "prevention" and "disclosure" for the outcomes "child maltreatment disclosure"; "child mental health"; "child knowledge and awareness".

Sometimes a study in a light cell is also in a heavy cell: again the list in the summary of that study of the cells in which that study appears helps to show this. (There are summaries of these studies because they appear in light cells.)

Some heavy cells contain one or more studies which also are in light cells, so there are summaries for those: the synthesis for the heavy cell states this.

Some studies contain other outcomes which are not on the EGM at all, and we do not include those (because we only include outcomes relevant to child protection).

The summary for each such study covers all outcomes which are relevant to the EGM. The summaries say at the top that the evidence is strong / weak, etc. "in relation to the outcomes of interest", meaning the outcomes that are on the EGM.

If a systematic review is empty in relation to some outcome (i.e., it looked for primary studies which reported on that outcome but didn't find any), then we report that in the summary.

Protocols

A protocol is a plan for a study, and therefore does not have results. We include them in the EGM because they show the concentration of research activity, and they are useful to find a completed study when we update EGM. But they are not useful for this Guidebook because the protocol does yet *say* anything. outcomes “child maltreatment disclosure”; “child mental health”; “child knowledge and awareness”. If a cell had only a protocol it would be a current gap which someone is in the process of filling.

Treatment of systematic reviews in cells which contain primary studies

Cells which have only SRs are regarded as empty, for the reason described. However, to prevent the guidebook having literally nothing about them, we have provided a table of them (**Error! Reference source not found.**), the cell(s) in which they appear, and the abstract as provided in the SR.

If a cell with 9+ studies has primary studies and SR(s), we did not include the SRs in the synthesis. This is because the relevant primaries in the SR(s) would have been included in their own right, and including the SR would risk (i) double-counting them, and (ii) including also findings from primaries that didn't meet our inclusion criteria (which normally means that they did not have a decent comparison group).

If a cell with 3-8 studies has primary studies and SR(s), we included the SRs in the synthesis. We did this, despite being somewhat inconsistent with other heavier cells, to strike a reasonable balance between, on one hand, providing helpful information about a cell, and on the other, the length of the Guidebook material. For cells that have 9+ studies, there is a lot of information from 'just' the primaries, but for cells with 3-8 studies, obviously there is less. For instance, one cell has one primary study (for which there is a summary) and three SRs, and those SRs would seem to have useful additional value for grant-managers and partners.

Treatment of cells with just systematic reviews and no primary studies

When making the EGM, we 'unzipped' the systematic reviews. This means that we looked at the studies analysed in the SR, and included on the EGM any primary studies that should be included in the map: consequently, if an SR in some cell contained primaries relevant to that cell, those primaries are in that cell. (An SR could be in a cell with no primaries if: the SR includes studies of different designs than we included on the EGM, such as pre- / post-studies; or the SR looked at the topic of that cell but didn't find any studies relevant to it, i.e., the SR was empty in relation to that cell.)

Hence that we judge a cell with no primary studies contains no evidence, even if it contains SRs - because any relevant primaries would have been found in the unzipping.

Therefore, for the purposes of determining whether a cell is light, heavy or empty, we discount instances where there are SRs but no primary studies. There are quite a few cells which only have SRs (total 11), which count as empty, under this definition: Hermenau (an SR) is on its own sometimes; Sherr (an SR) is also alone in the cell response /recidivism.

Appendix 2 lists the SRs which appear in cells with no primary studies.

Section 2: Syntheses of studies in cells which contain 3+ studies

For cells with multiple studies (which we defined as having three or more), this Guidebook has a synthesis of all those studies. These syntheses are presented in the order of the cells as one would be reading it (top-bottom, left-right. So, prevention ones are first, then response, then treatment, etc.).

Some studies appear in heavy cells (for which there is a synthesis here) and also in light cells. For those studies (and any which appears in a light cell), there is a summary in Section 3.

Prevention / Institutional Safeguarding Practice: Operations

The effect of prevention-focused interventions to improve institutional operations to safeguard children is uncertain.

Evidence status	High risk of bias	Unclear impact of prevention interventions focussed on institutional operations to safeguard children.
------------------------	--------------------------	--

The summary in brief

Training people who look after children - such as teachers and day care workers – might protect children from sexual abuse. Training programmes can improve the knowledge, attitudes, and practices of these caregivers to prevent abuse.

This synthesis includes three completed studies and two protocols of primary studies. In all three completed primary studies, teachers received training on preventing child sexual abuse. One large scale study from Spain also trained other professionals such as social workers, police officers and hospital staff. That latter study (*Cerezo 2004*) reported increased detection rates of child maltreatment after training of childcare professionals from different disciplines. Results from the other primary studies suggested modest improvements in knowledge, attitudes, and some behaviours (compared to controls) although these were usually only in the short term. Results need to be interpreted with caution because of the high risk of bias/low quality of studies.

The cell has three completed studies (*Cerezo 2004*; *Gushwa 2018*; *Rheingold 2014*) and two protocols for new studies (Perez 2021; Wangamati 2022).

Studies in this cell

Completed Primary Studies	
<i>Cerezo 2004</i> QED, high risk of bias	Spain (Balearic Islands). Professionals such as teachers, social workers, hospital staff, child protection services (CPS) staff and police. Evaluation of a large-scale training programme to improve detection of child maltreatment by professionals who worked with children
<i>Gushwa 2018</i> <i>RCT, high risk of bias</i>	USA (3 states – 2 Midwest, 1 Northeast). K-12 teachers from 3 public school districts. Evaluation of the <i>Enough! Preventing Child Sexual Abuse in My School</i> programme

<i>Rheingold 2015</i> <i>RCT, moderate risk of bias</i>	USA (three sites: one each in Atlanta, GA; Beaufort, SC; Bend, OR). Caregivers of children in day care, churches, schools Evaluation of the <i>Stewards of Children</i> programme to prevent CSA
Protocols for Primary studies	
<i>Perez 2021</i> <i>Study protocol for RCT</i>	USA: California Testing implementation strategy for "ACEs Aware" policy that provides Medicaid reimbursement for Adverse childhood experiences (ACEs) screening annually for child primary care visits to low-income families in Southern California
<i>Wangamati 2022</i> <i>Study protocol for QED</i>	Kenya and Tanzania Quasi experimental pilot to assess the delivery and potential changes in knowledge, attitudes, behaviours and violence against children (VAC) prevalence and incidence in and around schools following the whole school approach (WSA) intervention implementation

PRIMARY STUDIES

The interventions

Programme	Country	Description
Large-scale programme to increase child maltreatment detection	Spain	Phase 1: Training frontline professionals such as social workers, paediatricians, police officers, psychologists, psychiatrists, and nurses who are involved in reporting cases of child maltreatment to Child Protective Services (CPS). Phase 2: Focused on training teachers, psychologists, and support staff in preschools and primary schools. The training for professionals in both phases included 16-20 hours training over 2-3 days, covering various aspects of child maltreatment and the protocol to follow to report cases to CPS. A form was developed for streamlined referral of potential cases to CPS. Support was provided to trained professionals by a Local Coordination Team (LCT) through a helpline and in-person visits.
Enough! Preventing Child Sexual Abuse in My School	USA	Online training course developed for schools to help them prevent child sexual abuse (CSA). Nature and scope of CSA is covered in the course for school staff. Specific actions that school personnel can take to prevent CSA or precursor activities like grooming. Training can be completed in 1 hour or in 20-minute increments. The training was developed as part of a campaign (the "Enough Abuse Campaign") which was a citizen education and community engagement initiative. A collaborative of child-centred agencies came together to launch the campaign.

Stewards of Children USA Brief training programme for childcare professionals (e.g., teachers, childcare personnel, clergy, counsellor, probation officer, day care worker, coaches) to improve their knowledge, attitudes, and response to child sexual abuse (CSA). Training is delivered in two modes: in-person in a 2.5-hour session, and via the internet over 2 weeks.

Do these interventions work in improving institutional operating practices to safeguard children?

Cerezo 2004: found that the intervention increased the number of cases reported to CPS from in both phases. The study reports a threefold increase in the number of cases detected after the intervention compared to before. The second phase, i.e., training teachers, led to detection of 2-3 children per 1000 as new cases after accounting for duplications from the first phase. The higher the proportion of professionals trained, the higher was the detection rate.

The first phase, i.e., training frontline workers, was sequentially implemented in three territories. Referrals increased before and later the intervention in the first two territories but not in the third. According to the authors, this might be due to the intervention's knowledge spreading to the third territory before implementation via mass media, professional networks and professionals moving territories for new jobs. A later comparison of the outcomes with a different region of the Balearic Islands found expected increase in referrals.

In **Gushwa 2018:** teachers who received the online training intervention had significantly higher scores on a CSA knowledge scale than the comparison group. The awareness program included aspects of CSA such as prevalence rates, types of behaviours, impact on children, signs and symptoms, veracity of children's reports, perpetrator backgrounds/behaviours backgrounds/behaviours, factors associated with CSA in schools including examples of boundary-violating behaviours, specific behaviours to support prevention, reporting responsibilities, and responses to suspected abuse. Teachers who received the online training intervention scored significantly higher on average (nearly 90%) than the control group did (approximately 75%). Additionally, most of the trained teachers reported an increase in their knowledge, awareness, likelihood of taking action to prevent CSA: many indicated a "great deal" or "somewhat" increase, compared to "a little" or "none, already knew".

Rheingold 2014 found that *Stewards of Children* training improved childcare professionals' knowledge and behaviours about CSA:

- **Knowledge** about CSA increased after the intervention but declined over the next three months. The control group also showed increased knowledge during this period, but not to the same extent as the intervention group.
- **Attitudes.** Participants' belief in CSA myths was low initially, leaving little room for improvement. After training, the control group had the better score but at three months there was no difference between groups.
- **Behaviours.** Participants who received the training reported improvement in their behaviors three months after training, as compared to the control group. The behaviours reported to have improved most were:
 - "Limiting the opportunity for older youth and younger youth to have one-to-one interaction". This is significant given that juveniles are offenders for over a third of CSA cases (Finkelhor et al. 2009). And
 - Participants in the intervention group reported an increase in behaviours such as "Sharing with another adult an article, brochure, or other information about CSA prevention" within three months of the training. The control group also showed improvement in these

behaviours during the same period, possibly due to the influence of the changed behavior of their trained colleagues or knowledge sharing.

In terms of the difference between being trained in-person vs online, the evaluation also found:

- Knowledge: The group trained in-person learned 'significantly' less about CSA (their knowledge had changed less) than the group trained on-line. Three months after training, however, there were no differences in CSA knowledge between the two groups.
- Attitudes: No difference between the group trained in-person vs. the group trained on-line.
- Behaviours: No difference between the group trained in-person vs. the group trained on-line.

The size of the impact of training in terms of implications for practice are unclear.

Protocols for new studies

Perez 2021 (US, California) proposes to develop a better implementation strategy to improve the awareness and uptake of "ACEs Aware" policy that provides Medicaid reimbursement for Adverse childhood experiences (ACEs) screening annually for child primary care visits to low-income families in Southern California. The study will test a multifaceted implementation strategy in partnership with a Federally Qualified Health Center (FQHC) system. The Exploration, Preparation, Implementation, Sustainment (EPIS) framework is a widely used implementation framework. This study plans to follow the EPIS framework for implementation mapping to refine implementation. The refined implementation strategy will include online training videos, a customized algorithm and use of technology to improve workflow efficiency, implementation training to internal FQHC personnel, clinic support and coaching, and written implementation protocols. This randomized trial with five primary care clinics will assess this implementation strategy for (a) fidelity to the ACE screening protocol, (b) reach, defined as the proportion of eligible children screened for ACEs, and (c) the impact of the ACE policy on child-level mental health referrals and symptom outcomes.

Wangamati 2022 (Kenya and Tanzania) proposes to evaluate the whole school approach (WSA) for reducing violence against children (VAC) in and around schools in Tanzania and Kenya. WSA is developed by Investing in Children and their Societies-Strengthening Families & Protecting Children (ICS-SP) for reducing VAC in and around schools. The study plans to (1) test intervention's feasibility, (2) assess the extent to which the WSA can help reduce prevalence and incidence of VAC (3) learn from insights into changes in stakeholders' knowledge, attitudes and practices in relation to VAC following intervention implementation and (4) to build evidence base for refining intervention content, delivery and theory of change (ToC). The study is a mixed-methods, controlled before-and-after, quasi experimental pilot. The pre-intervention phase will involve school safety audits and stakeholder enhancement of the WSA ToC and surveys of teaching and non-teaching staff; parents (knowledge, attitude and practices); and pupils (VAC incidents and school climate). There will be process evaluation and random school visits. afterwards, there will be end-line surveys similarly to those at baseline. Focus groups are also planned, and in-depth interviews with ICS-SP staff, training facilitators, teachers, parents, and pupils to gain insights into acceptability, delivery and intervention effects.

Have the interventions been implemented at scale?

Cerezo 2004 was implemented on a large scale in the Balearic Islands, an autonomous community of Spain, with 161,287 children under 18 at the time of the intervention. The intervention was designed to include professionals from all frontline agencies, preschools, and primary schools that served the children living in this area.

The programs in *Rheingold 2014* and *Gushwa 2018* might have been implemented in many sites but not much detail is provided in the papers.

Which type of organisation delivered the intervention?

For the intervention in Spain (*Cerezo 2004*), a local coordinator was appointed (unclear by whom or what the selection criteria were) and they worked with two professionals from CPS and two school professionals. This Local Coordination Team (LCT) was responsible for coordinating intervention activities with the various agencies involved.

Stewards of Children (Rheingold 2014) was developed and delivered by a US NGO, Darkness to Light.

Enough! Preventing Child Sexual Abuse in My School (Gushwa 2018) was developed in Massachusetts as part of the Enough Abuse Campaign (EAC), a citizen education and community engagement initiative with the aim of preventing CSA.” The EAC was created by a collaboration of child-centered organisations with a shared purpose of preventing CSA.

What do the interventions cost?

None of the studies report any cost data.

How is the programme meant to work? The theory of change

Cerezo 2014 does not mention any theory in particular for its programme. Instead, the intervention was based on the principles of motivational interviewing to overcome barriers and change attitudes towards reporting potential cases of maltreatment.

Stewards of Children is not a theory-based prevention programme, but its principles are in line with Finkelhor’s theory, which suggests that certain conditions need to exist for CSA to occur. They include: an individual’s tendency to abuse, absence of internal or external inhibitions for the offender, and having access to the child. Preventing one or more of these preconditions should reduce the likelihood of CSA. This program aims to reduce access to children by offenders and to increase external barriers for them improving the knowledge, attitudes, and response of adults responsible for childcare.

Enough! Preventing Child Sexual Abuse in My School was developed based on adult learning theory and how educators learn best.

Are the results generalisable?

The results from the US studies might be generalisable to school settings that are like the US. The large study from Spain has issues with risk of bias which affects its generalisability.

How reliable is the evidence?

Not very.

Cerezo 2004 and *Gushwa 2018* are rated as having a *serious risk of bias*. *Rheingold 2014* is rated as having a *moderate risk of bias*.

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Cerezo 2014</i>	High Risk of Bias	Serious	Moderate	Moderate	Serious	Low	Moderate	Serious

What else is known from other studies about prevention interventions and institutional safeguarding practices?

The availability of evidence for prevention interventions in terms of changing institutional safeguarding practices is quite sparse. We have another cell on this outcome in the EGM for response interventions. Overall, training caregivers of children seems a promising intervention to improve their knowledge, attitudes, and behaviours of protecting children from sexual abuse. However, we need more robust studies from diverse settings with long-term outcomes to get a better picture on whether these trainings truly work.

Prevention / Institutional Safeguarding Practice: Environment

Prevention-focused interventions to improve institutional environments in schools to safeguard children have positive effects, though this finding is based on only a few studies.

Evidence status	Low risk of bias	Clear impact of prevention-focused interventions to improve institutional environments to safeguard children (based on four RCTs).
------------------------	-------------------------	--

The summary in brief

Institutions that care for children, such as children’s homes, orphanages, schools, daycares, foster homes and hospitals, can implement various preventive interventions and policies to safeguard the children in their care.

Interventions to improve the environment to protect children in institutions include training for staff (e.g., teacher or staff training programs). The studies in this summary are of interventions that trained caregivers and professionals to improve their knowledge, attitudes, and practices/behaviors when working with children, identifying maltreatments, and avoiding violent behaviors. These changes are intended to improve the organization’s environment. All four completed primary studies in this cell were in schools, and all found improvements in the quality of practitioner’s or caregiver’s response as a result of the training.

There are six studies in this cell: four papers from three studies, all recent (Baker-Henningham 2019; Baker-Henningham 2021; Kim 2019; and Nickerson 2019); one protocol (Perez 2021); and one systematic review (Lo and Cho 2021). Our confidence in the findings of studies is assessed to be high for primary studies but low for the systematic review.

Studies in this cell

A. Completed Primary Studies	
Baker-Henningham 2019 RCT low risk of bias	Jamaica, primary schools School teachers Evaluation of the effect of a school-based violence prevention programme on (1) teachers’ use of violence against children and (2) class-wide child aggression.
Baker-Henningham 2021 RTC low risk of bias	Jamaica, primary schools School teachers Evaluation of Irie Classroom Toolbox, a violence-prevention teacher-training programme to see its impact on violence against children (including physical violence and psychological aggression) by teachers and class-wide child aggression measured at post-intervention and 1-year follow-up.
Kim 2019 RCT, low risk of bias	USA, elementary schools School teachers

	Evaluation of the Second Step Child Protection Unit (CPU) to assess the impact on teacher outcomes of knowledge, attitude, and relationships as well as any interaction effect before and after the implementation of intervention.
Nickerson 2019 RCT, low risk of bias	USA, eight elementary schools from a large suburban school district in the Northeast School children Evaluation of the Second Step Child Protection Unit (CPU) to assess the impact on students' CSA prevention knowledge, ability to recognize, report, and refuse unsafe touches, and perceptions of teacher-student relations. The study also investigated the moderating role of age and gender on program effectiveness.
B. Protocols for Primary studies	
Perez 2021 Protocol for RCT	USA: California Testing the implementation strategy for the "ACEs Aware" policy that provides Medicaid reimbursement for adverse childhood experiences (ACEs) screening annually for child primary care visits to low-income families in Southern California.
C. Systematic Review	
Lo and Cho 2021 Systematic Review high risk of bias	To review the evidence available on the impacts of community-based interventions on the reduction of child maltreatment and to identify the core components of the interventions.

A. PRIMARY STUDIES

The interventions

Baker-Henningham 2019 and Baker-Henningham 2021: Both studies are of a violence prevention program named the IRIE Classroom Toolbox. This is a school-based teacher training programme for teachers of young children. The programme is designed for use in low- and middle-income countries (LMICs) to reduce violence against children by teachers and prevent the development of antisocial behaviour in children aged 3–8 years. The Toolbox also aims to improve the quality of the classroom environment and promote child mental health, self-regulation, and prosocial skills.

Both studies evaluated the effectiveness of the Irie Classroom Toolbox in Jamaican preschools. The major difference between the two studies is the scale of implementation. The first study (Baker-Henningham 2019) had a smaller sample (14 primary schools, 55 teachers, and 220 children), whereas the later study (Baker-Henningham 2021) was much larger (76 preschools, 229 teachers, and 3,993 children).

The Irie Classroom Toolbox has four modules: (1) creating an emotionally supportive classroom environment; (2) preventing and managing child behaviour problems; (3) teaching social and emotional skills; and (4) individual and class-wide behaviour planning. Training materials for teachers include (1) a

tools book which gives simple guidelines on how to use each strategy and the rationale behind that; (2) an activity book of songs, games, activities, lesson plans; (3) three sets of picture cards to help teachers teach classroom rules, friendship skills, and understanding emotions; and (4) a story book with 14 pictorial stories showing common classroom problems that children face in school, and suggesting how children tackle these problems.

Kim 2019 and Nickerson 2019: Two papers (Kim 2019 and Nickerson 2019) are from the same study (i.e., the same set of children) on an intervention called the Second Step Child Protection Unit (CPU).

CPU addresses child sexual abuse (CSA) through (a) school policies and procedures, (b) staff training, (c) student lessons, and (d) family education. CPU makes all school staff (e.g., teachers, office staff, monitors) complete a 75–90-min online module to prepare them to recognize indicators of CSA, responding better, and reporting abuse. After completing training, teachers and/or other staff provide six lessons to students. The lessons involve multiple delivery strategies (e.g., didactic instruction, songs, videos, case scenarios, role-plays) to address the topics of safety, asking adults, unsafe and unwanted touches, rules about private body parts, and practicing of rules to stay safe.

Kim 2019 evaluated CPU to assess its impact on teachers' knowledge, attitude, and relationships. Nickerson 2019 evaluated the effect on students' CSA prevention knowledge, ability to recognize, report, and refuse unsafe touches, and perceptions of teacher-student relations. Nickerson 2019 also investigated the role of age and gender on program effectiveness.

Do these interventions work in improving institutional safeguarding practices?

Baker-Henningham 2019 and Baker-Henningham 2021:

Both studies found that the programme reduced teacher violence against children and provided a more emotionally supportive classroom environment compared to the control group. But the training had no effects on class-wide child aggression, pro-social behaviour, teacher wellbeing, or child mental health.

The intervention improved children's early learning skills, like oral language and self-regulation skills, but had less effect on math calculation, reading and spelling.

Overall, the intervention seemed to reduce violence against children by teachers and raise the quality of the classroom environment.

Kim 2019 found that the CPU teacher training improved teacher knowledge, attitude, and student-relationship. Teachers' acceptance of the teacher training was associated with stronger improvements in knowledge, attitudes, and student-relationships.

Nickerson 2019 found that CPU students had higher scores on all outcomes than students who didn't take part in CPU. Children in younger grades experienced greater gains from the program. Whilst both boys and girls made significant gains in CSA knowledge and the ability to recognize, refuse, and report unsafe touches, the effect was stronger for girls than boys.

Have the interventions been implemented at scale?

None of the studies are implemented at scale.

Which type of organisation delivered the intervention?

The Irie Classroom Toolbox was delivered by female staff members. Among them, senior staff facilitated teacher-training workshops, and some facilitated in-class support. These facilitators were trained and supported by the principal investigator and joined weekly group supervision meetings with facilitators working on another similar intervention in Jamaican preschools.

The Second Step CPU is a program by The Committee for Children (2014) (CPU; www.cfchildren.org/child-protection). CPU teacher training is delivered as a 75–90-min online module, and after completing the CPU online training modules, these teachers delivered the CPU student curriculum in their classrooms.

What do the interventions cost?

None of the studies report cost data.

How is the programme meant to work? The theory of change

The Irie Classroom Toolbox studies followed a conceptual model that suggests the toolbox will reduce violence against children by helping teachers to gain skills, motivation, and the opportunity to use positive discipline techniques.

Kim 2019 did not present a theory of change but had hypotheses to guide their study. They hypothesized that: (1) Participation in the CPU will improve teacher outcomes in the intervention group as compared to the control group. (2) The CPU will interact with teachers’ prior knowledge, attitudes, and relationships to improve their outcomes. (3) A higher level of treatment acceptability will show better teacher outcomes for teachers who participated in the CPU.

Nickerson 2019 presented The Second Step CPU logic model (see Fig. 1) suggests the student lesson and reinforcement activity components of the intervention will increase student knowledge of personal safety rules and the ability to recognize, report, and refuse unsafe situations and touches. The authors thus hypothesized that participation in the CPU lessons would raise abuse prevention knowledge as well as an increased ability to recognize, report, and refuse unsafe and sexually abusive situations. The logic model also emphasizes intervention components that foster positive relationships with students through safe and supportive environment strategies; thus, the authors hypothesized that the intervention would improve teacher-student relations.

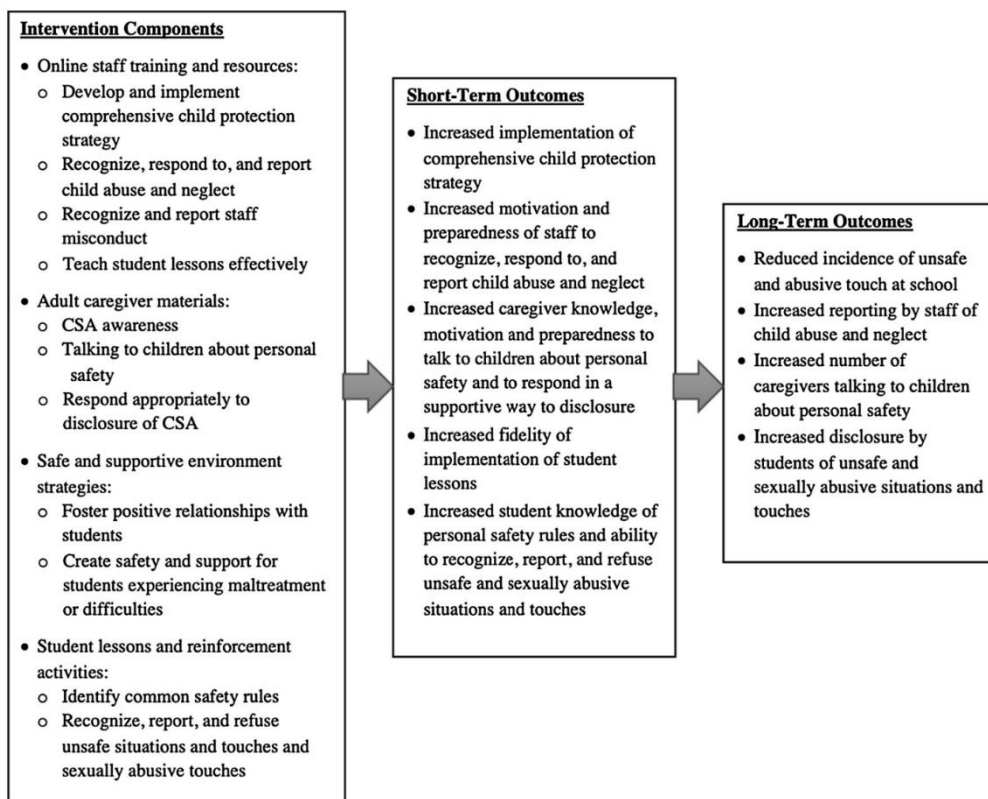


Fig. 1. Logic Model for the *Second Step* Child Protection Unit Curriculum.

How reliable is the evidence?

Very. All four RCTs are rated to have a low risk of bias.

B. RCT PROTOCOL

Perez 2021 (US, California) proposes to develop a better implementation strategy to improve the awareness and uptake of the "ACEs Aware" policy that provides Medicaid reimbursement for adverse childhood experiences (ACEs) screening annually for child primary care visits to low-income families in Southern California.

The study will test a multifaceted implementation strategy in partnership with a Federally Qualified Health Center (FQHC) system. The Exploration, Preparation, Implementation, Sustainment (EPIS) framework is a widely used implementation framework⁴. This study plans to follow the EPIS framework for implementation mapping to refine implementation. The refined implementation strategy will include online training videos, a customized algorithm and use of technology to improve workflow efficiency, implementation training for internal FQHC personnel, clinic support and coaching, and written implementation protocols. This randomized trial with five primary care clinics will assess this implementation strategy for (a) fidelity to the ACE screening protocol, (b) reach, defined as the proportion of eligible children screened for ACEs, and (c) the impact of the ACE policy on child-level mental health referrals and symptom outcomes.

C. SYSTEMATIC REVIEW

The cell has one systematic review (Lo and Cho 2021). Details are below:

The intervention: community-based interventions.

Inclusion Criteria:

- Those published in English before January 2020,
- Focused on the prevention of child maltreatment,
- Placed emphasis on modifying the community's environments and processes,
- Evaluated the actual change in child maltreatment at the community level as one of the outcomes,
- Primary studies reporting original data.

Studies Included: Four studies were included in this review.

Results: All four included studies are from the USA. The four community-based child maltreatment programs included were Strong Communities for Children, the Durham Family Initiative (DFI), the Enough Abuse Campaign, and Prevent Child Abuse Georgia. There were two interventions aimed at preventing physical child abuse and neglect in children between the ages of 0 and 6, and two interventions focused specifically on addressing child sexual abuse (CSA). These interventions incorporated various components of primary and secondary prevention strategies. Out of the four interventions, three focused on community-level interventions that involved modifying community environments and processes, while one intervention simultaneously targeted multiple ecological levels.

How reliable are the findings:

Not very. The systematic review is rated to have an overall **high** risk of bias.

Prevention / Adult Institutional Caregiver

School and institution-based training improve teachers' and childcare providers' knowledge and awareness on child abuse prevention and reducing corporal punishment.

Evidence status	Low risk of bias	Strong evidence that teacher and childcare provider training leads to improved knowledge and attitudes to recognizing signs of child sexual abuse and to reduce use of corporal punishment.
------------------------	-------------------------	---

The summary in brief

This synthesis is of a cell that has more than twenty studies; most of them recently published. The cell has 15 primary studies, two systematic reviews, and four protocols. Studies are from a mix of high-income and low- and middle-income countries, and in one case, teachers were trained in a refugee camp.

The evidence shows that:

- school-based training for teachers and childcare providers (in one case an e-training for healthcare professionals and in another training for orphanage personnel) can improve their knowledge, attitudes, and sense of preparedness in responding to signs of sexual abuse among their students.
- where corporal punishment is common, teacher training can help change attitudes of teachers away from harsh disciplinary methods.

However, it is uncertain whether these improvements directly translate into reductions in child maltreatment (few studies report this outcome).

Contents of the cell

This cell includes four protocols (*Wangamati 2022, Lopez Garcia, 2021 Scharpf 2021, Masath 2020*), 15 primary studies (*McElearney 2021, Baker-Henningham 2019, Baker-Henningham 2021, Fabbri 2021, Hecker 2021, Jewkes 2019, Kim 2019, Konig 2020, Nkuba 2018, Rheingold 2014, Ssenyonga 2022, Kolko 1987, Kolk 1989, MacIntyre 1999, and Martin 2020*) and two systematic reviews (*Lo 2021 and Russell 2020*).

Four of the primary studies come from the US, three from Tanzania, two from Jamaica, and one each from Germany, the UK, South Africa, Uganda, Iran, and Ireland. There are no studies from South Asia, East Asia, or South America. For the protocols, one is for a study in Haiti. The other three span multiple African countries: Tanzania (in three protocols), Kenya, Ghana, and Uganda.

The two systematic reviews are wide in scope and cover a range of interventions to prevent child maltreatment that are not specifically focussed on institutional care providers. *Lo 2021* assesses community-based interventions to reduce child maltreatment and identifies four major elements: involving community members, establishing community partnerships with institutions, promoting multidisciplinary collaborations, and responsiveness to community needs. *Russell 2020* looks at the efficacy of child sexual abuse prevention interventions in “developing countries” and finds that most evaluations are in educational settings.

The interventions

The interventions in this cell vary in their approach to child protection. Teachers (in most cases), childcare providers, and healthcare providers are engaged to improve their knowledge of child maltreatment, so they are better prepared to identify signs of it. Additionally, some of these interventions aim to reduce corporal punishment by teachers.

One approach involves training teachers, childcare professionals, or healthcare professionals on child sexual abuse so that they are better able to recognize and respond to potential abuse. Some trainings are delivered in-person while others are through e-learning. Most of these studies come from the US with one each from Germany and Iran.

The second approach aims to reduce corporal punishment by teachers in the classroom and in one case by institutional care staff at orphanages in Dar es Salaam, Tanzania. All studies included in-person group training for teachers to move away from physical or emotional violence as a form of discipline. Two studies are from Jamaica and the rest are from sub-Saharan Africa, including one in a refugee camp.

The third approach is in-classroom training for young children, such as the 'Red Flag/Green Flag' programme. It teaches children about inappropriate physical touching and trains teachers and parents to improve their knowledge and ability to recognize sexual abuse.

The fourth approach is a whole-school approach that engages children, teachers, parents, school administrators to create processes and a culture against violence in the school. One study from Northern Ireland (UK) using this approach focussed on all types of abuse and bullying. In another study, the same approach was used to prevent gender-based violence among eighth graders in South Africa.

Who delivers the intervention?

The interventions are delivered by researchers (who have developed the intervention), personnel from community-based organisations, or teachers in the classroom.

Have the interventions been implemented at scale?

Examples of large-scale programmes include *Stay Safe (MacIntyre 1999)*, which has been delivered in most primary schools in Ireland, and *Keeping Safe (McElearney 2021)*, a whole-school approach that teaches young children to identify abusive behaviours included 64 primary schools in Northern Ireland (UK). An ongoing study evaluating the *Interaction Competencies with Children – for Teachers ICC-T (Scharpf 2021)* has 72 schools enrolled across three African countries. However, most studies are small with few schools or classrooms.

What do the interventions cost?

Only three studies provided information on programme costs. In the *Irie Classroom Toolbox* intervention in Jamaica (*Baker-Henningham 2021*), teachers received lunch and a small stipend to cover transportation (USD 4 per workshop). In *Skhokho* (South Africa), a whole-school approach to reduce gender-based violence among eighth graders (*Jewkes 2019*), caregivers received Rand 50 and teenage participants received Rand 20 for transportation costs for each session. Similar compensation for travel (USD 2.17) and free food and beverages were available to institutional caregivers in Tanzania in the *ICC-T* intervention (*Hecker 2021*).

How are the programmes meant to work? The theory of change

Most of the programmes are theory-informed, drawing from multiple behavioural theories rather than entirely depending on one theory. These theories include the transtheoretical model, stages of change, social learning, and attachment theories.

Do the interventions work in improving adult institutional care provider outcomes?

Seems like it. Most studies report improvements in the knowledge, awareness, and attitudes after the programme of institutional care providers (teachers in most cases). Teachers reported feeling more

prepared to recognize and respond to signs of child sexual abuse among their students. Other teachers reported that they reduced their use of corporal punishment to discipline students. However, it is important to keep in mind that these are intermediate outcomes and we do not know if these actually and reliably lead to sustained reductions in child maltreatment.

Are the results generalisable?

Yes. The number of studies and the consistency of results – particularly given the number and diversity of countries in which the studies are run – indicate that results are highly generalisable.

How reliable is the evidence?

Pretty reliable. Almost half of the studies are rated ‘low risk of bias.’ The number of studies and the consistency of findings suggests that the evidence is quite reliable.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Baker-Henningham 2019</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
<i>Baker-Henningham 2021</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
<i>Fabbri 2021</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
<i>Kim 2019</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
<i>Hecker 2021</i>	Some concerns	Some concerns	Low risk	Some concerns	Some concerns	Low risk
<i>Jewkes 2019</i>	Some concerns	Some concerns	Low risk	Low risk	Low risk	Some concerns
<i>Konig 2020</i>	Some concerns	Some concerns	Low risk	Low risk	Some concerns	Low risk
<i>McElearney 2021</i>	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Low risk

Rheingold 2014	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns
Nkuba 2018	High risk	Some concerns	High risk	High risk	Some concerns	Some concerns
Ssenyonga 2022	High risk	Some concerns	Low risk	Some concerns	Low risk	High risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Kolko 1987	Low risk of bias	Low	Low	Low	Low	Moderate	Moderate	Low
Kolko 1989	Low risk of bias	Low	Low	Low	Low	Moderate	Moderate	Low
MacIntyre 1999	Low risk of bias	Low	Low	Low	Low	Low	Moderate	Low
Martin 2020	Some concerns	Low	Low	Low	Low	Low	Low	Moderate

Prevention / Disclosure: Disclosure rates

School-based interventions to tackle abuse can increase disclosure.

Evidence status	Some concerns	Evidence of positive impact, with need for testing to establish best practice
------------------------	----------------------	---

The summary in brief

School-based interventions to reduce child abuse usually involve a few sessions to increase children’s awareness of physical and sexual abuse and train them in what to do. The interventions often also include teachers and parents. Training on appropriate courses of action increases likelihood of disclosure in response to hypothetical situations. The interventions - and data collection for the evaluation - provide a good opportunity for disclosure of actual abuse, sometimes on a substantial scale. The studies suggest that recurrent activities result in more disclosure than one-off interventions. No studies found evidence of adverse side effects. Findings need to be treated with caution, as the studies mostly have a medium to high risk of bias.

All studies in this cell from high-income countries except one from Uganda (low-income country).

The interventions

All the interventions are school based, ranging from kindergarten to high school students aged 15-17. All interventions directly involve pupils other than the on-line programme for teachers, *Enough! Preventing Child Sexual Abuse in My School*.

Most of the programmes are targeted at protecting young children from abuse. The exception is the workshop by the agency *Viol Secours*, which is for teenagers as potential perpetrators as well as victims, and for people to whom abuse may be reported.

Most interventions are of limited duration, ranging from a 30-minute play with 15-minute discussion (Project Trust) to a one-hour online course (*Enough! Preventing Child Sexual Abuse in My School*) to a seven-session classroom course (*IGEL*). The exception is the *Good School Toolkit*, which is a comprehensive school-wide approach. The Good School Toolkit is also an exception, as its primary focus is the prevention of physical and sexual abuse by teachers in the school.

Table 1 provides an overview of the interventions included in the studies in this summary.

Table 1 Overview of the interventions in this summary

Programme (study)	Country	Description
<i>Tweenees</i> (Barron, 2013)	UK	Four classroom-based sessions of 50 mins on bullying, sexual assault etc.
<i>IGEL</i> (Czerwinski, 2018)	Germany	Seven teacher-delivered school sessions. In each session, the children participate in various (interactive and experiential) exercises and exchange views.
<i>Viol-secours workshop</i> (Daigneault, 2015)	Canada	One 75-minute workshop concerning sexual violence, date rape, common myths and misconceptions etc. for high school students aged 15-17.
<i>Prevención de abusos sexuales a menores</i>	Spain	School-based prevention programme for children aged 8-12.

<i>(del Campo Sanchez, 2006)</i>		
<i>Good School Toolkit (Devries, 2015)</i>	Uganda	School-wide intervention for primary schools implemented over 18 months. Schools are provided with booklets, posters, and facilitation guides for over 60 Toolkit activities. The activities, e.g. student discussions, debates, and booklet clubs, are mostly for a group setting.
<i>Think First and Stay Safe (Elfreich, 2020)</i>	USA	Elementary school training for students grades 1-6 on how to avoid abusive situations and to disclose to a trusted adult if experiencing abuse.
<i>Enough! Preventing Child Sexual Abuse in My School (Gushwa, 2018)</i>	USA	One-hour online training course for teachers, administrators, counselors, coaches, office personnel, and support staff for K-12 students.
<i>Feeling Yes, Feeling No (Hazzard, 1991)</i>	USA	Three sessions for third and fourth graders, including film, covering recognition of and response to sexual abuse. One session 'booster shot'.
<i>Red Flag, Green Flag People (Kolko, 1987 & 1989)</i>	USA	Two sessions for elementary school students based on colouring book and film.
<i>Project Trust (Oldfield, 1996)</i>	USA	30-minute play (Trust) followed by 15-minute Q&A for grades 1-6.
<i>In Situ Skills Training (IST) + Learn to be safe with Emmy and friends (White, 2019)</i>	Australia	Five one-hour weekly sessions delivered by trained facilitators to children in grade 1 + simulated risk scenario where child is asked by a stranger to go with them (in a controlled setting).

Who delivers the intervention?

The interventions were all developed either by university research teams (e.g., *Prevención de abusos sexuales a menores* from the University of Salamanca) or non-governmental agencies, such as Red Flag, Green Flag People from the Rape and Abuse Crisis Center of Fargo-Moorhead.

Delivery may be by agency workers – e.g., two facilitators from Viol-secours deliver their workshop – or teachers who receive training for this purpose. In one case – *Tweenees* – some sessions were run by an abuse survivor aged under 18.

Several of the interventions also involve parents, although this is usually limited to one session to explain the programme, and possibly to encourage engagement with the programme materials at home.

In *The Good School Toolkit*, students are directly responsible for some activities.

Have the interventions been implemented at scale?

School-based prevention programmes are a widely adopted approach, although they vary in duration and intensity. The *Red Flag / Green Flag People* colouring book – and similar books such as *Good Touch, Bad Touch* – are common, though the two studies of effectiveness were carried out on a small sample. The film *Feeling yes, Feeling no* was produced and distributed by the National Film Board of Canada. A video of the same name is used in Scotland (UK), but it is unclear if it is the same video.

The online course *Enough! Preventing Child Sexual Abuse in My School* is marketed by the Enough Abuse Campaign. It is not possible to determine from the website the scale of implementation, but it has been more widely adopted than in the evaluation setting.

The *Good School Toolkit* is promoted by the NGO Raising Voices, which has implemented it in primary and secondary schools in Uganda and is being tested in Tanzania. The scale has been limited to date.

IGEL in Germany was a Ministry of Education project funded from 2015-17. It was offered to other schools, but take-up is unclear.

The Spanish intervention was a one-off undertaken for the purposes of the study to assess prevention programmes in a Spanish setting. Likewise, the Viol-securus workshop was undertaken for the purposes of the study.

The Think First and Stay Safe programme started off as a pilot in 3 schools in the US and expanded to 66 schools over ten years.

No additional information could be found for Project Trust or the study from Australia, suggesting they are not operating at scale.

What do the interventions cost?

The study for only one programme – *Tweeness* – reports data on costs. The programme consists of four classroom-based sessions. Total costs were estimated at just over \$1,100 per school, or \$11 per student, which was equivalent to \$60 per disclosure. The major cost was the presenter.

For the other programmes, based on the programme descriptions, these seem to be similarly low cost, on account of their limited duration and the fact that they are mainly delivered by teachers.

For example, it is clear that *Red Flag / Green Flag* is relatively low-cost. Part of its stated rationale is that preventive interventions are a cost-effective alternative to taking children at risk of abuse into care.

Although it is an online programme and so may sound low-cost, *Enough! Preventing Child Sexual Abuse in My School*, is a branded programme which requires a licence fee.

The likely exception is *The Good School Toolkit*, which is a more intensive intervention and so has a higher cost. But it has more ambitious objectives in settings in which physical and sexual abuse in schools is relatively common.

How are the programmes meant to work? The theory of change

The core idea of prevention programmes is to establish a clear understanding in children to identify what constitutes unacceptable behaviours and feel empowered to report it.

The logic of personal safety prevention programmes targeted at young children is that children are active agents in child sexual abuse. If children are aware of what constitutes abuse (e.g., good touch versus bad touch), they are likely to recognize it, object to it, walk away from it, and report it, all of which reduce the likelihood of abuse occurring. Programmes teach children skills for managing and reporting abusive situations.

This process is supported by parent and teacher training so they can reinforce the approach and listen to children when they discuss these issues, including reporting abuse.

Although not stated as part of the theory of change, the intervention and evaluation activities provide an opportunity for disclosure.

The Good School Toolkit has a six-step process based on the Transtheoretical Model of Behavior Change. This model first makes people aware of the problem of physical and sexual abuse and then supports the planning and implementation of behaviours to deal with the problem. The final stage is when new behaviours have become the norm.

Do the interventions work in improving child disclosure of maltreatment?

Overall school-based prevention programmes are effective in increasing disclosure.

Direct disclosure – telling a teacher, counselor, or other project worker, or disclosing abuse during data collection – was measured in seven studies. These results show clearly that school-based prevention activities create an opportunity for disclosure. All the studies show that children disclose abuse during, or in association with, the intervention sessions (Table 2). There are usually zero disclosures in the control group over the same period. Even simply asking children about their experiences of abuse can result in substantial disclosure, notably in the case of Uganda, in which 434 cases were referred to child protection services after the follow-up survey.

Table 2 Direct disclosure during intervention or data collection (number of cases)

Programme	Disclosure in	
	Treatment group	Control group
<i>Tweenees</i>	<i>Grade 6</i> 18U presenter: 62 Teacher presenter: 3 <i>Grade 7/8</i> 18U presenter: 5 Teacher presenter: 0 (4 observed on video but not reported) Telephone helpline*: 44	0 0 28 (all other areas of city not just control)
<i>Prevención de abusos sexuales a menores</i>	8	2
<i>Good School Toolkit</i>	434 children referred to child protective services because of what they disclosed in the follow-up survey	
<i>Feeling yes, Feeling no</i>	Sexual abuse: 28 (8 on-going/20 past) Physical abuse: 6 (5 on-going/1 past) Teachers reported observing two incidents of children using prevention skills for strangers in cars	
<i>Red Flag, Green Flag</i>		
<i>1987 study</i>	20 (revealed in data collection of which 18 had been revealed to adult)	0
<i>1989 study</i>	20	0
<i>Project Trust</i>	4	1

Note: *Refers to the two weeks after intervention. No calls received in three weeks before intervention

In the study of *Tweenees*, some sessions were presented by a survivor of child sex abuse aged under 18 (18U). These presentations were associated with higher rates of disclosure than the classes presented by teachers. For grade 6 students, the four classes held by the 18U presenter resulted in 65 disclosures, compared to just 3 disclosures in the X classes with teacher presentations. For grades 7 and 8, just one session was given by an 18U presenter, resulting in five disclosures, compared to none in any of the teacher presented classes.

Several studies report ‘hypothetical disclosure’, sometimes called the course of action the child should take in the event of attempted abuse. In some cases, the child is presented with a vignette and asked how they or the child in the vignette should respond. For example, for the evaluation of IGEL, the children were told the story of Jona, whose guitar teacher behaves inappropriately. Other studies ask the child directly how they would respond to inappropriate behaviour. The study of Red Flag / Green Flag asked children, “Would you talk to an adult in order to get help if you were touched in a way that made you feel uncomfortable?”.

All studies reporting hypothetical disclosure found a significant effect, though it was usually only a small to moderate effect and weaker at follow-up than immediately after the intervention.

Prevention may also include activities with teachers, so they respond appropriately to disclosure. However, the effects of this component are not commonly evaluated. Three studies report findings about teacher preparedness: two found a positive effect and one found no effect, so no conclusion can be drawn about their effectiveness.

The evidence of increased disclosure is supported by the findings along the causal chain. The studies mostly report improving children’s awareness of potentially abusive situations and how to respond. For example, the Viol-secours workshop has only a small effect on knowledge and awareness and therefore a small effect on disclosure.

All findings need to be treated with caution, as the studies mostly have a medium to high risk of bias.

Evidence of adverse effects

There are concerns that prevention interventions may have adverse side effects on children exposed to discussions of sexual abuse or data collection which includes simulated situations such as abductions. Several studies collect data from parents on child mental health, such as anxiety or concerns about sex. No evidence is found of adverse effects from the interventions.

How good is the evidence?

Medium-weak. All studies are rated as overall medium (some concerns) or high risk of bias, except one RCT (*White 2019*) and the two studies of Red Flag and Green Flag which are low risk of bias (Table 3). However, these latter two studies are not RCTs, which generally give a higher quality of causal evidence.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
-------------------------	----------------------	--------------------	--	----------------------	----------------------------	----------------------------------

Barron 2013	Some concerns	Some concerns	Low risk	Low risk	Low risk	Some concerns
Daigneault 2015	High risk of bias	Some concerns	Some concerns	High risk	High risk	Some concerns
Del Campo Sanchez 2006	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Devries 2015	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
Gushwa 2018	High risk of bias	Some concerns	Low risk	High risk	Some concerns	Some concerns
Hazzard 1991	Some concerns	Some concerns	Some concerns	Low risk	Some concerns	Some concerns
Oldfield 1996	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
White 2019	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Czerwinski 2018	Moderate risk of bias	Moderate risk	Low risk	Low risk	Moderate risk	Moderate risk	Moderate risk	Low risk
Elfreich 2020	Moderate risk of bias	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk
Kolko 1987	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk
Kolko 1989	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk

Prevention / Child Safety: Maltreatment Occurrence/Recurrence

School-based interventions can reduce use of violence by teachers against students. Results need to be interpreted with caution due to multiple studies with high risk of bias. Inconsistent results for gender-based violence and peer violence among adolescents.

Evidence status	High risk of bias	Weak strength evidence to reduce violence by teachers against students. Inconsistent findings for gender-based violence and peer violence.
------------------------	--------------------------	--

The summary in brief

There are various types of school-based interventions to prevent child maltreatment. This synthesis features three categories: 1) programmes focused on preventing the use of violence as punishment by teachers within schools, particularly in settings where it is widely prevalent; 2) programmes for younger children delivered in the classroom to prepare them to recognise, avoid, and report instances of sexual abuse; and 3) prevention programmes for older children (middle school and high school) focused on interpersonal violence, dating violence, and bystander action. This category includes lessons for children aimed at preventing victimization and discouraging children from perpetrating offenses against others.

This cell includes studies that report on reductions in children reporting violence against them.

Training teachers seem to succeed in reducing violence against children. Results need to be considered with caution due to the high risk of bias in many studies.

Studies from the latter two categories are from US programmes. Results are inconsistent on reducing actual abusive incidents across studies.

There are many evaluations of school-based programmes in the EGM but very few studies report on actual incidence of child maltreatment. The majority report on intermediate outcomes such as knowledge, awareness, and attitudes.

Contents of the cell

The cell has four protocols (*Wangamati 2022, Knight 2020, Scharpf 2021, Masath 2020*) and 19 studies (*Devries 2015, Devries 2017, Devries 2018, Edwards 2019, Edwards 2021, Knight 2018, Merrill 2018, Nkuba 2018, Taylor 2010, Kolko 1987, Kolko 1989, Baker-Henningham 2019, Baker-Henningham 2021, Fabbri 2021, Hecker 2021, Jewkes 2019, and Ssenyonga 2022*).

Six studies are from the US; five papers are from the evaluation of the Good School Toolkit in Uganda, and three studies used the *Interaction Competencies with Children (ICC)* approach for teachers and institutional caregivers in Tanzania and Uganda. Jamaica (two studies), Tanzania (a refugee camp), South Africa, and Pakistan are the sites of the remaining studies. All protocols are from studies happening in various countries in sub-Saharan Africa.

The interventions

All the interventions are school based. There are three main themes to the types of interventions in this cell.

1. Reducing violence by teachers: All studies from Africa (except one from South Africa) and the two from Jamaica focussed on this aim. The core approach to achieving this was through planned

training programmes for teachers. Training took place over multiple sessions away from the classroom (or at least students) and aimed at changing attitudes regarding violence as punishment and providing teachers with alternate methods of disciplining students. Participating teachers taught in preschool, primary, and secondary schools. Some interventions only include the training programme but in others, such as the *Good School Toolkit*, a whole school approach is used, involving students, teachers, school administrators, parents, and caregivers. Typically, in these whole school approaches, children are empowered to lead programme coordination and planning (with the support of the implementers).

2. Gender-based violence prevention and peer violence prevention: These programmes include a specific focus on addressing sexual violence against girls and violence by peers. The participants in these interventions are middle schoolers and high schoolers, the age when children are transitioning into adulthood. One programme from South Africa called *Skhokho* (Jewkes 2019), much like the *Good School Toolkit*, uses a holistic approach involving the entire school to prevent gender-based violence. Another example is a programme that teaches self-defence against sexual assault to American Indian girls living on a reservation (Edwards 2021). A third programme, *Bringing in the Bystander – High School Curriculum*, aims to empower adolescents to say something when witnessing interpersonal violence among their peers. Another programme from Pakistan (Karmaliani 2020) uses a structured play-based life skills programme to reduce peer violence among 6th graders. A coach leads programme activities, which are followed by reflective discussions in the group.
3. School-based education for young children: *Red Flag/Green Flag* (Kolko 1987, Kolko 1989) involves teaching young children about inappropriate touching and what they can do if someone does that to them. This is the most common form of child abuse prevention programme implemented globally. An evaluation of *Early Head Start* in the US (Green 2020), which is an early childhood prevention programme available until age three to low-income families, found that the benefits to parents and families (by reducing stress and conflict) contribute to reduced abuse and neglect for many years in the child’s life.

Who delivers the intervention?

The *Good School Toolkit (GST)* developed by a Ugandan non-profit, Raising Voices – actively engages students, teachers, and school staff in various intervention activities. The *ICC* workshop is delivered by a psychologist. *The Red Flag/Green Flag programme* was chosen for use by an organization that develops community services for local school-age children. *Bringing in the Bystander* was originally developed by university researchers and has since been widely adopted in colleges, universities, and the military. *BITB-HSC* (Edwards 2019) is an adapted version for high-schoolers that is typically delivered by one person who identifies as a man and one as a woman. Other curricula are taught by experienced professionals from a local sexual assault centre, in most cases, or by regular classroom teachers.

Have the interventions been implemented at scale?

Most studies are small enough for researchers to test out various curricula. The *Good School Toolkit* is promoted by the NGO Raising Voices, which has implemented it in primary schools in one district and is planning to implement it in secondary schools in Uganda. It is also being adapted for Tanzania. The scale has been limited to date. *ICC*, which has different versions for teachers (*ICC-T*) and does not seem to have been widely implemented. It has been tested in institutional care (orphanages) and in primary schools albeit in small samples. The *Red Flag/Green Flag People* programme is common in the US, though the two studies were carried out in a small sample of schools. *BITB* has been implemented

widely in colleges, universities, and the military but it is not clear whether this specific adaptation, *BITB-HSC*, has been implemented at scale. *Early Head Start* is a well-established programme in the US available to low-income families throughout the country.

What do the interventions cost?

A separate study by Greco et al found GST to be cost-effective. Implementing GST over 18 months in 21 schools cost close to \$400,000. Monitoring and evaluation add another \$50,000 to the costs. The annual cost to run GST was approximately \$7,500 per school and \$15 per student. It costs close to \$250 to prevent a case of violence and approximately \$100 in annual implementation costs (excluding the initial implementation costs) for every prevented case.

Other programmes: While detailed cost information is not explicitly provided, programmes like *Red Flag/Green Flag People* or *BITB-HSC* are considered low-cost since they are incorporated within the regular school schedule and are of short duration.

How are the programmes meant to work? The theory of change

Most of the programmes are theory-informed, drawing from multiple behavioural theories rather than entirely depending on one specific behavioural change theory.

Do the interventions work in reducing child maltreatment occurrence and recurrence?

- **To prevent violence by teachers in schools:** Multiple studies reported a reduced risk of experiencing violence for children (GST and ICC-T). In most studies, teachers self-reported a reduced use of violence as punishment. However, in a few studies, this was not corroborated by children reporting less violence from teachers.
- **To prevent child abuse:** In the evaluation of the *Red Flag/Green Flag People* programme, in one study, the proportion of children who said they were touched inappropriately by an adult went from about 19% before the programme to about 11% immediately after the programme and approximately 6% six months later. The respective proportions for the control group were approximately 6%, 0%, and 7%. Importantly, the number of children in this study is very small which makes the reported effect sizes more pronounced than they probably are. In the other study, no significant differences were seen between groups in abusive encounters with adults immediately after the intervention. However, six months later, the intervention group children reported more abusive encounters than the control group, who reported none. It is hard to figure out whether the programme is effective given these results.
- **To prevent children from becoming victims and perpetrators of violence:** There are mixed results across studies. In some studies, there is a statistically significant decrease in victimization of various types, of violence especially sexual violence and intimate partner violence but in others, there is no statistical significance.

Are the results generalisable?

The corporal punishment prevention studies mostly come from Uganda and Tanzania are implemented on a small scale. School-based education programmes such as *Red Flag/Green Flag* are widely generalisable to other settings. Inconsistent findings for other outcomes limit generalizability.

How reliable is the evidence?

The studies are a mix of different quality in terms of risk of bias.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Baker-Henningham 2019</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
<i>Baker-Henningham 2021</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
<i>Devries 2015</i>	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
<i>Devries 2017</i>	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
<i>Devries 2018</i>	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
<i>Edwards 2019</i>	High risk of bias	Some concerns	Some concerns	High risk	High risk	Some concerns
<i>Fabbri 2021</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
<i>Green 2020</i>	Some concerns	Some concerns	Low risk	Low risk	Low risk	Low risk
<i>Hecker 2021</i>	Some concerns	Some concerns	Low risk	Some concerns	Some concerns	Low risk
<i>Jewkes 2019</i>	Some concerns	Some concerns	Low risk	Low risk	Low risk	Some concerns
<i>Karmaliani 2020</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

<i>Knight 2018</i>	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
<i>Merrill 2018</i>	High risk of bias	Low risk	Some concerns	Some concerns	High risk	Some concerns
<i>Nkuba 2018</i>	High risk of bias	Some concerns	High risk	High risk	Some concerns	Some concerns
<i>Ssenyonga 2022</i>	High risk of bias	Some concerns	Low risk	Some concerns	Low risk	High risk
<i>Taylor 2010</i>	High risk of bias	Low risk	Some concerns	High risk	High risk	Some concerns

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Kolko 1987</i>	Low risk of bias	Low	Low	Low	Low	Low	Low	Low
<i>Kolko 1989</i>	Low risk of bias	Low	Low	Low	Low	Low	Low	Low
<i>Edwards 2021</i>	Some concerns	Low	Low	Low	Low	Moderate	Moderate	Low

Prevention / Child wellbeing: cognitive functioning

Training preschool teachers and involving parents/caregivers early can improve cognitive function outcomes for young children. There are too few studies in primary schools to assess effectiveness.

Evidence status	Low risk of bias	Strong evidence that early childhood teacher training and parent/caregiver engagement can improve cognitive outcomes for under-5s.
------------------------	-------------------------	--

The summary in brief

This cell has studies from low- and middle-income countries, except for one study from the US. The evidence demonstrates that preschool is a good time to train teachers and engage with parents and caregivers. Children in these programmes have better cognitive development outcomes and academic outcomes than their peers not in these programmes. The interventions were helping teachers better manage their preschool classrooms, reduce parental stress, and reduce violence against children.

Similar improvements are not seen for primary school students, albeit from only one study.

Contents of the cell

This cell includes six studies, all recent except one from 2015, (*Baker-Henningham 2019, Baker-Henningham 2021, Baker-Henningham 2021a, Francis 2021, Devries 2015, and Green 2020*) and three protocols (*Knight 2020, Lopez Garcia 2021, Scharpf 2021*).

Four of the completed primary studies come from Jamaica and were conducted by the same group of researchers (*Baker-Henningham 2019, 2021, 2021, and Francis 2021*). The other studies and protocols are from Uganda, Tanzania, Haiti, Ghana, (one protocol includes multiple countries) and the US.

The interventions

All the interventions examined in this cell aim to address either violence by teachers or engage with parents of young children to reduce stress and prevent violence against children.

1. The *Irie Classroom Toolbox* (*Baker-Henningham 2021*) and other interventions like it (*Baker-Henningham 2021a, and Baker-Henningham 2019*) target training of preschool and grade 1 teachers. Training is delivered via group sessions by trained facilitators. Teachers are taught better classroom management and how to avoid using violence as a form of punishment. Additionally, teachers also receive in-classroom support. The *Irie Homes Toolbox* (*Francis 2021*) is a version of the same intervention but targeted at parents of preschoolers. Parents are recruited from their children's preschool and invited to weekly group sessions over two school terms. The sessions are meant to motivate parents away from harsh disciplinary methods and teach them techniques to improve the quality of engagement with their children.
2. The *Good School Toolkit* for adolescents in various primary schools in a district in Uganda where violence against children by teachers is common. This is a whole-school programme to prevent violence against children, which means it engages multiple stakeholders such as teachers, school administrators, parents, and caregivers. Students are empowered to coordinate and plan many of the activities of the programme to reduce violence by teachers and other adults. One ongoing study (*Knight 2020*) will look at the long-term outcomes of this programme.
3. *Early Head Start* is an American federally supported programme for low-income households with young children (a few months to five years). This programme includes day-care services but can also incorporate home visits to better engage parents and caregivers.

The remaining protocols are on teacher training programmes to reduce teacher violence against students in Haiti (*Lopez Garcia 2021*) and Ghana, Tanzania, and Uganda (*Scharpf 2021*).

Who delivers the intervention?

Programmes examined in this cell were delivered by either researchers, NGO staff (*Good School Toolkit*), or programme staff (*Head Start*).

Have the interventions been implemented at scale?

Head Start is widely available in the US. The other interventions are small scale.

What do the interventions cost?

GST: A separate study by Greco et al found *GST* to be cost-effective. Implementing *GST* over 18 months in 21 schools cost close to \$400,000. Monitoring and evaluation add another \$50,000 to the cost. The annual cost to run *GST* was approximately \$7,500 per school and \$15 per student. It costs close to \$250 to prevent a case of violence (total costs including producing the materials for *GST*) and approximately \$100 in annual implementation costs (to keep the programme running excluding initial costs to launch the programme) for every prevented case.

No cost information is available for the other interventions.

How are the programmes meant to work? The theory of change

Programmes are informed by various behavioural theories, but no specific theory is highlighted.

Do the interventions work in improving cognitive function?

For young children, yes. These interventions seem to improve outcomes such as reading, oral language skills, and scores on standardized cognitive development tests.

However, these findings did not extend to primary school children in the *GST* programme. All educational outcomes such as word recognition, reading comprehension, spelling, and written numeracy were not significantly different for students in *GST* compared to those who were not.

How reliable is the evidence?

Pretty reliable. Most studies were rated 'low risk of bias.'

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Baker-Henningham 2019</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Baker-Henningham 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Baker-Henningham 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Devries 2015	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
Francis 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Green 2020	Some concerns	Some concerns	Low risk	Low risk	Low risk	Low risk

Prevention / Child Wellbeing: Knowledge and Awareness

School-based sexual abuse prevention interventions improve children’s knowledge and awareness on prevention concepts and actions.

Evidence status	Low risk of bias	Strong evidence that sexual abuse prevention interventions increase children’s knowledge and awareness on sexual abuse prevention concepts and actions. Even though most of the studies in the cell are rated as ‘moderate risk of bias,’ the number of studies and consistency of results warrants high confidence in the findings.
------------------------	------------------	--

The summary in brief

This synthesis is of the cell with the largest number of studies (by far) in the EGM. It shows clearly that the bulk of evidence for preventing child maltreatment is school-based interventions delivered in the classroom, mostly about sexual abuse and mainly to young children. The aim is to teach them the concepts of sexual abuse prevention and empower them to recognise, avoid, and report sexual abuse (if it happens).

These programmes are usually short in duration (typically a few weeks) and delivered during regular school hours by trained personnel (often teachers). The studies of these programmes measure knowledge and awareness about sexual abuse prevention gained by the children after training - compared to a control group that did not receive the training.

Most studies show improvements in knowledge and awareness immediately after training, but these gains can fade in the long term.

Contents of the cell

This cell includes one 63 primary studies (and 1 paper) conducted over a long period (Blumberg 1991, Bustamante 2019, Daigneault 2015, Edwards 2019, Feldmann 2018, Krahe 2009, Ratto 1990, Taylor 2010, Wurtele 1992a,b, Barron 2013, Cecen-Erogul 2013, Chen 2012, Citak 2017, Conte 1985, Crowley 1989, Daigneault 2012, Dake 2003, del Campo Sanchez 2006, Grendel 1991, Harvey 1988, Hazzard 1991, Jin 2017, Oldfield 1996, Pulido 2015, Saslawsky 1985, Telljohann 1997, Tutty 1997, White 2018, Wolfe 1986, Wurtele 1986, Zhang 2014, Baker 2012, Czerwinski 2018, Dhooper 1995, Dryden 2014, Hebert 2001, Hillenbrand-Gunn 2012, Weatherley 2012, Kenny 2012, Kraizer 1991, Neherta 2017, Kolko 1987, Kolko 1989, MacIntyre 1991, Snyder 1986, Taal 1997, Warden 1997, White 2019, Jewkes 2019, Nickerson 2019, Jones 2020, Huang 2020, Juarez Diaz 2021, Thompson 2021, Weeks 2021, McElearney 2021, Bright 2022, Kang 2020, Urbann 2020, Kiziltepe 2021, and Edwards 2021).

In some cases, the same intervention was tested in different populations. For example, the effects of the ESPACE sexual abuse prevention workshop were tested in three different populations in various cities of Quebec, Canada (Hebert 2001, Daigneault 2012, Daigneault 2015).

This type of intervention, i.e., a sexual (or other abuse) prevention programme taught in the classroom for a limited time, is by far the most prevalent in the EGM. More than half of the studies in this cell come from the US with the rest from Canada, the UK, Europe, Australia, Central America, Africa, and East Asia. There are no studies from South Asia or South America. Earlier studies are almost exclusively from high-income countries while recent studies also have representation from low- and middle-income studies.

The interventions

All the interventions in this cell seem to be school-based prevention programmes, and about sexual abuse specifically. Some studies also focused on preventing other forms of abuse such as verbal abuse, physical abuse, and emotional abuse.

Most programmes target students in middle childhood (6-11 years) although a few also included early childhood students (2-5 years) and adolescents. When adolescents are targeted, the programme is

adjusted to include topics relevant to that age group like dating violence or promoting bystander action (on witnessing inappropriate behaviour).

Teachers are often part of the intervention: often the intervention is delivered by them in their classrooms. In some cases, parents were also involved.

The interventions typically consist of classroom-based education aimed at increasing the knowledge of young children on prevention concepts (related to sexual abuse) and providing them with skills to identify, avoid and respond to sexual abuse (and other forms of abuse). The interventions are generally offered over a short duration (a few weeks to a few months). Intervention activities are tailored to the children's age: programmes for younger children often include roleplay and simulations of inappropriate and appropriate interactions with adults.

Who delivers the intervention?

The interventions are delivered by researchers (who often have developed the programme and are looking to test whether it works or not), by personnel from community-based organisations; teachers in the classroom; or in some cases by older students (high school students teaching elementary school students).

Have the interventions been implemented at scale?

There are some examples of large-scale programmes. For example, *Stay Safe* (MacIntyre 1999) has been delivered in most primary schools in Ireland. However, most studies are evaluations from a small number of schools or classrooms.

What do the interventions cost?

While cost information is not explicitly provided in most studies (only one study seems to have provided this – see below), classroom-based prevention programmes are generally considered to be low-cost since they are of short duration and can be incorporated within regular school scheduling.

The *Kids Learning About Safety (KLAS)* programme for Latino preschoolers and their families in South Florida (Kenny 2012) reported an operating budget of \$150,000 for one year. The programme included 100 families (one child and one parent participant per family) but some families attended very few sessions. "The cost per group (averaging 10 child and 10 adult participants) is estimated to be \$7,000 (including indirect costs and salaries for the principal investigator, counsellors, and research assistant, as well as program materials, supplies, and incentive/milestone gifts). The cost per participant was approximately \$350."

Do the interventions work in improving child knowledge and awareness?

Absolutely. Almost all studies report that children's knowledge and awareness of concepts and actions on sexual abuse prevention significantly improved after participation in the programme compared to the control group. This is the most robust and consistent finding we have in the EGM for any outcome.

A few caveats need to be kept in mind:

1. Increased knowledge and awareness are intermediate outcomes and we do not know if increasing knowledge and awareness reduces *actual* incidence of child abuse / maltreatment.
2. Studies use a wide range of scales (some validated, some not) to measure knowledge and awareness. Since we have not conducted a meta-analysis in the EGM, i.e., a statistical method to pool effect sizes from different scales to get a common effect size, we are not able to quantify the size of the effect.

3. These interventions rather put the burden of prevention on children themselves. They aim to train and prepare children so that they can recognise, avoid, and report abusive situations. Training childcare providers or school administrators to shore up institutional responses to protect children are also important to help children avoid abuse.
4. Students' knowledge and awareness improves in the short-term with these programmes, but the effect probably fades over time. Repeated or refresher trainings might be needed to maintain knowledge and awareness.

Are the results generalisable?

Yes. The number of studies and the consistency of results indicate that results may be generalisable.

How reliable is the evidence?

Pretty reliable. While most studies are rated as 'moderate risk of bias,' the number of studies and the consistency of findings suggests that the evidence is quite reliable.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Blumberg 1991	High risk of bias	Some concerns	High risk	High risk	High risk	Some concerns
Bustamante 2019	High risk of bias	Low risk	High risk	High risk	Some concerns	Some concerns
Daignault 2015	High Risk of Bias	Some concerns	Some concerns	High risk	High risk	Some concerns
Edwards 2019	High risk of bias	Some concerns	Some concerns	High risk	High risk	Some concerns
Feldmann 2018	High risk of bias	Some concerns	Low risk	Low risk	High risk	Some concerns
Jones 2020	High risk of bias	High risk	Low risk	Some concerns	Some concerns	Low risk
Juarez Diaz 2021	High risk of bias	Some concerns	High risk	Some concerns	Some concerns	Low risk

Krahe 2009	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
Ratto 1990	High risk of bias	Some concerns	High risk	Some concerns	High risk	Some concerns
Taylor 2010	High risk of bias	Low risk	Some concerns	High risk	High risk	Some concerns
Tutty 1997	High risk of bias	Some concerns	Low risk	Low risk	High risk	Some concerns
Wurtele 1992b	High risk of bias	Some concerns	High Risk	High risk	Low risk	Some concerns
Kraizer 1988	High risk of bias	Some concerns	High risk	High risk	High risk	High risk
Barron 2013	Some concerns	Some Concerns	Low risk	Low risk	Low risk	Some concerns
Bright 2022	Some concerns	Some concerns	Low risk	Low risk	Some concerns	Low risk
Cecen-Erogul 2013	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Chen 2012	Some concerns	Some concerns	Some concerns	Low risk	Some concerns	Some concerns
Citak 2018	Some concerns	Some concerns	Low risk	Low risk	Low risk	Some concerns
Conte 1985	Some concerns	Low risk	High risk	Low risk	Low risk	Some concerns
Crowley 1989	Some concerns	Some concerns	Low risk	Some concerns	Some concerns	Some concerns
Daigneault 2012	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns

Dake 2003	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns
Del Campo Sanchez 2006	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Fryer 1987	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns
Grendel 1991	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns
Harvey 1988	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Hazzard 1991	Some concerns	Some concerns	Some concerns	Low risk	Some concerns	Some concerns
Jewkes 2019	Some concerns	Some concerns	Low risk	Low risk	Some concerns	Low risk
Jin 2017	Some concerns	Some concerns	Some concerns	Low risk	Some concerns	Some concerns
McElearn 2021	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Low risk
Oldfield 1996	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Pulido 2015	Some concerns	Some concerns	Low risk	Low risk	Low risk	Some concerns
Saslawsky 1976	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns
Telljohann 1997	Some concerns	Some concerns	Some concerns	Some concerns	Low risk	Some concerns

Thompson 2021	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Low risk
Weeks 2021	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Low risk
White 2018	Some concerns	Low risk	Low risk	Low risk	Low risk	Some concerns
Wolfe 1986	Some concerns	Some concerns	Some Concerns	Low risk	Low risk	Some concerns
Wurtele 1986	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns
Wurtele 1992a	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns
Zhang 2014	Some concerns	Some concerns	Some concerns	Low risk	Some Concerns	Some concerns
White 2019	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Nickerson 2019	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Huang, Cui 2020	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Baker 2012	Moderate risk	Moderate	Low	Low	Low	Low	Moderate	Low

	of bias							
Czerwinski 2018	Moderate risk of bias	Moderate	Low	Low	Moderate	Moderate	Moderate	Low
Dhooper 1995	Moderate risk of bias	Low	Low	Low	Low	Moderate	Moderate	Low
Dryden 2014	Moderate risk of bias	Low	Low	Moderate	Low	Low	Moderate	Low
Edwards 2021	Moderate risk of bias	Low	Low	Low	Low	Moderate	Moderate	Low
Hebert 2001	Moderate risk of bias	Low	Low	Low	Low	Moderate	Moderate	Low
Hillenbrand-Gunn 2012	Moderate risk of bias	Low	Low	Low	Low	Moderate	Moderate	Low
Kang 2020	Moderate risk of bias	Moderate	Low	Low	Low	Low	Moderate	Low
Urbann 2020	Moderate risk of bias	Low	Moderate	Moderate	Low	Moderate	Low	Low
Weathery 2012	Moderate risk of bias	Low	Low	Low	Moderate	Low	Moderate	Low

Kenny 2012	Serious risk of bias	Serious	Low	Low	Low	Moderate	Moderate	Low
Kraizer 1991	Serious risk of bias	Serious	Serious	Low	Moderate	Low	Moderate	Low
Neherta 2017	Serious risk of bias	Moderate	Serious	Moderate	Moderate	No info.	No info.	Low
Kiziltepe 2021	Low risk of bias	Low	Low	Low	Low	Low	Low	Low
Kolko 1987	Low risk of bias	Low	Low	Low	Low	Low	Low	Low
Kolko 1989	Low risk of bias	Low	Low	Low	Low	Low	Low	Low
MacIntyre 1991	Low risk of bias	Low	Low	Low	Low	Low	Low	Low
Synder 1986	Low risk of bias	Low	Low	Low	Low	Low	Low	Low
Taal 1997	Low risk of bias	Low	Low	Low	Low	Low	Low	Low
Ward 1997	Low risk of bias	Low	Low	Low	Low	Low	Low	Low

Prevention / Child Wellbeing: Mental Health

School-based sexual abuse prevention interventions do not increase anxiety in young children: a possible harm. Mixed findings on other mental health outcomes.

Evidence status	Moderate risk of bias	Moderate strength evidence that sexual abuse prevention interventions do not increase anxiety among young children. Unclear effects on other mental health outcomes.
------------------------	------------------------------	--

The summary in brief

This cell includes studies that report on mental health outcomes for children participating in various types of abuse-prevention programmes. They include: (1) classroom-based programmes for younger children to prepare them to recognize, avoid, and report sexual abuse; (2) programmes focused on preventing the use of violence as punishment by teachers in school in settings where it is widely prevalent; and (3) prevention programmes for older children (adolescents) to deter sexual harassment and prevent violence. The bulk of the evidence in this cell is from the first category.

The evidence suggests that abuse-prevention programmes do not lead to increased anxiety, which is often considered a potential “side effect” of sexual abuse prevention interventions. For other mental health outcomes, the results are mixed across studies.

Contents of the cell

This cell includes 25 primary studies (in 27 papers) conducted over a long time period (Devries 2015, Knight 2018, Ratto 1990, van Lieshout 2019, del Campo Sanchez 2006, Fryer 1987, Kraizer 1988, Grendel 1991, Hazzard 1991, Oldfield 1996, White 2018, Wurtele 1992a, Czerwinski 2018, Weatherley 2012, MacIntyre 1991, Taal 1997, Baker-Henningham 2019, White 2019, Jewkes 2019, Green 2020, Karmaliani 2020, Baker-Henningham 2021, Fabbri 2021, Baker-Henningham 2021a, Francis 2021, Urbann 2020). Six are QEDs. Additionally, there are protocols for three new RCTs (Masath 2020, Lopez Garcia 2021, Scharpf 2021).

Except for one intervention that took place in a residential care setting for boys, all interventions in this cell are school based.

There is a full summary of Karmaliani 2020 in the guidebook.

The interventions

The interventions can be classified into four broad categories:

1. Child abuse prevention programmes:

- Number of programmes: 15 programmes were assessed, with one programme reported in two papers.
- Countries: Eight programmes are from the US, two from Australia, and one each from Spain, Germany, the Netherlands, Malaysia, Australia, and the Republic of Ireland.
- Participants: Most programmes target students in middle childhood (6-11 years), although a few also include early childhood students (2-5 years). Teachers play an active role in the intervention, often delivering the curriculum in their classroom. In some cases, parents are also involved. One study specifically included deaf and hard-of-hearing children in Germany (Urbann 2020).

- Interventions: The interventions typically consist of classroom-based education to increase knowledge and awareness of young children on prevention concepts related to sexual abuse and to provide them with skills to identify, avoid, and respond to sexual abuse. Intervention activities are tailored to younger children and often include roleplay and simulations of inappropriate and appropriate interactions with adults. The interventions are generally offered over a short period (a few weeks to a few months).

2. Preventing violence by teachers:

- Number of programmes: Five programmes with one programme reported in two papers.
- Countries: Jamaica (three programmes by the same researchers), Uganda and Tanzania (one programme each)
- Participants: Pre-schoolers to adolescents, along with teachers, school administrators, and parents/caregivers. The programme in Tanzania took place in Nyarugusu refugee camp.
- Interventions: The Jamaican studies are based on an intervention called the 'Irie Classroom Toolbox'. The intervention is for teachers of preschoolers and grade 1 students. Trained facilitators offer multiple sessions of training to teachers with the goals of reducing their use of violence against children and improving their classroom management skills. In addition to the training sessions, teachers also receive in-classroom support. The study in Uganda is of a whole-school approach to reducing violence by teachers called the 'Good School Toolkit'. This is a school-wide programme involving teachers, students, school administrators, parents, and community members. The goal is to change the culture and environment of schools to move away from violence and improve relationships among various stakeholders, i.e., students, teachers, parents, and community members. The Tanzanian study is based on an intervention called 'EmpaTeach' that is focussed on reducing impulsive violence by teachers against students. This intervention is implemented in the Nyarugusu refugee camp in Tanzania that houses refugees from the Congo and Burundi. The intervention, like the others in this category, targets teachers' knowledge, attitudes, beliefs, and behaviours regarding the use of violence as a form of punishment. Teachers participate in self-guided groups inspired by the principles of cognitive behavioural therapy.
- All three RCT protocols in this cell also target preventing violence by teachers and are based on the same curriculum 'Interaction Competencies with Children – for Teachers (ICC-T).' The studies are taking place in various sites in sub-Saharan Africa and in Haiti.

3. Programme focused on preventing adolescents from harassing or being violent against their peers:

- Number of programmes: Three
- Countries: The Netherlands, Pakistan, and South Africa
- Participants: Adolescent boys and girls (11-17 years) attending primary or secondary school and living in youth residential care settings (one study).
- Interventions: For the boys in residential care, the intervention is a sexual harassment prevention programme delivered by freelance trainers via weekly sessions over two months. The main goals are respectful relationships and preventing sexual harassment. The other two interventions are school-based programmes. One in South Africa 'Skhokho' uses a life skills curriculum, teacher training, and workshops for teenagers and caregivers to prevent gender-based violence against girl students. The other from Pakistan uses a play-based intervention

delivered by trained coaches in school. Children are encouraged to reflect on how the activities made them feel and to relate them to their daily lives at school and home. The goals here are to prevent peer violence and reduce depression although these are not explicitly stated to participants.

4. Programme focused on parents/caregivers:

- Number of programmes: One
- Country: Jamaica
- Participants: Parents and caregivers of preschoolers
- Interventions: The intervention called the 'Irie Homes Toolbox' is inspired by the classroom version described earlier. The target participants for this intervention are parents/caregivers of preschoolers. Parents received multiple training sessions over two semesters of the school year at their children's school. The goal is to reduce harsh punishment and improve engagement with their children.

Who delivers the intervention?

The *Good School Toolkit* actively engages students, teachers, and school staff in various intervention activities. It was developed by a Ugandan non-profit, Raising Voices, and though the papers don't specify clearly, they imply that it was delivered by them.

The other programmes in this cell are delivered by researchers (who often have developed the programme), by personnel from community-based organisations, by teachers in the classroom, or by older students (high school students teaching elementary school students).

Have the interventions been implemented at scale?

There are some examples of large-scale programmes. For example, *Stay Safe* (MacIntyre 1999) has been delivered in most primary schools in Ireland. However, most studies evaluate interventions conducted in a few schools or classrooms.

What do the interventions cost?

GST: A separate study by Greco et al found *GST* to be cost-effective. Implementing *GST* over 18 months in 21 schools cost close to \$400,000. Monitoring and evaluation add another \$50,000 to the cost. The annual cost to run *GST* was approximately \$7,500 per school and \$15 per student. It costs close to \$250 to prevent a case of violence (total costs including producing the materials for *GST*) and approximately \$100 in annual implementation costs (to keep the programme running excluding initial costs to launch the programme) for every prevented case.

Other programmes: While cost information is not explicitly provided, classroom-based prevention programmes are generally considered to be low-cost since they are of short duration and can be incorporated into regular school scheduling.

Do the interventions work in improving mental health?

1. **Anxiety:** School-based prevention curricula on sexual abuse prevention for young children can cause anxiety given the nature of the topic and the content of the programme (e.g., "good touch" vs. "bad touch"). Therefore, studies on these programmes assess anxiety as a "side effect". The goal is to avoid increasing anxiety levels. Approximately half of the studies on school-based sexual

abuse prevention programmes reported on anxiety and, *in general, found that anxiety levels did not increase for children* in the programme or compared to the control group. The size of the impact is not as important here since the main issue is to see if anxiety increases after exposure to the programme.

2. **Self-esteem and pro-social behaviours:** Children with higher self-esteem before participating in sexual abuse prevention interventions tended to learn programme content and skills better. Self-esteem and pro-social behaviours (such as “helping peers”) improved for intervention group children compared to controls albeit, in a few studies while other studies showed no differences for these outcomes between intervention participants and those in the control group. However, the one study on sexual harassment prevention did not find any difference in self-esteem, self-efficacy, or emotional intelligence between interventions and controls.

3. **Wellbeing:** Mental health and wellbeing status remained mostly the same after intervention and were similar for the intervention and control groups. Children with mental health difficulties were likely to be least exposed to the intervention (the papers are not clear about why), which meant they were most vulnerable to experiencing violence from staff.

Are the results generalisable?

On anxiety and self-esteem, the results seem generalisable because there are several studies on school-based sexual abuse prevention interventions. For the other categories of interventions (preventing violence by teachers; preventing sexual harassment), the small number of studies and mixed results on mental health outcomes limits generalisability.

How reliable is the evidence?

Moderately so: see below. The risk of bias in the studies is mixed.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Baker-Henningham 2019	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Baker-Henningham 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Baker-Henningham 2021a	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Devries 2015	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns

Fabbri 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Francis 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Green 2020	Some concerns	Some concerns	Low risk	Low risk	Low risk	Low risk
Jewkes 2019	Some concerns	Some concerns	Low risk	Low risk	Low risk	Some concerns
Karmaliani 2020	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Knight 2018	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
Ratto 1990	High risk of bias	Some concerns	High risk	Some concerns	High risk	Some concerns
vanLieshout 2019	High risk of bias	Some concerns	High risk	High risk	Some concerns	Some concerns
del Campo Sanchez 2006	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Fryer 1987	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns
Grendel 1991	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns
Hazzard 1991	Some concerns	Some concerns	Some concerns	Low risk	Some concerns	Some concerns
Kraizer 1988	High risk of bias	Some concerns	High risk	High risk	High risk	High risk
Oldfield 1996	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns

White 2018	Some concerns	Low risk	Low risk	Low risk	Low risk	Some concerns
White 2019	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Wurtele 1992a	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Czerwinski 2018	Moderate risk of bias	Moderate	Low	Low	Moderate	Moderate	Moderate	Low
Weatherley 2012	Moderate risk of bias	Low	Low	Low	Moderate	Low	Moderate	Low
Kraizer 1991	Serious risk of bias	Serious	Serious	Low	Moderate	Low	Moderate	Low
MacIntyre 1991	Low risk of bias	Low	Low	Low	Low	Low	Moderate	Low
Taal 1997	Low risk of bias	Low	Low	Low	Low	Low	Moderate	Low
Urbann 2020	Some concerns	Low	Moderate	Moderate	Low	Moderate	Low	Low

Prevention / Child Wellbeing: Social-Emotional Functioning

School-based interventions to prevent child abuse have mixed results on children’s social-emotional functioning.

Evidence status	Moderate risk of bias	Moderate strength evidence but results are mixed for social-emotional functioning.
------------------------	------------------------------	--

The summary in brief

School-based interventions to prevent child abuse typically involve classroom-based courses and workshops for primary school children. The trainings aim to provide knowledge and skills to children to help them avoid becoming victims of child abuse and be empowered to disclose abuse when it happens. One outcome studied is the impact of these trainings on children’s social-emotional functioning. Another type of school-based intervention is to train teachers on better classroom management so that they are less violent towards children.

This cell has six primary studies; five are school-based and one is in a youth care institution. Two studies found that positive behaviours – including better conflict resolution skills, being more communicative and assertiveness – increase because of the training, as observed by parents and teachers. They did not find harms (e.g., increases in inability to sleep, fear of known and unknown adults). The other two studies found no such change after intervention.

The fifth study was to prevent adolescent boys living in residential homes from sexually harassing others. This study found no effect of the intervention.

The teacher training study did not find any increase in pro-social behaviours among children, although the intervention reduced teacher violence.

The completed primary studies vary vastly in age. They are all from high-income countries, plus Jamaica.

The one systematic review in this cell found that child abuse prevention programmes in schools incorporating social-emotional skills improve children’s knowledge about abuse but not self-protective skills.

Contents of the cell

A. Primary Studies

The cell has six primary studies (one of which is 25 years old) and two protocols (both in low- and middle-income countries):

Baker-Henningham 2019 RCT, low risk of bias.	Jamaica (seven primary schools with 27 teachers). Children in first grade: Middle childhood (6-11 years). Evaluation of a teacher training programme to improve classroom management.
Hebert 2001 QED, moderate risk of bias.	Canada (two schools in Quebec City). Children in first and third grades: Middle childhood (6-11 years). Evaluation of ESPACE child sexual abuse (and other abuse) prevention programme.
Daigneault 2012	Canada (three schools in Montreal). Children in first and third grades: Middle childhood (6-11 years).

RCT, moderate risk of bias.	Evaluation of <i>ESPACE</i> + booster (two years later) on violence prevention.
Taal 1997 QED, low risk of bias.	The Netherlands (Amsterdam elementary schools). Children in middle childhood (6-11 years). Evaluation of <i>Right to Security</i> sexual abuse prevention programme.
van Lieshout 2019 RCT, high risk of bias	The Netherlands (20 youth care institutions). Adolescent boys (12-17 years of age). Testing <i>the Make a Move</i> programme to prevent sexual harassment and promote respectful relationships for at-risk boys living in residential care.
del Campo Sanchez 2006 RCT, moderate risk of bias	Spain (primary schools). Primary school children: Middle childhood (6-11 years). Evaluation of <i>Prevención de abusos sexuales a menores</i> , a child sexual abuse prevention programme.
Lopez Garcia 2021 RCT protocol, ongoing study	Haiti (primary schools). Interaction Competencies with Children for Teachers (ICC-T). a training programme to reduce corporal punishment by schoolteachers
Scharpf 2021 RCT protocol, ongoing study	Ghana, Tanzania, Uganda (primary and secondary schools). Interaction Competencies with Children for Teachers (ICC-T). a training programme to reduce corporal punishment by schoolteachers

The interventions

Programme	Country	Description
<i>ESPACE</i> (Hebert 2001)	Canada	Quebec's adaptation of the <i>American Child Assault Prevention (CAP)</i> program. Classroom workshops (60-75 minutes) delivered by specialised community workers to first and third graders. Role-play, guided discussion, behaviour modelling, and rehearsal are strategies used in workshop to teach children basic prevention concepts and skills and make them aware of their personal rights. Children are taught to be assertive, to yell in self-defence and are to reach out to friends and a trusted adult if any abuse occurs. The programme's focus is expansive, covering sexual abuse, verbal abuse, physical abuse, and bullying. After the workshop, children can meet individually with community workers. Parents / guardians are also invited to a meeting and can also attend the workshop.
<i>ESPACE</i> + booster (Daigneault 2012)	Canada	The same <i>ESPACE</i> programme as described above was implemented in three low socioeconomic status Montreal public schools. Two years after the <i>ESPACE</i> programme, children currently in third and fourth grade (first and second grade during <i>ESPACE</i>) received either a complete <i>ESPACE</i> booster (same intervention again) or a brief version of it (to save time and money). Children currently in fifth and sixth grade (third and fourth grade during <i>ESPACE</i>) received a comprehensive intervention of <i>ESPACE</i> plus a general violence

		prevention workshop called <i>Confidence, Solidarity, Respect (CSR)</i> . <i>CSR</i> is meant to build on <i>ESPACE</i> and is more suited for older children.
<i>Right to Security</i> (Taal 1997)	The Netherlands	The programme is an initiative of the Amsterdam Prevention Council for Sexual Violence. The Programme is delivered to students in grades six to eight. The themes include ‘yes’ or ‘no’ feelings; the right to refuse unwanted sexual behaviours; and the right to seek help if abuse occurs. The main goal is to empower the child to decide what is right and wrong. The programme includes eight lessons – three delivered by actors simulating various situations and the rest by teachers (who received training for this). The lessons are meant to facilitate the realisation of the goals of the programme. Parents are invited to an informative meeting before the programme is implemented.
<i>Make a Move</i> (van Lieshout 2019)	The Netherlands	A sexual harassment prevention programme developed by the Dutch organisation Rutgers (Center for Sexual and Reproductive Health and Rights) for boys (aged 12-17) living in Dutch youth care. It is delivered by freelance trainers. The main goals are respectful relationships and preventing sexual harassment. It has a cognitive-behavioural approach and tackles determinants of sexual harassment such as attitudes and socio relational skills. It consists of eight themed weekly meetings lasting 90 minutes each. The themes are “men, image, girls, sex, flirting, dating, pleasurable sex, and the future.” Roleplay, discussion, and video clips are used to engage participants. Credits are earned for participation and good manners, with the final winner getting a prize.
<i>Prevención de abusos sexuales a menores</i> (del Campo Sanchez 2006)	Spain	The first of its kind, it was a school-based sexual abuse prevention programme in Spain. The Programme was delivered in classrooms to children aged 8-12 years. Teachers and parents received some training and orientation before the programme.
Teacher training programme for grade 1 teachers	Jamaica	School-based violence prevention programme for grade 1 teachers in various Jamaican primary schools. Teachers participated in workshops and received in-class support. The purpose of the programme was to help teachers improve classroom management with the twin goals of reducing violence against children and child aggression in the classroom.

Do the interventions work in improving children’s social-emotional functioning?

The results are mixed.

Some findings were positive:

- Hebert 2001 reported that, two weeks after participating in *ESPACE*, most children did not show any negative side effects (sleeping problems; being socially isolated; fear of known / unknown

adults; being disobedient; being clingy; being aggressive to siblings / peers) as observed by their parents. Some parents said their children seemed to be a little more afraid of strangers (25%) or clingier (13%). Note that the categories were “no change; a little; a lot.” More aggressiveness was reported towards peers (14% a little; 1% a lot), siblings (20% a little; 9% a lot), and disobedience (21% a little; 8% a lot). Very few parents said that any of these were a problem (aggression towards siblings was highest at close to 10%). For positive effects, more than half the parents said their children were more forthcoming about what they liked (57%) and did not like (53%) and they showed more self-confidence (54%). Parents also said that their children dealt better with conflict situations (46%), were more assertive (47%), and showed greater autonomy (42%). These measures were not reported for the control group.

- Del Campo Sanchez 2006 found that after the programme, parents reported increases in children’s positive behaviours such as asking more about sexual abuse (35%) and sexuality (27%); more open about their feelings (56%); and improving conflict management skills (42%). Teachers too observed increases in supporting peers (71%); being more assertive (43%); self-confidence, and conflict management skills (both 86%). The negative behaviours observed were minimal. Comparison group parents and teachers were not surveyed for these measures.
- Baker-Henningham 2019 found no benefits for pro-social behaviour, although teachers were less violent against children and the classroom environment improved.

Some studies found no effects:

- Daigneault 2012 found no significant effects on self-efficacy and empathy of the *ESPACE + booster* programme.
- Taal 1997 reported that subtest scores for “Relationships with Teacher” and “Relationships with Classmates” were unchanged one week and six weeks, after the programme, respectively. Social anxiety was unchanged at one week, but at six weeks a larger proportion of children rated their fear as ‘low’ (48% vs. 38%). This increase was in sixth and seventh graders since most eighth graders (52%) already considered themselves fearless!
- Van Lieshout 2019 did not find any significant effects on any of the 14 measures in the study. Those included self-efficacy, self-esteem, social norms, and empathy.

Have the interventions been implemented at scale?

ESPACE appears to have been implemented on a large scale, reaching hundreds of thousands of Quebecers over the years. However, the other studies are based on smaller samples. *Make a Move* seems to have been developed recently (so it is probably not implemented at scale) as a prevention intervention for sexual harassment targeting at-risk youth living in residential care. The extent of the implementation scale is unclear for either *Right to Security* or *Prevención de abusos sexuales a menores*. The teacher training programme in Jamaica was a small-scale research study.

Which type of organisation delivered the intervention?

ESPACE, *Make a Move*, and *Right to Security* are delivered by community-based organisations. The details for *Prevención de abusos sexuales a menores* are not clear.

What do the interventions cost?

None of the studies report the detailed costs of the intervention.

How are the programmes meant to work? The theory of change

None of the studies mention a specific theory on which they are based.

How reliable is the evidence?

Pretty reliable.

Taal 1997 is rated as having a *low risk of bias*; Hebert 2001, del Campo Sanchez 2006, and Daigneault 2012 are rated as having a *moderate risk of bias*; and van Lieshout 2019 is rated as a *high risk of bias*.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Baker-Henningham 2019	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
del Campo Sanchez 2006	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Daigneault 2012	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns
van Lieshout 2019	High risk of bias	Some concerns	High risk	High risk	Some concerns	Some concerns

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Taal 1997	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk	Moderate risk	Low risk
Hebert 2001	Moderate risk of bias	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk

B. Systematic Review (Gubbels 2021)

What is the systematic review about?

Gubbels 2021 is a systematic review with multiple meta-analyses to understand key components of effective school-based abuse prevention programmes.

What are the findings on children's social and emotional functioning outcomes?

Programmes that focussed on the social-emotional skills of children were more effective in improving knowledge about abuse than those that didn't. However, this was not the case for self-protection skills.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

How reliable is the evidence?

The systematic review is rated as 'high risk of bias'. This means that it has at least one major flaw that affects our confidence in the finding.

Prevention / Adult Perpetrator: Desistance

Training programmes for teachers, parents, and other adult care providers can reduce violence by caregivers against young children.

Evidence status	Low risk of bias	Structured training programmes are effective interventions to reduce violence against children by adult caregivers such as teachers and parents.
------------------------	-------------------------	--

The summary in brief

This cell has four primary studies and one protocol for an RCT. All studies are on training adult caregivers of children such as teachers, parents, and orphanage care providers to not use violence as a form of discipline against children. The children in these studies were in preschool, first grade or older children living in orphanages (average age: 9 years).

Teachers are taught techniques to improve classroom management. The training focuses on changing attitudes towards reduced violence against children.

Adult caregivers report that they use less violence after participating in training. However, this is self-reported and is often hard to verify. The findings across studies are consistent and this approach seems promising as a strategy.

Contents of the cell

The cell has four completed studies (*Baker-Henningham 2019, Baker-Henningham 2021, Francis 2021, Hecker 2021*) and one ongoing RCT protocol (*Lopez Garcia 2021*).

<i>Baker-Henningham 2021</i> RCT, low risk of bias.	Jamaica. Preschoolers. Evaluation of the Irie Classroom Toolbox, a programme aimed at improving classroom management by teachers and socioemotional development of children.
<i>Baker-Henningham 2019</i> RCT, low risk of bias.	Jamaica. Children in Grade 1. Evaluation of a violence prevention teacher training programme aimed at reducing corporal punishment and improving classroom management. Adapted from the Irie Classroom Toolbox.
<i>Francis 2021</i> RCT, low risk of bias.	Jamaica. Preschoolers. Evaluation of the Irie Homes Toolbox, a parenting programme delivered through schools aimed at reducing harsh punishment by parents and improved engagement with children.
<i>Hecker 2021</i> RCT, some concerns about possible bias.	Tanzania. Children average age 9 years. Evaluation of Interaction Competencies with Children – for Caregivers (ICC-C), a training programme for care providers at various orphanages in Dar es Salaam.
<i>Lopez Garcia 2021</i> RCT protocol, ongoing study	Haiti. Primary school children. Evaluation of Interaction Competencies with Children – for Teachers (ICC-T), a teacher training programme for primary school teachers to prevent violence against children.

The interventions

All the interventions examined in this cell are training programmes for adults who engage closely with children. In most programmes, teachers are the target of training (Baker-Henningham 2019, Baker Henningham 2021, Lopez Garcia 2021). There is also one study on a parent-focused programme delivered via preschools (Francis 2021) and another on training care providers in orphanages (Hecker 2021).

Three of the interventions were implemented by the same group of researchers albeit in different populations, i.e., first grade teachers, preschool teachers (Irie Classroom Toolbox), and parents of preschoolers (Irie Home Toolbox). The two other interventions are implemented by another group of researchers again in different populations, i.e., orphanage care providers (ICC-C) and primary school teachers (ICC-T).

Baker-Henningham (2019) and Baker-Henningham (2021) are both based on the Irie Classroom Toolbox, a universal violence prevention intervention for Grade 1 and preschool teachers. Most schools in the study are in urban, low-income areas. The goals of the programme are to reduce violence by teachers against students and aggressive behaviour by children. To achieve this, teachers participate in eight full-day workshops and four 1-hour sessions of in-class support delivered over six months. Teachers also get educational materials (books, manipulatives, Play-doh), picture cards, positive notes home, stickers, stamps, puppets, behaviour planning forms and other accessories.

The same researchers developed another version of the intervention called the Irie Homes Toolbox, targeted at parents/caregivers of young children (Francis 2021). The goals here are to prevent harsh punishment by parents, increase their involvement in their child's life, and to reduce behavioural difficulties for children. The intervention is delivered via teachers at the same preschools that participated in the Classroom Toolbox version. Parents participate in 90-minute sessions weekly over two successive school terms (six months). The content includes tips on managing child problem behaviours, managing emotions, positive parenting techniques and helping with schoolwork. It is not clear from the paper who delivered the training but ideally teachers trained in the classroom version of the Toolbox would deliver the Homes Toolbox programme to parents.

Hecker 2021 tested a programme called Interaction Competencies with Children – for Caregivers (ICC-C) targeted at orphanage care providers in various orphanages of Dar es Salaam, Tanzania. The main goals are to reduce child maltreatment and to improve the relationship between children and care providers. Multiple sessions on topics such as child development, caregiver-child relationship, effective caregiving strategies, maltreatment prevention, supporting children, child-centred institutional care, and teamwork and supervision are delivered over two weeks. Professionally trained psychologists deliver training, customizing the programme based on the participants needs. The average age of children cared for by orphanage care providers in this programme is 9 years.

The protocol in this cell (Lopez Garcia 2021) is for an RCT that is testing a version for teachers called Interaction Competencies with Children – for Teachers (ICC-T) in various primary schools in Haiti. The goals and format of the programme are very similar with the content for teachers in this intervention.

Do the interventions work in improving child maltreatment disclosure?

Yes. In all studies, adult caregivers in the intervention group said that they reduced their use of harsh punishment against children. The reductions were statistically significantly.

In one study (Baker-Henningham 2019), teachers were violent against their Grade 1 students an average four times a day in the intervention group compared to twelve times a day for teachers who did not take part in the programme (the control group). Preschool teachers in the Irie Classroom Toolbox intervention

(Baker-Henningham 2021) had 67% fewer counts of violence against children compared to the control group immediately after the programme ended and 54% less one year after the programme began. Parents in the Irie Home Toolbox group (Francis 2021) reported a 14% reduction in use of violence after doing the intervention compared to no change in the comparison group. However, in the study from Tanzania (Hecker 2021), adult care providers in orphanages said that they used less violence against children (after participating in the programme) but when children were independently asked, they did not think the harsh discipline had reduced.

Have the interventions been implemented at scale?

Not quite. All of these are relatively small research studies implemented in multiple schools and institutions but not truly scaled up.

Which type of organisation delivered the interventions?

Trained facilitators or the researchers conducted the training for adult caregivers. For the Irie Homes Toolbox, ideally teachers trained with the Irie Classroom Toolbox deliver the programme to parents, but it is unclear if that actually happened in the study.

What do the interventions cost?

Detailed cost analyses are not available, but the researchers claim that these interventions could be kept low-cost.

How are the programmes meant to work? The theory of change

The ICC-C programme is based on “attachment, behavioural, and social learning theories” as well as other programmes and research in this area. The Irie Toolbox interventions are based on the ‘Incredible Years Teacher Training Programme,’ an evidence-based approach to addressing behavioural issues in young children by training teachers and parents but no specific theory is cited in the studies.

How reliable is the evidence?

Quite reliable. Three out of four completed RCTs are rated as low risk of bias.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Baker-Henningham 2019</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
<i>Baker-Henningham 2021</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Francis 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Hecker 2021	Some concerns	Some concerns	Low risk	Some concerns	Some concerns	Low risk

Prevention / Child or Youth Offender: Maltreatment Behaviour

There is an unclear effect of school-based programmes in preventing peer violence among middle schoolers (12-14 years).

Evidence status	Some concerns	There are mixed results from studies of school-based interventions in preventing peer violence and gender-based violence.
------------------------	----------------------	---

The summary in brief

This cell has three primary studies. All examine school-based prevention programmes for middle-schoolers. The programmes come from various geographical regions: the US, Pakistan, and South Africa. The approach adopted in each is different: one is curriculum-based, another is life skills and holistically implemented, and the third is play-based.

The results from the studies are mixed. Whilst there are some indications that violence experienced and committed by students who participated in the interventions was reduced, many of the findings are not statistically significant, meaning the effect is only small or based on a small sample and so is not reliable.

Contents of the cell

The cell has three studies (*Taylor 2010, Jewkes 2019, Karmaliani 2020*).

There is a full summary of Karmaliani 2020 in the guidebook.

<i>Taylor 2010</i> RCT, high risk of bias.	USA. 6 th and 7 th graders. Evaluation of a curriculum-based gender violence and sexual harassment prevention intervention among adolescents.
<i>Jewkes 2019</i> RCT, some concerns.	South Africa. 8 th graders. Evaluation of Skhokho, a multi-pronged approach to prevent gender-based violence involving students, teachers, school personnel, and parents/caregivers.
<i>Karmaliani 2020</i> RCT, low risk of bias.	Pakistan. 6 th to 8 th graders. Evaluation of Right to Play’s structured play-based intervention with trained coaches to reduce peer violence and improve outcomes for depression.

The interventions

All interventions in this cell are school-based programmes for middle schoolers (12-14 years).

The intervention in Taylor 2010 aimed to prevent gender violence and sexual harassment among 6th and 7th graders attending suburban schools in a US city. Two curricula on gender violence, harassment and “dating violence” prevention were implemented. One curriculum was interaction-based and focused on “setting and communicating boundaries in relationships; the formation of deliberate relationships and friendships and the continuum between friendship and intimacy; the determination of wanted and unwanted behaviors; and the role of the bystander as intervener.” The lessons were intended to challenge students to think about these issues. The other curriculum took a “law and justice” approach. It included information on “laws, definitions, information, and data about penalties for sexual assault and sexual harassment, as well as imparting results from research about the consequences for perpetrators of GV/SH”. The interaction-based curriculum was meant to promote understanding and conversations about GV/SH, while the law and justice curriculum presented facts, data, and legal consequences.

Skhokho (Jewkes 2019) examined a holistic gender-based violence prevention programme. It targets various aspects of gender-based violence, such as promoting gender equality, challenging harmful gender norms, enhancing critical thinking skills, fostering positive relationships, and providing support systems for students. It is implemented in schools, using a combination of classroom-based activities, teacher training, peer education, and community engagement. It includes multiple interventions, such as a life-orientation curriculum workbook for eighth grade students, training for teachers, and a workshop for parents/caregivers and teenagers, which was rolled into student clubs in the second year of the programme.

Right to Play, an international NGO, implemented a play-based intervention for middle-schoolers in multiple schools in a large Pakistani city (Karmaliani 2020). The intervention included trained coaches who led students through manualized play-based activities (more than 100 of them over two years of the programme) followed by a period of reflection on the activity and encouragement for students to relate them to their lives in school and at home. The programme also included periodic information sessions for parents and caregivers on children's rights, gender equity, and positive discipline. Additionally, teachers received training on these topics.

Do the interventions work in improving child or youth offender maltreatment behaviour?

The results, based on self-reports of violence by children - were mixed. Right to Play's intervention resulted in less peer violence, both as perpetrator and victim, for those in the intervention than those in the control group. In the Skhokho intervention, there was a small effect on violence protection but it was not statistically significant. The US study also found no statistically significant effect on peer violence from the intervention.

Have the interventions been implemented at scale?

Not quite. While the Right to Play intervention covered many schools in one populous district in Pakistan, all the studies were designed as research studies that could be scaled up in the future.

Which type of organization delivered the interventions?

Trained facilitators and coaches, or researchers usually conduct intervention activities.

What do the interventions cost?

Detailed cost analyses are not available.

How are the programmes meant to work? The theory of change

The curricula in Taylor 2010 are based on the Theory of Reasoned Action (TRA). "TRA is based on research that demonstrates that intentions to behave are immediate predecessors to specific actions and proposes that attitudes towards and perceived norms about the desired behavior facilitate the intention to change, modify, or adopt a particular behavior" i.e., improved beliefs, attitudes (interaction-based curriculum), and knowledge (law and justice curriculum) related to gender violence and sexual harassment can reduce that behaviour.

A theory of change model was developed for Skhokho (Jewkes 2019) which was the conceptual basis for the programme. It includes drivers of gender-based violence such as patriarchal gender norms and a culture of acceptance of violence manifested through factors such as gender norms at home, at school, among peers, poor communication, harsh punishment at home and school, and witnessing intimate

partner violence at home. Finally, intervention targets for each of the factors are identified, such as students, teachers, or parents or caregivers.

How reliable is the evidence?

It's mixed. Each of the three RCTs has a different risk of bias rating.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Taylor 2010</i>	High risk of bias	Low risk	Some concerns	High risk	High risk	Some concerns
<i>Jewkes 2019</i>	Some concerns	Some concerns	Low risk	Low risk	Some concerns	Low risk
<i>Karmaliani 2020</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Prevention / Parent or Caregiver: Knowledge and Awareness

School-based interventions involving parents to tackle abuse can improve parent and caregiver knowledge and attitudes toward physical and sexual abuse.

Evidence status	Moderate risk of bias	School-based prevention interventions that involve parents can improve their knowledge regarding abuse and attitudes on violence against children.
------------------------	------------------------------	--

The summary in brief

School-based interventions to reduce child abuse and violence against children include training for teachers and parents. These interventions generally improve parent knowledge regarding abuse and attitudes toward the use of harsh discipline. The impact on parent or caregiver knowledge and attitudes varies, possibly on account of the varying intensity and duration of the intervention.

Contents of the cell

The cell has eight studies (Kolko 1987, Kolko 1989, MacIntyre 1991, Merrill 2018, McElearney 2021, Green 2020, Jewkes 2019, Francis 2021), and one protocol (Wangamati 2021). There are no systematic reviews.

A full summary of Merrill 2018 is available in the guidebook.

The interventions

All seven programmes covered in the eight studies are school-based, being delivered to preschoolers, primary and secondary school students, and their teachers and parents. They are from a diverse set of countries.

Programme	Country	Description
<i>Good School Toolkit</i>	Uganda	School-wide intervention implemented over 18 months. Schools are provided with booklets, posters, and facilitation guides for over 60 Toolkit activities. The activities, e.g., student discussions, debates, and booklet clubs, are mostly in a group setting. They address mutual respect, power relations, non-violent discipline techniques, classroom management strategies, and behaviour-change techniques, such as setting goals, making action plans, implementing rewards and reinforcement, and creating social support for change.
<i>Red Flag / Green Flag People</i> (2 studies)	USA	Staff and parent sessions are followed by two classroom training sessions of one and a half hours each. <i>The Green Flag / Red Flag</i> colouring book promotes behavioural strategies such as (1) how to say no to an adult, (2) how to get away from a perpetrator, and (3) how to tell someone about the experience of an actual abusive incident. A film entitled "Better Safe than Sorry II" is also presented. Parents were directly involved in the programme in one study (Kolko 1987) but not in the other (Kolko 1989).
<i>Stay Safe</i>	Republic of Ireland	Teacher and parent training sessions are followed by classroom implementation of 10-12 sessions of 30-40 minutes each. Teacher (two sessions) and parent training (one session) include sexual abuse prevention concepts such as definitions, myths, realities, prevalence, and information on typical victim and offender characteristics. It also

		<p>includes tips on how to identify victims and support them with disclosure and appropriate referrals to legal and social services. Community healthcare workers (doctors, nurses) also attended the parent training session.</p> <p>Pupil sessions cover five topics: feeling safe and unsafe; bullying; wanted and unwanted touches; telling adults about negative interactions with victimisers and bullies; and dealing with strangers.</p>
<i>Keeping Safe</i>	Northern Ireland, UK	<p>The 'whole-school' programme for children aged 4-11 on how to keep safe from any type of maltreatment. Teaching and learning resources are incorporated into regular curricula and the culture of the school. Classroom teaching covers three themes (healthy relationships; my body; and being safe) and 63 lesson plans for children as they progress from grade one to grade seven. School leaders and parents are also actively involved in the programme. Training and support are available for teachers and school staff.</p>
<i>Skhokho</i>	South Africa	<p>A holistic approach for 7th graders in multiple schools with a focus on preventing gender-based violence. Multiple interventions are part of the programme, including a life orientation curriculum for students, training for teachers, a workshop for parents/caregivers and teenagers, and the formation of clubs for children in year two of the programme to sustain momentum.</p>
<i>Early Head Start</i>	US	<p>Federally supported early childhood intervention for low-income households that includes day care services and home visits to provide resources to reduce parent/caregiver stress and improve the household environment.</p>
<i>Irie Homes Toolbox</i>	Jamaica	<p>A universal violence prevention programme targeted at parents of preschoolers. Training sessions are delivered via schools by preschool teachers. The aims of the programme are to move away from harsh punishment by parents and improve engagement with their children.</p>

Do the interventions work in improving parent or caregiver knowledge and attitudes?

Caregivers involved with the *Good School Toolkit* programme were surveyed two years after the programme, which ran for 18 months. They expressed reduced acceptability for the use of physical discipline at home (0.7 lower on a scale of 0-low to 12-high) and in school (0.8 lower on the same scale) compared to caregivers not in the programme, although the average difference was quite small. This most likely means a reasonable change in some parents but no change in others.

No impact was seen on their view of sexual abuse (sexual relationship between teacher and pupil), but that was probably because the acceptability was already low (approximately 0.5 on a scale of 0-low to 3-high). Caregivers were not surveyed before the programme, so we do not know whether or how their beliefs changed over time.

The evaluation of *Stay Safe* used a 38-item Parents' Knowledge and Attitudes Questionnaire. Significant improvements were recorded for eight of the eighteen items, reflecting belief in children's statements, more positive attitudes towards prevention programs, and knowledge about help-seeking. Parents' knowledge and attitude changes were not related to their age or gender.

Assessing the impact of the Red Flag/Green Flag People is challenging because Kolko's 1987 study did not measure parents' actual knowledge or awareness of child sexual abuse. Instead, it relied on self-reported perceptions of knowledge and awareness. Six months after intervention, parents' ratings on "seriousness of abuse, personal knowledge of abuse, confidence in identifying abuse and preparedness to deal with abuse" were similar across intervention and control groups. Intervention group parents' ratings on "information learned" from the programme (note that this is their feeling on how much they learned but not a test of their actual knowledge) were significantly higher than those not in the programme.

In addition "there were no differences in parents' ratings of their awareness of the problem of child sexual abuse over the previous six-month period, as all three groups indicated that their awareness had very much increased since then."

Finally, intervention group parents also reported significantly higher discussions on sexual abuse at home compared to controls. In summary, it is not clear from the study what impact this programme had on parent knowledge and attitudes.

Kolko 1989 (where the parents were not directly exposed to the programme), reported that parents whose children were in the programme reported higher scores for 'awareness', 'preparedness', 'utility' (of the programme) and 'change in understanding' compared to control group parents after training and at six-month follow-up. However, only 'change in understanding' improved after training compared to controls, whereas the other scores mostly stayed the same for the intervention group and decreased for the control group (which made the intervention group scores appear better).

In the Skhokho and Irie Homes Toolbox, parents reported better communication and engagement with their children and knowledge of positive parenting. Parents also reported less negative parenting, i.e., harsh punishments against children. Early Head Start parents tended to be more supportive of their children's development and made efforts to create a positive environment at home.

Have the interventions been implemented at scale?

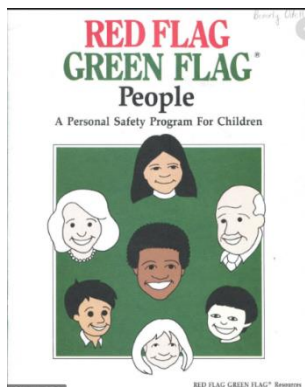
Personal safety programmes such as the *Red Flag /Green Flag People* colouring book are widely adopted, although they vary in duration and intensity. However, this study had only a small sample.

Early Head Start is a truly scaled-up programme, available across the US.

The other programmes are designed as research studies implemented in various schools and classrooms but not scaled up.

Which type of organisation delivered the intervention?

The Good School Toolkit was developed by Raising Voices, a Ugandan non-profit committed to preventing violence against women and children. They also seemed to be involved in the delivery and evaluation of



effects

The *Red Flag / Green Flag* programme was chosen for use by an organisation that developed community services for local school-age children “in light of its appropriate content and objectives, availability, inclusion of audio-visual materials and a workbook, limited cost, and short duration of presentation.”

The delivery organisation for *Stay Safe* is not mentioned, but the first author of the study is from the Eastern Health Board and Child Abuse Prevention Programme (Dublin, Ireland). The study mentions that *Stay Safe* is implemented in nearly all primary schools in the Republic of Ireland and “has the full support of the Department of Education, the Irish Government, and leaders of the major religious traditions in the country.”

The Irie Homes Toolbox is delivered by trained preschool teachers. Early Head Start employs child care workers and social workers to engage with families. Skhokho was facilitated by researchers involved in the design and implementation of the programme.

What do the interventions cost?

Cost data for the *Good School Toolkit* is reported in a study by Greco et al. Implementing it over 18 months in 21 schools was cost close to \$400,000. Monitoring and evaluation add another \$50,000 (for monitoring and evaluation for the whole programme). The annual cost to run the programme was approximately \$7,500 per school and \$15 per student. It cost close to \$250 to prevent a case of violence (total implementation costs including producing the toolkit materials) and approximately \$100 in annual implementation costs for every prevented case. *GST* was found to be cost-effective.

The *Red Flag / Green Flag* programme is noted in the study to be low cost, but actual cost data are not provided. Similarly, the authors claim that Irie Homes Toolbox and Skhokho can be delivered at low-cost, but actual costs are not provided.

How are the programmes meant to work? The theory of change

The logic of personal safety programmes is that children are active agents in child sexual abuse. If children are aware of what constitutes abuse (e.g., good touch versus bad touch), they are likely to recognise it, object to it, walk away from it, and report it, all of which reduce the likelihood of abuse occurring. Programmes teach children the skills for managing and reporting abusive situations. This process is supported by parent and teacher training so they can reinforce the approach and listen to children when they discuss these issues, including reporting abuse.

The *Good School Toolkit* has a six-step process based on the Transtheoretical Model of Behaviour Change. This model first makes people aware of the problem of physical and sexual abuse, and then supports the planning and implementation of behaviours to deal with the problem. The final stage is when new behaviours have become the norm.

Other programmes mention they are theory-informed but do not specify a particular theory they are based on. It is likely that programmes derive different elements from different theories to create a conceptual approach for a given intervention.

Will the results translate elsewhere?

Early childhood intervention and engagement with parents could potentially be implemented in other settings. Good touch/Bad touch-style prevention programmes are widely implemented in Western countries. Results from these programmes are likely transferrable globally (see the synthesis on *Prevention X Child Knowledge and Awareness*).

How reliable is the evidence?

Moderately so.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Merrill 2018	High risk of bias	Low risk	Some concerns	Some concerns	High risk of bias	Some concerns
Green 2020	Some concerns	Some concerns	Low risk	Low risk	Low risk	Low risk
McElearney 2021	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Low risk
Jewkes 2019	Some concerns	Some concerns	Low risk	Low risk	Some concerns	Low risk
Francis 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Kolko 1987	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk
Kolko 1989	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk

MacIntyre 1991	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk	Moderate risk	Low risk
-----------------------	------------------	----------	----------	----------	----------	----------	---------------	----------

What else is known from other studies about school-based programmes to prevent child sex abuse?

A Cochrane systematic review of 24 studies of school-based programmes that deliver information about child sex abuse and strategies to help children avoid it and encourage them to report abuse found that the programmes increase children’s knowledge and skills for dealing with abuse. But there is no impact on child mental health or sexual abuse. The review did not report on parent / caregiver outcomes.

Disclosure / Child Maltreatment Disclosure

Effects of disclosure-focused programmes to reduce child maltreatment are unclear.

Evidence status	Some concerns	A small number of studies suggest that disclosure-focused programmes can increase the likelihood of children self-disclosing potential abuse. More studies are needed.
------------------------	----------------------	--

The summary in brief

This cell has three primary studies and two systematic reviews – all in high-income countries. Two of the primary studies are school-based and focus on preparing students to identify dangerous situations and how to avoid them. One of these studies is specifically focussed on creating a safe environment for children to disclose potential abuse. The third study is inspired by Child Advocacy Centers (CACs) in the US and creates a multiagency team to work collaboratively in responding to alleged child abuse cases.

Two studies did report disclosure. The results are mixed. These interventions, especially the school-based one, can increase children’s willingness and confidence to disclose. However, the studies in this cell are still too few to provide really solid and reliable answers about the effectiveness of interventions to encourage disclosure.

Contents of the cell

The cell has three studies (*Herbert 2021*, *Hoefnagels 2021*, *White 2019*) and one systematic review (*Radford 2017*).

Full summaries of *Herbert 2021* and *Hoefnagels 2021* are available in the guidebook.

<i>White 2019</i> RCT, low risk of bias.	Australia (Gold Coast: eight government schools). Children in Year 1 (5-7 years). Evaluation of the effectiveness of <i>Learn to be safe with Emmy and friends</i> combined with an in-situ skills training programme to protect against potential child abuse situations
<i>Herbert 2021</i> QED, moderate risk of bias	Australia (city of Perth). Children of all ages. Evaluation of the <i>Multi-Agency Investigation & Support Team (MIST)</i> program.
<i>Hoefnagels 2021</i> QED, high risk of bias.	Netherlands (twelve primary schools) Children aged 10-12 years. Evaluation of a programme for primary school students to help them identify and talk about child abuse and neglect.
<i>Radford 2017</i> Rapid evidence assessment, low quality	UK - A rapid evidence synthesis commissioned by the Independent Inquiry into Child Sexual Abuse.

A. PRIMARY STUDIES

The interventions

The interventions from the three primary studies in the cell are summarized below.

White 2019 assessed a school-based child abuse prevention programme called ‘*Learn to be safe with Emmy and friends.*’ The programme, conducted in eight primary schools in the Gold Coast region of Australia, has five hour-long weekly sessions for year 1 students (age 5-7 years) delivered by trained facilitators in school premises. The facilitators are external and use ‘Emmy’, the programme mascot, to deliver key messages on how to recognize and respond to dangerous situations. This was combined with a simulated abduction scenario where a stranger asks a child to come with them.

MIST, or the *Multi-Agency Investigation & Support Team* (Herbert 2021), is a collaborative approach to responding to alleged cases of child abuse. Instead of having multiple responses individually from different agencies, *MIST* creates a team with representatives from different agencies located in the same physical location. Detectives, social workers, child family advocates, and therapists all work together on the same case.

The third programme was implemented in various primary schools in the Netherlands. Its goal is to create an environment where child abuse and neglect can be safely disclosed. The intervention includes four interactive sessions using video clips, TV content, and other educational material.

Do the interventions work in improving child maltreatment disclosure?

Unclear. The only RCT (*White 2019*) found increased disclosure intentions and confidence, but the study did not report actual disclosure rates. *MIST* reported lower disclosure rates during child interviews compared to usual practice, but the rate of substantiated offences was not different for the two groups. Teachers in the Dutch programme had more children disclose to them (five) than the control group (one). However, the researchers in this last case were not sure if this difference could be attributed to their programme.

Have the interventions been implemented at scale?

Both the school-based programmes were small-scale, with a small number of schools and classrooms. *MIST* received over a third of reported cases in Perth during the study period.

Which type of organisation delivered the interventions?

Trained facilitators delivered *Learn to be safe with Emmy and friends* in school buildings. In the Dutch study, regular classroom teachers, if interested, received training to deliver the curriculum in their classrooms. *MIST* included multiple public agencies involved in investigating child abuse cases.

What do the interventions cost?

None of the studies report costs.

How are the programmes meant to work? The theory of change

Two theories mentioned in the papers are social cognitive theory and the theory of planned behaviour. *MIST* is modelled on Child Advocacy Centers (CACs) in the US.

How reliable is the evidence?

There are some concerns. There are only a few studies. Two of the primary studies, including the one RCT, are rated as ‘low risk of bias,’ but there are serious concerns about the Dutch study.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
White 2019	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Herbert 2021	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Hoefnagels 2021	High risk of bias	Low risk	Moderate risk	Serious risk of bias	Low risk	Low risk	Low risk	Low risk

B. SYSTEMATIC REVIEW (Radford 2017)

What is the systematic review about (Radford 2017)?

Radford 2017 is a rapid review commissioned by the Independent Inquiry into Child Sexual Abuse in England and Wales (IICSA). It investigates whether public and non-state institutions have adequately protected children and young people from CSA and exploitation. The rapid review sought to learn how institutions (state and non-state) outside of England and Wales have prevented and responded to CSA and exploitation.

What are the findings on disclosure?

The review included 21 papers, both primary studies and systematic reviews, almost all on disclosure, identification, and reporting on child sexual abuse. All individual studies came from high-income countries, mostly the US. Different approaches at the institution and community-levels are described in the studies. Some of these are “*proactive outreach and engagement with minority communities, training those who work with children to be alert to the signs of sexual abuse and exploitation, co-located multi-disciplinary investigation and response models, protocols and best practice approaches for investigative interviewing, improved assessment methods and training for professionals (Radford 2017).*”

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

How reliable is the evidence?

Not very. The quality of the systematic reviews is rated as **low**. This means that there is at least one major flaw in how it was conducted.

Disclosure / Child Wellbeing: Child Knowledge and Awareness

There are too few studies to have a clear idea of the effect of school-based disclosure-focussed programmes on child knowledge and awareness regarding abuse.

Evidence status	Some concerns	Too few studies of disclosure-focussed programmes to know impact on child knowledge and awareness.
------------------------	----------------------	--

The summary in brief

This cell has two primary studies (both recent) and two systematic reviews, all in high-income countries.

Both primary studies are school-based and focus on preparing students to identify dangerous situations and how to avoid them. One of these studies is specifically focussed on creating a safe environment for children to disclose potential abuse.

Results show somewhat mixed results on increase in knowledge and awareness among children about how to recognize abuse and what they should do when they face a dangerous situation. These interventions can also increase children’s willingness and confidence to disclose. There are numerous prevention-focussed programmes conducted in schools that can increase knowledge and awareness, but only few studies focus on disclosure (in this cell).

Contents of the cell

The cell has two primary studies (*Hoefnagels 2021, White 2019*) and two systematic reviews (*Quadara 2015, Radford 2017*).

Full summaries of *Hoefnagels 2021* and *White 2019* are available in the guidebook.

<i>Hoefnagels 2021</i> QED, high risk of bias.	Netherlands (twelve primary schools) Children aged 10-12 years. Evaluation of a programme for primary school students to help them identify and talk about child abuse and neglect.
<i>White 2019</i> RCT, low risk of bias.	Australia (Gold Coast; eight government schools). Children in Year 1 (5-7 years). Evaluation of the effectiveness of ‘ <i>Learn to be safe with Emmy and friends</i> ’ combined with an in-situ skills training programme to protect against potential child abuse situations
<i>Quadara 2015</i> Scoping review, low quality.	Australia: A broad systematic review to identify the prevalence of child sex abuse, risk factors, and intervention approaches.
<i>Radford 2017</i> Rapid evidence assessment, low quality.	UK: A rapid evidence synthesis commissioned by the Independent Inquiry into Child Sexual Abuse.

A. PRIMARY STUDIES

The interventions

The interventions from the two primary studies in the cell are summarized below.

Hoefnagels 2021 evaluated a disclosure-focussed educational programme implemented in various primary schools in the Netherlands. The programme aimed to create an environment where child abuse and neglect can be safely talked about. The intervention includes an initial lesson followed by four interactive sessions in the classroom using video clips, TV content, and other educational materials. The introductory lesson is on moments of joy the children experienced with their parents or caregivers.

An informational flyer is sent to parents, a poster is placed in the classroom and teachers receive curriculum information. Teachers are encouraged to think about their own view on child abuse and neglect. Further, they are provided with guidance on where to seek assistance both within and outside the school and how to respond when children come to them with a maltreatment disclosure. Subsequent lessons are on neglect and child abuse.

White 2019 assessed a combination of two school-based interventions. One, a child abuse prevention programme called '*Learn to be safe with Emmy and friends.*' The programme, conducted in eight primary schools, in the Gold Coast region of Australia, has five hour-long weekly sessions for year 1 students (age 5-7 years) delivered by trained facilitators on school premises. The facilitators are external and used 'Emmy', the programme mascot, to deliver key messages on how to recognize and respond to potentially dangerous situations.

The second intervention is disclosure-focussed. The Observed Protection Behaviours Test (OPBT) in-person training is a single session to promote disclosure of abuse by a child. This is referred to as in-situ training (IST) in the published article. The IST interviewer is alone with a child and pretends to forget some documents and leaves the child alone in the interview room. A male stranger enters the room and urges the child to leave with him (this is meant to create a real-life risky situation). The child has ten seconds to respond. If they say 'no' or do not respond the man leaves but asks the child not to tell anyone about their interaction. If the child says 'yes', the man again asks the child to keep it secret and to remain in the room. The interviewer returns in a few minutes, and the child is prompted to disclose the encounter with the stranger every few minutes till they either talk about it or the interview ends.

Do the interventions work in improving children's knowledge and awareness?

The results are mixed. In the Dutch intervention (*Hoefnagels 2021*), children in the intervention improved at recognizing child abuse and neglect. However, in the Australian programme (*White 2019*), though disclosure confidence and intentions improved for children who did the programme, there was no improvement in their ability to identify unsafe situations, i.e., safety identification skills.

Children in the Dutch intervention (*Hoefnagels 2021*) were more than three times likely to recognize child abuse and neglect (from vignettes shown) than students who did not do the intervention - the likelihood was higher for physical and emotion abuse than neglect. Children of Dutch origin were twice as likely to spot child abuse and neglect than children of non-Dutch origin.

Children in the in-situ disclosure-focussed Australian programme, whether alone or in combined with the prevention-focussed education programme, showed statistically significantly higher increases in disclosure confidence and disclosure intentions than children in the control group (*White 2019*). The combination boosted effectiveness for disclosure intentions compared to the prevention programme alone but not for disclosure confidence or safety identification skills.

Have the interventions been implemented at scale?

No. Both programmes were school-based and small-scale, with a small number of schools and classrooms involved.

Which type of organisation delivered the interventions?

Trained facilitators delivered *Learn to be safe with Emmy and friends* in school buildings. In the Dutch study, regular classroom teachers, if interested, received training to deliver the curriculum in their classrooms.

What do the interventions cost?

None of the studies report costs.

How are the programmes meant to work? The theory of change

Two theories mentioned in the papers are social cognitive theory and the theory of planned behaviour.

How reliable is the evidence?

There are some concerns. The RCT has a low risk of bias, but the QED study has a high risk of bias.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>White 2019</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Hoefnagels 2021</i>	High risk of bias	Low risk	Moderate risk	Serious risk of bias	Low risk	Low risk	Low risk	Low risk

B. SYSTEMATIC REVIEWS

What are the systematic reviews about (Quadara 2015 and Radford 2017)?

Quadara 2015 was a broad scoping review to understand the prevalence of child sexual abuse in the Australian context, identify risk factors for abuse, and look at available interventions to address abuse.

Radford 2017 is a rapid review commissioned by the Independent Inquiry into Child Sexual Abuse in England and Wales (IICSA). It investigates whether public and non-state institutions have adequately protected children and young people from CSA and exploitation. The rapid review sought to learn how institutions (state and non-state) outside of England and Wales have prevented and responded to CSA and exploitation.

What are the findings on child knowledge and awareness?

There are not many disclosure-focussed interventions available. From the available studies, children's knowledge and awareness of recognizing abuse increase with school- or classroom-based programmes that are short-term conducted in one or more sessions delivered by teachers or trained facilitators.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

How reliable is the evidence?

Not very. The quality of both systematic reviews is rated as '**high risk of bias**'. This means that there is at least one major flaw in how they were conducted.

Response / Institutional Safeguarding Practice: Operations

The effect of response-focused interventions to improve institutional operations to safeguard children is mixed.

Evidence status	Some concerns	Unclear impact of response interventions on institutional operations to safeguard children.
------------------------	----------------------	---

The summary in brief

Institutions that care for children such as children’s homes, orphanages, schools, day cares, foster homes, and hospitals, can implement various safeguarding interventions and policies. These may include staff training or structural interventions such as improving the caregiver-to-child ratio.

These studies examine interventions that trained caregivers and professionals to improve their knowledge, attitudes, and practices/behaviours on working with children, identifying maltreatments, and responding better. The settings included children’s homes, schools, day care, hospitals, paediatric care units, and residential care. The training led to moderate improvements in the quality of practitioners’/caregivers’ response. The studies have medium to low confidence, results need to be interpreted with caution.

There are many recent and some even ongoing studies in the cell. The cell contains six completed primary studies (Rheingold 2015, Konijnendijk 2019, Hymel 2021, Cerezo 2004, Herbert 2021, and Johnson-Motoyama 2022), one systematic review (Hermenau 2017) and two study protocols (Perez 2021 and Taylor 2021).

Table 1: Contents of the cell

A. Completed Primary Studies	
<i>Cerezo 2004</i> QED, high risk of bias.	Spain (Balearic Islands). Professionals such as teachers, social workers, hospital staff, child protection services (CPS) staff and police. Evaluation of a large-scale training programme to improve detection of child maltreatment by professionals who worked with children.
<i>Rheingold 2015</i> RCT, moderate risk of bias	USA (three sites in different geographical regions – Atlanta, GA; Beaufort, SC; Bend, OR). Caregivers of children in day care, churches, schools. Evaluation of the Stewards of Children programme to prevent child sexual abuse.
Konijnendijk 2019 RCT, high risk of bias	The Netherlands. Child healthcare professionals. Assessment of the effects of a computerised support tool on child healthcare professionals' adherence to guidelines on child protection.
Hymel 2021 RCT, moderate risk of bias	USA. Paediatric intensive care unit and child abuse physicians. Evaluation of the impact of the PediBIRN (Pediatric Brain Injury Research Network) 4-variable clinical decision rule (CDR) on abuse evaluations and missed abusive head trauma.
Herbert 2021 QED, low risk of bias.	Australia (Perth). A multi-agency group of government and community support agencies providing combined support services with investigations, called the Multi-agency Investigation & Support Team (MIST).

	Evaluation of the fidelity with which this intervention was delivered, and whether it improved criminal justice, child protection, and service outcomes.
Johnson-Motoyama 2022 <i>QED, medium risk of bias.</i>	USA (data from the National Child Abuse and Neglect Data System) Evaluation of differential response (DR), a system policy that seeks to serve families of low-to moderate-risk for child maltreatment through family engagement, diversion from formal child protective services investigations, and service provision.
B. Study Protocols for Primary studies	
Perez 2021 <i>Study protocol for RCT</i>	USA (California) Testing implementation strategy for "ACEs Aware" policy that provides medicaid reimbursement for Adverse Childhood Experiences (ACEs) screening annually for child primary care visits to low-income families in Southern California.
Taylor 2021 <i>Study protocol for RCT</i>	USA. Implementation and evaluation of a web-based and consultative training for Victim Advocates to enhance early engagement in services (E3 training).
C. Systematic Review	
Hermenau 2017 <i>Systematic review, High Risk of Bias</i>	Assessment of impact of interventions in institutional care settings on children's development. The review assessed structural changes, caregiver training and enriched environments as interventions in institutional care settings. Nine studies included in systematic review for this EGM cell; studies from Tanzania, Chile, El Salvador, Turkey, Russia, and Romania.

A. PRIMARY STUDIES

There are eight primary studies in this cell (3 RCTs + 3 QEDs + 2 protocols)

Countries

Of eight primary studies, five are from the USA; the others are one each from The Netherlands, Australia, and Spain.

The interventions

Two of the QEDs (Herbert 2021 and Johnson-Motoyama 2022) assess interventions aimed to improve criminal justice, child protection, and service outcomes. Herbert 2021 studied the implementation of a Multi-agency Investigation & Support Team (MIST) program in Perth, Australia. MIST is similar to the prominent Children's Advocacy Centre approach where a group of government and community support agencies provide combined support services with investigations. Details of Herbert 2021 are given as individual summary separately.

Johnson-Motoyama 2022 studied the impact of an intervention called Differential Response (DR). DR policy serves families at low-to-moderate risk of child maltreatment, through family engagement, diversion from formal child protective services investigations, and service provision.

Two of the RCTs evaluated interventions in hospitals to see whether intervention can improve the practices of healthcare professionals in managing child maltreatment. Konijnendijk 2019 studied how a computerised support tool can help healthcare professions adhere to Child Abuse and Neglect (CAN) guidelines. The Dutch Centre for Child Health issued a clinical guideline on early detection and response to suspected CAN ("the CAN guideline") specifically for doctors and nurses in preventive child healthcare (CHC). Despite the potential benefits, the adherence to guidelines was poor. Konijnendijk 2019 developed a computerised guideline support tool that was integrated in the electronic health record used in Dutch preventive CHC, presenting CAN guidelines in a concise and easily accessible manner.

Hymel (2021) conducted another hospital-based intervention. This study evaluated the impact of the PediBIRN (Pediatric Brain Injury Research Network) 4-variable clinical decision rule (CDR) on abuse evaluations and missed abusive head trauma in pediatric intensive care settings.

Cerezo 2004 describes a large-scale intervention to improve child maltreatment detection in the Balearic Islands, Spain. The intervention consisted of two phases. Phase 1 focussed on training frontline professionals such as social workers, paediatricians, police officers, psychologists, psychiatrists, and nurses who are involved in reporting cases of child maltreatment to Child Protective Services (CPS). Phase 2 focussed on training teachers, psychologists, and support staff in preschools and primary schools. Rheingold 2014 assessed the effectiveness of Stewards of Children, a Child Sexual Abuse (CSA) prevention programme for childcare professionals. The programme was developed by the US non-profit Darkness to Light (D2L). It involved a 2.5-hour workshop to train adults in childcare settings on how to prevent, recognise and respond to CSA.

Both study protocols (Perez 2021 and Taylor 2021) propose to conduct RCTs in the USA.

Perez 2021 (US, California) proposes to develop a better implementation strategy to improve the awareness and uptake of "ACEs Aware" policy that provides Medicaid reimbursement for Adverse Childhood Experiences (ACEs) screening annually for child primary care visits to low-income families in Southern California. The study will test a multifaceted implementation strategy in partnership with a Federally Qualified Health Center (FQHC) system. The Exploration, Preparation, Implementation, Sustainment (EPIS) framework is a widely used implementation framework. This study plans to follow the EPIS framework for implementation mapping to refine implementation. The refined implementation strategy will include online training videos, a customized algorithm and use of technology to improve workflow efficiency, implementation training to internal FQHC personnel, clinic support and coaching, and written implementation protocols. This randomized trial with five primary care clinics will assess this implementation strategy for (a) fidelity to the ACE screening protocol, (b) reach, defined as the proportion of eligible children screened for ACEs, and (c) the impact of the ACE policy on child-level mental health referrals and symptom outcomes.

Taylor 2021 proposes a feasibility RCT to test the implementation of a web-based and consultative training for victim advocates at Children's Advocacy Centers. It aims to enhance mental health services provided to maltreated children by CACs, which serve as family navigators that connect children impacted by maltreatment to appropriate Evidence-Based Practices in the USA. The interactive web-based training is based on three key targets of knowledge and skills related to 1) family engagement, 2) trauma, and 3) Evidence-Based Practice services. The study proposes to assess the feasibility of the training program and the effectiveness and costs for web-based and consultative training.

Do these interventions work in improving institutional operating practices to safeguard children?

Herbert 2021 found that MIST was delivered with reasonable fidelity to its intended plan. Benefits of program were: high levels of caregiver satisfaction with the response; and high rates of children's engagement with therapy. However, workload for staff was a challenge. A quasi-experimental comparison between MIST (n = 126) and Practice as Usual (n = 276) found that MIST was significantly faster throughout the criminal justice and child protection processes, but the rate of arrest or child protection actions were not very different between MIST and usual practice. Despite these results, MIST still is promising, because of the uptake of therapeutic services and parental satisfaction.

Johnson-Motoyama 2022 found that US states with DR programs had approximately 19% fewer substantiated reports, 25% fewer children who were shown to suffer from neglect, and a 17% reduction in use of foster care services than states without DR programs. The study suggests additional research to better understand DR programs and their effects geographically.

Konijnendijk 2019 did not observe enough differences in guideline adherence between the intervention and control groups to draw conclusions. The results regarding use of the tool were inconclusive – though, in contrast to expectations, performance of guideline activities was high in both groups.

Hymel 2021 found that doctors in the intervention group evaluated more high-risk patients thoroughly (81% vs 73%) and missed fewer potential cases of abusive head trauma (21% vs 32%). We conclude that PediBIRN-4 CDR application enabled changes in evaluations of abuse, reducing potential cases of missed abusive head trauma in PICU settings.

Cerezo 2004 found that the intervention increased the number of cases reported to CPS from in both phases. The study reports a threefold increase in the number of cases detected after the intervention compared to before. The second phase (training teachers) led to detection of 2-3 children per 1000 as new cases after accounting for duplications from the first phase. The more professionals were trained, the more cases were detected.

The first phase (training frontline workers) was implemented in three territories. Referrals increased before and after the intervention in the first two territories but not in the third. This might be due to knowledge about the intervention spreading to the third territory before implementation - via mass media, professional networks and professionals moving territories for new jobs. A later comparison of the outcomes with various region of the Balearic Islands found expected increase in referrals.

Rheingold 2014 found that Stewards of Children training improved childcare professionals' knowledge and behaviours about CSA:

- **A: Knowledge** about CSA increased after the intervention but declined over the next three months. The control group also showed increased knowledge during this period, but not to the same extent as the intervention group.
- **B: Attitudes.** Participants' belief in CSA myths was low initially, leaving little room for improvement. After training, the control group had the better score but at three months there was no difference between groups.
- **C: Behaviours.** Participants who received the training reported improvement in their behaviors three months after training, as compared to the control group. The behaviours reported to have improved most were:
 - "Limiting the opportunity for older youth and younger youth to have one-to-one interaction". This is significant given that juveniles are offenders for over a third of CSA cases (Finkelhor et al. 2009). And
 - Participants in the intervention group reported an increase in behaviours such as "Sharing with another adult an article, brochure, or other information about CSA prevention" within

three months of the training. The control group also showed improvement in these behaviours during the same period, possibly due to the influence of the changed behavior of their trained colleagues or knowledge sharing.

In terms of the difference between being trained in-person vs online, the evaluation also found:

- Knowledge: The group trained in-person learned ‘significantly’ less about CSA (their knowledge had changed less) than the group trained on-line. Three months after training, however, there were no differences in CSA knowledge between the two groups.
- Attitudes: No difference between the group trained in-person vs. the group trained on-line.
- Behaviours: No difference between the group trained in-person vs. the group trained on-line.

The size of the impact of training in terms of implications for practice are unclear.

Have the interventions been implemented at scale?

Some yes, some no. The scale of intervention across studies varied from local to national:

- PediBIRN-4 CDR application was tried in 8 US PICUs.
- integrated CAN guidelines in a large Dutch organisation.
- MIST was run across Perth, western Australia.
- Differential Response was implemented across many states in the USA.
- Cerezo 2004 was implemented on a large scale in the Balearic Islands, Spain, with 161,287 children.
- Rheingold 2014 was implemented at three sites in different geographical regions in the USA – Atlanta, GA; Beaufort, SC; Bend, OR.

What do the interventions cost?

Cost data is not available in any of these studies.

How is the programme meant to work? The theory of change

There isn’t any specific theory for change given in these studies.

How reliable is the evidence?

Not very. Herbert 2021 is most reliable as it has only low risk of bias. Rheingold 2014, Hymel 2021, Johnson-Motoyama 2022 are having some concerns, so findings from these should be considered cautiously. However, we are less confident about two studies as they have high risk of bias (Konijnendijk 2019 and Cerezo 2014).

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Rheingold 2014	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns

Konijnendijk 2019	High Risk	Some concerns	Low risk	High Risk	High Risk	Some concerns
Hymel 2021	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Low risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

	Overall Risk of Bias	Bias due to Confounding	Selection bias	Bias due to Intervention classification	deviations from intended interventions	Bias due to missing data	Bias due to measurement of outcomes	selection of the reported result
Cerezo 2014	Serious	Serious	Moderate	Moderate	Serious	Low	Moderate	Serious
Herbert (2021)	Low	Low	Low	Low	Low	Low	Low	Low
Johnson - Motoyama 2022	Moderate	Low	Low	Moderate	Low	Moderate	Low	Low

B. SYSTEMATIC REVIEW (Hermenau 2017)

What is the systematic review about?

Hermenau 2017 looked at the impact of interventions in institutional environments on children’s development. The review assessed structural changes, caregiver training and enriched environments in institutional care settings. It looks particularly at orphans – in orphanages and foster homes.

What are the findings on institutional operations to safeguard children?

The review included 24 studies from 15 studies on 5 continents. Approximately 40% were from high-income countries with the others evenly distributed between upper-middle income and low or lower-middle income countries. The range of children included was wide – from less than 4 weeks to 16 years; close to two-thirds were infants or toddlers. Both state-run and private institutions were in the mix with the publication period of studies ranging over six decades. Finally, the review included a diversity of study designs that were classified into two groups – “dependent designs” i.e., studies with matched controls or repeated measures, and “independent designs” i.e., studies without matched controls or randomised controls (includes RCTs). Independent designs would be the more robust study design in terms of studying the effects of an intervention. Results for studies from both groups of designs were presented separately.

A total of nine studies reported on the intervention’s impact on either caregiving quality/institutional environment or on attachment (as an indicator for the bond between caregivers and children).

Four studies (from Tanzania, Chile, El Salvador) found a wide range of effect sizes which the systematic review authors describe as “very large” to “small”. Two studies reported on the quality of caregiving; one on whether children continued to experience physical maltreatment (it was an environment where violence as a form of discipline was common); one on the quality of attachment (results were not statistically significant). Three studies had caregiver training as the only intervention while the fourth study also had structural changes to improve institutional quality implemented (although the specific changes are not mentioned). All four were theory-based interventions.

Five studies (from Turkey, Russia, Romania) reported on various outcomes. Three studies examined had caregiver training and structural changes; one examined only structural changes and one only had caregiver training. Two studies did not report data that could be used for the systematic review's analysis although one study reported that care quality was better in the intervention group after intervention compared to controls (the type of data they reported was not usable for the review's chosen method of analysis). Two other studies of the same intervention found improvements in caregiving quality compared to controls, after the intervention had run for three years. One assessed an intervention that included both caregiver training and structural changes and found a larger, significant impact than the other that only included caregiver training (not significant). The fifth study improved the child-caregiver ratio (an example of a structural improvement) and found better attachment outcomes compared to the control group. Only three of the interventions were theory-based.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

Are results generalisable?

Probably. There were nine studies from various diverse countries. Almost all used caregiver training as the intervention with mostly positive results (although results of some were not statistically significant). Results are probably generalisable to institutions that care for children without families.

How reliable is the evidence?

Not very. The quality of the systematic review is rated as '**high risk of bias**'. This means that there is at least one major flaw in how it was conducted which reduces our confidence in the findings.

The systematic review adopted a method which combined RCTs and QEDs in the same statistical analysis which is not the convention. The study also depended on a separate meta-analysis paper for some of the calculations which could add errors to the estimates.

Response / Institutional Safeguarding Practice: Environment

The effectiveness of response-focused interventions to improve institutional environments to safeguard children is promising.

Cell: Intervention = Response; Outcome = Institutional Safeguarding Practice: Environment

Evidence status	Low risk of bias	Clear impact of response-focused interventions to improve institutional environments to safeguard children (based on three RCTs).
------------------------	-------------------------	---

The summary in brief

This synthesis includes five recent studies, with 4 primary studies and one systematic review. Out of four primary studies, three have been completed, while one is a protocol.

Institutions that care for children, such as children’s homes, orphanages, schools, daycares, foster homes, and hospitals, can implement various interventions and policies to safeguard the children under their care. Typically, interventions to improve operations to protect children in institutions include training for staff or structural interventions such as, improving caregiver to child ratio. These studies are based on interventions that trained caregivers and professionals to improve their knowledge, attitudes, and practices or behaviors when working with children, identifying instances of maltreatments, and enhancing their ability to respond effectively. The settings for the studies in this cell included schools and residential care. Results suggested improvements in the quality of practitioner’s/caregiver’s response (compared to controls), and the confidence in the findings of studies is assessed to be high for primary studies but low for the systematic review. Results should be interpreted accordingly.

The cell has three completed studies (Kim 2019; Nickerson 2019; and Santos 2022), one protocol (Perez 2021), and one systematic review (Lo and Cho 2021).

Studies in this cell

A. Completed Primary Studies	
Kim 2019 RCT, low risk of bias	USA, elementary schools School teachers Evaluation of the Second Step Child Protection Unit (CPU) to assess the impact on teacher knowledge, attitude, and relationships as well as any before and after the implementation of intervention.
Nickerson 2019 RCT, low risk of bias	USA, eight elementary schools from a large suburban school district in the Northeast School children Evaluation of the Second Step Child Protection Unit (CPU) to assess the impact on students’ CSA prevention knowledge, ability to recognize, report, and refuse unsafe touches, and perceptions of teacher-student relations. The study also investigated the moderating role of age and gender on program effectiveness.
Santos 2022	Portugal 11 Portuguese residential care homes

RCT, low risk of bias	Caregivers in residential care homes Evaluation of the effects of the CMT-Care Homes on self-reported compassion-related and emotional climate/social safeness variables, analyzing preliminary evidence on intervention.
B. Protocols for Primary studies	
Perez 2021 Study protocol for RCT	USA: California Testing implementation strategy for “ACEs Aware” policy that provides Medicaid reimbursement for Adverse Childhood Experiences (ACEs) screening annually for child primary care visits to low-income families in Southern California.
C. Systematic Review	
Lo and Cho 2021 Systematic Review high risk of bias	To review the evidence available on impacts of community-based interventions on reduction of child maltreatment and to identify the core components of the interventions.

A. PRIMARY STUDIES

Three complete primary studies are Kim 2019; Nickerson 2019; and Santos 2022.

The interventions

Two of the RCTs (Kim 2019 and Nickerson 2019) in this cell assess the same intervention, called the Second Step Child Protection Unit (CPU). CPU addresses child sexual abuse through (a) school policies and procedures, (b) staff training, (c) student lessons, and (d) family education. CPU makes all school staff such as teachers and office staff complete a 75–90-minute online module to prepare them to recognizing indicators of child sexual abuse , responding effectively, and report instances of abuse. After completing training, teachers and/or other staff provide six lessons to students, using multiple delivery strategies such as didactic instruction, songs, videos, case scenarios and role-plays. The lessons address the topics of safety, asking adults, unsafe and unwanted touches, rules about private body parts, and practice of rules to stay safe. Kim 2019 evaluated CPU to assess its effect on teachers’ knowledge, attitude, and relationships, before and after the intervention.

Same intervention was assessed by Nickerson 2019 to see how the intervention was helping in students’ knowledge about child sexual abuse prevention, their ability to recognize, report, and refuse unsafe touches, and how students perceive teacher-student relations. Nickerson 2019 also investigated if age and gender has to do anything with the program effectiveness.

Santos 2022 is based on Compassionate Mind Training program for caregivers (CMT-Care Homes). The CMT-Care Homes program was designed to cultivate a compassionate caregiving mindset among care

givers, both in self-to-self relationship and in their interaction with others. The programme aims to create a safe and secure environment in residential care facilities (RCF). Based on a 12-session structured and manualized program, CMT-Care Homes is delivered in a group format. The CMT-Care Homes is organized across 3 sequential modules: 1) Our mind according to a compassion-based approach (to provide insight into the evolved and socially shaped mind and the affect regulation systems); 2) Compassionate mind training (understanding and cultivating the attributes and competencies of compassion in its three flows, and addressing its fears); and 3) Final session (revising key information/practices, and its application into the RCH settings). This study is actually part of a larger study, to examine the efficacy of the CMT-Care Homes program. This study aimed to investigate the effects of the CMT-Care Homes on self-reported compassion-related and emotional climate/social safety variables, analyzing preliminary evidence.

Do these interventions work in improving institutional safeguarding practices?

Kim 2019 found that the CPU teacher training was effective in improving teacher's knowledge, attitude, and student-relationships. The study also assessed if teachers accept taking the training, will it affect their knowledge. and they found that teachers with higher levels of acceptability of the teacher training had better outcomes in terms of their knowledge, attitudes, and student-relationships.

Nickerson 2019 found that students who got the programme scored significantly higher afterwards than did students who did not get it.

younger children gained most from the program. On gender, girls had better knowledge than boys about sexual abuse, including the ability to recognize, refuse, and report unsafe touches, but both boys and girls made significant gains after the training.

Santos 2022 found that at the end of CMT-Care Homes delivery, professionals who received the intervention had lesser fears of receiving compassion from others and fears of self-compassion as compared to the professionals who did not receive the intervention.

Which type of organisation delivered the intervention?

The Second Step CPU is a program by an organization called The Committee for Children (2014) (CPU; www.cfchildren.org/child-protection). CPU teacher training is delivered as a 75–90-min online module, and after completing the CPU online training modules, these teachers delivered the CPU student curriculum in their classrooms.

The CMT-Care Homes program was delivered in residential care homes by the study's lead author (Laura Santos), who is a clinical psychologist trained in cognitive-behavioral interventions and compassionate approaches with experience in residential youth care. The CMT-Care Homes program was delivered in accordance with the handbook, in face-to-face weekly sessions of 2.5 hours in each RCH, to a group of 6–10 participants, over 3–4 months, from October 2019 to February 2020.

What do the interventions cost?

None of the studies report any cost data.

How is the programme meant to work? The theory of change

Not specifically a theory of change, but Kim 2019 had hypotheses to guide their study. They hypothesized that: (1) Participation in the CPU will improve teacher outcomes in the intervention group compared to the control group. (2) The CPU will interact with teachers' prior knowledge, attitudes, and relationships to

improve their outcomes. (3) A higher level of acceptability for the training will help teachers gain more from the training.

Nickerson 2019 had the Second Step CPU logic model (see Fig. 1) that suggests the student lesson and reinforcement activity components of the intervention should increase student knowledge of personal safety rules and the ability to recognize, report, and refuse unsafe situations and touches. The authors hypothesized that the CPU lessons would improve students' knowledge about abuse prevention, as well as ability to recognize, report, and refuse unsafe and sexually abusive situations. The logic model also emphasizes intervention components that foster positive relationships with students through safe and supportive environment strategies; the authors hypothesized that the programme would improve teacher-student relations.

There was no study specific theory of change but the intervention in Santos 2022 has some adaptations from the Social Mentalities Theory (Gilbert, 2015b, 2019). That suggests the mentality of caregiving or affiliation stems from the mammalian ability to form attachments, involving inclination to provide care, exhibit empathy, and act altruistically.

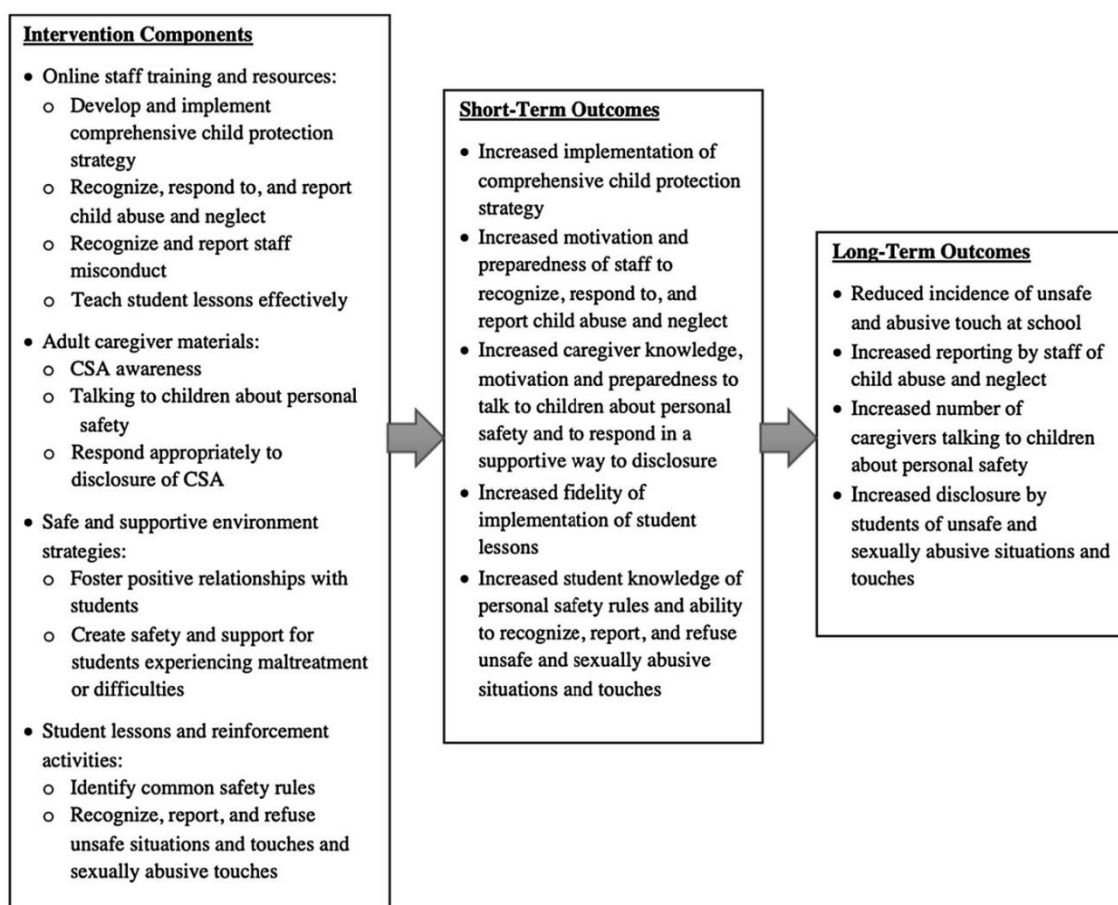


Fig. 1. Logic Model for the *Second Step* Child Protection Unit Curriculum.

Are the results generalisable?

Kim 2019 and Nickerson 2019 were conducted in a suburban area, and the teachers were mostly white, so there is a possibility that findings may not be generalizable to urban or rural settings or to individuals from other racial and ethnic backgrounds. We can't say much about Santos 2022, but compassion as such is a generic attribute that one can assume has a positive impact.

How reliable is the evidence?

Quite a lot.

All three RCTs are rated to have a low risk of bias so we can be confident about the findings.

PROTOCOL

Perez 2021 (US, California) proposes to develop a better implementation strategy to improve the awareness and uptake of the "ACEs Aware" policy that provides Medicaid reimbursement for Adverse Childhood Experiences (ACEs) screening annually for child primary care visits to low-income families in Southern California. The study will test a multifaceted implementation strategy in partnership with a Federally Qualified Health Center (FQHC) system. The Exploration, Preparation, Implementation, and Sustainment (EPIS) framework is a widely used implementation framework². This study plans to follow the EPIS framework for implementation mapping to refine implementation. The refined implementation strategy will include online training videos, a customized algorithm and use of technology to improve workflow efficiency, implementation training for internal FQHC personnel, clinic support and coaching, and written implementation protocols. This randomized trial with five primary care clinics will assess this implementation strategy for (a) fidelity to the ACE screening protocol, (b) reach, defined as the proportion of eligible children screened for ACEs, and (c) the impact of the ACE policy on child-level mental health referrals and symptom outcomes.

B. SYSTEMATIC REVIEW

The cell has one systematic review (Lo and Cho 2021). Details of the review are given below:

The intervention: community-based interventions.

Data Sources: Medline, PsycInfo, and Web of Science.

Inclusion Criteria:

- those published in English before January 2020,
- focused on the prevention of child maltreatment,
- placed emphasis on modifying the community environments and processes,
- evaluated the actual change in child maltreatment at the community level as one of the outcomes,
- primary studies reporting original data.

Studies Included: four studies were included in this review.

Results: All four studies included were from the USA. The four community-based child maltreatment programs included were Strong Communities for Children, the Durham Family Initiative (DFI), the Enough Abuse Campaign, and Prevent Child Abuse Georgia. There were two interventions aimed at preventing physical child abuse and neglect in children between the ages of 0 and 6, and two interventions focused specifically on addressing child sexual abuse (CSA). These interventions incorporated various components of prevention strategies. Of the four interventions, three focused on community-level interventions that involved modifying community environments and processes, while one intervention simultaneously targeted multiple ecological levels.

How reliable are the findings?

The review is rated to have an overall **high risk of bias**.

Response / Adult institutional caregiver: behaviour, knowledge, attitudes

Training for childcare professionals can improve knowledge, attitudes, and perceptions of readiness to respond to child maltreatment.

Evidence status	Some concerns	Moderate evidence of impact on teacher, childcare provider, healthcare personnel knowledge, attitudes, and perceptions of readiness to respond to potential child maltreatment.
------------------------	----------------------	---

The summary in brief

Training interventions – both online and in-person – for adult professionals who work closely with children can be successful in improving knowledge, attitudes, and readiness to respond to signs of child abuse. The cell has nine primary studies, one protocol for a primary study, and one systematic review. Professionals in the studies include teachers, early childhood educators, physicians, nurses, psychotherapists, residential care staff, victim advocates and counsellors. The components of the training include multiple modules spaced out over a few weeks or months that provide information on child maltreatment, the signs to look for, and how professionals can respond. One study for paediatric intensive care unit (PICU) doctors in the US combined training with a clinical decision-making tool to better identify abuse-related head trauma in children. Other settings for training included day care centres, churches, residential care for youth, and primary school classrooms.

Most primary studies in the cell are recently published. Five studies are from the US, two from Turkey, and one each from Germany, Portugal, and Iran.

The impact on childcare professional knowledge, attitudes, and behaviours suggests that improvements through training are possible, but the long-term effects on reducing child maltreatment are unclear because few studies measure them. Most studies were small, short-term, and run by researchers, which limits generalizability to other settings and contexts.

Contents of the cell

The cell includes eight studies (*Hymel 2021, Humphreys 2021, Kim 2019, Konig 2020, Martin 2020, Peker 2020, Rheingold 2014, Santos 2022, Turan 2022*), one protocol (*Taylor 2021*) and one systematic review (*Lo 2021*).

<i>Humphreys 2021</i> RCT, moderate risk of bias	US (one state). Evaluation of <i>iLookOut</i> , an online child abuse identification and reporting training for early childhood professionals.
<i>Hymel 2021</i> RCT, moderate risk of bias	US (eight paediatric intensive care units across the country). Evaluation of a clinical decision rule and training for physicians to reduce missed abusive head trauma in PICUs.
<i>Kim 2019</i> RCT, low risk of bias.	US (eight schools in one state). 161 teachers in study. Evaluation of <i>Second Step Child Protection Unit</i> , a multipronged approach to child sexual abuse prevention and response.
<i>Konig 2020</i> RCT, moderate risk of bias	Germany (nationwide recruitment). 80-262 participants in the training arm for the three courses. Evaluation of E-learning training for healthcare professionals (physicians, psychotherapists, nurses) called “Child protection in medicine, Safeguarding

	standards in medicine, and Basic knowledge of child protection in institutions.” The last course was for managers.
<i>Rheingold 2014</i> RCT, moderate risk of bias	US (three sites in three states – Atlanta, GA; Beaufort, SC; Bend, OR). Caregivers of children in day care, churches, schools. Evaluation of the Stewards of Children programme to prevent child sexual abuse.
<i>Santos 2022</i> RCT, low risk of bias	Portugal (11 residential care homes for at-risk adolescents). Residential youth care (RYC) staff. Evaluation of a compassionate mind training program for caregivers of RYC (CMT-Care Homes).
<i>Taylor 2021</i> RCT protocol, results forthcoming	US. 81 victim advocates. Evaluation of E3, a webinar- based training for victim advocates to improve engagement with families in mental health services and referral of children to Children’s Advocacy Centers (CACs).
<i>Martin 2020</i> QED, moderate risk of bias	Iran. 80 preschool teachers. Evaluation of a teaching intervention for preschool teachers on sex education for preschoolers including recognizing signs of child sexual abuse.
<i>Peker 2020</i> QED, high risk of bias	Turkey. 16 counselors. Evaluation of a psychoeducation program to improve reporting of child sexual abuse by counselor teachers.
<i>Turan 2022</i> QED, low risk of bias	Turkey. 62 student nurses. Evaluation of an online educational course a nursing students’ attitudes to child abuse disclosure.
<i>Lo 2021</i> Systematic Review, high risk of bias	Systematic review on community-based interventions to reduce child maltreatment.

The interventions

The interventions involve training for various professionals who work with children including healthcare providers (physicians, nurses, psychotherapists), early childhood educators, residential care providers, primary school teachers, school administrators, victim advocates, and counselors.

The training includes online or in-person modules for professionals increase their knowledge on the signs of child maltreatment and provides them with tools and skills to respond to potential cases in their practice and classrooms. For example, PICU doctors in multiple US hospitals were trained to use a tool to assess whether head trauma in the children they treated could be from abuse (*Hymel 2021*). In other examples, early childhood educators who work in day cares, churches, schools were trained to identify and respond to possible cases of child abuse. One study in residential care settings in Portugal, provided caregivers with a compassion-based programme. An ongoing study (*Taylor 2021*) is offering webinar-based training for victim advocates to improve rates of screening, referral, and access to evidence-based practices for child victims. In two instances, one, the *Second Step Child Protection Unit* (*Kim 2019*) in primary schools and second, an e-learning programme in healthcare settings in Germany, managers and administrators are also trained on institutional standards and responses against child maltreatment.

The RCT evidence mostly comes from high income countries primarily the US, with one study each from Germany and Portugal. The QED evidence is from Iran and Turkey, both middle-income economies.

The cell also has one systematic review (*Lo 2021*) that examines community-based interventions to reduce child maltreatment and identifies four key elements to be effective: involving community members, establishing community partnerships with institutions, promoting multidisciplinary collaborations, and responsiveness to community needs.

Do the interventions work in improving adult institutional caregiver outcomes?

In general, the trainings increase knowledge and attitudes of institutional caregivers of children. Participants reported feeling better prepared to respond to signs of child maltreatment. However, whether these results translate into increased disclosure by students and reduced child maltreatment was not reported in these studies.

In the PICU study, high risk cases were evaluated more thoroughly, and the number of potential cases of missed abusive head trauma fell. However, the number of estimated cases of missed abusive head trauma did not differ statistically significantly from the control group.

Have the interventions been implemented at scale?

Does not seem like it. Most studies were small, and run by researchers assessing whether a programme works or not.

Which type of organisation delivered the intervention?

In most cases, the interventions were delivered by the researchers who developed it and are testing it out.

Stewards of Children (Rheingold 2014) was developed and delivered by a US NGO, Darkness to Light.

What do the interventions cost?

None of the studies report costs.

How are the programmes meant to work? The theory of change

Studies did not explicitly mention a theory on which they were based except the compassion-based training (*Santos 2022*) which was based on the *Social Mentalities* theory.

How reliable is the evidence?

Moderately so. Most studies are of ‘moderate quality’ indicating some concerns.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Humphreys 2021</i>	Some concerns	Some concerns	Low risk	Low risk	Low risk	Low risk

Hymel 2021	Some concerns	Some concerns	Low risk	Low risk	Low risk	Low risk
Kim 2019	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk
Konig 2020	Some concerns	Some concerns	Low risk	Low risk	Some concerns	Low risk
Rheingold 2014	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns
Santos 2022	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Martin 2020	Some concerns	Low risk	Low risk	Low risk	Low risk	Low risk	Moderate risk	Low risk
Peker 2020	High risk of bias	Low risk	Serious risk	Low risk	Low risk	Low risk	Moderate risk	Low risk
Turan 2022	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk

Response / Disclosure: Disclosure rates

Effectiveness on reforming response-approaches to potential child maltreatment unclear.

Evidence status	Some concerns	Mixed results suggest reforming the response to alleged child abuse can encourage disclosure that is substantiated.
------------------------	----------------------	---

The summary in brief

This cell has two primary studies and one review. Both primary studies - one from Australia and the other from the US – are about reorganizing the response to alleged cases of child abuse. One is called *Differential Response (DR)*, a policy approach which engages families considered low- to moderate-risk for child abuse and guides them away from formal investigations. DR is followed in more than 20 US states. The other study is on *MIST*, based on the US Child Advocacy Centers (CACs) model. Multiple agencies come together to form a team and respond to a specific case in a coordinated and collaborative manner.

The results on actual disclosure are mixed. More evaluations of these programmes are needed.

Contents of the cell

The cell has two studies (*Herbert 2021, Johnson-Motoyama 2022*) and one systematic review (*Radford 2017*)

A full summary of *Herbert 2021* is available in the guidebook.

<i>Herbert 2021</i> QED, moderate risk of bias	Australia (city of Perth). Children of all ages. Evaluation of the <i>Multi-Agency Investigation & Support Team (MIST)</i> programme.
<i>Johnson-Motoyama 2022</i> QED, moderate risk of bias	US. Children of all ages. Evaluation of <i>Differential Response (DR)</i> , a policy for families with low- to moderate-risk for child treatment to receive family engagement, services, and diversion for child protection investigations.
Radford 2017 Rapid evidence assessment, low quality	A rapid evidence synthesis commissioned by the <i>Independent Inquiry into Child Sexual Abuse (UK)</i> .

A. Primary Studies (Herbert 2021, Johnson-Motoyama 2022)

The interventions

MIST or Multi-Agency Investigation & Support Team (Herbert 2021) is a collaborative response to child abuse allegations. By bringing together representatives from different agencies in one location, MIST reduces the number of individual responses from different agencies. A case involves detectives, social workers, child family advocates, and therapists working together.

Johnson-Motoyama (2022) assesses a policy approach called *Differential Response (DR)*. This is an alternative approach to child protection that focuses on providing a range of responses based on the needs and risks of each individual case, as opposed to a one-size-fits-all approach. The goal of this program is to provide early intervention and support to low- to moderate-risk families to prevent the occurrence or

recurrence of child maltreatment. Families receive engagement, access to various services, and diversion from child protection investigations.

Do the interventions work in improving child maltreatment disclosure?

Unclear. *MIST* reported lower disclosure rates during child interviews compared to usual practice, but the rate of substantiated offences was not different for the two groups. However, the US states with DR programmes had 19% fewer substantiated reports for any abuse and 25% fewer for neglect. Additionally, foster care utilization in DR states was lower (by 16%).

Have the interventions been implemented at scale?

Yes. *MIST* received over a third of reported cases in Perth during the study period. More than 20 US states used DR during 2004-2017.

Which type of organisation delivered the intervention?

MIST included multiple public agencies involved in investigating child abuse cases. DR is a different approach or system to respond to allegations of child abuse. There is no deliverer per se.

What do the interventions cost?

Neither study reports costs.

How are the programmes meant to work? The theory of change

No specific theories are mentioned. *MIST* is modelled on Child Advocacy Centers (CACs) from the US.

Are the results generalisable?

Likely. The DR programme results are likely generalizable to high-income county settings with access to child protection and family engagement services. *MIST* is based on the US model and could be replicated in other high-income countries.

How reliable is the evidence?

Not very. There are only two studies, both QEDs with one rated as ‘low risk of bias’ and the other as ‘moderate risk of bias/some concerns.’

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selecti on bias	Bias in interventi on classificati on	Deviation from intended interventi on	Missing outcom e data	Measurem ent of the outcome	Selecti on of the report ed result
<i>Herbert 2021</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk

Johnson-Motoma 2022	Some concerns	Low risk	Low risk	Moderate risk	Low risk	Moderate risk	Low risk	Low risk
----------------------------	----------------------	----------	----------	---------------	----------	---------------	----------	----------

B. Systematic Review (Radford 2017)

What is the systematic review about?

Radford 2017 is a rapid review commissioned by the Independent Inquiry into Child Sexual Abuse in England and Wales (IICSA). It investigates whether public and non-state institutions have adequately protected children and young people from CSA and exploitation. The rapid review sought to learn how institutions (state and non-state) outside of England and Wales have prevented and responded to CSA and exploitation.

What are the findings on disclosure?

The review included 21 papers, both primary studies and systematic reviews, almost all on disclosure, identification, and reporting on child sexual abuse. All individual studies came from high-income countries, mostly the US. Different approaches at the institution and community-levels are described in the studies. Some of these are “*proactive outreach and engagement with minority communities, training those who work with children to be alert to the signs of sexual abuse and exploitation, co-located multi-disciplinary investigation and response models, protocols and best practice approaches for investigative interviewing, improved assessment methods and training for professionals (Radford 2017).*”

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

Are results generalisable?

Probably to high-income country settings, but some of the intervention approaches, such as training for professionals and those who work with children, should be transferable to low- and middle-income countries.

How reliable is the evidence?

Not very. The quality of the systematic review is rated as **low**. This means that there is at least one major flaw in how it was conducted, which reduces our confidence in the findings.

Response / Child maltreatment occurrence/recurrence

Evidence on response-focused interventions reporting to child maltreatment is scarce.

Evidence status	Some concerns	Insufficient evidence that response-focused interventions can reduce child maltreatment.
------------------------	----------------------	--

The summary in brief

The evidence on institutional response interventions is sparse, especially from low- and middle-income countries. The available evidence shows mixed findings without clear patterns about how to reduce child maltreatment.

Contents of the cell

All studies in this cell are recent :-)

The cell has two primary studies (*Herbert 2021, Hymel 2021*) and three systematic reviews (*Hermenau 2017, Lo 2021, Sherr 2017*).

Herbert 2021 has an individual summary in the guidebook.

A. Primary Studies (Herbert 2021, Hymel 2021)

The interventions

MIST, or the Multi-Agency Investigation & Support Team (Herbert 2021: a QED), is a collaborative approach to investigating allegations of child abuse in an Australian city. MIST creates a team with representatives from different agencies located at the same location, rather than having individual responses from each agency. A case team involves detectives, social workers, child family advocates, and therapists.

The other study evaluated a training programme combined with a clinical decision tool for paediatric intensive care unit (PICU) doctors in multiple US hospitals. The goal was to improve assessments of whether head trauma observed in children could result from abuse (*Hymel 2021: an RCT*). Training comprised an initial 15-minute online session, a booster session, access to a clinical head abuse probability calculator, and additional information sessions. Physicians were also provided progress reports every six months.

Do the interventions work in reducing child maltreatment occurrence or recurrence?

Whether these interventions reduce child maltreatment is unclear.

In the PICU study, high risk cases were evaluated more thoroughly (than usual clinical practice), and the possibility of missed abusive head trauma decreased. However, there was no statistical difference between the estimated number of missed abusive head traumas (by PICU doctors) in the intervention and control group.

During child interviews, MIST reported lower disclosure rates than usual, but the rate of substantiated crimes was not different than the usual practice/control group.

Have the interventions been implemented at scale?

MIST investigated a third of the cases of alleged child abuse in a large Australian city.

The PICU study involved eight hospital sites across the US.

Which type of organisation delivered the intervention?

In MIST, there is no specific deliverer.

In the PICU study, training was delivered online with additional check-ins and information sessions by the research team.

What do the interventions cost?

Neither study reported costs.

How are the programmes meant to work? The theory of change

Studies did not explicitly mention a theory on which they were based. MIST simulates the Child Advocacy Centers (CACs) in the US.

How reliable is the evidence?

Moderately. There are some concerns with the RCT (the MIST study), but the QED study (the PICU study) is rated as low risk of bias.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Hymel 2021 (MIST: Australia)</i>	Some concerns	Some concerns	Low risk	Low risk	Low risk	Low risk

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Herbert 2021 (PICU: US)</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk

B. Systematic Reviews (Hermenau 2017, Lo 2021, Sherr 2017)

What are the systematic reviews about?

Hermenau 2017 assessed structural changes, caregiver training, and enriched environments in institutional care settings and their subsequent impact on child development. It looked particularly at orphans – in orphanages and foster homes.

Lo 2021 is a wide-spanning systematic review assessing community-based approaches to prevent child maltreatment.

Sherr 2017 is a series of systematic reviews that looked at multiple aspects of child violence in institutional care, including interventions to reduce abuse in these settings.

What are the findings on child maltreatment occurrence and recurrence?

In general, the evidence on interventions in institutions or even in the community in response to child maltreatment is sparse. Simple training or more complex structures to place children in conducive alternative environments (or to avoid institutionalised placements in the first place) seem to be the main approaches to intervention to address abuse.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

How reliable is the evidence?

Not very. The risk of bias in all three systematic reviews is rated as high. This means that there is at least one major flaw in how they were conducted, which reduces our confidence in the findings.

Response / Child Wellbeing: Mental Health

Impact of response-focused interventions to improve mental health is unclear.

Evidence status	High risk of bias	Unclear impact of response interventions on children's mental health
------------------------	--------------------------	--

The summary in brief

Child maltreatment can lead to adverse mental health outcomes for children. Even participating in an intervention (such as a sexual abuse prevention education intervention in school) might frighten children and make them more anxious.

Three studies included in two of the systematic reviews in this cell are from institutional care settings. Two focused on training caregivers of children living in orphanages in Tanzania. Effects on mental health outcomes such as depressive symptoms, internalising ("being withdrawn") and externalising ("acting out") behaviours were mixed. In a study from Portugal, more than half of children in institutional care reported a suicide attempt.

Hence no firm conclusions can yet be reached about the effect(s) of response interventions in institutional care settings on mental health.

The cell has three systematic reviews (Hermenau 2017, Sherr 2017, Radford 2017) and one protocol for a new RCT (Taylor 2021).

Contents of the cell

A. Systematic Reviews (Hermenau 2017, Radford 2017, Sherr 2017)

Systematic Reviews	
Hermenau 2017 Systematic Review, low rating	Assessment of impact of interventions in institutional care settings on children's development. The review assessed structural changes, caregiver training and enriched environments as interventions in institutional care settings. Two studies included in systematic review: both from Tanzania
Sherr 2017 Systematic Review, low rating	Series of systematic reviews on child maltreatment in institutional care examining prevalence of abuse and peer violence in institutions, interventions to reduce abuse, and measures of children's cognitive and social development. One study from Portugal included in systematic review.
Radford 2017 Systematic Review, low rating	Rapid review commissioned by independent panel investigating institutional failures in England and Wales to protect children from CSA and exploitation. The rapid review aimed to find effective interventions that institutions in countries outside England and Wales have implemented.

What are the systematic reviews about?

Hermenau 2017 and Sherr 2017 are peer-reviewed publications. Radford 2017 is a commissioned report, so it was probably not peer reviewed.

Hermenau 2017 looked at the impact of interventions in institutional environments on children's development. The review assessed structural changes, caregiver training and enriched environments in institutional care settings.

Sherr 2017 was a series of systematic reviews that looked at multiple aspects of child violence in institutional care: (i) the prevalence of maltreatment of children in institutional care (ii) interventions to reduce abuse in these settings (iii) peer violence in institutions and (iv) on the cognitive and social development of children in institutional care.

Radford 2017 was a rapid review commissioned by the Independent Inquiry into Child Sexual Abuse in England and Wales (IICSA) which is investigating if public and non-state institutions have done enough to protect children and young people from CSA and exploitation. The focus of the rapid review was to learn how institutions (state and non-state) outside of England and Wales have prevented and responded to CSA and exploitation.

What are the findings on children's mental health outcomes?

Three studies included in Hermenau 2017 and Sherr 2017 (two studies were in both reviews and a third study was included in Sherr 2017) reported mental health outcomes (two studies were included in both reviews). Two of the three studies from Tanzania were by the same author group. Caregivers in various orphanages attended a two-week training workshop aimed at improving their practices and the quality of their relationship with children. Neither study had a control group. Outcomes were compared before and after the workshop. Results on mental health outcomes were mixed. One study reported a moderate reduction in PTSD symptoms but no effect on depressive symptoms, internalising ("being withdrawn") and externalising ("acting out") behaviours six months after the workshop. The second study found a large drop in depressive symptoms, a moderate drop in internalising and externalising behaviours and a large drop in aggressive behaviours at three months. The third study from Portugal compared outcomes for children who lived at home vs in institutions after a Child Protective Services (CPS) investigation (they also compared both groups to a third group which did not receive any intervention). No differences were found on overall risk behaviours, but individual risk behaviours varied between groups. Soberingly, more than half the children in institutional care had attempted suicide compared to about a third in those that continued to live at home.

Radford 2017 provided results on mental health outcomes for treatments such as cognitive behavioural therapy in children who had been abused or exploited. The review did not report on mental health outcomes for response interventions.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness in any of the studies.

How reliable is the evidence?

Not very. The quality of each of the three systematic reviews is rated as **high risk of bias**. This means that there is at least one major flaw in how they were conducted which reduces our confidence in the findings.

Hermenau 2017 adopted a method which combined RCTs and QEDs in the same statistical analysis: this is unusual because it can be problematic. They also depended on a different meta-analysis for some of their calculations which could add errors to their estimates. Sherr 2017 did not provide details on the quality of the primary studies included in their reviews. Radford 2017 conducted a quality assessment of their included primary studies, but it is unclear why they did not provide any information on the size of impact.

B. RCT Protocol (Taylor 2021)

This is a protocol for a new RCT called "Enhancing Early Engagement (E3) in Mental Health Services Training." The trial aims to improve the training provided to victim advocates at Children's Advocacy Centers (CACs) in the US.

Children's Advocacy Centers are organizations that support children who have experienced abuse or trauma. Victim advocates play a crucial role in providing emotional support and helping these children navigate the legal and mental health systems. The E3 trial seeks to enhance the training of victim

advocates by incorporating evidence-based practices related to mental health services. The researchers plan an RCT, meaning that participants, i.e., victim advocates, will be randomly assigned to either a control group or an intervention group.

The control group will receive the existing standard training provided by the CAC, while the intervention group will receive an enhanced training program that incorporates additional strategies and techniques for engaging with children and families in need of mental health services.

The feasibility of implementing the E3 trial will be assessed through several measures, such as recruitment and retention of participants, adherence to the intervention protocol, and data collection processes.

The goal of the E3 trial is to determine the effectiveness of the enhanced training program in improving early engagement with mental health services for children and families receiving support from CACs.

Response / Child Wellbeing: Child Knowledge and Awareness

Interventions to improve response to abuse increase child knowledge and awareness about abuse, though this finding is based on only few studies.

Evidence status	Some concerns	The evidence is thin, but response-focused programmes can increase children’s knowledge and awareness regarding abuse.
------------------------	----------------------	--

The summary in brief

This cell has only one primary study and four systematic reviews.

The primary study is a school-based programme with (i) training for teachers to respond better to potential abuse situations with their students and (ii) teaching students to recognize abuse when it happens and to talk to a trusted adult.

These interventions increase knowledge and awareness among children about how to recognize abuse and what they should do when they face a dangerous situation. There are numerous prevention-focused programmes conducted in schools that can increase knowledge and awareness, but only few studies focus on response (i.e., relevant to this cell).

Contents of the cell

The cell has one study (*Nickerson 2019*) and four systematic reviews (*Quadara 2015, Radford 2017, Ricardo 2011, Sherr 2017*).

<i>Nickerson 2019</i> RCT, low risk of bias.	US (eight schools in one state). 161 teachers in study. Evaluation of <i>Second Step Child Protection Unit</i> , a multipronged approach to child sexual abuse prevention and response.
<i>Quadara 2015</i> Scoping review, low quality	Australia: A broad review to identify the prevalence of child sex abuse, risk factors, and intervention approaches.
<i>Radford 2017</i> Rapid evidence assessment, low quality	UK: A rapid evidence synthesis commissioned by the Independent Inquiry into Child Sexual Abuse.
<i>Ricardo 2011</i> Systematic review, low quality	Global: A systematic review of interventions to prevent boys and young men from committing sexual violence. The interventions could look to address gender inequitable attitudes, bystander intentions, and other related behaviours.
<i>Sherr 2017</i> Rapid evidence assessment, low quality	A series of systematic reviews on child maltreatment in institutional care, examining the prevalence of abuse and peer violence in institutions, interventions to reduce abuse, and measures of children’s cognitive and social development.

A. PRIMARY STUDIES

The intervention

The Second Step Child Protection Unit (CPU) curriculum aims to empower students to address false assumptions regarding child sexual abuse. The programme provides teachers with tools so that they can identify and respond to abuse situations and teaches children about personal safety, not keeping secrets, and good/bad touch. School staff are trained via online sessions with simulated scenarios. Teachers also receive support to teach the lessons to their students. CPU lessons are delivered over six weeks in children’s regular classrooms. Younger children (pre-K and kindergarten) get short, daily lessons, while older students (grades 1-4) receive longer weekly sessions. Teachers use discussion, pictures, videos, puppets, and vignettes to teach prevention concepts.

Does the intervention work in improving child knowledge and awareness?

The CPU program was effective in improving child sexual abuse prevention knowledge and self-efficacy among elementary school students. The programme was more effective for younger students (lower grades) than for older students (higher grades). Girls gained more knowledge and self-efficacy than did boys.

Has the intervention been implemented at scale?

No.

Which type of organisation delivered the interventions?

Teachers receive online training and some resources before they deliver the curriculum in their regular classrooms.

What do the interventions cost?

No information on costs.

How are the programmes meant to work? The theory of change

No specific theory is mentioned in the paper.

How reliable is the evidence?

The one primary study in the cell is judged to have a ‘low risk of bias’.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
<i>Nickerson 2019</i>	Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

B. SYSTEMATIC REVIEWS

What are the systematic reviews about?

Quadara 2015 is a broad scoping review to understand the prevalence of child sexual abuse in the Australian context, to identify risk factors for abuse, and look at available interventions to address abuse.

Radford 2017 is a rapid review commissioned by the Independent Inquiry into Child Sexual Abuse in England and Wales (IICSA). It investigates whether public and non-state institutions have adequately protected children and young people from CSA and exploitation. The rapid review sought to learn how institutions (state and non-state) outside of England and Wales have prevented and responded to CSA and exploitation.

Ricardo 2011 is a systematic review that examines the global evidence on effective interventions to reduce the use of sexual violence by boys and young men. The review looked at engaging with boys and men directly in the community and in schools to intervene on various risk factors for sexual violence. Interventions could target gender equitable attitudes, bystander intentions, and other behaviours.

Sherr 2017 is a series of systematic reviews on multiple aspects of child violence in institutional care: (i) the prevalence of maltreatment of children in institutional care; (ii) interventions to reduce abuse in these settings; (iii) peer violence in institutions and (iv) the cognitive and social development of children in institutional care.

What are the findings on child knowledge and awareness?

There are not many response-focussed interventions available. In general, interventions looking to reduce child maltreatment can improve children's knowledge and awareness of how to recognize abuse and be inclined to report it to a trusted adult. However, most interventions have a prevention focus rather than a response or disclosure focus.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

How reliable is the evidence?

Not very. The quality of all systematic reviews is rated as **high risk of bias**. This means that there is at least one major flaw in how they were conducted.

Treatment / Child Wellbeing: Cognitive Functioning

A foster care programme for young children (less than three years old) which permanently removed them from terrible institutional care in Bucharest, Romania, was effective in improving cognitive development in early childhood. However, by adolescence, foster care children’s cognitive functional outcomes were still lagging children who had never been in institutional care.

Evidence status	Moderate risk of bias	Moderate evidence of impact on cognitive outcomes
------------------------	------------------------------	---

The summary in brief

After the Ceausescu political regime ended in 1989, Romania was left with many children living in terrible orphanages. To address these issues, *The Bucharest Early Intervention Project (BEIP)*, a novel foster care programme, was introduced in 2000. At the time of launch, foster care was very uncommon in Romania.

The cell has five papers (all on *BEIP*) that report cognitive functioning outcomes – Johnson 2010, Smyke 2010, Wade 2018, Wade 2020, and Colich 2021. A full summary of two *BEIP* papers (Johnson 2010 and Bick 2015) is available in the guidebook. There are eleven papers in total of this RCT in the EGM all in the ‘treatment’ row across multiple outcomes.

Before their third birthday, children in orphanages (institutional care) were randomised to either move to foster homes (i.e., receive *BEIP*) or to remain in the orphanages (institutions). The aim was to study the effects of reasonably early intervention for children who experience adverse situations after birth. The domains studied included growth, brain development, cognition, and behaviours. Various outcomes are measured at various times in each of three groups of children: a group moved from institutional care to foster care (FC: *BEIP*); a group which remained in institutional care (care as usual, CAU); and a comparison group who were never in care (NIC).

Foster caregivers received regular support from trained social workers. Social workers aimed to facilitate the establishment of a strong bond between children and their foster carers. At 42-54 months, cognitive development in foster care children was associated with developmental quotient (DQ) scores at the start of the study (lower the initial DQ, higher the improvement in DQ and IQ), by standardised change-in-height scores (every unit height increase led to a 12 point increase in verbal IQ) and birth weight (low birth weight and entering foster care after age two combined for poor cognitive development). Foster care children demonstrated higher positive attachment to their caregiver (such as being secure) than those that remained in institutional care but lower than for children who had always lived at home. Every month spent in foster care and every unit of improvement in DQ added up to better attachment outcomes for foster care children.

Cognitive functions such as memory, learning and problem-solving were tested between ages eight and 16. Children who had never been in institutional care (always lived at home) performed better on almost all outcomes at age eight than both the groups in foster care or institutional care, and they consolidated their lead by age 16 (i.e., the others never caught up). Surprisingly, the foster care group and their peers in institutional care performed similarly on almost any cognitive function measure (one would expect foster care children to perform better).

The cell also includes one systematic review of cognitive behavioural therapy for adolescent sexual offenders. There were no differences in cognitive outcomes between CBT and usual treatment groups in residential facilities.

Contents of the cell

The papers are all from the same RCT, and all have moderate risk of bias.

Johnson 2010	Romania. Children in institutional care (orphanages) in Bucharest.
--------------	--

Smyke 2010	Evaluation of the <i>Bucharest Early Intervention Project (BEIP)</i> , a foster care programme for children in institutional care.
Wade 2018	Three studies from <i>BEIP</i> report on predictors of cognitive development and association between cognitive development and attachment in the short-term (42-54 months) and cognitive functional outcomes in the long-term (8 years-16 years).
Wade 2020	
Colich 2021	

A. Primary Studies

The intervention

Under the Ceausescu dictatorship in Romania (until 1989), abandoned children lived in dreadful orphanages. Bucharest had six institutional care centres (“orphanages”), which were characterised by terrible environments for children to grow physically, mentally, socially, or emotionally. Foster care was practically non-existent in Romania during this time.

In the year 2000, *The Bucharest Early Intervention Project (BEIP)* was created to provide foster care. *BEIP* established 56 foster families that could take in children from institutions. Randomisation of children to *BEIP* or continuing institutional care was rationalised since, without *BEIP*, all children would continue to live in awful conditions. This was a chance to identify an effective intervention that could be used to address this issue.

Foster care recruitment and training were standardised and relevant to the local context. Three social workers supported foster caregivers on a regular basis. Social worker roles focused on monitoring the relationship between children and their foster caregivers, promoting parent-child attachment relationships, providing support for behavioural management as needed, and serving as a resource for foster caregivers on the special needs of their children. Social workers were trained and received ongoing support from US-based mental health practitioners every week. Overall, social workers promoted a committed relationship between foster caregivers and the children.

Children entered foster care between five and 31 months of age. An assessment conducted when a child was four and a half years old showed that most *BEIP* children were still with their foster family. *BEIP* was not directly supported by the local government initially, but after a few years, the local government in Bucharest provided financial and administrative support for foster families and children.

Does the intervention work in improving children’s cognitive functioning?

Effects on cognitive functioning were measured in the short term (42-54 months) and in the longer term (age 8-16).

Short-term (42-54 months):

Three factors were statistically significant predictors of DQ and IQ for children who entered foster care:

1. *Baseline DQ* (when the study began) for all DQ and most IQ measures (except for performance IQ) at 54 months. The lower the baseline DQ (i.e., the child was worse off developmentally) the greater the increase in DQ and IQ measures at 42 and 54 months.

2. *Change in height z score* (a standardised measure for height change in children) for DQ at 42 months and verbal IQ at 54 months. Each unit increase in the z score meant an average increase of about 12 verbal IQ points at 54 months.
3. *Birthweight* for full IQ and performance IQ at 54 months. The impact of low birth weight and delayed placement in foster care is clear. Low birth weight children (<2.5kg) placed in foster care after age two had statistically significantly lower IQ scores than children with normal birth weight who were placed in foster care before their second birthday (average IQ score of 67.7 vs. 91.1 points at 54 months). Low birth weight infants in institutional care are especially vulnerable to cognitive deficits. The earlier the intervention (foster care), the better the chances of preventing this.

Cognitive development was also studied as a predictor for attachment in children at 42 months of age. DQ scores were highest for children who had always lived at home with their birth families, followed by foster care children and children who remained in orphanages. Across groups, children with organised attachment (a set of positive indicators of attachment) and secure attachment (a specific positive indicator of attachment) had higher DQ scores at 42 months.

Each month spent in foster care increased the odds of the child demonstrating organised attachment by approximately 27%. For children in 'Care As Usual' (CAU), DQ was a predictor of organised attachment. Each unit increase in DQ scores improved the odds of organised attachment by close to 7%. DQ scores were not predictive of organised attachment for foster care children. Secure attachment, one specific indicator within organised attachment, was associated with DQ scores for foster care children (nearly 6% increased odds of secure attachment with every unit increase in DQ). This was also seen in children who had always lived at home with their birth families (14% increase in odds) but not in CAU children.

Long-term (8-16 years):

Memory and 'executive functioning' (EF) were tested at ages eight through 16 for foster care children (FC), children who remained in institutional care (care as usual: CAU) and children who had always lived with their families at home (never in care: NIC). EF "is an umbrella term for a group of skills involved in goal-directed action and problem solving, including working memory, cognitive flexibility, response inhibition, and attentional control." These cognitive abilities can predict future educational attainment, mental health, income, and psychosocial wellness in adulthood. The Cambridge Neuropsychological Test Assessment Battery (CANTAB) was used to measure these cognitive domains.

1. *Attention and short-term visual memory*: HC performed statistically significantly better than both FC and IC at age eight and at age 16 (all groups seemed to have improved their performance at the same rate from age eight to 16).
2. *Spatial planning (e.g., being able to accurately describe their environment and orient themselves in new surroundings) and problem solving*: At age eight, all three groups did equally well, i.e., there were no statistically significant differences among groups. However, at age 16, HC did statistically significantly better than the other two groups.
3. *Spatial working memory*: HC were able to strategize statistically significantly better and make fewer mistakes at age eight compared to the other two groups. All three groups improved their performance on this domain as they got older, but the HC group improved statistically significantly more than the other two. This meant that by age 16, the HC group had considerably widened their lead.
4. *Visual-spatial memory and new learning*: At age eight, FC had statistically significantly more errors and needed more attempts than the other two groups. However, they were able to catch up with the other groups by age 16.

Being placed in foster care before age two did not seem to make much difference, like it did for short-term outcomes (discussed above).

The findings from long-term assessments are surprising. Though it is plausible that children who have always lived at home (NIC) will have an advantage compared to children who have ever been in institutional care (FC or CAU), one would expect better outcomes for FC compared to CAU (which was not the case here).

Have the interventions been implemented at scale?

Not really. This was a unique study in Bucharest in a situation that was quite out of the ordinary.

Which type of organisation delivered the intervention?

The intervention and the associated RCT were designed by researchers from various US universities. The researchers partnered with a local NGO (SERA Romania) to implement *BEIP*. The team also collaborated with local authorities at the Ministry of Health and the Directorates of Child Protection.

What do the interventions cost?

The study does not report cost data.

How is the programme meant to work? The theory of change

The study does not mention a specific theory on which the programme is based.

Are the results generalisable?

BEIP covered the entire Bucharest area since children in all six institutional care facilities were included in the programme. It was implemented in Bucharest at a time when foster care was uncommon in Romania. Findings from this trial are probably generalisable to jurisdictions looking to ramp up support for fostering children - because *BEIP* was a newly created foster programme.

How reliable is the evidence?

Moderately reliable.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Johnson 2010	Some concerns	Low risk	Some concerns	Low risk	Low risk	Some concerns
Smyke 2010	Some concerns	Some concerns	Some concerns	Some concern	Low risk	Some concerns

Wade 2018	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Some concerns
Wade 2020	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Some concerns
Colich 2021	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Some concerns

What else do we know about the *Bucharest Early Intervention Project (BEIP)*?

The *Bucharest Early Intervention Project (BEIP)* – which is the subject of eleven papers on the EGM - placed children from terrible orphanages in Romania into foster care. Foster care children, predictably, did much better than institutional care children on most developmental, physical, mental, emotional, and cognitive health outcomes. They were also able to catch up with their peers, who had always lived at home with their birth families, on many of these outcomes, especially in the short term. The earlier the intervention, the more beneficial the intervention. *BEIP* is classified as a treatment intervention on the EGM since foster care was considered a treatment after living in extremely poor conditions (in orphanages).

B. Systematic Review (Sneddon 2020)

What is the systematic review about?

Sneddon 2020 is a systematic review assessing the effectiveness of cognitive behavioural therapy (CBT) for adolescents (10-18 years) who show ‘problematic or harmful sexual behaviour.’ Usually, these adolescents have received a formal reprimand or conviction for their behaviour and are receiving treatment in residential facilities.

What are the findings on children’s cognitive functional outcomes?

Three were US studies and one from South Africa. There was no difference between CBT and regular treatment on cognitive outcomes such as overall distortions on harmful sexual behaviour or specifically on distortions about rape, i.e., that rape or sexual assault is wrong.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

Are results generalisable?

Not applicable. Findings were not significantly different for CBT vs. usual treatment.

How reliable is the evidence?

The systematic review was judged to be **low risk of bias**. This means that it was conducted well, which increases our confidence in the findings.

Treatment / Child Wellbeing: Social Emotional Functioning

A foster care intervention in Romania improved the social emotional deficits of children who spent early years neglected in orphanages.

Evidence status	Moderate risk of bias	Moderate evidence of impact on social and emotional functional outcomes
------------------------	------------------------------	---

The summary in brief

After the Ceausescu political regime ended in 1989, Romania was left with many children living in terrible orphanages. To address these issues, *The Bucharest Early Intervention Project (BEIP)*, a novel foster care programme, was introduced in 2000. At that time, foster care was very uncommon in Romania.

The cell has four primary papers (all on *BEIP*) that report social and emotional functioning outcomes: Myke 2010, Wade 2020, Tang 2021, and Colich 2021. A full summary of two *BEIP* papers (Johnson 2010 and Bick 2015) is available in the guidebook. There are eleven papers in total of this RCT in the EGM.

Before their third birthday, children in orphanages (institutional care) were randomised to either move to foster homes (i.e., receive *BEIP*) or to remain in the orphanages (institutions). The aim was to study the effects of reasonably early intervention for children who experience severe neglect in early years. Foster caregivers received regular support from trained social workers, who aimed to facilitate a strong bond between children and their foster carers.

Children who entered foster care (in *BEIP*) before they turned two (children were eligible to enter *BEIP* till their third birthday) tended to form secure attachments with their caregivers while institutionalized children generally had more disorganized attachments with their caregivers. Even in adolescence, *BEIP* foster care children were able to form good friendships like children who had always lived at their birth homes. Institutionalized children continued to struggle with social communication and social relationships well into adolescence. The evidence shows that early intervention can reverse many of the adverse effects of the orphanages.

The cell also includes two systematic reviews. One looks at cognitive behavioural therapy for young sexual offenders and the other at systemic efforts to provide trauma-informed care to children in foster care. There is some indication that social emotional functioning outcomes improve through these interventions, but the quality of the available evidence is low.

Contents of the cell

A. Primary Studies

The papers are all from the same RCT, and all have moderate risk of bias.

Smyke 2010	Romania. Children in institutional care (orphanages) in Bucharest.
Wade 2020	Evaluation of the <i>Bucharest Early Intervention Project (BEIP)</i> , a foster care programme for children in institutional care.
Tang 2021	The papers are reporting on different social emotional functions of children from early childhood to adolescence.
Colich 2021	

The intervention

Under the Ceausescu dictatorship in Romania (until 1989), abandoned children lived in dreadful orphanages. Bucharest had six institutional care centres (“orphanages”), which were characterised by terrible environments for children to grow physically, mentally, socially, or emotionally. Foster care was practically non-existent in Romania during this time.

In the year 2000, *The Bucharest Early Intervention Project (BEIP)* was created to provide foster care. *BEIP* established 56 foster families that could take in children from institutions. Randomisation of children to *BEIP* or continuing institutional care was rationalised since, without *BEIP*, all children would continue to live in awful conditions. This was a chance to identify an effective intervention that could be used to address this issue.

Foster care recruitment and training were standardised and relevant to the local context. Three social workers supported foster caregivers on a regular basis. Social worker roles focused on monitoring the relationship between children and their foster caregivers, promoting parent-child attachment relationships, providing support for behavioural management as needed, and serving as a resource for foster caregivers on the special needs of their children. Social workers were trained and received ongoing support from US-based mental health practitioners every week. Overall, social workers promoted a committed relationship between foster caregivers and the children.

Children entered foster care between five and 31 months of age. An assessment conducted when a child was four and a half years old showed that most *BEIP* children were still with their foster family. *BEIP* was not directly supported by the local government initially, but after a few years, the local government in Bucharest provided financial and administrative support for foster families and children.

More details on *BEIP* are provided in other summaries in the guidebook (Bick 2015 and Johnson 2010; and the syntheses of Treatment X Child Cognitive Functioning and Treatment X Mental Health).

Does the intervention work in improving children’s social emotional functioning?

Children who went into foster care before they turned two (children were entered into *BEIP* until their third birthday) tended to develop strong emotional bonds with their caregivers, i.e., secure attachment, when assessed at 42 months. Children who continued to live in the orphanages typically showed more tenuous bonds with their caregivers, i.e., disorganized, or insecure attachment.

This secure attachment persisted in adolescence (age 16 when assessed), with fostered children having more friends and better relationships with friends than the institutionalized children. In fact, fostered children had similar friendships as children who had always lived with their birth families. Social communication problems persist into adolescence for institutionalized children, and early intervention (foster care programme) can mitigate these issues to a large extent.

Have the interventions been implemented at scale?

Not really. This was a unique study in Bucharest in a situation that was quite out of the ordinary.

Which type of organisation delivered the intervention?

The intervention and the associated RCT were designed by researchers from various US universities. The researchers partnered with a local NGO (SERA Romania) to implement *BEIP*. The team also collaborated with local authorities at the Ministry of Health and the Directorates of Child Protection.

What do the interventions cost?

The study does not report cost data.

How is the programme meant to work? The theory of change

The study does not mention a specific theory on which the programme is based.

Are the results generalisable?

BEIP covered the entire Bucharest area since children in all six institutional care facilities were included in the programme. It was implemented in Bucharest at a time when foster care was uncommon in Romania. Findings from this trial are probably generalisable to jurisdictions looking to ramp up support for fostering children - because *BEIP* was a newly created foster programme.

How reliable is the evidence?

Moderately reliable.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomised process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Smyke 2010	Some concerns	Some concerns	Some concerns	Some concern	Low risk	Some concerns
Tang 2021	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Some concerns
Wade 2020	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Some concerns
Colich 2021	Some concerns	Low risk	Some concerns	Some concerns	Low risk	Some concerns

What else do we know about the *Bucharest Early Intervention Project (BEIP)*?

The *Bucharest Early Intervention Project (BEIP)*, which is the subject of eleven papers on the EGM, placed children from terrible orphanages in Romania into foster care. Foster care children, predictably, did much better than institutional care children on most developmental, physical, mental, emotional, and cognitive health outcomes. They were also able to catch up with their peers, who had always lived at home with their birth families, on many of these outcomes, especially in the short term. The earlier the intervention, the more beneficial the intervention. *BEIP* is classified as a treatment intervention on the EGM since foster care was considered a treatment after living in extremely poor conditions (in orphanages).

B. Systematic Reviews (Sneddon 2020, Bailey 2019)

What are the systematic reviews about?

Sneddon 2020 is a systematic review on the effectiveness of cognitive behavioural therapy (CBT) for young people who show ‘problematic or harmful sexual behaviour.’ Typically, these children have received a formal reprimand or conviction for their behaviour and are receiving treatment in residential facilities.

Bailey 2019 is another systematic review examining trauma-informed care models for children and youth living in out-of-home care.

What are the findings on children’s cognitive functional outcomes?

Only one study reported a small increase in victim empathy with CBT compared to no treatment (Sneddon 2020).

Bailey 2019 found few studies on organizational-wide trauma-informed care models. All studies were from the US. Three models of care were identified – Attachment Regulation and Competency framework (ARC), the Children and Residential Experiences programme (CARE), and The Sanctuary Model. Studies reported different measures of social and emotional function and the results were generally better for children in these programmes compared to usual care. However, most studies in the review were rated as having a high risk of bias meaning low confidence in the “positive” findings for outcomes.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

Are results generalisable?

Not applicable.

How reliable is the evidence?

It’s mixed. The quality of one systematic review is rated as low risk of bias (Sneddon 2020) and the other as high risk of bias (Bailey 2019).

Treatment / Child Wellbeing: Mental Health

High-quality foster care which removes children from terrible institutional care and cognitive behavioural therapy (CBT) for sexually abused minors are two effective interventions which improve mental health for children who have suffered adverse experiences.

Evidence status	Some concerns	Moderate evidence of impact on mental health outcomes
------------------------	----------------------	---

The summary in brief

This cell includes two primary studies (which are studied in seven separate research papers) and three systematic reviews. The interventions in both primary studies started / were given over 20 years ago. They assess the effects of interventions to improve mental health in children who have experienced extremely adverse conditions such as living in dreadful institutions or being sexually abused and exploited or both. The cell also contains one protocol for a new study (in the UK).

Six papers are written about *The Bucharest Early Intervention Project (BEIP)*, a (then) novel foster care programme introduced in 2000 to address the aftermath of the Ceausescu political regime which left Romania with many children living in terrible orphanages. At the time of launch, foster care was very uncommon in Romania. Researchers took ‘advantage’ of that scarcity to create an RCT: children in the orphanages (yet to reach their third birthday) were randomly assigned to either move to foster homes (i.e., receive *BEIP*) or to remain in the orphanages (institutions). In the foster care intervention, foster caregivers received regular support from trained social workers, who aimed to facilitate a strong bond between children and their foster carers. Both groups (foster care and treatment-as-usual in the orphanages) were compared with children who had always lived at home with their birth families.

Mental health outcomes were measured when children were about eight years old, again when they were 12 years old, and recently at 16 years of age. They suggest that children in the foster care group had better outcomes than children who remained in orphanages on many (though not all) outcomes. Children who lived at home with their birth families had consistently better mental health outcomes than children who were ever in institutional care. Early entry into foster care and longer length of stable foster care led to improved mental health outcomes compared to remaining in orphanage care.

The other primary study in this cell examined psychotherapy for children who had suffered sexual abuse. A year after receiving therapy many mental health outcomes improved but others were not different for the therapy group and the group that did not get it.

Two systematic reviews in the cell found only a small number of RCTs of therapy interventions for children who are victims abuse or for perpetrators of abuse (both adults and children) against children. Cognitive behavioural therapy has been tested in the few trials that exist with mixed results. The third systematic review was on trauma-informed out-of-home care models which found a small number of studies most of which had a high risk of bias.

Contents of the cell

There are 11 papers in this cell:

- two primary studies about *BEIP* (six papers: Troller-Renfree 2015, Humphreys 2015, Wade 2019, Wade 2020, Wade 2020a, Colich 2021)
- Sullivan 1992, a QED in the US
- three systematic reviews (Radford 2017, Bailey 2019, Sneddon 2020) and
- one protocol for an RCT (Hiller 2021).

On geography: both completed RCTs are from US and Romania, ie., middle-high income countries. The systematic reviews also cover high-income countries plus South Africa (which has the highest inequality in the world, so is like a high-income country co-located with a low-income one).

A. Primary Studies	
<p>Troller-Renfree 2015, Humphreys 2015 Wade 2019 Wade 2020 Wade 2020a Colich 2021 RCT, moderate risk of bias</p>	<p>Romania. Children in institutional care (orphanages) in Bucharest. Evaluation of the <i>Bucharest Early Intervention Project (BEIP)</i>, a foster care programme for children in institutional care. Six papers from <i>BEIP</i> reporting on mental health outcomes in the long-term (age 8-16 years).</p>
<p>Sullivan 1992 QED, moderate risk of bias</p>	<p>USA. Sexually abused children in residential care (for deaf children) Evaluation of psychotherapy compared to no treatment for sexually abused children from one residential school.</p>
<p>Hiller 2021 RCT protocol, results not yet reported</p>	<p>UK. Children 10-17 years in out-of-home care with high post-traumatic stress symptoms due to experiencing or witnessing maltreatment. Evaluation of an online low-intensity group therapy programme called ‘Teaching Recovery Techniques (TRT).’</p>
B. Systematic Reviews	
<p>Radford 2017 Systematic Review, low quality rating</p>	<p>Rapid review commissioned by independent panel investigating institutional failures in England and Wales to protect children from CSA and exploitation. The rapid review aimed to find effective interventions that institutions in countries outside England and Wales have implemented. Included studies in review relevant to this cell are from the US and Canada.</p>
<p>Bailey 2019 Systematic Review, low quality rating</p>	<p>Systematic review of organisation-wide, trauma-informed care models in out-of-home care (OoHC) settings. Included studies were all from the US.</p>
<p>Sneddon 2020 Systematic Review, high quality rating</p>	<p>Systematic review of Cognitive Behavioural Therapy (CBT) for adolescents (10-18 years) with harmful sexual behaviour. Included studies are from the US (three) and one from South Africa.</p>

A. Primary Studies

The interventions

Under the Ceausescu dictatorship in Romania (until 1989), abandoned children lived in dreadful orphanages. Bucharest had six institutional care centres (“orphanages”), which were characterised by

terrible environments for children to grow physically, mentally, socially, or emotionally. Foster care was practically non-existent in Romania during this time.

In the year 2000, *The Bucharest Early Intervention Project (BEIP)* was created to provide foster care. *BEIP* established 56 foster families that could take in children from institutions. Randomisation of children to *BEIP* or continuing institutional care was rationalised since, without *BEIP*, all children would continue to live in awful conditions. This was a chance to identify an effective intervention that could be used to address this issue.

Foster care recruitment and training were standardised and relevant to the local context. Three social workers supported foster caregivers on a regular basis. Social worker roles focused on monitoring the relationship between children and their foster caregivers, promoting parent-child attachment relationships, providing support for behavioural management as needed, and serving as a resource for foster caregivers on the special needs of their children. Social workers were trained and received ongoing support from US-based mental health practitioners every week. Overall, social workers promoted a committed relationship between foster caregivers and the children.

Children entered foster care between five and 31 months of age. An assessment conducted when a child was four and a half years old showed that most *BEIP* children were still with their foster family. *BEIP* was not directly supported by the local government initially, but after a few years, the local government in Bucharest provided financial and administrative support for foster families and children.

More details on *BEIP* are provided in other summaries in the guidebook (Bick 2015 and Johnson 2010; and the syntheses of Treatment X Child Cognitive Functioning and Treatment X Mental Health).

Troller-Renfree 2015 assessed children's attention biases (a tendency to focus on certain things while ignoring others – and a potential sign of future mental illness) when the children were eight to nine years old, and Humphreys 2015 reported mental illness-related symptoms at 12-13 years of age. The more recent papers about *BEIP* reported on externalizing symptoms (“acting out”) and general psychopathology after the children reached 16 years of age.

Sullivan 1992 reported the impact of psychotherapy on sexually abused children (ages 12-16) living in a residential school for deaf children in the US. The type of abuse suffered ranged from witnessing sexual abuse to being victims of sexual violence (most children in the study experienced severe abuse). All children were offered psychotherapy but only some parents accepted the offer. Other parents refused and their children did not receive therapy elsewhere either. The investigation of the abuse in this school was very public and many parents and even some school staff strongly denied that the children were sexually abused.

Hiller 2021 is a protocol of an ongoing RCT in the UK. Adolescents in care settings without a biological parent and with high levels of post-traumatic stress usually due to maltreatment at home are offered low-intensity group therapy delivered virtually. The goal of this programme is to provide skills to children to deal with their post-traumatic stress symptoms.

Do these interventions work in improving children's mental health?

The papers about *BEIP* studied various aspects of mental health. Troller-Renfree 2015 looked at attention bias which is considered a precursor to mental illnesses, such as anxiety disorder and depressive disorders. The primary outcome in Humphreys 2015 was the symptom counts for mental illnesses such as internalising disorders (anxiety, depression), externalising disorders (oppositional defiant disorder and conduct disorder) and attention deficit hyperactivity disorder (ADHD).

Troller-Renfree 2015 used a validated test called the 'dot-probe task' to assess attention bias. In summary, this test presents a pair of emotional faces (some combination of angry, happy, and neutral) followed immediately by a symbol (+) behind one of the images. The children are meant to very quickly indicate which side the symbol is on. Bias scores are calculated by subtracting the reaction time when the symbol was behind an emotion face (angry or happy) from that when it was behind a neutral face. A positive bias score indicates a bias towards threat or positive stimuli and negative scores are the converse, i.e., bias away from threat or positive stimuli. A zero score means no bias was shown.

Children in the orphanage group (n=50, i.e., the group had 50 people) showed a significant bias towards threat stimuli while those that went into foster care (n=55) showed that towards positive stimuli, but when all three groups were compared to each other there were no significant differences. Children who were always at home with their biological parents (n=52) had no bias to either positive stimuli or threats. The size of the positive bias was associated with fewer externalizing problems (acting out, aggression), better prosocial behaviour and engagement and fewer signs of being emotionally withdrawn. However, the size of the threat bias was not significantly associated with any social outcomes. Entering foster care (whether *BEIP* or government foster care) at a younger age was related to a large positive attention bias (and therefore, better social outcomes).

Humphreys 2015 used the Diagnostic Interview Schedule for Children, 4th edition (DISC-IV) to interview caregivers (parent or institutional caregiver) to get information on "symptom levels, duration or persistence, age of onset and functional impairment" when the children were close to 12 years of age. 44% of the orphanage group (n=55) fulfilled criteria for "any psychiatric disorder," 27% for those in "stable foster care" i.e., children who continued in *BEIP* foster care, 43% for those in "disrupted foster care" i.e., where the *BEIP* foster care arrangement was changed (for e.g., placement into government foster care, readmitted to institutional care) and 16% for children always at home (n=49). The prevalence of "any psychiatric disorder" was 39% for children who were ever in institutional care.

Internalising symptoms, externalising symptoms, ADHD symptoms were statistically significantly lower for children always at home than children who had ever been in institutional care. Results were similar for both girls and boys - except for internalising symptoms in boys which was still lower for the never-institutionalized group but not statistically significant. Institutionalized children (specifically boys in this group) had statistically significantly higher externalising symptoms than the foster care group but internalising symptoms and ADHD symptoms were similar between groups.

"Disrupted foster care" children had higher symptoms across all categories compared to "stable foster care" children suggesting that such stability is an important factor in mental health outcomes.

Both papers use data gathered quite a while after *BEIP* began: they followed up with children in the long term, when they were age 8-12. In fact, the papers cover a period of four years. The children in the two 'treatment' groups (*BEIP* vs institutional care) experienced many life changes over the years that could affect our understanding of the effects of *BEIP*. Of the children moved into *BEIP* foster care, only about half stayed there: some went into government foster care while others returned to their biological families or were lost to follow-up. Similarly, of the 68 children in the study who stayed in institutional care, only 43 were available at age eight: only about a quarter of them were still in institutional care, whereas others had moved to foster care or returned to their families. This pattern continued when another assessment was done at age 12.

The more recent studies of *BEIP* (Wade 2019, Wade 2020, Wade 2020a, Colich 2021) assessed outcomes up to age 16 to understand how early adversity and the foster care intervention affected mental health in adolescence. In general, they found that externalizing symptoms ("acting out") and psychopathology in general were lower for foster care children compared to their peers who remained in orphanages.

Sullivan 1992 reported on 72 children who had been sexually abused at a residential school for deaf children. The severity of abuse was ranked from 1-4 (with 4 being most severe) and close to 80% were in categories 3 and 4. Psychotherapy was delivered by three therapists (Master's level) with supervision by a

psychologist and a psychiatrist. 35 children (21 boys, 14 girls) received therapy and 37 children (30 boys, seven girls) whose parents refused therapy were controls. Treatment goals were: “alleviation of guilt; treatment of depression; learning to express anger relevant to the event; basic information on normal human sexuality and interpersonal relationships; sexual preference and homosexual issues; maltreatment issues; self-protection techniques; affective vocabulary for emotions and feelings; emotional independence; establishment of a meaningful and stable identity; personal value system; and a capacity for lasting relationships.” Each child received a two-hour weekly session of therapy for 36 weeks.

One year after therapy began, boys in the therapy group had significantly lower scores (lower is better) as reported by their “house parents” (presumably guardians assigned to each child in the residential school) than the control group on behaviour scales for – total, internal, external, somatic, uncommunicative, immature, hostile, delinquent, aggressive, and hyperactive. No differences were seen for schizoid and obsessive scales. For girls, those that got therapy did better than controls on total, external, depressed, aggressive, and cruel. There were no differences between groups in internal, anxious, schizoid, immature, somatic, and delinquent scales.

Have the interventions been implemented at scale?

Not really. BEIP was a unique study in Bucharest in a situation that was quite out of the ordinary. The therapy intervention described in Sullivan 1992 was implemented for children in only one residential school.

Which type of organisation delivered the intervention?

For *BEIP*, the intervention and the associated RCT was designed by researchers from various US universities. The investigators partnered with a local NGO (SERA Romania) to implement various intervention activities. The team also collaborated with local authorities at the Ministry of Health and the Directorates of Child Protection.

The implementing organisation is not mentioned in Sullivan 1992.

What do the interventions cost?

No cost data is reported.

How is the programme meant to work? The theory of change

No specific theory is mentioned in either study.

Are the results generalisable?

BEIP covered the entire Bucharest area since children in all six institutional care facilities were included in the programme. It was implemented in Bucharest at a time when foster care was uncommon in Romania. Findings from this trial are probably generalisable to jurisdictions looking to ramp up support for fostering children - because *BEIP* was a newly created foster programme.

We can't say much about generalisability on the findings from Sullivan 1992 as it was implemented only in one residential school.

How reliable is the evidence?

Moderately reliable.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Troller-Renfree 2015	Some concerns	Some concerns	Some concerns	Some concerns	Low risk	Some concerns
Humphreys 2015	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns
Wade 2019	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns
Wade 2020	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns
Wade 2020a	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns
Colich 2021	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study	Overall risk of bias	Confounding	Selection bias	Bias in intervention classification	Deviation from intended intervention	Missing outcome data	Measurement of the outcome	Selection of the reported result
Sullivan 1992	Moderate	Moderate	Low	Low	Low	Moderate	Moderate	Low

B. Systematic Reviews (Radford 2017, Bailey 2019, Sneddon 2020)

What are the systematic reviews about?

Radford 2017 was a rapid review commissioned by the *Independent Inquiry into Child Sexual Abuse in England and Wales (IICSA)* which investigated whether public and non-state institutions have done enough to protect children and young people from child sexual abuse and exploitation. The focus of the rapid review was to learn how institutions (state and non-state) outside England and Wales prevented and responded to child sex abuse and exploitation.

Sneddon 2020 is a systematic review on the effectiveness of cognitive behavioural therapy (CBT) for young people who show ‘problematic or harmful sexual behaviour.’ Typically, these children have received a formal reprimand or conviction for their behaviour and are receiving treatment in residential facilities.

Bailey 2019 is another systematic review examining trauma-informed care models for children and youth living in out-of-home care.

What are the findings on children’s mental health outcomes?

Radford 2017 included studies on treatment for sexually abuse and exploited children such as cognitive behavioural therapy (CBT). Fifteen studies – seven systematic reviews (one review was an update) and eight primary studies (six from the US and two from Canada) – examined the effectiveness of various treatment programmes for victims and survivors of child sexual abuse.

Their findings are as follows:

- The overall evidence is poor with few RCTs.
- CBT with a trauma focus is a promising treatment to overcome the adverse effects of sexual abuse for minors.
- Other promising therapeutic approaches are drama-based therapy, Eye Movement Desensitization and Reprocessing - EMDR (which uses eye movements to reduce the emotional impact of past trauma and adverse events) and Modular Approaches to treatment and support (uses a menu of evidence-based treatment modules for different issues such as depression, anxiety, trauma and conduct problems)
- Interventions to help children who have suffered online abuse is an evidence gap, i.e., more studies are needed.
- Benefit is increased by: longer duration of therapy, working with older children, and tailoring therapy to an individual’s specific needs.

From Sneddon 2020, no studies reported on mental health outcomes such as self-harm and suicidality. Only one study reported on sexual aggression and found no difference between those who got CBT and those who didn’t.

Bailey 2019 found only a few studies on organizational-wide trauma-informed care models. All studies were from the US. Three models of care were identified – Attachment Regulation and Competency framework (ARC), the Children and Residential Experiences programme (CARE), and The Sanctuary Model. Studies reported different measures of mental health and the results were generally better for children in these programmes compared to usual care. However, most studies in the review were rated as having a high risk of bias meaning low confidence in the “positive” findings for outcomes.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

Are results generalisable?

Unclear. The inconsistent findings limit generalizability. CBT approaches to reduce the adverse effects of sexual abuse in minors is likely applicable to settings that can provide resources and professionals to deliver it.

How reliable is the evidence?

Unclear. On quality, one systematic review is rated as low, and the other as high.

Radford 2017 did not combine effect sizes statistically, so the overall impact is not clear, nor is whether results were consistent across included studies.

Section 3: Summaries of individual studies which appear in cells with only one or two studies

These summaries are presented in alphabetical order of their lead author, to enable the reader to find them easily.

Summary of three studies from one RCT (Bucharest Early Intervention Project): Johnson (2010), Smyke (2010), Bick (2015) (Treatment / child physical health, and child social-emotional functioning)

Light cells in which this study appears:

Intervention = Treatment; Outcome = Child Wellbeing: Physical Health (2 studies)

Intervention = Treatment; Outcome = Child Wellbeing: Social-Emotional Functioning (1 study)

Summary: A foster care programme for young children (two to three years) permanently removed from terrible institutional care in Bucharest, Romania, was effective in helping them catch up with their peers (who had always lived at home with their birth families) on height, weight, brain, and cognitive development.

This was one RCT, about which multiple papers have been written (studying various aspects of the children’s development and at various times.). The EGM has six papers about this trial. This summary is about three of them (the three which appear in cells with just one or two studies):

- **Johnson et al. (2010)**⁷ ‘Growth and associations between auxology, caregiving environment, and cognition in socially deprived Romanian children to randomised to foster vs. ongoing institutional care,’
- **Smyke at al. (2010)**⁸ ‘Placement in foster care enhance quality of attachment among young, institutionalised children’ and
- **Bick et al. (2015)**⁹ ‘Effect of early institutionalisation and foster care on long-term white matter development. A randomised clinical trial.’

Evidence status	Moderate risk of bias	Moderate evidence of impact on child wellbeing: physical health, social-emotional functioning, cognitive functioning
------------------------	-----------------------	--

The summary in brief

⁷ Johnson DE, Guthrie D, Smyke AT, et al. Growth and associations between auxology, caregiving environment, and cognition in socially deprived Romanian children randomized to foster vs ongoing institutional care. *Arch Pediatr Adolesc Med.* 2010;164(6):507-516. doi:10.1001/archpediatrics.2010.56

⁸ Smyke AT, Zeanah CH, Fox NA, Nelson CA, Guthrie D. Placement in foster care enhances quality of attachment among young institutionalized children. *Child Dev.* 2010;81(1):212-223. doi:10.1111/j.1467-8624.2009.01390.x

⁹ Bick J, Zhu T, Stamoulis C, Fox NA, Zeanah C, Nelson CA. Effect of Early Institutionalization and Foster Care on Long-term White Matter Development: A Randomized Clinical Trial. *JAMA Pediatr.* 2015;169(3):211–219. doi:10.1001/jamapediatrics.2014.3212

After the Ceausescu political regime ended in 1989, Romania was left with many children living in terrible orphanages. Conditions in these institutions were severely detrimental to the children's growth and development. To address these issues, *The Bucharest Early Intervention Project (BEIP)*, a novel foster care programme, was introduced in 2000. At the time of launch, foster care was very uncommon in Romania. Researchers took advantage of that shortage to create an RCT. Before their third birthday, children in orphanages (institutional care) were randomised to either move to foster homes (i.e., receive *BEIP*) or to remain in the orphanages (institutions). Foster caregivers received regular support from trained social workers. Social workers aimed to facilitate the establishment of a strong bond between children and their foster carers.

There are six papers¹⁰ of this RCT in the EGM. This summary includes three papers that appear in "Light cells." These three studies focus on growth, emotional, cognitive, brain development and behaviour outcomes.

Early intervention through *BEIP* improved outcomes compared to institutional care children. *BEIP* children caught up with children who had never been in institutional care on many growth outcomes such as height and weight (but not head circumference) and on brain development. Children who were younger, had higher growth impairments or better caregiver quality gained the most. Cognitive development depended on height catch-up, birthweight, and initial development scores. *BEIP* children also did much better on attachment outcomes compared to institutional care children but not at the level of children who had always lived at home with their birth families. Children who went into foster care earlier benefitted more.

Most children continued to live with their foster caregivers four to five years later.

Type of study: RCT. Peer-reviewed journal articles.

The intervention

Under the Ceausescu dictatorship in Romania (till 1989), abandoned children lived in dreadful orphanages¹¹. Bucharest had six institutional care centres ("orphanages") which were characterised by terrible environments for children to grow physically, mentally, socially, or emotionally. Foster care was practically non-existent in Romania during this time.

In the year 2000, *The Bucharest Early Intervention Project (BEIP)* was created to provide foster care. *BEIP* established 56 foster families that could take in children from institutions. Randomisation of children to *BEIP* or continuing institutional care was rationalised since without *BEIP* all children would continue to live in awful conditions. This was a chance to identify an effective intervention that could be used to address this issue.

Foster care recruitment and training was standardised relevant to the local context. Three social workers supported foster caregivers on a regular basis. Social worker roles focused on monitoring the relationship between children and their foster caregivers, promoting parent-child attachment relationships, providing support for behavioural management as needed and serving as resource for foster caregivers on the special needs of their children. Social workers were trained and received ongoing support from US-based

¹⁰ The three other studies on this trial (six total) in the EGM are: Humphries 2015, Troller-Renfree 2015 and Wade 2018)

¹¹ Weir, K: American Psychological Association. (June, 2014). The lasting impact of neglect. Retrieved from <https://www.apa.org/monitor/2014/06/neglect>

mental health practitioners every week. Overall, social workers promoted a committed relationship between foster caregivers and the children.

Children entered foster care between five and 31 months of age. An assessment conducted when a child was four and half years old showed that most *BEIP* children were still with their foster family. *BEIP* was not initially supported by the local government, but after a few years, the local government in Bucharest provided financial and administrative support for foster families and children.

How is the programme meant to work? The theory of change

The study does not mention a specific theory on which the programme is based. However, the aim was to study the effects of reasonably early intervention for children who experience adverse situations after birth. The domains studied included growth, brain development, cognition, and behaviours.

Has the intervention been implemented at scale?

Yes. *BEIP* introduced a foster care programme in Bucharest, Romania in the year 2000 when none existed. It covered all six institutions providing care to abandoned children in Bucharest.

What does the intervention cost?

The studies do not report any data on cost.

The trial

Participants: The trial involved 136 children living in institutional care in Bucharest. The mean age was 21 months with half female, 55% of Romanian ethnicity and the rest of Roma, other or unknown background. Children had spent about two-thirds of their lives in institutional care and close to a quarter had a low birth weight.

Study design: Any child less than 32 months of age living in institutional care was eligible for randomisation except if they (i) were scheduled for adoption (ii) had serious handicapping conditions such as signs of alcohol exposure while in the womb, cerebral palsy or hearing loss. Recruitment to the trial happened in the first half of 2001.

Children were randomly assigned to one of two groups: (a) *BEIP* (b) continued institutional care. A third group of children who had never been in institutional care but were born in the same hospitals and had a similar distribution of age and gender as the children in institutional care was also recruited.

Outcomes (reported in one or more studies):

- Physical development. The children were tested for height, weight, birth weight and head size (occipital frontal circumference). The measures were taken at: (i) when children were randomised (average age 21 months), (ii) 30 months of age, and (iii) 42 months of age. Head size was measured monthly, at baseline and 42 months of age.
- Cognitive development. Developmental quotients (DQs) were measured at baseline (the start) and when children were 30 and 42 months old. This was based on the Bayley Scales of Infant Development II, Mental Developmental Index. IQ was measured using the Wechsler Preschool Primary Scale of Intelligence II, at 54 months.

- Quality of caregiving. This was measured to investigate factors that might lead to improved physical growth. Researchers video-taped 90 minutes of the caregiving and coded the activities. The factors coded were:
 - a. Detachment: caregiver is emotionally uninvolved, disengaged, and unaware of infant's needs
 - b. Flat affect: caregiver expresses no emotion or animation
 - c. Positive regard for the child, caregiver expresses positive feelings in interactions with the child
 - d. Sensitivity, caregiver responds to the infant's social gestures and is attuned to the infant's needs and moods
 - e. Stimulation of cognitive development, caregiver engages in activities that can facilitate the infant's learning
- Attachment. This was measured using Ainsworth's Strange Situation procedure (SSP), an established tool. Children were seen with their caregiver. Their behaviour was coded (by two independent coders) for five behaviours indicative of attachment:
 - a. Secure. This involves demonstrating a positive, engaged, and open style of verbal and nonverbal interaction but less proximity seeking than in infancy.
 - b. Avoidant. "Children whose attachment is classified as avoidant have reduced involvement with the caregiver and do not seek contact when distressed. In addition, they say little to the caregiver, particularly about feelings, and they remain affectively neutral."
 - c. Ambivalent. "Children whose attachment is ambivalent-dependent display passivity, helplessness, immaturity, or petulance and resistance in their interactions with the caregiver."
 - d. Disorganised-controlling. "Some preschool children whose attachment is classified as disorganised display behaviours similar to toddlers classified as disorganised, but others exhibit efforts to control the behaviour of the caregiver, particularly following the stressful separation episodes of the SSP. Such behaviour tends to follow one of two patterns: controlling-caregiving or controlling-punitive."
 - e. Insecure-other. "Such children have not evolved a reliable strategy for managing their distress by seeking and receiving comfort and reassurance from their caregivers. Although behaviour in interaction with their caregivers may vary, the underlying attachment representation for children who are categorised controlling or insecure-other, may be characterised by 'themes of fear, confusion, chaos, and disorganisation'".

The first three items above were combined into a score for 'organised' behaviour, and the latter two combined into 'atypical' behaviour.

The results were analysed according to the age of the child at placement.

The following were measured to test for whether they affected attachment (i.e., as input variables, rather than outcomes):

- Brain white matter development. This current study looks specifically at “the organisation of white matter microstructure” in the brain, and thereby “the specific white matter tracts that may contribute to the global improvements in white matter changes”. A previous study had provided evidence that moving children from institution to foster care improves the total volume of brain white matter. There are also previous studies demonstrating that “caregiving-based early intervention programs can support more normalised white matter development among children who are exposed to prenatal risk”.

Did the intervention work?

- ✓ **Growth:** The children in institutions were all smaller in all measures than never-institutionalised children.

Children in the foster group grew significantly faster in **height** and **weight** than those who remained in institutions. 12 months after the randomisation (i.e., the start of foster care), “100% of the foster care group was in the normal range for height, 90% for weight, and 94% for weight-for-height.” In other words, the effect of foster care on height and weight was more-or-less complete recovery and was fast. (Reminder: the children were on average only two when the intervention started.) Children with the most significant growth impairments at baseline showed greater catch-up.

This finding corroborates that of other research that height catch-up improved if children were placed in foster care prior to 12 months of age. The study attributes some of the benefit of foster care to diet / malnutrition, and this is known to be particularly important in the early years.

No significant change in any parameter occurred in the six months after that. At 42 months, there was no differences between boys and girls.

However, growth in **head circumference** did not differ between the groups who remained in institutions vs. who moved to foster care.

Factors affecting height and weight gain. The children in foster care who gained most were those who:

- Had lower height, weight and head-size to start with
- Better **caregiving quality** (i.e., quality matters) as measured by “caregiving quality scores” developed by the study. Specifically, “positive regard for the child and sensitivity” were positively correlated with height and weight catchup. Conversely, “caregiver detachment” i.e., being disengaged or uninvolved was related to poor weight and height outcomes.

- ✓ **Cognitive development:** These authors had previously found that placement into foster care prior to the age of 24 months led to better cognitive recovery.

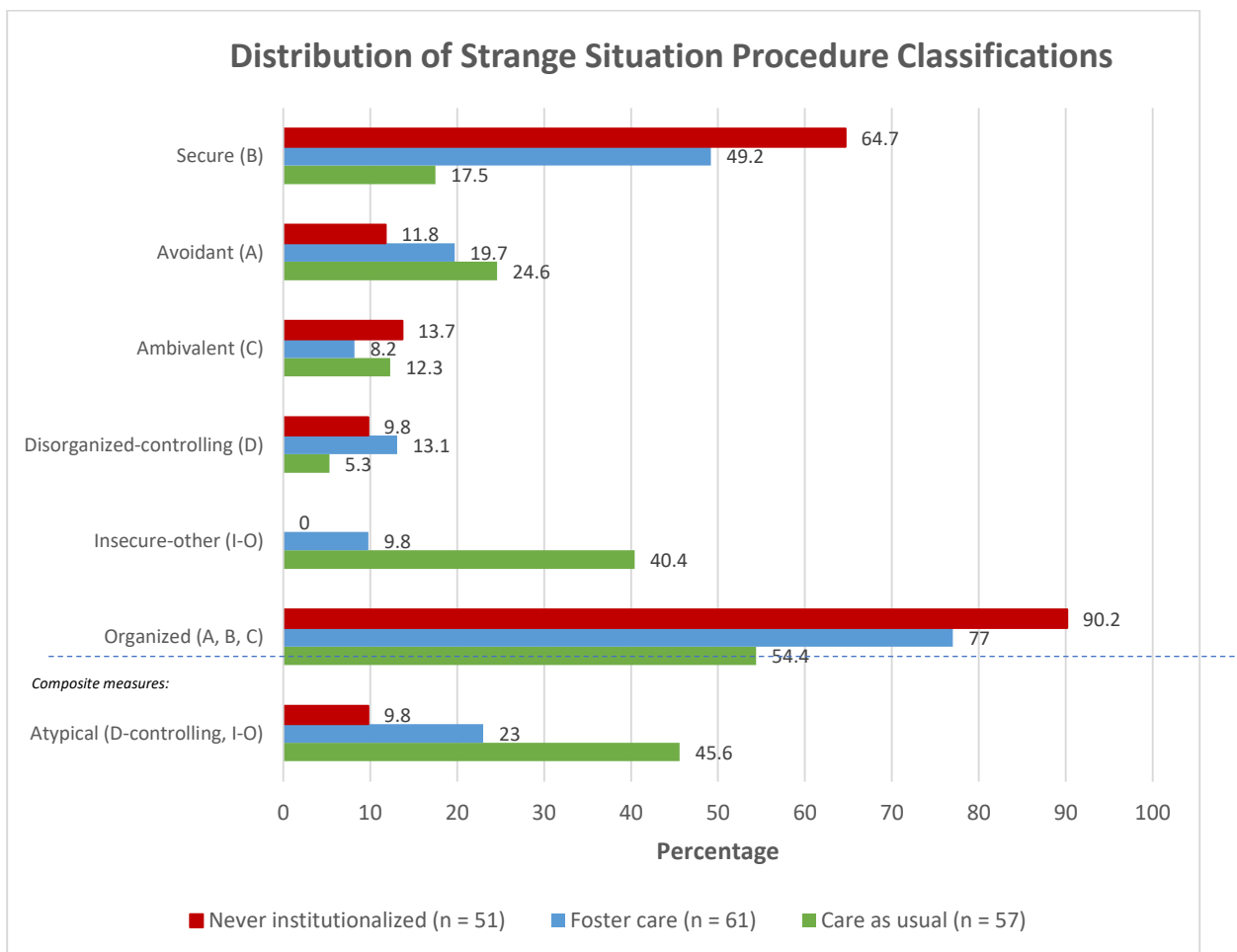
They found that baseline DQ (i.e., when the study began) was the sole significant predictor of all DQ and IQ measures at 54 months, other than performance IQ. Change in height score was a significant predictor of DQ at 42 months and verbal IQ at 54 months. Birth weight was a significant unique predictor of full IQ at 54 months and performance IQ at 54 months. Changes in DQ and IQ between baseline and 42 and 54 months were inversely related to initial DQ. (i.e., children with low development at the beginning developed less later). The extreme cognitive

vulnerability of children with low birth weight (LBW) is highlighted when the risk factors of birth weight and delayed placement into foster care are both considered.

- ✓ **Attachment:** Young children placed into foster care after early institutional rearing may experience significant recovery regarding attachment. This study, when the children were 42 months, supports similar previous findings by the authors when the children were 21 months old. This is one of several studies (by various authors) that have shown that changing caregiving environments can change children’s level of attachment.

Positive indicators for attachment were best for never-institutionalised children (90%) followed by foster care children (77%) and children in institutions (54%). Negative indicators too suggested a similar picture (graph). Foster care children had much better attachment outcomes than children in institutions but did not quite catch up to children who had always lived at home with their birth families.

Figure 1: Distribution of Strange Situation Procedure Classifications



Factors influencing attachment:

Three factors were analysed for their effect on two aspects of attachment: organised attachment, and secure attachment. The sole factor which was reported which affected attachment was cognitive development (NI = never institutionalised, i.e., raised by family):

Table 1: Factors influencing attachment

	Organised attachment			Secure attachment		
Group:	Care as usual	Never institutionalized (family care)	Foster care	Care as usual	Not institution (family)	Foster care
Potential predictor:						
Quality of caregiving	x	x	x	x	Not reported	
Gender	x	x	x	x		
Cognitive development	✓	x	x	x	✓	✓

Girls were more responsive to *BEIP* than boys in terms of attachment outcomes. For the group remaining in institutional care, the secure / insecure status was not affected by caregiving quality, gender, nor cognitive development. However, for the groups in foster care, and with their families, cognitive development at 42 months of age was associated with an increase in the odds of a child having a secure attachment.

Age of placement

The age at which children were placed in foster care was strongly related to recovery of attachment. This was evident in both analyses of organised attachments (secure, avoidant, and ambivalent) and of secure attachment. The younger a child was placed in foster care, the more likely the child would develop an organised attachment at 42 months. This indicates decreasing plasticity of attachment with increasing age.

Brain white matter development

Unsurprisingly, neglect in early life is very detrimental to brain development. However, by the time of the study, the group placed into foster care has pretty much caught up with children who always lived with their families {except in “the body of the corpus callosum and superior corona radiata”.} The study suggests albeit cautiously that a potential for recovery in children exposed to extreme adverse conditions just after birth, and that early intervention *may support* recovery in the long-term.

The children were about two years old when *BEIP* started, i.e., when the foster care started. This result of catching up may only apply if the poor-quality institutional care ceases when the child is that young.

Adverse outcomes: None was reported.

Generalisability

BEIP covered the entire Bucharest area since children in all six institutional care facilities were included in the programme. It was implemented in Bucharest at a time when foster care was uncommon in Romania. Findings from this trial probably do translate to jurisdictions looking to ramp up support for fostering children.

What else is known about the intervention from other studies?

A recent study¹² which includes three meta-analyses found that family foster care led to consistently improved outcomes for internalizing behaviours (e.g., being withdrawn), externalising behaviours (e.g., acting out) and perception of care (i.e., attitudes towards care received) compared to residential care. Data from *BEIP* is also included in this study.

How reliable is the evidence?

These studies are rated as having a moderate risk of bias, with concerns on multiple domains:

Overall risk of bias	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns

Cells in which these studies appear:

The study relates to both treatment and child well-being:

- Intervention = Treatment; Outcome = Child Wellbeing: Physical Health (2 studies)
- Intervention = Treatment; Outcome = Child Wellbeing: Social-emotional Functioning (1 study)
- Intervention = Treatment; Outcome = Child wellbeing: Cognitive Functioning (3 studies)

¹² Li, D., Chng, G. S., & Chu, C. M. (2019). Comparing Long-Term Placement Outcomes of Residential and Family Foster Care: A Meta-Analysis. *Trauma, Violence, & Abuse, 20*(5), 653–664. <https://doi.org/10.1177/1524838017726427>

Edwards (2019) (Prevention / adult maltreatment behaviour)

Light cell(s) in which this study appears:

Intervention = Prevention; Outcome =Adult Perpetrator or Offender: Maltreatment Behaviour (1 study)

Summary: High school bystander programme improves students’ knowledge and attitudes on interpersonal violence but is inconclusive on bystander behaviour and violence prevention.

The summary is based on **Edwards et al. (2019)**¹³ ‘Evaluation of a bystander-focused interpersonal violence prevention program with high school students.’

Evidence status	High risk of bias	Weak evidence of impact on maltreatment behaviour (adult perpetrator and child safety) and knowledge/awareness (child wellbeing)
------------------------	--------------------------	--

The summary in brief

Bringing in the Bystander-High School Curriculum (BITB-HSC) is a classroom-based intervention, received by all students to improve knowledge and attitudes toward interpersonal violence and so reduce interpersonal violence and promote better bystander behaviour (i.e., so people can recognise and respond to aggression that they witness).

This study randomly assigned 25 schools in New England, USA into treatment and control groups, with 2,403 students taking part. The students were aged 13-19 with a mean age of 15.8 years. Half of students were female, and most were white.

The programme improved knowledge and some measures of attitudes. But there was no effect on most measures of bystander behaviour or interpersonal violence, and the effects that were observed mostly were not sustained when measured again a year later.

It is the only primary study on prevention of maltreatment behaviour by adult offenders on our EGM.

Type of study: RCT. Peer-reviewed journal article.

The intervention

Bringing in the Bystander (BITB) is a universal (i.e., not targeted) intervention developed to tackle interpersonal violence in colleges. The intervention studied here is an adaptation of the programme to a high school setting: *BITP-HSC*. The course comprises seven sessions of 45 minutes each delivered by a team of one man and one woman. The seven sessions include lectures, large and small group discussions, hands-on and experiential exercises, skill building activities, and video which covers stalking, sexual harassment, sexual assault, and dating violence. The programme also covers the bystander, framework, participants’ roles in creating a healthy community, how to recognise interpersonal violence, and how to intervene safely and effectively. In addition to student classes, *BITB-HSC* includes a 60-minute School Personnel Workshop that trains teachers and other school staff skills to be positive bystanders in situations of adolescent interpersonal violence.

¹³ Edwards KM, Banyard VL, Sessarego SN, Waterman EA, Mitchell KJ, Chang H. Evaluation of a Bystander-Focused Interpersonal Violence Prevention Program with High School Students. *Prev Sci.* 2019;20(4):488-498. doi:10.1007/s11121-019-01000-w

How is the programme meant to work? The theory of change

Bringing in the Bystander—High School Curriculum is grounded in several different behaviour change theories: the health belief model, transtheoretical model of change, theory of planned behaviour and theory of how innovation diffuses (spreads) across populations.

The core of the approach is to create knowledge and awareness of the nature of interpersonal violence, and change attitudes by challenging rape myth and other false beliefs about sexual assault victims and perpetrators. Once the problem is acknowledged, participants are helped to achieve 'bystander readiness' to support positive behaviours, such as reactive behaviours (e.g., speaking up for someone) and proactive behaviours such as talking about prevention or using social media proactively (e.g., blogging about one's unsupportiveness of violence).

Has the intervention been implemented at scale?

BITB has been implemented amongst young adults at hundreds of colleges and universities in the United States, as well as being adapted for other populations such as the military. This study is the first trial of *BITB* for high school students based on pilot research. The study does not say if the programme is delivered by teachers from the school or not, which is the likely approach to take *BITB* to scale, but which may be less effective than if delivered by members of the *BITB* team.

What does the intervention cost?

The study does not report any data on cost.

The trial

Participants: Randomisation to *BITB-HSC* or control was at the school level and not at the individual student level. 25 schools in the US (New England region) participated in the trial. 2403 high school students (grades ninth to 12th) with a mean age of 15.8 years were included. Students were mostly white (85%), mostly heterosexual (84.5%) and half were girls.

Study design: Details on study design are sparse. One group of schools received the *BITB-HSC* intervention while the other group served as controls. The number of schools in each group is unclear.

Outcomes: The outcomes measured were:

- Violence victimisation – being a victim of interpersonal violence
- Violence perpetration – causing interpersonal violence to someone else
- Students' knowledge – knowledge on different aspects of interpersonal violence
- Rape myth acceptance – to assess students' agreement with rape myths
- Relationship media literacy – to measure students' discomfort with media portrayals related to interpersonal violence
- Bystander behaviour – to assess students' actions during or after witnessing a situation with interpersonal violence

- Bystander readiness – to measure students’ agreement with various statements denying that they could play a role in preventing interpersonal violence as bystanders
- Barriers and facilitators of bystander helping – to understand what the barriers and facilitators for bystanders to intervene are
- Victim empathy – to measure the level of concern and understanding of a victim’s situation

The study measured outcomes at four times: before the programme started, a few days after implementation, a few months later and more than a year later. Due to differences in school calendars and weather-related disruptions, there was a wide range for when each of these measurements happened across schools. For example, the average duration for the measurement a few days after programme implementation was 44 days but it ranged from 21 days to 109 days across schools. By the time of the final measurement (more than a year after the programme), many students had dropped out of the study.

Did the intervention work?

The intervention had positive effects on knowledge and attitudes, but much less on bystander behaviour and interpersonal violence. Specifically:

- ✓ **Interpersonal violence:** No effect was found on sexual assault and dating violence from the perspective of either victim or perpetrator, when measured two months or one year after the programme. No effect was found on sexual harassment and stalking from the point of view of the victim when measured two months or one year after the programme. There was a significant reduction in perpetration when measured two months afterwards but that was not sustained when measured a year later.
- ✓ **Reactive bystander behaviour:** There were no significant effects on participants stopping harassment, speaking against blame or excuses, talking to an upset person, or getting help for a friend at either two months or one year. There was an improvement in talking to a hurt friend at two months, but the effect disappeared after one year.
- ✓ **Proactive bystander behaviour:** There were no significant effects on prevention talk or talking about safety after two months or one year. Use of social media in recent months improved after one year, which was not observed two months after the intervention.
- ✓ **Knowledge and attitudes:** There were significant improvements in knowledge, rape denial, being bothered by the media’s portrayal or relationship abuse and sexual assault, and accepting a possible role for bystanders in preventing abuse and assault after two months which was still observed after one year. Improvements in positive attitudes to helping, bystander barriers, and victim empathy were seen after two months but not after one year. There was no improvement in traditional gender expectations at either two months or a year.
- ✓ **Rape myth acceptance:** Rape denial significantly decreased for intervention students compared to control. However, traditional gender expectations were not different between groups.
- ✓ **Media literacy:** Media literacy significantly improved over time for intervention students over control group students.
- ✓ **Bystander readiness:** Intervention group students demonstrated a significant sustained decrease in denial on bystanders’ role in preventing interpersonal violence compared to the control group.

- ✓ Barriers and facilitators of bystander helping: Positive attitudes on helping increased significantly for the intervention group students compared to controls a few months after intervention. However, when measured more than a year later, the difference between the groups was not significant.
- ✓ Victim Empathy: Initially, victim empathy increased significantly for the intervention group compared to controls, but the difference was not sustained after a year.

Adverse outcomes: None was reported.

Generalisability

BITB has been widely implemented in the US in college and university settings. This trial aimed to pilot this approach for high school students. We cannot say much about generalising these findings for high school settings because so few details are provided about implementation.

What else is known from other studies about the intervention?

A Campbell systematic review¹⁴ on bystander programmes summarises evidence from 27 studies of programmes for adolescents and college students. There are positive effects on some but not all measures of student knowledge of and attitudes to sexual assault. There may be an effect from bystander programmes on bystander’s intervening, but there is no effect on sexual assault by perpetrators

How reliable is the evidence?

The study is rated as having a high risk of bias, with concerns on many domains:

Overall risk of bias	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
High Risk	Some concerns	Some concerns	High risk	High risk	Some concerns

Cells in which this study appears:

The study relates to both prevention and response:

- Intervention = Prevention; Outcome = Child Safety: Maltreatment Behaviour (13 studies + 2 Systematic Reviews)
- Intervention = Prevention; Outcome = Child Wellbeing: Knowledge or Awareness (51 studies + 10 Systematic Reviews)
- Intervention = Prevention; Outcome = Child Perpetrator or Offender: Maltreatment Behaviour (1 study)

¹⁴ Kettrey, HH, Marx, RA, Tanner-Smith, EE. Effects of bystander programs on the prevention of sexual assault among adolescents and college students: A systematic review. *Campbell Systematic Reviews*. 2019; 15:e1013. <https://doi.org/10.4073/csr.2019.1>

Herbert, Bromfield (2021) (Disclosure; Operational Practice)

Skinny cell(s) in which this study appears:

Interventions = Disclosure, Response, Treatment; Outcomes =Operational Practice, Child Maltreatment Occurrence, Adult Maltreatment Behaviours

Summary: Primary school-based universal child sexual abuse prevention programme, which improves children’s knowledge, response, and recognition of CSA.

The summary is based on **Herbert and Bromfield (2021)** ‘A quasi-experimental study of the Multi-Agency Investigation & Support Team (MIST): A collaborative response to child sexual abuse’.

Evidence status	Some concerns	Moderate evidence of impact on reducing investigation and other process times but rates of disclosure and interviewing potential abuser low
------------------------	----------------------	---

The summary in brief

MIST (Multi-Agency Investigation & Support Team) is a cross-agency approach to responding to child abuse, especially sexual abuse happening within families. It is based on the Child Advocacy Centre (CAC) models seen in the US. The team is composed of different professionals who typically respond to abuse reports such as police, child protection workers, therapists, child and family advocates and specialist interviewers. The team works on each case from the same location after an initial strategy meeting.

MIST was implemented to process all cases during the study period from one suburb of the Australian city of Perth. Cases from the remaining part of the city went through the usual practice of an inter-agency response with different agency professionals launching their own investigations and interventions without necessarily conferring with each other.

The findings suggest that MIST offers modest benefits. It reduced the processing time, especially police investigation time. However, there was no effect on the number offenses substantiated. Disclosure rates and rates of interviews with a person of interest were higher in the usual practice group. Caregivers expressed a high level of satisfaction with MIST but only a small percentage completed the survey.

Type of study: QED. Peer-reviewed journal article.

The intervention

MIST (Multi-Agency Investigation & Support Team) is a type of “planned response” to child abuse cases with simultaneous police and child protection investigations. The aim is to create a response that is collaborative and based on inter-agency planning and teamwork. MIST was established in Armadale, a suburb of Perth Australia.

A team of detectives, interviewers, a child protection worker, two Child and Family Advocates, and a therapy team was brought together. The team was “co-located” in a facility that was specifically built to investigate and support child abuse cases. This collaborative approach allows for better support of families and coordination of the investigation. The same personnel in the team continue working with the family as the case progresses. The usual practice included many of these same professionals, but the response was fragmented with different departments engaging with the case and the family at different times. Access to support services also depends on the resources available to the family.

The essential differences between MIST and usual practice were:

“(a) the co-location of interviewers, detectives, child protection workers, support workers and therapists in a built for purpose facility to work as a multi-agency team;

(b) the availability of a Child and Family Advocate to provide ongoing support to children and families and connection to therapeutic services;

(c) Multi-Disciplinary Team (MDT) case review meetings between team members to plan responses to children/families.

(d) a co-located therapy team resourced to provide extended trauma therapy to MIST cases.”

How is the programme meant to work? The theory of change

The study does not mention a specific theory on which the programme is based. MIST is generally based on the Children’s Advocacy Centre (CAC) approach used widely in the US that includes “Multi-Disciplinary Team (MDT) case review meetings, forensic interviewing, and child advocacy” that is widely prevalent in the US.

Has the intervention been implemented at scale?

MIST was available to all cases in the city of Perth, Australia. It is not clear if MIST continued after the study.

What does the intervention cost?

The study does not report any data on cost.

The study

Participants: All cases in Perth that met the criteria for a specialist police/child protection response between October 2015 and July 2016 were included in the analysis.

Study design: The study used a quasi-experimental design to compare two groups:

1. MIST (126 cases) and
2. practice as usual (276 cases).

The authors used a ‘follow-forward’ approach to study the effects of the two interventions. Depending on where participants lived in Perth, they either received MIST (one suburb) or practice as usual (rest of the city).

Objectives: The study objectives were:

1. *“Was MIST delivered with fidelity to the agreed operating procedure:*
 - a. *Was the agreed process reflected in the investigative response?*
 - b. *Was the agreed process reflected in the support and therapeutic response?*
 - c. *Was the MIST team resourced to respond to the volume of cases they saw?*
2. *Were caregivers satisfied with the response they received from the MIST?*

3. Did MIST significantly differ from Practice as Usual on:

- a. Rate of disclosures, interviews with persons of interest, and arrests?
- b. Number of offences and number of contact offences charged?
- c. Referral for a forensic medical examination?
- d. Days from intake to a strategy meeting, to the allocation of a detective, to a victim interview, to an interview with a person of interest, and to the point of case closure?
- e. Rate of Child Protection Investigations, Substantiations, and Further Child Protection action?
- f. Days from intake to a strategy meeting, opening a Child Protection Investigation, and closure of a Child Protection Investigation?"

Did the intervention work?

MIST was not better than usual practice except for a few outcome measures.

- The rate of disclosure (51.6% vs. 64.9%) and interview with a person of interest (57.1% and 68.8%) were both lower for MIST.
- The proportion of offences substantiated was similar for both groups (32.5% vs. 34.3%).
- Child protection measures, i.e., assessing safety, wellbeing and offering support, were not different.
- The number of offences charged was also not significantly different between groups.
- Usual practice had a higher, albeit small, rate of referral to forensic medical examination vs. MIST (9% vs. 2%).

MIST was significantly faster in the timeliness of response across many stages, especially the police investigation.

- Days from intake to strategy meeting: MIST 1 day vs. usual practice 1 day (*not statistically significant*)
- Days from intake to allocation: MIST 2 days vs. usual practice 14 days
- Days from intake to specialist child interview: MIST 5 days vs. usual practice 10 days
- Days from intake to person of interest interview: MIST 28 days vs. usual practice 86.5 days
- Days from intake to police investigation close: MIST 44 days vs. usual practice 86 days
- Days investigation was open: MIST 31.5 days vs. usual practice 71 days
- Days from child protection interaction to intake to a safety and wellbeing assessment: MIST 1 day vs. usual practice 3 days (*not significant*)
- Days from child protection interaction to safety and wellbeing assessment closed: MIST 81 days vs. usual practice 89 days (*not significant*)
- Days safety and wellbeing assessment was open: MIST 76 days vs. usual practice 77 days (*not significant*)

Caregivers generally had high levels of satisfaction with MIST although few responded to the satisfaction survey (24%). The items receiving the highest satisfaction scores were staff behaviour, children feeling safe, and having their questions answered by the MIST team. Getting updates on the case were rated the lowest although most respondents were reasonably satisfied.

How reliable is the evidence?

The study is rated as having a low risk of bias, with few concerns:

Overall risk of bias	Confounding	Selection bias	Intervention classification	Deviation in intervention	Missing data	Outcome measurement	Reporting bias
Low	Low	Low	Low	Low	Low	Low	Low

Cells in which this study appears:

The study relates to disclosure, response, and treatment:

- Intervention = Disclosure; Outcome = Operational practice
- Intervention = Response; Outcome = Operational practice
- Intervention = Treatment; Outcome = Operational practice
- Intervention = Disclosure; Outcome = Child Maltreatment Disclosure
- Intervention = Response; Outcome = Child Maltreatment Disclosure
- Intervention = Treatment; Outcome = Child Maltreatment Disclosure
- Intervention = Disclosure; Outcome = Child Maltreatment Occurrence/Recurrence
- Intervention = Response; Outcome = Child Maltreatment Occurrence/Recurrence
- Intervention = Treatment; Outcome = Child Maltreatment Occurrence/Recurrence
- Intervention = Disclosure; Outcome = Adult Perpetrator: Maltreatment Behaviours
- Intervention = Response; Outcome = Adult Perpetrator: Maltreatment Behaviours
- Intervention = Treatment; Outcome = Adult Perpetrator: Maltreatment Behaviours

Hoefnagels (2021) (Disclosure; Adult Institutional Care Provider)

Skinny cell(s) in which this study appears:

Intervention = Disclosure; Outcome = Adult Institutional Care Provider

Summary: Disclosure-focussed intervention improves primary school students’ recognition and talkability on child abuse and neglect, but study rated high risk of bias.

The summary is based on **Hoefnagels et al. (2021)** ‘Changing the classroom climate to lower the threshold for child abuse and neglect self-disclosure: a non-randomized cluster-controlled trial.’

Evidence status	<i>High risk of bias</i>	Child recognition and talkability on abuse and neglect and teacher reporting self-disclosure by students improves but study has high risk of bias
------------------------	--------------------------	---

The summary in brief

The intervention aimed to set the conditions for children to disclose abuse and encourage / enable other children to do so. It targeted primary school children (10-12 years) in the Netherlands and aimed to improve (a) children’s recognition of child abuse and neglect (CAN) and (b) “talkability” on CAN (i.e., children’s ability and willingness to talk about it). The idea was that encouraging students to talk about CAN would eventually lead to increased disclosure and hence reduced child maltreatment.

The intervention was delivered by the students’ regular primary school teachers. It targeted schools where parental educational levels were low. Teachers were trained in the intervention, which consisted of four interactive lessons, along with video clips and TV content on CAN. The lessons focussed on various types of abuse: physical, emotional and neglect. The nine schools in the control group continued their usual practices.

The intervention significantly improved both students’ recognition of CAN and talkability. Children from Dutch origin families were likely to do better at recognizing CAN than those from non-Western migrant backgrounds. Students of teachers that had experience talking to parents about CAN also were more likely to recognize CAN.

Talkability was more likely when children perceived it as a social norm and when they used social coping strategies to process their feelings.

The study also recorded actual disclosures. Teachers in the intervention group reported more disclosures by children during the intervention month (five) compared to the control group (one).

Type of study: QED (non-randomised cluster-controlled trial, i.e., the ‘unit’ was schools (who either got the intervention or didn’t) rather than individual classes or students). Peer-reviewed journal article.

The intervention

The intervention ran in 2013. An instructional program was delivered by the students’ regular teacher. It featured a preliminary session and four CAN lessons, along with materials such as an informational flyer for parents, posters in classrooms, and curriculum information for teachers. Teachers were encouraged to be well prepared, due to the delicate subject matter, the potential responses from students, and their own emotional responses to CAN. They were also given access to resources for assistance within and outside their schools and on steps on how to act if any of their students disclose CAN. Teachers also had access to written materials on the topic to improve their knowledge and awareness.

The lessons for children included:

- A preparatory lesson on “personal moments of happiness” with their parents or caregivers
- Lesson one was on neglect
- Lesson two on physical abuse
- Lesson three on seeking help
- Lesson four involved the development of a logo on the theme “Child maltreatment: you can do something about it”.

Lessons were interactive, with video clips and depicted child maltreatment situations (by actors). Teachers paused at multiple points to discuss with their students on how they felt.

How is the programme meant to work? The theory of change

The authors mention that based on social cognitive theory when disclosure about CAN is considered as positive from a social norms perspective this can contribute to lowering the threshold on ‘talkability’ of these delicate topics. Children are more likely to disclose when they feel they will be believed.

Has the intervention been implemented at scale?

No. Only 21 of nearly 7000 primary schools in the Netherlands were part of the study.

What does the intervention cost?

The study does not report any data on cost.

The study

Participants: Primary school students aged 10-12 years, across the Netherlands. The schools (a) were selected to have low educational levels of parents and (b) opted in. Recruitment was stopped when 35 schools agreed to participate. In the end, 21 schools were in the study.

Study design: The study had two groups:

- a. Nine schools were the control group and did not implement the intervention.
- b. The remaining schools ran the intervention.

Outcomes: The main outcomes measured were:

1. CAN recognition: Students were shown vignettes of children in a bad situation and asked “Do you recognize this as maltreatment?” with “Yes, No or Maybe” as the possible answers.
2. CAN talkability: Children were asked “Did you talk about child maltreatment after school in the last month?” with a “Yes or No” response solicited.

The study measured outcomes before the intervention began and immediately after it ended.

The number of actual disclosures during the period was also recorded. At the beginning of the study, nearly half of the children were not sure if other children were comfortable talking about CAN.

Did the intervention work?

The intervention improved both students’ recognition of CAN and talkability. It appears also to have increased disclosures.

- **A: Recognition of CAN:**

- The odds of children in the intervention group recognizing the vignettes as related to CAN were more than three times those for children in the control group.
 - The intervention group children had a higher chance (2-3 times) of recognizing physical and emotional abuse compared to neglect.
 - Children of Dutch origin were twice as likely to recognize CAN than were children from non-Western immigrant families.
 - Many factors did not affect students’ recognition of CAN. They include:
 - Children’s views on whether the child was to blame, views on whether the teacher would believe their disclosure, whether another student had disclosed to them previously, whether the child would inform someone if they saw what was happening in the vignette happening to another child, perceived social norms, level of parental education, children’s sex, and ethnicity.
 - The teacher’s beliefs on CAN, the teacher’s experience dealing with CAN allegations, or previous interaction with youth medical services regarding CAN.
 - However, students of teachers who had more contact with parents regarding CAN were more likely to recognize CAN.
- **B: Talkability of CAN:**
- Similar to recognition, children who received the intervention were nearly three times more likely to talk about CAN than children in the control group.
 - Children who perceived talkability as a social norm and used coping skills (measured using the ‘*Children’s Coping Strategies Checklist*’) to address their feelings, had a higher odds of talking about CAN. The other variables assessed were not significant.
 - Only one teacher variable seemed to affect children’s talkability on CAN: “whether they (teachers) had contacts with youth medical services regarding CAN”. The authors also report “tentative evidence” that talkability promotes self-disclosure.
 - On actual disclosures, five teachers in the intervention group reported children confiding in them on CAN during the study period compared to only one teacher in the control group.

Generalisability

No information is provided about this, and it is not discussed in the study report.

How reliable is the evidence?

The study is rated as having a high (serious) risk of bias, with concerns on multiple domains:

Overall risk of bias	Confounding	Selection bias	Intervention classification	Deviation in intervention	Missing data	Outcome measurement	Reporting bias
Serious	Low	Moderate	Serious	Low	Low	Low	Low

Cells in which this study appears:

The study relates to disclosure:

- Intervention = Disclosure; Outcome = Disclosure: Child Maltreatment Disclosure
- Intervention = Disclosure; Outcome = Child Knowledge and Awareness

Karmaliani (2020) (Prevention; Adult Perpetrator: Maltreatment Behaviours)

Skinny cell(s) in which this study appears:

Intervention = Prevention; Outcome = Adult Perpetrator: Maltreatment Behaviours

Summary: Play-based intervention delivered in secondary schools can lead to less violent discipline against children.

The summary is based on Karmaliani et al. (2020) ‘Right to Play’s intervention to reduce peer violence among children in public schools in Pakistan: a cluster-randomized controlled trial.’

Evidence status	Low risk of bias	A manualized play-based intervention delivered by coaches in school can reduce the incidence of harsh punishment experienced by children at school and home.
------------------------	------------------	--

The summary in brief

A play-based intervention using a standardised manual delivered by trained coaches in secondary schools in Pakistan is evaluated in this RCT. The intervention is developed by Right to Play, an international NGO that focuses on play-based programmes to improve outcomes for children. After each play activity session, children are encouraged to reflect on how the activity makes them feel and to relate it to their lives at school and home. The intervention in this case was delivered to 20 public middle schools in a large district in Pakistan; 20 other schools which did not receive the intervention were the control group.

Intervention activities lasted two years, during which over 120 sessions were delivered to students. Parents and teachers received regular information sessions on child rights, gender equity, and positive discipline.

Children in the programme said they experienced less violent discipline by their teachers and at home than did children in the control group. The reductions were more pronounced for girls. Peer violence went down for intervention participants – both committing violence and being a victim. Depression outcomes also improved for students in the intervention compared to the control group.

Type of study: RCT. Peer-reviewed journal article.

The intervention

Right To Play is an international non-profit organization that uses sports and play-based programmes to empower children and youth facing adversity. The organization aims to promote holistic development, including physical, cognitive, social, and emotional skills, through engaging activities and games. Right To Play operates in various countries around the world, working with local communities and partners to deliver programs that help children learn essential life skills, build resilience, and improve their overall well-being. Their interventions often focus on areas such as education, child protection, gender equality, and health promotion.

In this study, coaches employed by Right to Play delivered the intervention to students in grades 6-8 (ages 12-14) attending public middle schools in the Hyderabad district of Sindh province in Pakistan. The intervention includes 103 play-based learning activities. Coaches pick an activity, each with its own goal, from a manual. Once the play is complete, a three-step discussion called 'Reflect-Connect-Apply' takes place. Students are encouraged to talk about how the activity made them feel and to apply it to other aspects of life.

The intervention was implemented over a two year period with 120 sessions for each class: two sessions on average every week. There were also information sessions for parents and training sessions for teachers on topics such as child rights, gender equality, and positive discipline.

The trial

Approximately 1,750 sixth, seventh, and eighth graders from 40 schools participated in the intervention or were in the control group. There were more girls than boys in each of the groups. Students were surveyed for outcomes before the intervention began, six months into the programme, one year, and at two years when the programme ended. Data from almost 80% of students were available at all the time points.

The researchers examined if the intervention reduced peer-violence (both victimization and perpetration) and depression among students. They also measured other outcomes such as how often corporal punishment and punishment at home happened (as reported by students).

How is the programme meant to work? The theory of change

No specific theory for this intervention is mentioned. However, the program uses various theories in areas such as social, cognitive, child development, and experiential learning. Its main emphasis is on promoting physical, cognitive, social, and emotional development through sports and games. The program is based on the belief that children learn by exploring and reinforcing new ideas and behaviours, and for long-lasting effects to occur, interventions need to be continuous. An important aspect of the program is its commitment to promoting gender equality, which is incorporated throughout and aims to empower girls by providing opportunities for participation and discussion.

Did the intervention work?

Yes. Students in the intervention group reported statistically significantly less corporal punishment (a secondary outcome in the study) at school compared to students in the control group. Girls reported a larger drop in experiencing corporal punishment. Harsh discipline at home (another secondary outcome) over the last few weeks before the survey was also reduced.

The primary outcomes were peer violence victimization, peer violence perpetration, and depression for boys and girls. The study shows girls benefitted far more than boys through the program. The reductions in peer violence victimisation in the intervention and control groups respectively were 33% against 28% for boys and 58% versus 21% for girls. In the intervention and control arms, the reductions in peer violence perpetration were 25% versus 11% for boys and 56% versus 28% for girls. The mean depression ratings decreased for girls (9.5% vs. 5.6% intervention and control) and boys (7.2% vs. 4.8% intervention and control).

Adverse outcomes: There were no reports of serious adverse events.

Has the intervention been implemented at scale?

Not quite. Although 40 out of 55 secondary schools in a major city district were enrolled, this was principally designed as a research study.

What does the intervention cost?

The study does not report any data on cost.

Generalisability

Likely generalisable to other secondary schools in South Asia, other LMICs, and even globally. This intervention is structured and manualized, so it could potentially be implemented in most secondary school settings around the world. The programme has already been translated and adapted for different settings. In Pakistan, where the NGO has worked since 2008, the material has been translated into Urdu and Sindhi. Although the programme has been tested and refined over 20 years across 18 countries, this was the first time that the programme was evaluated for long-term impact on children, i.e., two-years.

How reliable is the evidence?

Very. The study is rated as having a **low risk of bias**.

Overall risk of bias	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Cells in which this study appears:

The study relates to both prevention and disclosure:

- Intervention = prevention; Outcome = Child Safely: child maltreatment occurrence/recurrence
- Intervention = prevention; Outcome = Child wellbeing: child mental health
- Intervention = prevention; Outcome = Adult perpetrator: maltreatment behaviours
- Intervention = prevention; Outcome = Child/youth offender: maltreatment behaviours

Merrill (2018) (Prevention / institutional culture)

Light cell(s) in which this study appears:

Intervention = Prevention; Outcome = Institutional Safeguarding Practice: Culture (2 studies + 1 protocol)

Summary: Primary school-based universal physical violence prevention in Uganda programme led to some improvements in school operational culture and normative beliefs for students. Results need to be interpreted with caution due to study's high risk of bias.

The summary is based on **Merrill et al. (2018)**¹⁵ 'Effects of a violence prevention intervention in schools and surrounding communities: Secondary analysis of a cluster-randomised controlled trial in Uganda.'

Evidence status	High risk of bias	Weak evidence of impact on maltreatment behaviour (child safety) knowledge or awareness (parent caregiver) and culture (institutional safeguarding practice).
------------------------	--------------------------	---

The summary in brief

The *Good School Toolkit (GST)*¹⁶ is a universal programme that aims to prevent physical violence from school staff towards primary school children (aged 11-14). Developed by Raising Voices, a non-profit in Uganda, *GST* is a comprehensive behavioural intervention that involves students, school staff, school administrators and parents. The toolkit outlines a six-step process for a school to achieve "Good School" status, i.e., a violence-free environment where students can thrive. A few staff members and students from each school are recruited to serve as "protagonists" and trained to lead their peers through the programme. *GST* involves several activities for schools such as student debates, painting murals and displaying codes of conduct in prominent areas. Schools develop action plans to achieve programme goals. Raising Voices team members provide ongoing support to "protagonists." Overall, *GST* aims to improve school culture.

21 primary schools from Luwero District in Uganda were randomly assigned to *GST* and 21 others were controls.

There are multiple papers¹⁷ of this RCT in the EGM. This paper focuses on school operational culture, normative beliefs and violence against children at home.

Some school operational culture outcomes improved for the *GST* group compared to controls such as: emotional support for students from their teachers and peers; reduced acceptance by both students and teachers of physical discipline in school; students identifying with their schools; both student and staff perceptions of involvement in school operations.

¹⁵ Merrill KG, Knight L, Namy S, Allen E, Naker D, Devries KM. Effects of a violence prevention intervention in schools and surrounding communities: Secondary analysis of a cluster randomised-controlled trial in Uganda. *Child Abuse Negl.* 2018;84:182-195. doi:10.1016/j.chiabu.2018.06.007

¹⁶ <http://raisingvoices.org/good-school/download-good-school-toolkit/>

¹⁷ Devries 2015, Devries 2017, Devries 2018, Knight 2018, Merrill 2018,

GST children indicated a reduced acceptance of physical discipline at school and at home compared to controls. However, no difference was seen in child-reported or caregiver-reported physical or emotional violence at home. The prevalence of corporal punishment at the end of the programme was still high (30%).

These findings warrant caution because many of the scales used to measure outcomes are not validated. Further, the study is categorised as high risk of bias reducing confidence in results.

Type of study: RCT. Peer-reviewed journal article.

The intervention

The *Good School Toolkit (GST)* was developed by Raising Voices, a Ugandan non-profit committed to preventing violence against women and children. Physical violence by school staff against children is common in Uganda. Data from one Ugandan district shows almost all children (aged 11-14) report some form of physical violence from school staff in their lifetime. Caning is the most prevalent form reported but extreme acts such as being choked, burned, stabbed and severely beaten up also happen (8% of students surveyed).

GST's purpose is to help create a "Good School: a school which aims to create a violence-free learning environment within which students develop their skills and confidence to grow into creative, constructive and thoughtful members of society."

GST's objectives are to (i) help teachers increase their students' confidence and success (ii) develop a learning environment that is safe and respectful and (iii) facilitate transparency and accountability among school administrators. *GST* includes six steps to achieve these objectives. The estimated completion time for all six steps is two years but schools set their own pace.

Raising Voices takes the lead on introducing *GST* to a school. Two staff members and two students are then recruited to serve as "protagonists" i.e., their role is to engage and motivate their colleagues, classmates, school administrators and parents to develop schoolwide action plans and goals. Raising Voices provides a three-day residential workshop for "protagonists" from different schools. One-on-one support for "protagonists" to help them implement school action plans continues throughout the program.

GST comes with posters, booklets, and facilitation guides for more than 60 group activities over six steps. Examples of activities include student debates, painting murals or hanging codes of conduct in a prominent place. The activities promote a better learning environment, mutual respect, empowering students to make decisions, insight into power relationships in the school setting, using nonviolent discipline, better classroom management and improved school governance.

Setting school-wide goals, developing action plans with specific dates and deliverables, promoting empathy, non-violent techniques for discipline and most importantly an opportunity to practice these new behaviours is at the core of *GST*. Children, school staff, administrators and parents all have roles to play for effective implementation.

Overall, *GST* aims to change culture to one of empathy, positivity, and non-violence.

How is the programme meant to work? The theory of change

GST is based on the Transtheoretical Model¹⁸ that theorises that behaviour change goes through “six stages of change: precontemplation, contemplation, preparation, action, maintenance and termination.” This model has been employed widely in other areas such as quitting smoking, reducing domestic violence and teacher behaviours in the classroom.

Has the intervention been implemented at scale?

No. *GST* was implemented in only one district in Uganda. A *Good School Toolkit Secondary* is currently being piloted for secondary schools.

What does the intervention cost?

Cost data is reported in a study by Greco et al¹⁹. Implementing *GST* over 18 months in 21 schools was close to \$400,000. Monitoring and evaluation add another \$50,000 to costs. The annual cost to run *GST* was approximately \$7500 per school and \$15 per student. It costs close to \$250 to prevent a case of violence and approximately \$100 in annual implementation costs for every prevented case. *GST* was found to be cost-effective.

The trial

Participants: 268 primary schools from Luwero district in Uganda were initially considered. Schools with less than 40 students or with an ongoing intervention were excluded. From the remaining schools, 42 were chosen randomly for the trial. These 42 schools were in turn randomised to either *GST* or control.

Most schools were in rural locations and more than half of students reported recent violence (“within the last week”) and eating two or fewer meals in the previous day. While entire schools participated in *GST* only students from classes 5th-7th (ages 11-14) were surveyed for outcome data along with all school staff. Over 1800 students in each arm were assessed. The mean age of students was 13 years with just over half girl students. 7% reported having some disability and 1 in 5 students reported a school absence in the previous week. 600 school staff were surveyed, the average age was mid-30s with close to 60% female. Approximately two-thirds belonged to the Baganda tribe. Nearly 800 caregivers were also surveyed but only after *GST* ended.

Study design: The trial had two groups:

- a) 21 schools got *GST* implemented over 18 months (September 2012 to April 2014)
- b) 21 schools did not receive *GST*

Students and school staff were surveyed a couple of months before *GST* was launched, and a few months after it ended. Caregivers were only surveyed at the latter time point.

¹⁸ Prochaska, J. O., & Velicer, W. F. (1997). The Transtheoretical Model of Health Behavior Change. *American Journal of Health Promotion*, 12(1), 38–48. <https://doi.org/10.4278/0890-1171-12.1.38>

¹⁹ Greco G, Knight L, Ssekadde W, et al Economic evaluation of the Good School Toolkit: an intervention for reducing violence in primary schools in Uganda. *BMJ Global Health* 2018;3:e000526.

Outcomes: Several papers assessing various outcomes from this one trial have been published. This paper (Merrill 2018) looks at three main outcome areas: school operational culture, normative beliefs, and violence against children at home.

The outcomes measured were:

- School operational culture was looked at across different domains using Likert scales, i.e., rating from low to high for each question which is then converted to a numeric score.
 - a. Relational: Includes emotional support for students from their teachers and peers; school staff’s perceived relationship with students, their colleagues, and caregivers; and caregivers’ perceived relationship with staff.
 - b. Psychological: Includes student and staff identification with their school; student and staff acceptance of physical discipline in school; and student acceptance of sexual violence from schoolteachers.
 - c. Structural: Includes student, caregiver and staff perceptions of involvement in school operations
- Normative beliefs were measured surveying caregivers using Likert scales on acceptability of physical discipline and sexual violence from teachers at school and physical discipline at home
- Violence against children at home: Student and caregiver self-reported physical and emotional violence at home in the past week measured based on The International Society for the Prevention of Child Abuse and Neglect Screening Tool – Child Institutional (ICAST-CI)²⁰ and the WHO Multi-Country Study on Women’s Health and Domestic Violence against Women (WHO MCS)²¹

Did the intervention work?

- ✓ **School operational culture outcomes for GST compared to control:**
 - Relational: Students felt significantly more emotional support from teachers and peers
 - Psychological: Student and staff acceptance of physical discipline in school reduced and student identification with their school improved significantly.
 - Structural: Student and staff perceptions of involvement in school operations improved significantly.
 - All other outcomes were similar between groups (not statistically different).
- ✓ **Normative beliefs:** GST caregivers’ acceptability of physical discipline at school and at home reduced significantly compared to controls. Acceptability of sexual violence from teachers at school did not improve for the GST group although the scores for both groups were similar (and acceptability was already low). Importantly, caregivers were not surveyed before the programme, so we do not really know how their beliefs changed over time.

²⁰ International Society for the Prevention of Child Abuse and Neglect. 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, 2006.

²¹ Garcia-Moreno, C., Ellsberg, H., Heise, L., & Watts, C. (2005). WHO multi-Country study on women’s health and domestic violence against women.

No difference seen for child-reported and caregiver-reported physical and emotional violence against children at home between the two groups.

It is difficult to quantify the size of the impact because different ranges of Likert scales were used for different outcomes. For example, the range for one outcome extends from 0-3 but for another it is 0-12. This makes it difficult to interpret the size of effect. Most of these scales have not been validated which further undermines results. Overall, flaws in the how the study was carried out reduce one’s confidence in the reported results.

Adverse outcomes: No adverse effects are reported in this study. However, as reported in Devries (2015) over 400 children were referred to child protective services based on their disclosures during the follow-up survey. This is something that can be expected when maltreatment prevention interventions are implemented. Participating in the intervention or even being asked questions on experiencing violence and abuse can empower children to disclose an ongoing or past abuse.

Generalisability

This paper does not discuss generalisability. Devries (2015) which is also written on this trial does state that results should be generalisable to other African settings. However, this trial was only conducted in one district of Uganda. More studies would be needed in other parts of Uganda and Africa to get a better grasp on generalisability. This intervention will likely be effective in settings where violence against students by teachers is common (as was the case in this trial).

What else is known about the intervention from other studies?

There seem to be no systematic reviews on the effectiveness of preventing corporal punishment by school staff. Systematic reviews of school-based violence prevention tend to focus on topics such as anti-bullying, sexual abuse prevention or intimate partner violence prevention.

The papers about this RCT describe it as the first study of an intervention to reduce violence by staff against pupils.

How reliable is the evidence?

The study is rated as having a high risk of bias, with concerns in many domains:

Overall risk of bias	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
High Risk	Low Risk	Some concerns	Some concerns	High risk	Some concerns

Cells in which this study appears:

The study relates to prevention:

- Intervention = Prevention; Outcome = Parent Caregiver: Knowledge or Awareness (4 studies + 1 protocol)
- Intervention = Prevention; Outcome = Institutional Safeguarding Practice: Culture (2 studies + 1 protocol)
- Intervention = Prevention; Outcome = Child Safety: Maltreatment Behaviour (10 studies + 2 systematic reviews + 3 protocols)

It is one of five studies on the *Good School Toolkit* RCT conducted in Uganda on our EGM.

Nkuba (2018) (Prevention / institutional culture)

Light cell(s) in which this study appears:

Intervention = Prevention; Outcome = Institutional Safeguarding Practice: Culture (2 studies + 1 protocol)

Summary: Training workshop for secondary school teachers on violence prevention in Tanzania reduced self-reported violence by teachers and improves teachers' attitudes away from corporal punishment. Results need to be interpreted with caution due to study's high risk of bias.

The summary is based on **Nkuba et al. (2018)**²² 'Reducing violence by teachers using the preventative intervention *Interaction Competencies with Children for Teachers (ICC-T)*: A cluster randomised controlled trial at public secondary schools in Tanzania.'

Evidence status	High risk of bias	Weak evidence of impact on maltreatment behaviour (child safety) behaviours, attitude, or knowledge (adult institutional caregiver) and culture (institutional safeguarding practice).
------------------------	--------------------------	--

The summary in brief

Interaction Competencies with Children for Teachers (ICC-T) is a training programme for teachers aimed at preventing use of corporal punishment by teachers and to improve teacher-student relationships. *ICC-T* has been successfully tried with approximately 30 institutional caregivers and 30 primary school teachers in Tanzania. In this study, *ICC-T* is evaluated for nearly 160 secondary school teachers.

ICC-T includes a week-long workshop for schoolteachers. Teacher participation is voluntary, and training is led by a psychologist. The focus is on improving teacher-student relationships, understanding the needs of students better, increasing awareness of non-violent methods of discipline and thinking about how to implement new skills at school. The workshop is meant to be interactive to encourage teachers to share their views and experiences on corporal punishment. Opportunities are also provided for teachers to practice new skills learned at training. School children are not involved in the workshop.

Four secondary schools from two regions of Tanzania were randomly assigned to *ICC-T* and four others were controls.

Teachers' self-reported use of violent discipline dropped for teachers who attended the workshop compared to controls three months after the workshop. Teachers' endorsement of corporal punishment also went down significantly compared to controls. For example, support for caning as a discipline method dropped by half in the *ICC-T* group. Students too reported reduced exposure to emotional and physical violence from their teachers. Teacher surveys characterised *ICC-T* training as highly relevant and acceptable to their daily work. Teachers expressed high levels of demand for training and satisfaction with the workshop.

Results need to be interpreted with caution since this study has a high risk of bias.

²² Nkuba M, Hermenau K, Goessmann K, Hecker T. 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*. 2018;13(8):e0201362. Published 2018 Aug 15. doi:10.1371/journal.pone.0201362

Type of study: RCT. Peer-reviewed journal article.

The intervention

Corporal punishment and physical violence against children by teachers are widely prevalent in Tanzania (and in some other countries but the setting for this intervention is Tanzania). *Interaction Competencies with Children – for Teachers (ICC-T)* is a week-long training intervention for teachers to (i) prevent violent discipline and (ii) improve teacher-student relationships.

The key principles on which the training is based are:

- a. Participative approach – teachers actively participate in training sessions
- b. Practice – combining theory and practice to help teachers use new skills at school
- c. Trustful atmosphere – safe space for teachers to share their views and experiences on corporal punishment
- d. Sustainability – various strategies to ensure gains from the training are not lost such as intense practice, repetition of content and creation of a peer-support network.

The training itself is comprised of sessions on:

- i. Teacher-student interaction – focused on fostering empathy, understanding student behaviour better, being a role model for students and on the importance of the teacher-student relationship.
- ii. Maltreatment prevention – raises awareness on the bad effects of corporal punishment and encourages teachers to reflect on their own experiences with corporal punishment
- iii. Effective discipline strategies – introduces non-violent discipline methods that teachers can learn about and ultimately use in their daily work
- iv. Identifying and supporting burdened students – provides information on common emotional and behavioural problems students face and ideas on how teachers can support students better
- v. Implementation – encourages teachers to apply the skills gained in their daily work at school

The training is delivered by a psychologist assisted by three facilitators. All training materials are in English while group discussions happen both in English and Swahili. Teachers are provided with food, beverages and money for travel. Participation is voluntary and teachers are free to leave at any time.

How is the programme meant to work? The theory of change

The study does not mention a specific theory as the basis for the intervention. However, from the description of the intervention it seems to be based on the Transtheoretical Model²³ that theorises that behaviour change goes through “six stages of change: precontemplation, contemplation, preparation, action, maintenance and termination.” This model has been employed widely in other areas such as quitting smoking, reducing domestic violence and teacher behaviours in the classroom. The *Good School Toolkit (GST)* trial in the EGM is also based on this model (Devries 2015)²⁴

²³ Prochaska, J. O., & Velicer, W. F. (1997). The Transtheoretical Model of Health Behavior Change. *American Journal of Health Promotion*, 12(1), 38–48. <https://doi.org/10.4278/0890-1171-12.1.38>

²⁴Devries KM, Knight L, Child JC, et al. The Good School Toolkit for reducing physical violence from school staff to primary school students: a cluster-randomised controlled trial in Uganda. *Lancet Glob Health*. 2015;3(7): e378-e386. doi:10.1016/S2214-109X(15)00060-1 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4928210/>

Has the intervention been implemented at scale?

No. Only four secondary schools from two regions in Tanzania (out of 25 regions) received *ICC-T*.

What does the intervention cost?

No information is provided on cost.

The trial

Participants: Four regions from Tanzania's 25 regions were randomly selected. Two regions (four secondary schools) received *ICC-T* and the other two regions (four other secondary schools) were controls. Each arm had two urban and two rural schools.

Approximately 160 teachers participated, the average age was early-30s with close to 60% female. Nearly 70% had a bachelor's degree, most of the rest had a teaching diploma. The average class size that each teacher managed was close to 60.

Nearly 500 secondary school students were surveyed. Note that students were not part of the actual intervention (their teachers participated in the training workshop). More than half were girls with an average age close to 16 years.

Study design: The trial had two groups:

- a) Four secondary schools (across two regions) sent their teachers to be trained (*ICC-T*)
- b) Four other secondary schools (from two different regions) were controls

Teachers and students were surveyed before the training and three months later.

Outcomes:

The study measured the following outcomes on efficacy, i.e., how well the intervention worked, using a modified version of the Conflict Tactics Scale (CTSPC).²⁵ CTSPC is a valid tool to measure maltreatment and neglect of children by parents. The study only used the 'physical violence' and 'emotional violence' sections of CTSPC which is not how it is meant to be used.

- Violent discipline by teachers: whether teachers continued to use violent discipline
- Teachers' attitude towards violent disciplining: whether they continued to approve of corporal punishment or not
- Students' exposure to school violence: students reporting on whether they experienced violent discipline from teachers

Additionally, teachers were also surveyed on demand for, applicability (relevance) and acceptability of *ICC-T* training.

²⁵ Straus MA, Hamby SL. Measuring physical & psychological maltreatment of children with the Conflict Tactics Scales. In: Kaufman Kantor G, Jasinski JL, editors. Out of the darkness Contemporary research perspectives on family violence. Thousand Oaks, CA: Sage; 1997.

Did the intervention work?

Yes. Teachers in the intervention arm reported statistically-significant reductions in use of both emotional and physical violence to discipline compared to the control group. Teachers in the control group reported reduced use of violent discipline too but it was not as pronounced. This same pattern was seen for positive attitudes towards use of emotional and physical violence as discipline. Specifically, teachers' positive attitude towards caning, a prevalent method for violent discipline, dropped by almost half when surveyed three months after training. The study characterised the size of this impact across outcomes on violence and attitudes to violence as moderate.

Students also reported significant reductions in exposure to emotional violence compared to the control group. Exposure to physical violence in the intervention group was significantly lower than control group before the intervention was implemented. However, even after this was adjusted for in the statistical analysis physical violence exposure reported by intervention group children was significantly lower compared to controls.

While these results suggest the intervention works, they need to be interpreted with caution since this study has a high risk of bias.

From the survey on how useful the teachers thought the programme was:

Demand: Before participating in the training, teachers strongly agreed that such programs were needed. For example, "I think this workshop as it is planned is highly needed for teachers in Tanzania (92% strongly agreed), "I am motivated to participate in the workshop" (91% strongly agreed), "the topics of the workshop are related to my daily work" (83% strongly agreed).

Applicability (relevance): Immediately after training, teachers strongly agreed that the training workshop and training content were relevant to their daily work and to Tanzanian teachers in general. These high levels of endorsement were evident three months later.

Acceptability: Teachers were highly satisfied with the workshop, its content and the trainers straight after the workshop. In fact, they wished that it had been longer. Satisfaction endured at three months later.

Adverse outcomes: No adverse effects are reported in the study.

Generalisability

Findings may not be generalisable because only a few schools were involved in the intervention even though many teachers were trained (nearly 160). More studies and longer-term outcomes might provide more generalisable findings. The intervention ran in a context where teacher-on-pupil violence is extremely high: a study reported a prevalence of violence by teachers of about 95% in Tanzania. Reducing it may be harder where it is lower.

What else is known about the intervention from other studies?

There seem to be no systematic reviews on the effectiveness of preventing corporal punishment by school staff. Systematic reviews of school-based violence prevention tend to focus on topics such as anti-bullying, sexual abuse prevention or intimate partner violence prevention.

The Good School Toolkit (GST) an RCT in the EGM was a comprehensive violence prevention programme that involved teachers, students, school administrators and parents. It was implemented over 18 months

in 21 schools within a district in Uganda. The trial found a large reduction in physical violence reported by students in the past week compared to controls in the main analysis (Devries 2015²⁶). Mental health and educational performance outcomes stayed mostly the same. This trial has five papers in the EGM.

How reliable is the evidence?

The study is rated as having a high risk of bias, with concerns in many domains:

Overall risk of bias	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
High Risk	Some concerns	High risk	High risk	Some concerns	Some concerns

Cells in which this study appears:

The study relates to prevention:

- Intervention = Prevention; Outcome = Institutional Safeguarding Practice: Culture (2 studies + 1 protocol)
- Intervention = Prevention; Outcome = Adult Institutional Caregiver: Behaviours, Attitudes or Knowledge (5 studies + 2 protocols)
- Intervention = Prevention; Outcome = Child Safety: Maltreatment Behaviour (10 studies + 2 systematic reviews + 3 protocols)

²⁶ Devries KM, Knight L, Child JC, et al. The Good School Toolkit for reducing physical violence from school staff to primary school students: a cluster-randomised controlled trial in Uganda. *Lancet Glob Health*. 2015;3(7): e378-e386. doi:10.1016/S2214-109X(15)00060-1 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4928210/>

White (2019) (Disclosure; Child Mental Health)

Skinny cell(s) in which this study appears:

Intervention = Disclosure; Outcome = Child Mental Health

Summary: Combining classroom-based abuse prevention programme with one-on-one simulated training on dangerous situations can improve the confidence and intentions for Grade 1 students (5-7 years) to disclose.

The summary is based on White et al. (2019) ‘Outcomes of In Situ Training for Disclosure as a Standalone and a Booster to a Child Protective Behaviors Education Program.’

Evidence status

Evidence status	<i>Low risk of bias</i>	Strong evidence of improving disclosure confidence and disclosure intentions.
------------------------	-------------------------	---

The summary in brief

In this study, two interventions were tested for children in grade 1 (5-7 years old). One is a conventional school-based sexual abuse prevention programme to recognize and respond to dangerous, potentially abusive situations. The other is a single session ‘In Situ Training’ (IST) for children where a potentially dangerous situation is simulated. In this scenario, during a routine interview as part of the study, the child is left alone for a few minutes when a stranger appears and asks the child to go with them. The stranger then asks the child not to tell anyone about their interaction. The interviewer returns and resumes the interview noting if the child discloses the encounter or not. Children thereby get a chance to ‘practice’ disclosing.

The research study tested each of the interventions alone and in combination compared to a waitlist control group.

Grade 1 students (5-7 years) from various schools in Queensland, Australia were randomized to one of four groups: three interventions (IST alone, group education programme, and combination of both) and a control - which was a waitlist who would get the intervention/s later. All intervention arms led to increased disclosure intentions and confidence (compared to waitlist) but there were no differences in safety identification skills across groups. The two groups which got the ‘practice’ improved more than the one which just did an education programme; and the group which got the combination improved the most.

Children’s anxiety (a possible harm) did not increase from participating in the interventions.

This study highlights the importance of combining comprehensive educational programs with in-person training to make child protective behaviours interventions more effective. Giving children the chance to practice disclosing abuse helps them to protect themselves. By using in-person training alone or as a booster alongside existing programs, educators and professionals can give children the skills and confidence they need to report abuse. This could lead to early intervention and prevention of further harm.

Type of study: RCT. Peer-reviewed journal article.

The intervention

This school-based research study looked at how teaching children to disclose potential abuse situations can be helpful on its own or when combined with a child protective behaviours education program. The goal was to see if in-person training in real-life situations could improve children's ability to tell someone about abuse and if it provided extra benefits when used with an educational program.

The Observed Protection Behaviours Test (OPBT) in-person training is a single session to promote disclosure (of abuse) by a child. The study calls this in-situ training (IST). The IST interviewer is alone with a child conducting an interview like how baseline data was collected. The interviewer pretends to forget some documents and leaves the child alone in the interview room. A male stranger enters the room and urges the child to leave with him (meant to simulate a real-life risky situation). The child has ten seconds to respond. If they say 'no' or do not respond, the man leaves but asks the child not to tell anyone about their interaction. If the child says 'yes', the man again asks the child to keep it secret and to remain in the room. The interviewer returns in a few minutes, and the child is prompted to disclose the encounter with the stranger every few minutes till they either talk about it or the interview ends.

The second is a classroom-based sexual abuse prevention educational programme called 'Learn to be safe with Emmy and friends,' a very commonly implemented type of child protective intervention for young children. Five, one-hour weekly sessions are delivered by trained facilitators at school. The sessions cover "emotion knowledge and regulation, early warning signs of danger, personal space, private body area, safe and unsafe secrets, disclosure and disclosure networks." "Emmy" is the programme mascot that is part of the curriculum that is delivered via didactics and group work. The goals are to help children recognize risky situations and to disclose.

The trial

Nearly 300 children 5-7 years old in Grade 1 from many schools in Australia's Queensland participated in the trial. They were divided at random into four groups: one group received in-person training only, another group received the educational program only, the third group received both and the fourth group were waitlisted, i.e., they served as the control group for the research study. In the in-person training sessions, children learned how to recognize abuse and tell a trusted adult about it. The educational program taught children about protecting themselves from abuse and how to prevent it.

The researchers measured the results by giving the children questionnaires before and after the interventions, and then again 3 and 6 months after the intervention ended. The questionnaires asked about their knowledge of protective behaviours, how they felt about telling someone about abuse, and if they could recognize potential abuse situations. The researchers also recorded how many children actually told someone about abuse using confidential reporting methods.

How is the programme meant to work? The theory of change

No specific theory is mentioned. The programme is built on behavioural skills training (BST) which is used in different forms (group, individual) for interpersonal harm reduction.

Did the intervention work?

The results showed that both the in-person training and the educational program helped children improve their knowledge of protective behaviours and their ability to recognize potential abuse situations. ***The in-person training alone led to more children telling someone about abuse compared to just the educational***

program. This suggests that the interactive nature of the in-person training sessions made children feel more comfortable talking about sensitive issues.

The group that received both the in-person training and the educational program had the best outcomes. These children showed the highest levels of knowledge, positive attitudes about telling someone about abuse, and a higher number of actual reports of abuse compared to the other groups. The in-person training seemed to reinforce the lessons learned in the educational program and made children more confident about disclosing potential abuse situations. However, safety identification skills weren't different between groups.

Adverse outcomes: Participation in the intervention did not increase children's anxiety.

Has the intervention been implemented at scale?

No. This is a small-scale research study.

What does the intervention cost?

The study does not report any data on cost although it is mentioned that web-based training could theoretically be cheaper than in-person training.

How reliable is the evidence?

The study is rated as having a low risk of bias.

Overall risk of bias	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported result
Low risk of bias	Low risk	Low risk	Low risk	Low risk	Low risk

Cells in which this study appears:

The study relates to both prevention and disclosure:

- Intervention = prevention; outcome = child maltreatment disclosure (5 studies)
- Intervention = prevention; outcome = child mental health
- Intervention = prevention; outcome = child knowledge and awareness
- Intervention = disclosure; outcome = child maltreatment disclosure
- Intervention = disclosure; outcome = child mental health
- Intervention = disclosure; outcome = child knowledge and awareness

Glossary

Some of these terms are explained in more detail in Section 1.

Baseline

Data gathered at baseline is gathered before the intervention starts. For example, the head circumference of children in the Bucharest Early Intervention Project (foster care) was measured before the foster care started. Those same data are often gathered again at mid-line (i.e., part-way through the interventions), end-line (when the intervention ends), and sometimes at follow-up points (i.e., after the intervention ends.)

Disclosureⁱⁱⁱ

Disclosure interventions were defined as any intervention that aimed to facilitate, support, or promote the disclosure of child maltreatment. This encompassed a range of universal interventions, such as traditional or social media campaigns, or child helplines, as well as therapeutic interventions for children that aimed to promote disclosure (e.g., play therapy). It included tertiary interventions relating to perpetrators, such as mandatory reporting, and also included any intervention that aimed to promote disclosure within an organisational context (e.g., staff training, organisational guidelines).

Effect size / effect estimate

The observed association between interventions and outcomes, or a statistic to summarise the strength of the observed associationⁱⁱ. It is the size of the effect that an intervention has. For instance, the Good Schools Toolkit reduced violence from being experienced by 80% of students in the previous term (clearly a giant amount) to being experienced by 'only' 60% of them: its effect is to reduce violence by 20 percentage points, which in this case is a quarter.

Effect sizes are only ever estimates for two reasons. First, one study can only ever report on the effect that it found within the specific place, time and people involved, and the intervention may have a different effect elsewhere. And second, even within that place, time and population, the intervention's apparent effect is tempered by the sample size: the smaller the sample, the larger the chance that the apparent result was the effect of chance rather than the intervention.

Endline

This is when the intervention ends. For example, the level of violence in schools in the Good Schools Toolkit trial was measured before the programme started (the base-line); and again at the end-line (when the intervention ends), and sometime. Sometimes, those same data are also gathered at mid-line (part-way through the interventions), and at follow-up points (after the intervention ends).

Follow-up

Points when data are gathered after the intervention has ended.

Preventionⁱⁱⁱ

Prevention interventions were defined as any intervention where the primary aim was to decrease the likelihood or risk of child maltreatment occurring or recurring in the future. This encompassed both interventions for any child / adult ('universal populations'), as well as interventions targeted at specific populations. Examples of types of prevention interventions that could be included were school-based safety programmes, organisational guidelines or practices, or perpetrator targeted interventions to reduce reoffending.

Primary study

This is a study which directly involves people, such as a survey or study involving training teachers. Primary studies (or primary research) are opposed to secondary studies (or secondary research) which are studies of studies, such as a literature review or systematic review.

Protocolⁱⁱ

A plan or set of steps that defines how something will be done. Before carrying out a research study, for example, the research protocol sets out what question is to be answered and how information will be collected and analysed.

Quasi-experimental design (QED) or quasi-experimental study

A study based on a true experimental design meets two criteria:

- it has two groups of people, one of which gets an intervention and the other doesn't. (Or there may be groups which get different interventions), and
- people are assigned to those groups at random.

A study with a quasi-experimental design uses the first criterion, but not the second because people are not randomly assigned to groups. This means a researcher cannot draw conclusions about 'cause and effect' as reliable as from an RCT. This design is frequently used when it is not feasible, or not ethical, to conduct a randomised controlled trial.

Randomised controlled trial (RCT) ⁱⁱⁱ

A study in which a number of similar people are randomly assigned to two or more groups to test an intervention. One group (the experimental group) has the intervention being tested, the other (the comparison or control group) has an alternative intervention, or no intervention at all. The groups are followed up to see how effective the experimental intervention was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically. This method is also used to reduce bias.

Responseⁱⁱⁱ

Response interventions were defined as any intervention that aimed to improve institutional responses to child maltreatment in relation to each of the target populations. Response interventions included enhancing safeguarding practices, legal and policy interventions, supporting the victim and/or family, working with child protection agencies, and providing training and crisis support to staff within organisations.

Systematic reviewⁱⁱ

A research report that summarises the evidence (ie., existing studies) on a clearly formulated question according to a predefined protocol, using systematic and explicit methods to identify, select and appraise relevant studies, and to extract, analyse, collate and report their findings. It may or may not use statistical techniques, such as meta-analysis. It is an example of secondary research.

Treatment^{iv}

Treatment interventions were defined as any intervention that aimed to provide a therapeutic response to a target population. This included therapeutic interventions provided to children who experienced child maltreatment in institutions, and interventions targeted at institutional perpetrators of child abuse. The Romania studies are included here, because foster care was provided as treatment for young children who spent their early lives in institutionalised care.

Appendix 1: The Visual Evidence and Gap Map

The interactive EGM can be found here. [add link](#)



Appendix 2: Systematic Reviews on the EGM which are in cells empty apart from themselves.

The AMSTAR2 rating is a rating of the reliability of systematic reviews. It is detailed in the full EGM report. Most SRs on the EGM get only low ratings.

Location of the cell (Intervention; outcome)	Systematic Review Name	AMSTAR 2 Rating	Abstract for the Systematic Review (or intro / summary / overview, where the SR has no abstract)
<p>1. Response; Institutional Safeguarding Practice: Culture</p> <p><i>Also in:</i></p> <p>2. Response; Institutional Safeguarding Practice: Operational practice</p> <p>3. Response; Child Safety: Maltreatment Behaviour</p> <p>4. Response; Child Wellbeing: Physical Health</p> <p>5. Response; Child Wellbeing: Mental Health</p> <p>6. Response; Child Wellbeing: Social and Emotional Functioning</p> <p>7. Response; Child Wellbeing: Cognitive Function</p>	<p>Hermenau 2017: Fostering Child Development by Improving Care Quality: A Systematic Review of the Effectiveness of Structural Interventions and Caregiver Trainings in Institutional Care</p>	<p>Low</p>	<p>Quality of child care has been shown to have a crucial impact on children’s development and psychological adjustment, particularly for orphans with a history of maltreatment and trauma. However, adequate care for orphans is often impacted by unfavourable caregiver–child ratios and poorly trained, overburdened personnel, especially in institutional care in countries with limited resources and large numbers of orphans. This systematic review investigated the effects of structural interventions and caregiver trainings on child development in institutional environments. The 24 intervention studies included in this systematic review reported beneficial effects on the children’s emotional, social, and cognitive development. Yet, few studies focused on effects of interventions on the child–caregiver relationship or the general institutional environment. Moreover, our review revealed that interventions aimed at improving institutional care settings have largely neglected violence and abuse prevention. Unfortunately, our findings are partially limited by constraints of study design and methodology. In sum, this systematic review sheds light on obstacles and possibilities for the improvement in institutional care. There must be greater efforts at preventing violence, abuse, and neglect of children living in institutional care. Therefore, we advocate for combining attachment theory-based models with maltreatment prevention approaches and then testing them using rigorous scientific standards. By using approaches grounded in the evidence, it could be possible to enable more children to grow up in supportive and nonviolent environments.</p>

<p>1. Prevention ; Disclosure Rates</p> <p><i>Also in:</i></p> <p>2. Disclosure: Disclosure Rates</p> <p>3. Treatment ; Disclosure Rates</p> <p>4. Response; Disclosure Rates</p> <p>5. Prevention ; Child Wellbeing: Knowledge or Awareness</p> <p>6. Disclosure: Child Wellbeing: Knowledge or Awareness</p> <p>7. Response; Child Wellbeing: Knowledge or Awareness</p> <p>8. Treatment ; Child Wellbeing: Knowledge or Awareness</p> <p>9. Prevention ; Child Wellbeing: Mental Health</p> <p>10. Disclosure; Child Wellbeing: Mental Health</p> <p>11. Response;</p>	<p>Radford 2017: Rapid Evidence Assessment: What can be learnt from other jurisdictions about preventing and responding to child sexual abuse</p>	<p>Low</p>	<p>This Rapid Evidence Assessment was commissioned by the Independent Inquiry into Child Sexual Abuse in England and Wales which is investigating whether public bodies and other non-state institutions have taken seriously their duties to care for and protect children and young people from child sexual abuse and exploitation. The question for the review was: What can be learnt from jurisdictions, outside of England and Wales, about the role of institutions, including accountable state and non-state organisations with responsibility for children in preventing and responding to child sexual abuse and exploitation?</p>
---	--	------------	--

<p>Child Wellbeing: Mental Health</p> <p>12. Treatment ; Child Wellbeing: Mental Health</p>			
<p>1. Response; Child Safety: Maltreatment Behaviour</p> <p>Also in</p> <p>2. Prevention ; Child Safety: Maltreatment Behaviour</p> <p>3. Prevention ; Child Wellbeing: Child Mental Health</p> <p>4. Prevention ; Child Wellbeing: Knowledge or Awareness</p> <p>5. Response; Child Wellbeing: Child Mental Health</p> <p>6. Response; Child Wellbeing: Knowledge or Awareness</p> <p>7. Prevention ; Child/youth offender:</p>	<p>Sherr 2017: Child violence experiences in institutionalised / orphanage care</p>	<p>Low</p>	<p>Institutions are not necessarily good environments for children. In the face of challenges such as HIV, Ebola, poverty, conflict and disaster the numbers have grown rather than reduced. Some countries have closed institutions down –driven by findings that cognitive developmental delay is associated with institutional care. Yet insight into abuse and violence within institutionalised settings is neglected. Maltreatment - violence and abuse - may be an issue. This systematic review series addresses violence and abuse experiences in institutionalised care, exploring firstly the frequency of abuse/violence in institutions, secondly any interventions to reduce such violence or abuse and thirdly the perpetrators of such violence or abuse. The final systematic review updates the findings on cognitive delay associated with institutionalised care. With a violence lens, cognitive delay may well be considered under the umbrella of neglect. Maltreatment and abuse may be a driver of cognitive delay. The keyword search covered several electronic databases and studies were included for data abstraction if they met adequacy criteria. Eight studies were identified on the prevalence of abuse in institutions and a further three studies reported on interventions. Only one study was identified documenting peer on peer violence in institutions. Sixty-six studies were identified examining cognitive development for institutionalised children. All but two of these record cognitive deficits associated with institutionalisation. Only two asked about violence or abuse which was found to be higher in institutionalised children. Overall the abuse experiences of children in institutions are poorly recorded, and in one study violence was associated with high suicidal attempts. The major intervention pathway for ameliorating cognitive challenge seems to be placement out of the institutions which shows benefits and redresses some cognitive outcomes – yet not a total panacea. The single study providing training and monitoring of harsh</p>

<p>Recidivism 8. Response; Child/youth offender: Recidivism</p>			<p>punishment and maltreatment showed immediate and decided reductions. This data suggest, despite the paucity of studies, violence and abuse, by commission or omission is prevalent in institutions, has an effect on child well-being and is amenable to intervention. Simple training or more complex structures to place children within conducive alternative environments (or to avoid institutionalised placements in the first place) seem to be the main pathway of intervention.</p>
<p>Treatment; Child Wellbeing: Mental health And Treatment; Child Wellbeing: Child social emotional functioning</p>	<p>Bailey 2019: systematic Review on trauma-informed care models in out-of-home care (OoHC) settings</p>	<p>Low</p>	<p>Trauma in early childhood has been shown to adversely affect children's social, emotional, and physical development. Children living in out-of-home care (OoHC) have better outcomes when care providers are present for children, physically, psychologically, and emotionally. Unfortunately, the high turnover of out-of-home carers, due to vicarious trauma (frequently resulting in burnout and exhaustion) can result in a child's trauma being re-enacted during their placement in OoHC. Organisation-wide therapeutic care models (encompassing the whole organisation, from the CEO to all workers including administration staff) that are trauma-informed have been developed to respond to the complex issues of abuse and neglect experienced by children who have been placed in OoHC. These models incorporate a range of therapeutic techniques, and provide an overarching approach and common language that is employed across all levels of the organisation. The aim of this study was to investigate the current empirical evidence for organisation-wide, trauma-informed therapeutic care models in OoHC. A systematic review searching leading databases was conducted for evidence of organisation-wide, trauma-informed, out-of-home care studies, between 2002 and 2017. Seven articles were identified covering three organisational models. Three of the articles assessed the Attachment Regulation and Competency framework (ARC), one study assessed the Children and Residential Experiences programme (CARE), and three studies assessed The Sanctuary Model. Risk of bias was high in six of the seven studies. Only limited information was provided on the effectiveness of the models identified through this systematic review, although the evidence did suggest that trauma-informed care models may have significantly positive outcomes for children in OoHC. Future research should focus on evaluating components of trauma-informed care models</p>

			and assessing the efficacy of the various organisational care models currently available.
<p>1. Prevention; Institutional (safeguarding) practice: Environment</p> <p>And also in</p> <p>2. Response; Institutional (safeguarding) practice: Environment</p> <p>3. Prevention; Care provider: Adult institutional care provider</p> <p>4. Response; Care provider: Adult institutional care provider</p> <p>5. Prevention; Child Safety: Child maltreatment occurrence/re occurrence</p>	<p>Lo and Cho 2021</p> <p>community-based interventions on the actual reduction of child maltreatment</p>	Low	<p>Purpose: This review seeks to summarize selected literature on existing findings on the impacts of community-based interventions on the actual reduction of child maltreatment and to identify the core components of the interventions.</p> <p>Methods: This study systematically searched electronic databases, including PsycInfo, Medline, and Web of Science. The findings of the selected studies were summarized using narrative synthesis. Results: A total of four studies met the inclusion criteria of this study. The studies showed declines in child maltreatment incidences reported by child protective services and hospitals during the study periods. Four major components and approaches were identified among the selected interventions, including (1) the involvement of community members, (2) partnerships with community institutions, (3) multidisciplinary collaboration, and (4) responsiveness to the needs of the communities involved.</p> <p>Conclusions: The results of this review support the need for further development of community-based interventions using a hybrid approach.</p>

<p>6. Response; Child Safety: Child maltreatment occurrence/re occurrence</p> <p>7. Prevention; Parent/caregiver: Parent/caregiver knowledge/awareness</p> <p>8. Response; Parent/caregiver: Parent/caregiver knowledge/awareness</p>			
<p>1. Treatment; Child Wellbeing: Child cognitive functioning</p> <p>And</p> <p>2. Treatment; Child Wellbeing: Child mental health</p> <p>3. Treatment; Child Wellbeing: Child social emotional</p>	<p>Sneddon 2020</p> <p>Cognitive-behavioural therapy (CBT) interventions for young people aged 10 to 18 with harmful sexual behaviour</p>	<p>High</p>	<p>OBJECTIVES: To evaluate the effects of CBT for young people aged 10 to 18 years who have exhibited HSB., SEARCH METHODS: In June 2019, we searched CENTRAL, MEDLINE, Embase, 12 other databases and three trials registers. We also examined relevant websites, checked reference lists and contacted authors of relevant articles., SELECTION CRITERIA: We included all relevant randomised controlled trials (RCTs) using parallel groups. We evaluated CBT treatments compared with no treatment, waiting list or standard care, irrespective of mode of delivery or setting, given to young people aged 10 to 18 years, who had been convicted of a sexual offence or who exhibited HSB., DATA COLLECTION AND ANALYSIS: We used standard methodological procedures expected by Cochrane., MAIN RESULTS: We found four eligible RCTs (115 participants). Participants in two studies were adolescent males aged 12 to 18 years old. In two studies participants were males simply described as "adolescents." Three studies took place in the USA and one</p>

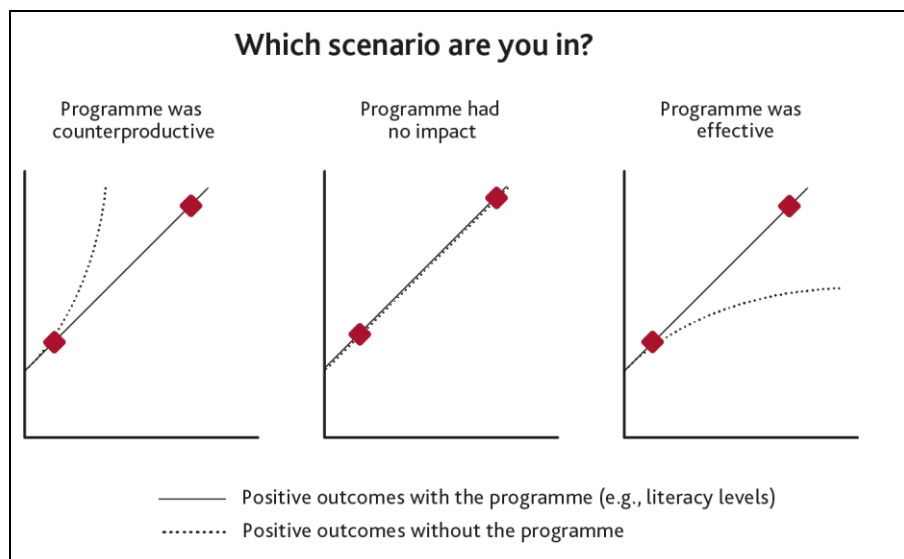
functioning			in South Africa.
-------------	--	--	------------------

Appendix 3: Guidance on reading social science studies, and using this Guidebook

Counterfactuals / comparison groups

A counterfactual shows what would have happened without the programme. It is important that interventions are tested in ‘fair tests’. Otherwise, if we see some improvement (such as rising literacy), we do not know whether that improvement is more than would have happened without the programme, exactly the same as it would have been without the programme, or possibly the improvement may have been greater without the programme.

Figure 2: The Importance of a Counterfactual For Establishing Impact



Having a fair test requires that there is a group which gets the intervention and an equivalent group which does not. This latter is the control group, which is used for comparison. The best way to ensure that the two groups are equivalent (i.e., similar in all other respects) is to take a group of people and use a random process to decide who gets the intervention and who doesn't. This is called a randomised controlled trial (RCT). It is the fairest type of test.

The EGM has many RCTs, and has some non-randomised experiments which we call quasi-experimental designs, QEDs. Not all of the QEDs explain why they did not use random assignment. A clear example of a non-equivalent control group is in Czerwinski, a QED run in Germany. The control group seems meaningfully different from the intervention group. “The children in the control group were more girls and more often reported having a foreign background. {34% in the control group vs 51% in the programme group.} Half speak Turkish at home (54.4%) and half languages from other European countries, the former USSR or other countries.” This may matter because “Besides the intervention effect, we observed a strong effect of foreign background on children’s knowledge related to child sexual abuse (CSA). In general, children from families of non-German descent scored lower on average across all three groups, and boys had lower values than girls.” On courses of action, “children of foreign background scored lower”. The study acknowledges that that difference in racial background could account for some or all of the effects: “The fact that children with foreign background scored lower on

positive outcome scales could be caused partly by language difficulties.” The non-equivalence of the comparison group creates risk of bias and reduces our confidence in the study’s findings: the sex and origin of the students may explain all or some of the results. These are called ‘confounding factors’. This is also an example in which the small number of schools involved made it difficult to achieve ‘balance’, i.e. similar characteristics in programme and non-programme schools.

Sometimes, the outcomes improve in the intervention group but also in the control group, and this sometimes turns out to be because people in the control group found a way of getting the intervention or benefiting from it. This happened in one programme in Spanish islands on the EGM (see Cerezo 2004), and researchers think that the teachers who got the programme passed the knowledge to their friends and colleagues who didn’t. This is a contamination effect.

Translating evidence to a difference context

For grant-managers and policy-makers, a key challenge is figuring out whether an intervention will get the same results in their context as it did where it was studied: whether the evidence ‘translates’ to a different context. For instance, whether a programme run in Europe will achieve the same results in Ghana. We therefore comment on this in the Guidebook where we can. A few comments on that.

First, few of the studies discuss whether their results likely apply elsewhere. That is not a criticism: rather, it is not really the purpose of a study report - which is to report just on the specific study done, and what was found in that place at that time.

Second, essential to assessing whether a finding will apply in (translate to) a different context is to understand the theory of change / mechanism by which the intervention worked when it was studied. In the little example in the box below, providing standpipes reduces water-borne diseases only when those diseases arise because of lack of access to clean water: the theory of change is around providing clean water where there was not enough before. If we can see the theory of change, we can make an informed guess about whether the mechanism in it will apply to a new context. Unfortunately, few of the studies discuss theory of change, or the theory on which the intervention is based. We have stated it where the study report states it, and have inferred it in some others.

Third, in writing this Guidebook, we cannot say whether a study finding will translate to your context: rather you (the person using the evidence) must assess that because you know your context. We have provided (i) some principles in this section for doing that assessment, and (ii) the relevant information from the studies, in the Guidebook.

The information in this Guidebook to help with this translatability assessment includes:

- What the study says about it, if anything.
- The theory of change / theory on which the intervention is based. Sometimes this is stated explicitly in the studies and sometimes it isn’t, but we recognised it nonetheless.
- The range of geographies, and groups of people, where the study was run, and whether it has yet been run at scale.
- Comment from us where we could be helpful. This is based on the general context and unusual characteristics of the circumstances in which a study was run. It is our opinion.

Clearly it is easier to assess transferability when the theory, context and results are all clear, which is certainly not universal amongst the studies on the EGM.

For example, the *BEIP* got good results in some respects. On one hand, that would seem likely to translate elsewhere, because it is about how children develop, which is pretty universal (especially the physical development). But, on the other hand, the orphanages in Romania were so awful that (a) it is probably not difficult to improve on them and (b) mercifully, few children in institutions elsewhere will experience neglect so severe, so the effect of removing them to foster care may be less pronounced. The *BEIP* was a very precise type of foster care, prescribed by the academics who set it up and monitored carefully, so its results may not generalise to all foster care.

Equally, the *GST* ran in a place where physical violence by school staff against students is very high (including even choking, burning and stabbing). Again, that intervention (or similar) probably wouldn't get the same dramatic results in places where violence is, happily, lower.

Fourth, the studies are spread across the EGM sparsely: many cells are empty, but even those which have studies, only have one, two or three. That is, few interventions (if any) have been tested multiple times in multiple places, which is what is really needed to understand whether and when a finding will translate to many other places.

The most tested interventions, by far, are programmes run in schools to prevent sexual abuse by educating children about the issue, how to spot it, how to avoid it and what to do about it. Those have been tested multiple times - though rarely in Africa, South Asia, or South America. In the studies on the EGM, they have always worked, in terms of increasing children's knowledge and awareness. That consistency implies (though does not prove) that the results will translate, i.e., these programmes would probably work anywhere. Research to see whether these programmes are successful in Africa, South Asia, or South America (low-income, high-population places) seems a clear priority. Moreover, as the experience of the control group in *GST* shows, simply collecting data can lead to widespread disclosure.

The issue of whether evidence translates to other contexts is also referred to as 'generalisability', i.e., whether the results apply 'generally' - in **all** places and all times - or only in specific time/s or place/s. Generalisability is obviously a higher bar. Normally, you don't care whether they are 'generally' (universally) true: rather, it only matters that they apply in the place and time in which you are interested.

Translating evidence to new contexts: a key issue for grant-managers and others in using existing research

Most people's behaviour is influenced by financial incentives. That is, if a study in Japan offers people a financial incentive if they do a particular thing, and finds that that incentive increases the proportion of people who do that thing, it's likely that a study in, say, Venezuela will find much the same. In other words, that finding translates elsewhere: it applies generally and not just in that one place in Japan where the study was done.

Other times, study findings are not generalisable, or are only generalisable to particular other places. Obviously if a study ran in the UK in 2016 had found that some experience was more likely to make people vote for their country to leave the European Union, that is not generalisable to other countries

(it wouldn't make *those people* vote for *their* country to leave the European Union) because one can only vote like that if there is a poll on that topic, which was unique to the UK. That finding is therefore not generalisable.

Some findings are somewhat generalisable, or generalisable to particular other places: for example, installing water pipes might improve sanitation and therefore reduce water-borne diseases, and that result might be generally true for places that have poor water and sanitation; but not true in places where clean water is already readily available: it won't achieve much in Paris, for example.

(By the way, the fact that a study doesn't generalise is not a criticism of the research. Rather, it's just a feature of the world, that different things happen in different places.)

A key issue in using research is figuring out whether the results that were obtained by that intervention in that place and that time will be obtained by running that intervention in a different place and/or a different time.

Measures

We are interested in a range of outcomes from knowledge and attitudes to the prevalence and disclosure of abuse. These are sensitive areas, and these outcomes not necessarily easy to measure. All studies rely on asking people questions to elicit this information rather than, say, direct observation, role play or asking people to keep diaries of incidences. The way in which these data are collected is called the measurement tool.

The choice of measurement tools matters. Good ones have been objectively validated (ie., have been tested and to produce correct answers), and are known to measure only what they purport to measure. For example, an IQ test will be valid if its results accurately reflect intelligence, and not other traits such as the ability to read. Moreover, tools have to be validated for the setting in which they are being used – what is appropriate in Indiana may not be appropriate in India. The extent of the practice of validation varies across the social sciences. It is most well established in psychology – indeed there is a field called psychometrics devoted to measurement and validation – and a large number of the papers in the map come from research in this discipline.

However, some studies on the EGM invented their own measurement tools (such as surveys, or tests). This can be a problem. First, newly invented tools have rarely been objectively validated, so often give unreliable answers. And second, even if a new tool is objectively validated, it is normally impossible to compare the answers that it gives to the results that other tools would find. For example, one way of measuring confidence may be quite different to another. This is a super-common and annoying feature of social science (and also medical science) because it prevents anybody from using the study to decide between two courses of action (examined in two different studies). One implication is that, when you are commissioning new research, always require that the researchers look for and use existing validated tools if possible and choose the one/s used most commonly. If you are commissioning new research, you may need to validate a tool (new or existing) to that context. Another is that, sadly, in quite a few studies on the EGM, it is surprisingly hard (and sometimes impossible) to identify the effect of the intervention(!).

When using a measure, it is necessary to both know the likely range of values which will be observed and what practitioners – who are often familiar with these tools – would consider to be a change of practical significance. Sometimes, the study authors themselves say that they are unsure of the practical significance of their results.

Some measures on the EGM seem somewhat strange. For example, a study (Kolko 1987) of the *Red Flag / Green Flag People* programme to raise awareness and prevention of child sexual victimisation did not assess parents' actual knowledge or awareness, e.g., through a test, but rather asked them how knowledgeable or aware they thought they were. A weakness here is that we can't say whether knowledge or awareness changed.

Other times, results come from regression analyses (i.e., statistical interrogation of the data). These can give results that the study authors describe as 'large' without being able to be more specific, e.g., to say that knowledge grew from 40% on some scale to 60% on that scale. Where this happens, the Guidebook reports what the study says, e.g., that the results were 'large'.

When results are measured

Most primary studies on the EGM measure some outcome (e.g., bed-wetting) at multiple time-points: before the intervention starts, immediately after it finishes, and some period thereafter. These are called, respectively, baseline or pre, end-line or post, and 'at (say) 12 months', or 'x month follow up' where x is typically 6 and 12 but rarely longer. If an effect had occurred by the end-line (say, an improvement in reading levels) and was still there three months later, some studies report this as the effect being "sustained at three months [follow up]".

Note that some studies measure time from when the programme started. *The Good School Toolkit*, for example, is 18 months long, so when those studies talk about 'at 24 months', they mean only six months after it ended.

Some studies (normally RCTs) do not take measurements at baseline, because measurement can be an intervention and therefore surveying people can influence their attitudes which will distort the apparent effect of the intervention. This approach is deemed acceptable for RCTs as the randomization process will usually ensure the equivalence of the programme and control groups. But non-experimental designs are expected to demonstrate this equivalence and so have to collect data.

Attenuation / effects which fade

Often, an effect evident at end-line (when the programme ends) fades over time. People learn things and then forget them. Their behaviour improves but later reverts. This is why it is important to measure outcomes not just at end-line but also later, to see whether and where any effects endure.

The EGM certainly has studies in which the effects attenuate. For example, the *Bringing in the Bystander* programme appeared to increase the behaviour of 'talking to a hurt friend' by two months after end-line, but that effect has disappeared by a year after end-line (Edwards 2019). Equally, sometimes the comparison group catches up. This happened in some outcomes measured in the *BEIP* children. It also happens in other studies, e.g., a study of the effect of cash transfers in Uganda²⁷ found that they helped

²⁷ <https://www.vox.com/2018/9/10/17827836/cash-basic-income-uganda-study-blattman-charity>

when measured four years later, but that by nine years later, the control group had caught up. These studies can be read in two different ways: one reading is that the effect 'disappears', i.e., that the long-term view is that the intervention doesn't work; the other is that it is successful in improving matters at least for a while. Those people in Uganda had a better life for at least four years (achieved pretty cheaply). Those students in Edwards' study helped people for at least two months more than did their peers. That will often be a win (depending on what the intervention is trying to achieve). After all, almost all of medicine - e.g., fixing a broken leg - isn't about creating lasting improvement, but rather about returning the patient to the level of their peers.

Attrition

Sometimes people drop out of a study. There are various reasons for this: it might be voluntary on their part (e.g., they decide not to complete the survey 12 months after the programme); it might be administrative (e.g., the Romania studies, some follow-ups were many years later, and the researchers no longer had contact details for some of the children); it might be because the participants were no longer eligible (in the Romania studies, some of the children eventually returned to their original families, so they ceased to be relevant to include in a study of children in orphanages or foster care); it may also be if participants have died.

This attrition can be a problem for study results and can introduce risk of bias. A severe example is that in some medical trials, an intervention is lethal to some people: obviously, if the study only measures outcomes of the people who are still there when the trial ends, it will omit all of those whom the trial has killed, so would miss that (obviously!) important finding. This is called 'survivor bias'. It means that apparent effects of an intervention may simply reflect the different characteristics of the remaining people in the trial, or omit potentially serious adverse effects. A less extreme form of survivor bias would be if some people really dislike the intervention - or dislike the researchers - so choose to disengage from it. That too could be biased: maybe many of the less educated people disliked the researchers, so the group remaining at end-line are mainly more educated, which is a bias.

The risk of bias is particularly serious when there is differential attrition, i.e., different rates of attrition between the treatment and comparison groups. This is very common, and mostly not reported. Most usually attrition is higher in the comparison group as they have less incentive to stay engaged. But sometimes there may be high attrition for programme participants for the reasons outlined above.

Some studies simply don't give the reason for attrition. That raises the possibility of a bias. There is an example in Czerwinski, a QED in Germany on the EGM. Data were available for 291 children and 328 parents before and immediately after the programme, and 292 children and 304 parents before and three months after the programme was delivered. However, authors note that data for all three time-points was only available for 256 students and 240 parents. No information is given about which students and parents were which, nor why the attrition. That counts towards the quality assessment for that study.

Programme participants

Some of the interventions studied on the EGM are targeted at particular groups, e.g., children in residential care or in orphanages. Most are not, and they are called 'universal' interventions. Examples are programmes run in normal schools for all the children.

Quality of studies / risk of bias

The reliability of the studies on the EGM is low, in general. This is discussed in the EGM report. As a reminder, the EGM only contains studies above a certain threshold in terms of their design: it only has primary studies which have a counterfactual (that is, randomised controlled trials or quasi-experimental designs which have some other counterfactual), and systematic reviews: it does not include any studies which simply describe behaviour nor which just compare the situation before the intervention with that afterwards. We graded all of the included studies on their reliability (e.g., the possibility that they may be biased, for instance by only reporting selected results) and in general found this to be low.

The Guidebook material talks about the quality of studies / the risk of them being biased: this is to enable the reader to beware of results which are particularly likely to be inaccurate.

As explained in the EGM documents, the assessments were:

- RCTs were assessed on their risk of bias (RoB) using the Cochrane tool²⁸
- Non-RCT primary studies (i.e., quasi-experimental designs, QEDs) were assessed using ROBINS II tool²⁹
- Systematic reviews were assessed using the AMSTAR2 tool³⁰.

Be aware that some aspects of a study which influence its RoB rating are within the researchers' control, and other features are not and are therefore not a reflection on them. For instance, if researchers randomised schools but somehow used a bad randomisation method (they vary in how random they actually are, because some can be influenced), that is normally within the researchers' control. But if lots of people drop out of the programme because they move away or don't like it or get other jobs with different time commitments - or some politician cancels half the programme: that would affect the RoB rating but be outside the researchers' control.

Risk of Bias ratings work on a 'weakest link' principle: the study is assessed on several domains where there may be bias (e.g., method of randomisation, attrition bias): and if any of them is high risk, then the study is deemed to be high risk of bias overall.

ⁱ Both available at www.giving-evidence.com/csa

ⁱⁱ <https://ies.ed.gov/ncee/wwc/>

ⁱⁱⁱ From or based on <https://www.nice.org.uk/glossary>

^{iv} From the EGM report, at www.giving-evidence.com/csa

²⁸ Higgins et al 2016, A revised tool for assessing risk of bias in randomized trials. In: Chandler J., McKenzie J., Boutron I., Welch V. (Eds.), *Cochrane Methods. Cochrane Database of Systematic Reviews*, Issue 10 (Suppl 1).

²⁹ Sterne et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919. doi:10.1136/bmj.i4919

³⁰ Shea et al (2017). *AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both*. *BMJ*, 21(358), 4008.