

Institutional Responses to Child Maltreatment: Guidebook to the Evidence

September 2020

Prepared with staff of the Campbell Collaboration Secretariat

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 draws – 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 funder, wanted to find the rigorous causal evidence around ‘what works’ in child abuse in institutional contexts. With the Centre for Evidence and Implementation, we searched for such studies, collated them, coded them, and presented them in the Evidence and Gap Map¹ (EGM). 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. Our² aim is to help funders, policy-makers, 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:</p> <p>Syntheses of the studies in cells which have three or more studies (which we call ‘heavy’ cells)</p>	<p>Twelve cells on the EGM have 3+ studies, and we have synthesised 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.</p> <p>These cells vary hugely in their number of studies: most have 3-8 studies; some 9-18; and one has 51.</p>
<p>Section 3:</p> <p>Summaries of individual studies which appear in cells on the EGM which have just one or two studies (‘light’ cells)</p>	<p>These are presented in alphabetical order. There are seven such studies, plus one protocol, plus one summary of three papers which concern a single study (in Romanian orphanages).</p>

¹ This EGM is based on a search conducted in early 2019, and therefore should contain all relevant studies published up to that point was published in July 2020. It is being published through the Campbell Collaboration, which has already published the protocol (<https://onlinelibrary.wiley.com/doi/full/10.1002/cl2.1039>). Non-academic versions of the report, and a summary of it, are at www.giving-evidence.com/csa There are some minor differences between versions because some pre-date and some post-date the peer review process: the overall picture is unchanged.

² Throughout, this document uses ‘we’ and ‘our’ to refer to the papers’ authors, not to the funder.

Contents

Introduction to the Guidebook.....	2
Main findings of the EGM: where there is evidence and where there isn't.....	5
Section 1a: Introduction and comment about the whole evidence-base	6
1. Overall findings in the Guidebook: what that evidence says.....	6
Effect sizes.....	7
2. Some comments on the studies on the EGM	9
Geography.....	9
Theories on which the interventions are based / theories of change.....	9
Age of the studies, and whole-school approaches	9
Outcomes which are measured	10
Length of trials, and follow-up periods.....	10
Cost information	11
Sample sizes in the studies	11
Who created and ran the interventions, and the theories on which they are based?	12
What else do we know about these topics?.....	13
Section 1b. What this Guidebook contains.....	14
Studies vs papers	14
Studies that appear in multiple cells and/or have multiple outcomes.....	14
Protocols	15
Treatment of systematic reviews in cells which contain primary studies	15
Treatment of cells with just systematic reviews and no primary studies	15
Section 2: Syntheses of studies in cells which contain 3+ studies.....	1
Prevention / Institutional Safeguarding Practice: Operations.....	1
Prevention / Adult Institutional Caregiver.....	6
Prevention / Disclosure: Disclosure rates	13
Prevention / Child Safety: Maltreatment Behaviour	19
Prevention / Child Wellbeing: Knowledge and Awareness	26

Prevention / Child Wellbeing: Mental Health.....	33
Prevention / Child Wellbeing: Social-Emotional Functioning.....	39
Prevention / Parent Caregiver: Knowledge and Awareness.....	44
Response / Institutional Safeguarding Practice: Operations.....	50
Response / Child Wellbeing: Mental Health.....	56
Treatment/ Child Wellbeing: Mental Health.....	59
Treatment / Child Wellbeing: Cognitive Functioning.....	66
Section 3: Summaries of individual studies which appear in cells with only one or two studies.....	71
Protocol: Baker-Henningham (2016) (Prevention / cognitive functioning, and adult desistance).....	71
Czerwinski (2018) (several heavy cells).....	76
Devries (2015) Good Schools Toolkit (Prevention / child wellbeing, and cognitive functioning).....	81
Edwards (2019) (Prevention / adult maltreatment behaviour).....	86
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).....	90
Merrill (2018) (Prevention / institutional culture).....	98
Nkuba (2018) (Prevention / institutional culture).....	104
Rheingold (2014) (Response / adult caregiver behaviours, knowledge or attitudes).....	109
Taylor (2010) (Prevention / child perpetrator maltreatment behaviour).....	113
Glossary.....	118
Appendix 1: The Visual Evidence and Gap Map.....	121
Appendix 2: Systematic Reviews on the EGM which are in cells empty apart from themselves.....	1
Appendix 3: Guidance on reading social science studies, and using this Guidebook.....	5
Counterfactuals / comparison groups.....	5
Translating evidence to a difference context.....	6
Measures.....	8
When results are measured.....	9
Attenuation / effects which fade.....	9
Attrition.....	10
Programme participants.....	10
Quality of studies / risk of bias.....	11

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.

- At the time of the work, we identified 82 relevant papers. There were **72 studies**³: 58 completed primary studies (of which 49 are RCTs), three protocols (i.e., plans) for three further RCTs which we assumed were underway, and 11 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 32 of the 58 completed studies. The EGM has no studies from India, only two from China, and only three studies from Africa. 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. Fully 53 of the 61 primary studies included in the EGM (including the three protocols for planned RCTs) examined these programmes.
 - The most commonly-studied interventions (42 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,** which was considered in 56 of the primary studies.
- **Most of the studies are about prevention,** which was examined in 59 primary studies (completed and on-going studies), and 10 systematic reviews. We found no primary studies of interventions specifically aiming to facilitate disclosure⁴(!)
- **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 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.
- **Only one completed study had assessed interventions with adults to stop them offending** within organisations (that is the Good Schools Toolkit).
- **Only four studies focused on children particularly at-risk.**
- **Almost all the studies have appreciable risk of bias**⁵. None of the RCTs was low risk of bias.
- **Only one study looked at the effect on educational attainment** (the *Good School Toolkit* in Uganda).
- **Institutional safeguarding practice was studied in seven primary studies.**
- **Very few studies came from practitioners and non-profits.**

³ Sometimes, multiple research papers are written about a single experiment, e.g., by collecting data from the participants at various junctures after the intervention finishes, or by reporting different outcomes in different papers. This constitutes one *study* but multiple *papers*. The EGM includes several papers about one study of the *Good School Toolkit* in Uganda; several papers about one study of the Bucharest Early Intervention Project in Romania; and one systematic review which is an update of a previous systematic review.

⁴ 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.

⁵ This Guidebook takes the assessments of risk of bias and study quality from the EGM report: the studies were not re-assessed for the Guidebook.

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

This section discusses the evidence on the EGM as a whole. 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 almost every cell on the EGM: almost every cell has too little evidence to enable a clear view of the effects of the programmes, and/or the evidence is too weak. 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 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, the *Bucharest Early Intervention Project (BEIP)* found that “placement into foster care does not significantly impact development / performance [in] executive functioning - i.e., memory and cognitive monitoring”⁶. 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. (Effect sizes are [explained here](#).) 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

⁶ Source: <http://www.bucharestearlyinterventionproject.org/About-Us.html>

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 of 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 scales⁷ 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.

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.

⁷ 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."

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 81% of the completed primary studies. By contrast, the EGM has zero studies from India, and only three studies from the whole of Africa. It has only two from China, and only five of the 61 primary studies (completed and planned) are from countries which have Muslim majorities. (As a reminder, our search specified no time period, so this represents all findable studies from any year.)

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 49 primary studies of them on the EGM. However, those studies are all from the US, Canada, the UK, Europe, Australia, Central America, and East Asia. In other words, nobody has yet 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

Many of the studies on the EGM are now quite old. The first one published in 1985 (now 35 years ago), and nearly half (46%) published before 2012. Given the delay between a study happening and being published, that means that nearly half the studies probably happened at least about ten years ago.

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.

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 foundations⁸ 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(!): perhaps three at most. (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.

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 costs⁹: 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

⁸ John D. and Catherine T. MacArthur Foundation, the Binder Family Foundation, the Sinneave Family Foundation.

⁹ 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.

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.

Researchers should undertake what are called '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

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 - but multiple 'papers' written about them: the EGM has six papers about that study.

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

We made a little exception for some of the *BEIP* papers. One cell has one *BEIP* paper (Smyke) and another has two (Bick and Johnson): we wrote one summary which covers all three of these papers - partly to avoid describing the intervention multiple times.

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, Hermenau (a systematic review, SR) appears in several light cells, in the EGM columns for: institutional safeguarding practice, culture; child physical health; child Social-Emotional Functioning; and child cognitive functioning. Radford, another SR, is in light cells for various outcomes. For that reason, the summaries of individual studies list the cells in which the study appears. 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.

For example, Rheingold, an RCT, is in a light cell for adult care-givers' attitudes and knowledge; and also more populous cells for other outcomes. The summary for Rheingold relates to all the outcomes that are on the EGM.

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, but they are not useful for this Guidebook because the protocol does yet *say* anything. But they are included in the EGM as one purpose of the map is to identify research gap. If a cell had only a protocol it would be a current gap which someone is in the process of filling.

Where a protocol is one of three studies in a cell, i.e., the cell is heavy if the protocol is included but light if it's excluded, we did not count it for that cell. The cell top-left on the EGM is one such: it is light if the protocol is removed, so we counted it as light and did summaries of the two completed studies in there.

We did a summary of the one protocol of this type - Baker-Henningham.

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, which count as empty, under this definition: Hermenau (an SR) is on its own sometimes; Radford (an SR) is also; the cell response / maltreatment behaviour has three SRs but no primary studies; one cell has just Quadara and Radford (both SRs).

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

Navigation table

Links to syntheses of a whole cell are in [blue](#). Links to summaries of individual studies are in [green](#).

xxx	Institutional Safeguarding Practice: operations	Adult ...								
Prevention	Prevention / Institutional Safeguarding Practice: Operations	Prevention / Adult Institutional Caregiver	Prevention / Disclosure: Disclosure rates Disclosure: Disclosure rates	Prevention / Child Safety: Maltreatment Behaviour	Prevention / Child Wellbeing: Knowledge and Awareness Devries (2015) Good Schools Toolkit (Prevention / child wellbeing, and cognitive functioning)	Prevention / Child Wellbeing: Mental Health Czerwinski (2018) (several heavy cells)	Prevention / Child Wellbeing: Social-Emotional Functioning	Protocol: Baker-Henningham (2016) (Prevention / cognitive functioning, and adult desistance)	Prevention / Parent Caregiver: Knowledge and Awareness	
Disclosure										
Treatment						Treatment/ Child Wellbeing: Mental Health		Treatment / Child Wellbeing: Cognitive Functioning		
Response	Response / Institutional Safeguarding Practice: Operations					Response / Child Wellbeing: Mental Health				

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

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

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

The summary in brief

Training caregivers of children such as teachers and day care workers is a potential method to protect children from sexual abuse. Training programmes can improve the knowledge, attitudes, and practices of caregivers to prevent abuse.

This synthesis includes three primary studies and one systematic review. In all three primary studies, teachers received training on preventing child sexual abuse. One large scale study from Spain also trained other professionals such as social workers, policemen and hospital staff. It 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 this was 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 sources of risk of bias are listed below.)

The systematic review was a scoping exercise that looked to identify broad themes on prevention of sexual abuse and exploitation of children living in residential care relevant to the Australian context. The review did not evaluate the effectiveness of interventions (i.e., how well various interventions work) but provided an overview of what is important from a prevention perspective.

The cell has three primary studies (Cerezo 2004¹⁰; Gushwa 2018¹¹; Rheingold 2014¹²). A full summary of Rheingold 2014 is available in the guidebook.

Contents of the cell

A. Primary Studies	
Cerezo 2004 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.
Gushwa 2018	USA (3 states – 2 Midwest, 1 Northeast). K-12 teachers from 3 public school districts

¹⁰ Cerezo MA, Pons-Salvador G. Improving child maltreatment detection systems: a large-scale case study involving health, social services, and school professionals. *Child Abuse Negl.* 2004;28(11):1153-1169. doi:10.1016/j.chiabu.2004.06.007

¹¹ Gushwa, M., Bernier, J., & Robinson, D. (2018). Advancing child sexual abuse prevention in schools: An exploration of the effectiveness of the enough! Online training program for K-12 teachers. *Journal of Child Sexual Abuse*, 1–16

¹² Rheingold, Alyssa & Zajac, Kristyn & Chapman, Jason & Patton, Meghan & de Arellano, Michael & Saunders, Benjamin & Kilpatrick, Dean. (2014). Child Sexual Abuse Prevention Training for Childcare Professionals: An Independent Multi-Site Randomized Controlled Trial of Stewards of Children. *Prevention Science : The Official Journal of the Society for Prevention Research*. 16. 10.1007/s11121-014-0499-6.

RCT, high risk of bias	Evaluation of the Enough! Preventing Child Sexual Abuse in My School programme
Rheingold 2015 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 CSA

A. Primary Studies (Cerezo 2004, Gushwa 2018, and Rheingold 2014)

The interventions

Programme	Country	Description
Large-scale programme to increase child maltreatment detection (Cerezo 2004)	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 two to three 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.
<i>Enough! Preventing Child Sexual Abuse in My School</i> (Gushwa 2018)	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 one 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.
<i>Stewards of Children</i> (Rheingold 2014)	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 two weeks.

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

The intervention in Cerezo 2004 led to an increase in the number of cases reported to CPS from both its phases. The study reports a tripling of detected cases after the intervention compared to before. The second phase, i.e., training teachers, led to detection of two to three 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. An increase in referral before and after intervention was seen for the first two territories but not for the third

territory. The authors thought that this might be due to knowledge of the intervention spreading to the third territory before it was implemented there via mass media, professional networks and professionals moving territories for new jobs. The outcomes were then compared with a different region of the Balearic Islands and the expected increase in referrals was seen.

From Gushwa 2018, teachers who received the online training intervention (*Enough! Preventing Child Sexual Abuse in My School*) had significantly higher scores compared to controls on a knowledge scale that included questions on various aspects of CSA including “prevalence rates, types of CSA behaviours, impact of CSA on children, signs and symptoms of CSA, the veracity of children’s reports, backgrounds / behaviours of perpetrators, factors associated with CSA in schools including examples of boundary-violating behaviours, specific behaviours in support of prevention, reporting responsibilities, and responses to suspected abuse.” Intervention group teachers answered nearly 90% of the questions correctly on average (after training) while control group teachers got approximately 75% of the questions right. Most teachers who received training also stated that their knowledge, awareness and likelihood of taking action to prevent CSA had increased “a great deal” or “somewhat” after training (as compared to “a little” or “none, already knew”).

From Rheingold 2014, the results demonstrate that *Stewards of Children* training improved knowledge and behaviours.

- ✓ **Knowledge** about CSA increased. That knowledge declines over the three months after the training (as one would expect), though oddly, the knowledge of the control (‘waitlist’) group increased during that time, but still not to as high as the trained group.
- ✓ **Attitudes.** Participants’ belief in CSA myths was low to begin with so there was little room for improvement. After training, the control group had the better score but at three months there was no difference between groups.
- ✓ **Behaviours.** This also improved, i.e., participants reported having done more of the behaviours three months after the training that did people in the control group. The behaviours most improved were:
 - “limiting the opportunity for other youth and younger youth to have one-to-one interaction”. This is significant because juveniles are offenders in more than a third of CSA (Finkelhor et al. 2009¹³) and
 - “sharing with another adult an article, brochure, or other information about CSA prevention”. Interestingly, the behaviour of people in the control group also improved during the three months after the training: quite possibly because colleagues who had received the training changed their behaviour (making it more normal) and they shared this information with colleagues who had not received the training.

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 had the group trained online. Three months after training, however, there were no differences between the two groups.

Attitudes: No difference between the group trained in-person vs the group trained online.

¹³ Finkelhor, D., Ormrod, R., & Chaffin, M. (2009). Juveniles who commit sex offenses against minors. Office of Juvenile Justice and Delinquency Prevention. <http://www.ncjrs.gov/pdffiles1/ojjdp/227763.pdf>.

Behaviours: No difference between the group trained in-person vs the group trained online.

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

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 (it is 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.¹⁴

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 reports any cost data.

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

Cerezo 2014 does not mention any specific theory for their programme. The approach adopted was based on the principles of motivational interviewing to overcome barriers and change attitudes on reporting potential cases of maltreatment.

Stewards of Children is not a theory-based prevention programme, but its principles are in line with Finkelhor's¹⁵ theory that for CSA to occur, certain preconditions must exist. They include: an individual's tendency to abuse, absence of internal or external inhibitions for the offender, and the offender 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 offenders by 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 best¹⁶.

Will the results translate elsewhere?

¹⁴ <https://www.d2l.org/get-trained/>

¹⁵ Finkelhor, D. (1984). *Child sexual abuse: New theory and research*. New York: Free Press.

¹⁶ Jarvis, P, & Jarvis, P. (2004). *Adult education and lifelong learning: Theory and practice*. London, UK: RoutledgeFalmer.

The results from the US studies might translate 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*.

Risk of Bias for Quasi-Experimental Designs (QEDs) using ROBINS-II¹⁷

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
Cerezo 2014	Serious (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 (two primary studies are common to both cells).

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.

¹⁷Sterne JA, Hernán MA, Reeves BC, 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

Prevention / Adult Institutional Caregiver

School-based interventions that train teachers can improve teacher knowledge, attitudes, and behaviours to prevent child sexual abuse and to reduce corporal punishment

Evidence status	Moderate risk of bias	Moderate evidence of impact on teacher knowledge, attitudes, and behaviours. Worth testing different approaches to achieving this outcome using strong study designs.
------------------------	------------------------------	---

The summary in brief

School-based interventions to reduce child abuse often include training for teachers. This cell has five primary studies and two protocols (both for RCTs). Four studies (and one protocol) include training for teachers to improve their knowledge, attitudes, and protective behaviours to prevent child sexual abuse (some of the interventions include other components as well such as educational activities for children). One study (and one protocol) implemented a training workshop for secondary school teachers to improve their attitudes and use of corporal punishment (for e.g., caning students as a form of discipline) in a setting where it was widely prevalent.

The impact on teacher knowledge, attitudes, and behaviours suggests that improvements through trainings are possible for both preventing child sexual abuse and reducing corporal punishment. Most interventions were one-time trainings which might limit their benefits in the long run.

Contents of the cell

The cell has five studies (Kolko 1987¹⁸, Kolko 1989¹⁹, MacIntyre 1991²⁰, Nkuba 2018²¹, Rheingold 2014²²) and two protocols (McElearney 2018²³, Ssenyonga 2018²⁴).

Full summaries of Nkuba 2018 and Rheingold 2014 are available in this Guidebook.

¹⁸ Kolko, D.J., Moser, J.T., Litz, J. *et al.* Promoting awareness and prevention of child sexual victimization using the Red Flag / Green Flag program: An evaluation with follow-up. *J Fam Viol* **2**, 11–35 (1987). <https://doi.org/10.1007/BF00976368>

¹⁹ Kolko, D.J., Moser, J.T. & Hughes, J. Classroom training in sexual victimization awareness and prevention skills: An extension of the Red Flag / Green Flag people program. *J Fam Viol* **4**, 25–45 (1989). <https://doi.org/10.1007/BF00985655>

²⁰ MacIntyre, D., & Carr, A. (1999). Evaluation of the effectiveness of the Stay Safe primary prevention programme for child sexual abuse. *Child Abuse & Neglect*, **23**, 1307-1325.

²¹ Nkuba M, Hermenau K, Goessmann K, Hecker T. Reducing violence by teachers using the preventive intervention Interaction Competencies with Children for Teachers (ICC-T): a cluster randomized controlled trial at secondary schools in Tanzania. *PLoS One*. <https://doi.org/10.1371/journal.pone.0201362>

²² Rheingold, Alyssa & Zajac, Kristyn & Chapman, Jason & Patton, Meghan & de Arellano, Michael & Saunders, Benjamin & Kilpatrick, Dean. (2014). Child Sexual Abuse Prevention Training for Childcare Professionals: An Independent Multi-Site Randomized Controlled Trial of Stewards of Children. *Prevention Science : The Official Journal of the Society for Prevention Research*. **16**. 10.1007/s11121-014-0499-6.

²³ McElearney, A., *et al.*, 2018. Cluster randomised controlled trial of ‘whole school’ child maltreatment prevention programme in primary schools in Northern Ireland: study protocol for keeping safe. *BMC public health*, **18** (1), 590. doi:10.1186/s12889-018-5492-8

²⁴ Ssenyonga J, *et al.* Reducing violence against children by implementing the preventative intervention interaction competencies with children for teachers (ICC-T): study protocol for a cluster randomized controlled trial in southwestern Uganda. *Trials*. 2018;19:435.

<p><i>Kolko 1987</i> QED, low risk of bias.</p>	<p>US (three schools in Pennsylvania). Children in third and fourth grades: Middle childhood (6-11 years) Evaluation of the <i>Red Flag / Green Flag Program</i> to raise awareness and prevention of child sexual victimisation</p>
<p><i>Kolko 1989</i> QED, low risk of bias.</p>	<p>US (six schools in Pennsylvania). Children in third and fourth grades: Middle childhood (6-11 years) Also implemented <i>Red Flag / Green Flag</i> but in more schools with a few changes (parents were not involved in the programme in this study but were in the previous study)</p>
<p><i>MacIntyre 1991</i> QED, low risk of bias.</p>	<p>Republic of Ireland (Dublin: five suburban schools). Children in middle childhood (6-11 years). Evaluation of the effectiveness of the <i>Stay Safe</i> primary prevention programme for child abuse</p>
<p><i>McElearney 2018</i> RCT protocol: the study is still ongoing, so no results reported</p>	<p>UK (Northern Ireland). Children between 4-11 years. Evaluation of a school-based child maltreatment prevention programme, <i>Keeping Safe</i>, designed by the NSPCC, a charity.</p>
<p><i>Nkuba 2018</i> RCT, high risk of bias</p>	<p>Tanzania (2 regions, 4 schools). Secondary school children. Testing the <i>Interaction Competencies with Children for Teachers (ICC-T)</i> programme to reduce use of violence as a form of punishment against children.</p>
<p><i>Ssenyonga 2018</i> RCT protocol: the study is still ongoing, so no results reported</p>	<p>Uganda (6 districts, 12 schools). Secondary school children. Implementing <i>ICC-T</i> programme to train secondary schoolteachers in one region of Uganda.</p>
<p><i>Rheingold 2014</i> RCT, moderate risk of bias</p>	<p>USA (three sites in different geographical regions – Atlanta, GA; Beaufort, SC; Bend, OR). Caregivers of children in day care, churches, schools <i>Evaluation of the Stewards of Children</i> programme to prevent child sexual abuse</p>

The interventions

Intervention overviews are in the following table.

There were four programmes from completed studies: *Red Flag / Green Flag People* (Kolko 1987; Kolko 1989); *Stay Safe* (MacIntyre 1999); *Interaction Competencies with Children for Teachers, ICC-T* (Nkuba 2018) and *Stewards of Children* (Rheingold 2014). Kolko 1987 and Kolko 1989 had one difference in the

interventions – parents were involved in training in the earlier study but not in the later study. McElearney 2018 is a protocol for *Keeping Safe* and Ssenyonga 2018 describes a protocol for a study implementing *ICC-T* in a different location (Uganda) than Nkuba 2018 (Tanzania).

Red Flag / Green Flag People; Stay Safe; and Keeping Safe are delivered to primary school students and their teachers. All are aimed at improving children’s awareness of the nature of child sexual abuse (CSA) and equip them with the skills to respond, although *Keeping Safe* is a more comprehensive ‘whole school’ approach to change school culture. *Stewards of Children* is also aimed at preventing child sexual abuse but by training child caregivers – mostly schoolteachers but also day care workers and church workers. Children are not directly involved in intervention activities in this programme.

ICC-T is a training programme for secondary school teachers in Tanzania and Uganda to reduce their use of violence as a form of discipline with their students. The programme was developed for a context where corporal punishment (such as caning) is widely prevalent.

Note that McElearney 2018 and Ssenyonga 2018 are protocols so no results were available.

Programme	Country	Description
<i>Red Flag / Green Flag People</i>	USA	Staff and parent sessions followed by two classroom training sessions of 1.5 hours each (<i>Kolko 1989</i> did not have parent sessions). Teachers received an “extensive in-service training program conducted by caseworkers from Children and Youth Services which involved didactic instruction, roleplaying, and group discussion.” Four sessions were conducted over three weeks and included sexual abuse prevention concepts, information on legal aspects and strategies to support listening to children and leading classroom discussions on sexual abuse prevention. <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.
<i>Stay Safe</i>	Republic of Ireland	Teacher and parent training sessions following by classroom implementation of 10-12 sessions of 30-40 minutes each. Teacher (2 sessions) and Parent training (1 session) includes sexual abuse prevention concepts such as definitions, myths, realities, prevalence and information on typical victim and offender characteristics. It also includes tips on how to identify victims and support them with disclosure and for appropriate referral to legal and social services. Pupil session 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.
<i>Keeping Safe</i>	UK (N. Ireland)	‘Whole-school’ programme for children aged 4-11 on how to keep safe from any type of maltreatment. Teaching and learning resources are incorporated within regular curricula and the culture of the school. Classroom teaching

(protocol only, no results available)		covers three themes (healthy relationships; my body; and being safe) and 63 lesson plans for children as they progress from grade 1 to grade 7. School leaders and parents are also actively involved in the programme. Training and support are available for teachers and school staff.
<i>Interaction Competencies with Children for Teachers (ICC-T)</i>	Tanzania (completed study) Uganda (protocol)	Week-long workshop for secondary school teachers led by a psychologist. Participation is voluntary. The focus is on improving teacher-student relationships, better understanding students' needs, 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.
<i>Stewards of Children</i>	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 the interventions work in improving parent / caregiver knowledge and attitudes?

Teachers participating in *Red Flag / Green Flag* did not report any significant differences in knowledge or awareness compared to controls (Kolko 1987) although the sample size was small. In the other study on this programme (Kolko 1989), teachers in the training group reported significantly higher 'utility' (of the training programme) and 'change in understanding' (on child sexual abuse) compared to control. 'Awareness' or 'preparedness' were not statistically different.

The evaluation of *Stay Safe* used a 38-item Knowledge and Attitudes Questionnaire for teachers. Significant improvements were recorded and maintained for 21 of the 38 items, these items reflected "factual knowledge about abuse; belief in children's statements; attributions of responsibility for abuse; attitudes towards prevention programs; knowledge about help-seeking and anxiety or confidence concerning the management of sexual abuse." Knowledge and attitudes did not differ by age or gender of teachers.

ICC-T teachers reported significant reductions in use of both emotional and physical violence to discipline students 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, had dropped by almost half when teachers were surveyed three months after training. The study characterised the size of this impact across outcomes on violence and attitudes to violence as 'moderate.' A large proportion of teachers (80-90%) agreed with the statements: "*Did the workshop change your understanding of student's problems in relation to their behavior?*" (immediately after training: much = 26%, very much = 68%; 3 months after training: much = 55%, very much = 42%) and "*Do you think this workshop will influence your previous strategies in dealing with disciplining students?*" (immediately after training: much = 35%, very much = 60%; 3 months after training: much = 35%, very much = 51%). 71% of teachers (3 months after training) said they frequently

used non-violent discipline and 40% stated they were more sensitive when communicating with their students. However, 11% of teachers opined that it would be difficult to eliminate corporal punishment in Tanzanian schools.

The evaluation of *Stewards of Children* demonstrates that the intervention improved knowledge and behaviours. CSA knowledge increased but it declined over the three months after the training (as one would expect), though oddly, the knowledge of the control ('waitlist') group increased during that time, but it was still not as high as the trained group. Participants' belief in CSA myths was low to begin with so there was little room for improvement. After training, the control group had the better score but at three months there was no difference between groups. CSA prevention behaviours also improved, i.e., participants reported having done more of the behaviours three months after the training that did people in the control group. The behaviours most improved were:

- "limiting the opportunity for other youth and younger youth to have one-to-one interaction". This is significant because juveniles are offenders in more than a third of CSA²⁵ and
- "sharing with another adult an article, brochure, or other information about CSA prevention". Interestingly, the behaviour of people in the control group also improved during the three months after the training: quite possibly because colleagues who had received the training changed their behaviour (making it more normal) and they shared this information with colleagues who had not received the training.

Results were similar for both in-person and online training. The size of the impact in terms of programme implications are unclear.

Have the interventions been implemented at scale?

Personal safety 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*²⁶ – are common, though this study of effectiveness was carried out in a small sample. Similarly, *Stewards of Children* is also available widely for training of caregivers, but this study covered only three sites.

The *Stay Safe* programmes delivered in most primary schools in the Republic of Ireland.

ICC-T was tested in a small number of schools in four schools from two regions of Tanzania.

Which type of organisation delivered the intervention?

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."

²⁵ Finkelhor, D., Ormrod, R., & Chaffin, M. (2009). Juveniles who commit sex offenses against minors. Office of Juvenile Justice and Delinquency Prevention. <http://www.ncjrs.gov/pdffiles1/ojjdp/227763.pdf>.

²⁶ https://books.google.ca/books/about/Bobby_and_Mandee_s_Good_Touch_Bad_Touch.html?id=aBPm5HuBwpYC&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false

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 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 organisation that delivered *ICC-T* is not mentioned.

Stewards of Children (Rheingold 2014) was developed and delivered by a US NGO, Darkness to Light.²⁷

What do the interventions cost?

None of the studies report costs.

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

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

Are the results generalisable?

Probably not. The number of schools included in the studies are quite small. Some of the programmes might have been implemented on a large scale but they have not been evaluated at that level.

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
Nkuba 2018	High risk of bias	Some concerns	High risk	High risk	Some concerns	Some concerns
Rheingold 2014	Some concerns	Low risk	Some concerns	Low risk	Some concerns	Some concerns

²⁷ <https://www.d2l.org/get-trained/>

Risk of Bias for Quasi-Experimental Designs (QEDs)

Study (Author and year)	Over all RoB	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	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk
Kolko 1989	Low	Low risk	Low risk	Low risk	Low risk	Moderate risk	Moderate risk	Low risk
MacIntyre 1991	Low	Low risk	Low risk	Low risk	Low risk	Low risk	Moderate risk	Low risk

Prevention / Disclosure: Disclosure rates

School-based interventions to tackle abuse can increase disclosure

Evidence status	Low risk of bias	Evidence of positive impact, with need for testing to establish best practice
------------------------	------------------	---

The summary in brief

School-based interventions to reduce child abuse are usually focused interventions carried out over a small number of sessions to increase children’s awareness of physical and sexual abuse and to train them in appropriate courses of action. The interventions often also include teachers and parents. Training on appropriate courses results in a greater 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. This finding suggests that recurrent activities will result in more disclosure than one-off interventions. One study shows that disclosure is far greater to a young abuse survivor than to teachers. No studies find evidence of adverse side effects. Findings need to be treated with caution as the studies mostly have a medium to high risk of bias.

The interventions

All the interventions are all school-based, ranging from kindergarten to high school students aged 15-17. All interventions directly involve pupils other than the online 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 also to 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) and 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 prevention of physical and sexual abuse by teachers in the school.

The table below provides an overview of the interventions included in the studies in this summary. (The Guidebook contains a summary of Czerwinski.)

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 (del Campo Sanchez, 2006)</i>	Spain	School-based prevention programme for children aged 8-12.

<i>Good School Toolkit</i> (Devries, 2015)	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>Enough! Preventing Child Sexual Abuse in My School</i> (Gushwa, 2018)	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</i> (Hazzard, 1991)		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</i> (Kolko, 1987 & 1989)	USA	Two sessions for elementary school students based on colouring book and film.
<i>Project Trust</i> (Oldfield, 1996)	USA	30-minute play (Trust) followed by 15-minute Q&A for grades 1-6.

Who delivers the intervention?

The interventions in all studies were 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 engagement 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 in 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, 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 was a Ministry of Education project funded from 2015-17. Whilst it was offered to other schools, 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-Secours workshop was undertaken for the purposes of the study.

No additional information could be found for Project Trust, suggesting that it is 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 as just over \$1,100 per school, or \$11 per student, which was equivalent to \$60 per disclosure. The most expensive cost was the presenter.

For the other programmes, based on the programme descriptions, it can be seen that these are mostly similarly low-cost interventions 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 for which a license fee must be paid.

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

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

The core idea of 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 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 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 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, counsellor 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 as seen in the table below. There are usually zero

disclosures in the control group over the same period. Even simply asking children about experience 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.

The interventions' disclosure rates

Programme	Disclosure in	
	Treatment group	Control group
Tweenees	<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)
Prevención de abusos sexuales a menores	8	2
Good School Toolkit	434 children referred to child protective services because of what they disclosed in the follow-up survey	
Feeling yes, Feeling no	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	
Red Flag / Green Flag		
1987 study	20 (revealed in data collection of which 18 had been revealed to adult)	0
1989 study	20	0
Project Trust	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 six students, the four classes held by the 18U presenter resulted in 65 disclosures, compared to just three disclosures in the X classes with teacher presents. For grades seven and eight, 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 German programme), 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 – and *Enough!* is solely for teachers. However, the effects of this component are not commonly evaluated. Three studies report findings of teacher preparedness, two finding a positive effect and one no effect, so no conclusion can be drawn about their effectiveness in the absence of meta-analysis and preferably additional studies.

The evidence in favour of a positive impact on disclosure is supported by the findings along the causal chain. The studies mostly report a positive impact on children’s awareness of potentially abusive situations and so how to respond. For example, *Viol-Secours* workshop has only a small effect on knowledge and awareness and so 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 situation such as abductions. A number of studies collect data from parents on child mental health, such as anxiety, or concerns about sex. No evidence is found of adverse effects from exposure to the intervention.

How good is the evidence?

All studies are rated as overall medium (some concerns) or high risk of bias, except the two studies of *Red Flag / Green Flag* which are low risk of bias as shown in the table below. 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	Randomisation 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

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
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

References

Barron I G, and Topping K J. (2013). Prevention Programs and Legal Considerations for Child Sexual Abuse Cases. , 22(8), pp.931-948.

Czerwinski F, Finne E, Alfes J, and Kolip P. (2018). Effectiveness of a school-based intervention to prevent child sexual abuse—Evaluation of the German IGEL program. , 86, pp.109-122.

Daigneault I, Hébert M, McDuff P, Michaud F, Vezina-Gagnon P, Henry A, and Porter-Vignola E. (2015). Effectiveness of a sexual assault awareness and prevention workshop for youth: A 3- month follow-up pragmatic cluster randomisation study. University of Toronto Press, 24(1), pp.19-30.

Devries KM, Knight L, Child J C, Mirembe A, Nakuti J, Jones R, Sturgess J, Allen E, Kyegombe N, Parkes J, and Walakira E. (2015). The Good School Toolkit for reducing physical violence from school staff to primary school students: a cluster-randomised controlled trial in Uganda. Science Direct, 3(7), pp.e378-e386.

Gushwa M, Bernier J, and Robinson D. (2018). Advancing Child Sexual Abuse Prevention in Schools: An Exploration of the Effectiveness of the Enough! Online Training Program for K-12 Teachers. Journal of Child Sexual Abuse, 28(2), pp.144-159.

Hazzard A, Webb C, Kleemeier C, Angert L, and Pohl J. (1991). Child sexual Abuse Prevention: Evaluation and one-year follow-up. Science Direct, 15(1-2), pp.123-138.

Kolko D J, Moser J T, and Hughes J. (1989). Classroom Training in Sexual Victimization Awareness and Prevention Skills: An Extension of the Red Flag / Green Flag People Program. Springer Link, 4(1), pp.25-45.

Kolko D J, Moser J T, Litz J, and Hughes J. (1987). Promoting Awareness and Prevention of Child Sexual Victimization Using the Red Flag / Green Flag Program: An Evaluation with Follow-up. Springer Link, 2(1), pp.11-35.

Oldfield D, Hays B J, and Megel M E. (1996). Evaluation of the effectiveness of Project Trust: An elementary school-based Victimization Prevention Strategy. Science Direct, 20(9), pp.821-832.

Prevention / Child Safety: Maltreatment Behaviour

School-based interventions can reduce use of corporal punishment. Results need to be interpreted with caution due to high risk of bias.

Evidence status	High risk of bias	Weak strength evidence for use of violence as discipline by teachers against students
------------------------	--------------------------	---

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 use of violence as punishment by teachers in school (corporal punishment) in settings where it is widely prevalent; 2) programmes for younger children delivered in the classroom to prepare them to recognise, avoid, and report sexual abuse; and 3) prevention programmes for older children (middle school and high school) focused on interpersonal violence, dating violence, and bystander action. This last category includes lessons for children aimed at preventing them being victimised as well from perpetrating offenses against others.

This cell includes studies that report on reductions in children disclosing violence.

On corporal punishment, interventions seem to succeed in reducing violence against children. Both RCTs in the cell come from sub-Saharan Africa, though their results need to be considered with caution due to a high risk of bias for both 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 (most report on intermediate outcomes such as knowledge / awareness and attitudes).

Contents of the cell

This cell includes ten papers about completed primary studies²⁸ (Devries 2015²⁹, Devries 2017, Devries 2018, Edwards 2019, Knight 2018, Merrill 2018, Nkuba 2018, Taylor 2010, Kolko 1987, Kolko 1989) and three protocols for RCTs (Baker-Henningham 2016, McElearney 2018, Ssenyonga 2018).

The guidebook has summaries for Devries 2015, Edwards 2019, Merrill 2018, Nkuba 2018, Taylor 2010 (all completed primary studies) and Baker-Henningham 2016.

Devries 2015, Devries 2017, Devries 2018, Knight 2018, Merrill 2018 RCT, high risk of bias	Uganda (21 primary schools in one district). Children in 5 th -7 th classes: Early adolescence (11-14 years) Evaluation of <i>The Good School Toolkit (GST)</i> – a whole-school approach to preventing the use of violence as discipline by teachers against students. The intervention ran for 18 months.
Nkuba 2018 RCT, high risk of bias	Tanzania (4 secondary schools). Early and late adolescence. (11-17 years) Evaluation of the <i>Interaction Competencies with Children for Teachers (ICC-T)</i> intervention – a training workshop for secondary school teachers to reduce their use of corporal punishment
Ssenyonga 2018 RCT protocol only	Uganda (6 secondary schools). Early and late adolescence. (11-17 years) Planned evaluation of <i>ICC-T</i> in southwestern Uganda.

²⁸ Full citations for studies are available at the end of this document

²⁹ Devries 2015, Devries 2017, Devries 2018, Knight 2018, Merrill 2018 are papers from one RCT on the Good School Toolkit in Uganda.

Baker-Henningham 2016 RCT protocol only	Jamaica (38 preschools). Early childhood (24 months-5 years) Planned evaluation of the <i>Irie Classroom Toolbox (ICT)</i> – a training programmes for preschool teachers to prevent use of violence.
Edwards 2019 RCT, high risk of bias	USA (25 high schools). Late adolescence (15-17 years) Evaluation of the <i>Bringing in the Bystander – High School Curriculum (BITB-HSC)</i> to reduce interpersonal violence and promote bystander action.
Taylor 2010 RCT, high risk of bias	USA (7 middle schools). Early adolescence (11-14 years) Evaluation of two curricula to prevent gender violence and sexual harassment (GV/SH) or “dating violence”.
Kolko 1987, QED, low risk of bias	US (three schools in Pennsylvania). Children in third and fourth grades: Middle childhood (6-11 years) Evaluation of the <i>Red Flag / Green Flag People Program</i> to raise awareness and prevention of child sexual victimisation.
Kolko 1989 QED, low risk of bias	US (six schools in Pennsylvania). Children in third and fourth grades: Middle childhood (6-11 years) Also implemented <i>Red Flag / Green Flag People</i> but in more schools with a few changes (parents were not involved in the programme in this study but were in the previous study).
McElearney 2018 RCT protocol only	UK (Northern Ireland). Early and middle childhood (4-11 years). Evaluation of a school-based child maltreatment prevention programme, <i>Keeping Safe</i> , designed by the NSPCC, a charity.

The interventions

Programme	Country	Description
<i>The Good School Toolkit (GST)</i>	Uganda	‘Whole school’ 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>Interaction Competencies with Children for Teachers (ICC-T)</i> (protocol)	Tanzania (completed study) Uganda	Week-long workshop for secondary school teachers 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. School children are not involved.
<i>Irie Classroom Toolbox (ICT)</i> (protocol)	Jamaica	Prevention programme that aims to reduce aggressive behaviours by preschoolers and violent discipline by their teachers. ICT training provides teachers with a low-cost toolkit comprising various options and strategies to address children’s behaviour and to teach pupils better social and emotional skills. Training includes five full-day workshops; monthly in-person coaching; and encouragement via text messages. The toolkit also has booklets, story cards and play cards.
<i>Red Flag / Green Flag People</i>	USA	Staff and parent sessions followed by two classroom training sessions of 1.5 hours each (Kolko 1989 did not have parent sessions). The <i>Green Flag / Red Flag People</i> colouring book promotes strategies such as how to say no to an adult, how to get away from a perpetrator,

		and how to tell someone about the experience of an actual abusive incident. A film entitled "Better Safe than Sorry II" is also presented.
<i>Keeping Safe (protocol)</i>	UK (N. Ireland)	'Whole-school' programme for children aged 4-11 on how to keep safe from any type of maltreatment. Teaching and learning resources are included within 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 from grade one to grade seven. School leaders and parents are also actively involved in the programme. Training and support provided for school staff.
<i>Bringing in the Bystander – High School Curriculum (BITB-HSC)</i>	USA	Classroom-based intervention for high school 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 intervene when they witness aggression).
<i>Unnamed gender-based violence/sexual harassment (GV/SH) prevention programme</i>	USA	Two separate curricula both with the aim of reducing GV/SH or "dating violence" among sixth and seventh graders. The first, an interaction-based curriculum, focused on emphasizing better relationships and behaviours. The second, a law and justice curriculum, presented facts and knowledge on GV/SH including definitions, relevant laws, and legal consequences.

All interventions are school-based. The primary aims of the interventions can be classified into three broad categories:

- 1) To prevent corporal punishment in schools, i.e., teachers use of violence to enforce discipline: *The Good School Toolkit* in Uganda (5 papers); *Interaction Competencies with Children for Teachers* in Tanzania (1 study) and Uganda (1 protocol); the *Irie Classroom Toolbox* in Jamaica (1 protocol).
- 2) To prevent child abuse: *Red Flag / Green Flag People* (2 studies); *Keeping Safe* (1 protocol).
- 3) To prevent children from becoming victims and perpetrators of violence: *Bringing in the Bystander-High School Curriculum* (1 study) and an intervention to prevent gender-based violence / sexual harassment among middle-schoolers (1 study).

Who delivers the intervention?

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

Have the interventions been implemented at scale?

GST is promoted by the NGO Raising Voices which has implemented it in primary schools of one district and is planning to implement it in secondary schools in Uganda and it is being adapted for Tanzania. The

scale has been limited to date. *ICC-T* does not seem to have been implemented widely. 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. The *GV/SH* intervention was developed by researchers in consultation with school personnel for testing within a small number of schools.

What do the interventions cost?

GST: A separate study by Greco et al³⁰ found *GST* to be cost-effective. 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.

Other programmes: While 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

GST 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 planning and implementation of behaviours to deal with the problem. The final stage is when new behaviours have become the norm.

ICC-T too probably has a similar conceptual approach although it is not specifically stated in the study.

BITB-HSC is based on multiple models such as the transtheoretical model, the health belief model, the theory of planned behaviour, and the diffusion of innovation theory. The theory of reasoned action forms the basis for the *GV/SH* prevention programme.

Do the interventions work in preventing child maltreatment?

1) To prevent corporal punishment in schools: Both studies here – *GST* and *ICC-T* – reported positive results. However, results need to be interpreted with caution since both studies have high risk of bias.

Students in the *GST* arm had a 42% reduction in the risk of experiencing physical violence in the past week from teachers compared to the control group. 31% in the intervention group reported past week physical violence from school staff, versus 49% in the control group (the baseline levels were >50% in both groups); 60% of the *GST* students reported past school term violence experienced compared to 81% in the control group; and just over 15% *GST* school teachers said that they used physical violence in the past week compared to 33% of the control group teachers.

Shown below are results from the *GST* study on different groups and situations.

- Gender: After the 18-month *GST* intervention, boys and girls in the intervention group reported less violence experienced in the past week (boys: 36% vs 58%; girls: 44% vs 56%) and past school terms (boys: 63% vs 87%; girls: 72% vs 84%) from either staff or their peers compared to the control

³⁰ 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.

group. Similarly, students reporting an injury inflicted by school staff in the past week (boys: 13% vs 27%; girls: 20% vs 30%) or past school term (boys: 59% vs 68%) was also lower for the GST group compared to controls. However, the proportion of girl students reporting an injury over the past school term was similar for both groups (approximately two-thirds of girl students). The results suggest that GST might have worked better for boy students than girls.

- Disabilities/functional difficulties: GST students with or without disabilities or functional difficulties reported less violence experienced at school in the past week or past term compared to controls.

Disability / Functional difficulties	Violence in past week		Violence in past term	
	GST	Control	GST	Control
Disability	59%	84%	81%	89%
Some functional difficulties in one domain	64%	44%	69%	90%
No functional difficulties	38%	54%	66%	84%

- Level of exposure to GST and schoolwide prevalence of violence: The level of exposure which students had to GST was related to physical violence by staff (teachers and others). The higher the exposure, the less the odds of violence. Prevalence of violence in GST schools ranged from 7% to 65% at the end of the intervention (the average baseline prevalence was over 50%).
- Students who experience other violence, with poor mental health and those having difficulty with self-care were identified to be at high risk of violence even after the intervention.
- Violence at home: There was no difference between the GST group and the control group on physical violence or emotional violence experienced at home (as reported by the children themselves and their caregivers).

Results in the ICC-T study were also similar. Teachers in the ICC-T arm and their students reported substantially less use of and less exposure to physical and emotional violence respectively compared to the control arm.

- 2) To prevent child abuse:** Two studies – both of the *Red Flag / Green Flag People* programmes. 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% (NB, the sample size is so small that this 6% is only one child), 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 on 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. Of course, that increased reporting may be because of more abuse, or, on the other hand, the heightened awareness in the treatment group means that they were more likely to recognise abuse and know (how to) report it. It is hard to figure out whether the programme is effective given these results.
- 3) To prevent children from becoming victims and perpetrators of violence:** *BITB-HSC* 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. For the GV/SH prevention intervention no significant differences were seen for dating violence victimisation or sexual harassment between intervention and control arms. Surprisingly, students in intervention arms reported committing significantly *more* violence against their dating partners than control. The study suggests that this might be due to increased sensitivity in the intervention groups due to the programme (and hence more reporting). Some of the measures of GV/SH attitudes and knowledge improved significantly compared to controls especially for students in the law and justice curriculum group but many of these effects faded with time.

Evidence of adverse effects

None of the studies reported any adverse effects from the interventions – beyond the ambiguous results discussed above.

Will the results translate to other places?

The studies on preventing corporal punishment are from Uganda and Tanzania. However, they were both implemented in a small number of schools limiting the generalisability of results. More such studies are planned, the results of which may have indications about generalisability.

The inconsistent results for the other studies prevent generalising findings to other settings.

How reliable is the evidence?

Not very. All the RCTs in the cell are rated as **high risk of bias**. Both QEDs 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
Devries 2015	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
Devries 2017	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
Devries 2018	High risk of bias	Low risk	Some concerns	High risk	High risk	Some concerns
Edwards 2019	High risk of bias	Some concerns	Some concerns	High risk	High risk	Some concerns
Knight 2018	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
Merrill 2018	High risk of bias	Low risk	Some concerns	Some concerns	High risk	Some concerns
Nkuba 2018	High risk of bias	Some concerns	High risk	High risk	Some concerns	Some concerns
Taylor 2010	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
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 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 RCT protocol (McElearney 2018), and 49 primary studies (and 1 paper) (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).

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

The guidebook has summaries for Taylor 2010, Czerwinski 2018.

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 come from the US with the rest from Canada, the UK, Europe, Australia, Central America, and East Asia. There are no studies from South Asia, Africa or South America.

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 also 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 PI, counselors, 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 levels will reduce 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 put the burden of prevention in many ways on the children themselves. The aim here is 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.

- Students' knowledge and awareness improves in the short-term with these programmes but the effect likely fades in the longer term. Regular trainings 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 certainly indicate that results are highly 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
Daigneault 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
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

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
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
Jin 2017	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
Pulido 2015	Some concerns	Some concerns	Low risk	Low risk	Low risk	Some concerns
Saslowsky 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
Tutty 1997	High risk of bias	Some concerns	Low risk	Low risk	High risk	Some concerns
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

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 of bias	Moderate	Low	Low	Low	Low	Moderate	Low
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	Serious	Low	Low	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
Weatherley 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
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 1991	Low risk of bias	Low	Low	Low	Low	Low	Moderate	Low

Synder 1986	Low risk of bias	Low	Low	Low	Low	Low	Moderate	Low
Taal 1997	Low risk of bias	Low	Low	Low	Low	Low	Moderate	Low
Warden 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

Evidence status	Moderate risk of bias	Moderate strength evidence that sexual abuse prevention interventions do not increase anxiety among young children.
------------------------	------------------------------	---

The summary in brief

This cell includes studies that report on mental health outcomes for children participating in various types of programme. They include studies of (1) programmes for younger children delivered in the classroom to prepare them to recognize, avoid, and report sexual abuse; (2) programmes focused on preventing use of violence as punishment by teachers in school (corporal punishment) in settings where it is widely prevalent and; (3) prevention programmes for older children (adolescents) to prevent them for sexually harassing others. The bulk of the evidence in this cell is from the first category (13 programmes) with only one each in the other two categories.

Most of the sexual abuse prevention programmes in these studies are from the US. The evidence suggests that these programmes do not increase anxiety in children (considered a “side effect” of the intervention) and could improve self-esteem of students.

Studies from the latter two categories are too few to draw any conclusions.

Contents of the cell

This cell includes 15 primary studies (and two papers) (Devries 2015, Knight 2018³¹, Ratto 1990, van Lieshout 2019, del Campo Sanchez 2006, Fryer 1987, Kraizer 1988³², Grendel 1991, Hazzard 1991, Oldfield 1996, White 2018, Wurtele 1992a, Czerwinski 2018, Weatherley 2012, MacIntyre 1991, Taal 1997) and three protocols for RCTs (Baker-Henningham 2016, McElearney 2018, Ssenyonga 2018).

The guidebook has summaries for Devries 2015, Czerwinski 2018 (both completed primary studies) and Baker-Henningham 2016 (a protocol).

Devries 2015, Knight 2018 RCT, high risk of bias	Uganda (21 primary schools in one district). Children in 5 th -7 th classes: Early adolescence (11-14 years) Evaluation of <i>The Good School Toolkit (GST)</i> – a whole-school approach to preventing the use of violence as discipline by teachers against students. The intervention ran for 18 months.
van Lieshout 2019 RCT, high risk of bias	The Netherlands (20 youth care institutions). Adolescent boys (12-17 years of age) <i>Testing the Make a Move</i> programme to prevent sexual harassment and promote respectful relationships for at-risk boys living in residential care
Ratto 1990, RCT, high risk of bias	Tanzania (four secondary schools). Early and late adolescence. (11-17 years) Evaluation of the <i>Interaction Competencies with Children for Teachers (ICC-T)</i> intervention – a training workshop for secondary school teachers to reduce their use of corporal punishment

³¹ Devries 2015 and Knight 2018 are papers from the same RCT on *The Good School Toolkit (GST)*

³² Fryer 1987 and Kraizer 1988 are papers from the same RCT on the *Children Need To Know Personal Safety Program*

del Campo Sanchez 2006, Fryer 1987, Kraizer 1988, Grendel 1991, Hazzard 1991, Oldfield 1996, White 2018, Wurtele 1992a. RCTs, moderate risk of bias	Seven intervention programmes from US with one each from Spain, Germany, the Netherlands, Malaysia, Australia and the Republic of Ireland. Middle childhood (6-11) and early childhood (2-5 years) Evaluation of various classroom-based sexual abuse prevention programmes.
Czerwinski 2018, Weatherley 2012 QED, moderate risk of bias	
Kraizer 1991 QED, high risk of bias	
MacIntyre 1991, Taal 1997 QED, low risk of bias	
Ssenyonga 2018 RCT protocol only, no results available	Uganda (six secondary schools). Early and late adolescence. (11-17 years) Planned evaluation of <i>ICC-T</i> in southwestern Uganda.
Baker-Henningham 2016 RCT protocol only, no results available	Jamaica (38 preschools). Early childhood (24 months-5 years) Planned evaluation of the <i>Irie Classroom Toolbox (ICT)</i> – a training programmes for preschool teachers to prevent use of violence
McElearney 2018 RCT protocol only, no results available	UK (Northern Ireland). Early and middle childhood (4-11 years). Evaluation of a school-based child maltreatment prevention programme, <i>Keeping Safe</i> , designed by the NSPCC, a charity.

The interventions

All but one of the interventions are school-based (one intervention took place in a residential care setting for boys). The primary aims of the interventions can be classified into three broad categories:

1) Child abuse prevention programmes:

- Number of programmes: 13 programmes were assessed (one programme has two papers).
- Countries: Seven programmes were from the US with one each from Spain, Germany, the Netherlands, Malaysia, Australia and the Republic of Ireland.
- Participants: Most programmes targeted students in middle childhood (6-11 years) although a few also included early childhood students (2-5 years). Teachers were also often part of the intervention (many times they were the ones who delivered the intervention in their classrooms). In some cases, parents were also involved.
- Interventions: The interventions typically consist of classroom-based education aimed at increasing the knowledge of young children on prevention concepts (related to sexual abuse) and also providing 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 duration (a few weeks to a few months).

2) Corporal punishment prevention programme:

- Number of programmes: One programme (*The Good School Toolkit*) with two papers
 - Country: Uganda
 - Participants: All primary school children, teachers, school administrators and parents were involved but the assessment of impact was limited to early adolescents (11-14 years).
 - Intervention: a universal programme that runs for almost two full school years. It aims to prevent physical violence from school staff towards primary school children. It is an intensive whole-school approach that engages students, teachers, administrators and parents in creating a culture that moves away from violence as a form of punishment.
- 3) **Programme focused on preventing adolescents from sexually harassing others:**
- Number of programmes: One programme (*Make a Move*)
 - Country: The Netherlands
 - Participants: Adolescent boys (12-17 years) living in youth residential care settings.
 - Interventions: 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. It has a cognitive-behavioural approach and tackles determinants of sexual harassment such as attitudes and socio-relational skills.

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, teachers in the classroom or in one case 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?

GST: A separate study by Greco et al³³ found *GST* to be cost-effective. 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.

³³ 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.

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 within 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 here 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”) were also improved for intervention group children compared to controls albeit from few studies. 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:** From the papers on *GST*, mental health and wellbeing status remained mostly the same after intervention and was 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 corporal punishment; preventing sexual harassment) the small number of studies means that we cannot be confident about generalisability.

How reliable is the evidence?

Moderately so: see below.

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
Devries 2015	High risk of bias	Low risk	Some concerns	Low risk	High risk	Some concerns
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
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

Prevention / Child Wellbeing: Social-Emotional Functioning

Mixed results for social-emotional functioning from school-based interventions on preventing child sexual abuse.

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 being empowered to disclose abuse when it happens. One outcome studied is the impact of these trainings on children’s social-emotional functioning.

This cell has five primary studies – four of which are school-based child abuse prevention interventions. Two studies found that positive behaviours increase after training (as observed by parents and teachers) such as better conflict resolution skills, being more communicative and assertive while negative behaviours did not increase (e.g., not able to sleep, fear of known and unknown adults). The other two studies found no such change after intervention. The fifth study was also a prevention programme but to prevent adolescent boys living in residential homes from sexually harassing others.

This study found no effect of the intervention compared to controls.

Contents of the cell

The cell has five studies:

Hebert 2001 ³⁴ 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 ³⁵ RCT, moderate risk of bias.	Canada (three schools in Montreal). Children in first and third grades: Middle childhood (6-11 years) Evaluation of <i>ESPACE</i> + booster (two years later) on violence prevention.
Taal 1997 ³⁶ 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.

³⁴ Hebert M, Lavoie F, Piche C, Poitras M. Proximate effects of a child sexual abuse prevention program in elementary school children. *Child Abuse & Neglect* 2001;25(4):505-22.

³⁵ Daigneault I, Hébert M, McDuff P et al. (2012) Evaluation of a sexual abuse prevention workshop in a multicultural, impoverished urban area. *Journal of Child Sexual Abuse*, 21(5): 521–542.

³⁶ Taal M, Edelaar M (1997), Positive and negative effects of a child sexual abuse prevention program. *Child Abuse Negl* 21:399–410

van Lieshout 2019 ³⁷ 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 ³⁸ 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.

The interventions

Programme	Country	Description
<i>ESPACE</i> (Hebert 2001)	Canada	Quebec adaptation of the <i>American Child Assault Prevention (CAP)</i> program ³⁹ . Classroom workshop (60-75 minutes) delivered by specialised community workers to first and third graders. Role-play, guided discussion, behaviour modelling and rehearsal are strategies used for workshop to teach children basic prevention concepts and skills and make them aware of their personal rights. Children are taught to be assertive, a self-defence yell and are encouraged to reach out to friends and a trusted adult if any abuse occurs. The programme 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 + booster</i> (Daigneault 2012)	Canada	Same <i>ESPACE</i> programme as described above 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 from <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. Programme is delivered to students in grades six to eight. The themes include 'yes' or 'no' feelings; right to refuse unwanted sexual behaviours; and 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

³⁷ van Lieshout, S., Mevissen, F. E. F., van Breukelen, G., Jonker, M., & Ruiter, R. A. C. (2019). Make a Move: A Comprehensive Effect Evaluation of a Sexual Harassment Prevention Program in Dutch Residential Youth Care. *Journal of Interpersonal Violence*, 34(9), 1772–1800. <https://doi.org/10.1177/0886260516654932>

³⁸ Del Campo, A. & López, F. (2006). Evaluación de un programa de prevención de abusos sexuales a menores en Educación Primaria. *Psicothema*, 18, 1-8.

³⁹ Cooper, S. J. (1991). *New strategies for free children: child abuse prevention for elementary school children*. Columbus, OH: The National Assault Prevention Center.

		facilitate 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 sociorelational skills. It consists of eight themed weekly meetings of 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	First of its kind, it was a school-based sexual abuse prevention programme in Spain. Programme was delivered in the classroom to children aged 8-12 years. Teachers and parents received some training / orientation before the programme.

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

Hebert 2001 reported that two weeks after participating in *ESPACE*, most children did not show any negative side effects (sleeping problems; being socially isolated; afraid of known / unknown adults; disobedient; clingy; aggressive to sibling / 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 (aggressive towards siblings was highest at close to 10%). For positive effects, more than half the parents said their children were more forthcoming on what they liked (57%) and did not like (53%) and 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.

Daigneault 2012 found no significant differences in self-efficacy and empathy scores between the *ESPACE* + *CSR* group and controls.

Taal 1997 reported that subtest scores for “Relationships with Teacher” and “Relationships with Classmates” were unchanged one week and six weeks after the programme. 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 differences between intervention and control groups for any of the 14 measures in the study. Examples of the measures included self-efficacy, self-esteem, social norms and empathy.

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 improved 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.

Have the interventions been implemented at scale?

ESPACE seems to have been implemented for hundreds of thousands of Quebecers over the years. However, the evaluation seen in the studies is in 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 targeted at at-risk youth living in residential care. It is not clear from the studies whether *Right to Security* or *Prevención de abusos sexuales a menores* have been implemented at scale.

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 costs.

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

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

Are the results generalisable?

Hard to say because the results are mixed across the five studies.

How reliable is the evidence?

Pretty reliable.

Taal 1997 is rated as *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 *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
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

Prevention / Parent Caregiver: Knowledge and Awareness

School-based interventions to tackle abuse can improve parent / caregiver knowledge and attitudes, but more RCTs are needed to establish effectiveness.

Evidence status	Moderate risk of bias	Some evidence of impact on parent knowledge and attitudes. Worth testing different approaches to achieving this outcome using strong study designs.
------------------------	------------------------------	---

The summary in brief

School-based interventions to reduce child abuse include training for teachers and parents. The impact on parent / caregiver knowledge and attitudes vary, possibly on account of the varying intensity and duration of the intervention.

Contents of the cell

The cell has four studies (Kolko 1987⁴⁰, Kolko 1989⁴¹, MacIntyre 1991⁴², Merrill 2018⁴³) and one protocol (McElearney 2018⁴⁴). A full summary of Merrill 2018 is available in the guidebook.

Kolko 1987 QED, low risk of bias.	US (three schools in Pennsylvania). Children in third and fourth grades: Middle childhood (6-11 years) Evaluation of the <i>Red Flag / Green Flag Program</i> to raise awareness and prevention of child sexual victimisation.
Kolko 1989 QED, low risk of bias.	US (six schools in Pennsylvania). Children in third and fourth grades: Middle childhood (6-11 years) Also implemented <i>Red Flag / Green Flag</i> but in more schools with a few changes (parents were not directly involved in the programme in this study but were in Kolko 1987)
MacIntyre 1991 QED, low risk of bias.	Republic of Ireland (Dublin: five suburban schools). Children in middle childhood (6-11 years). Evaluation of the effectiveness of the <i>Stay Safe</i> primary prevention programme for child abuse.
McElearney 2018	UK (Northern Ireland). Children between 4-11 years.

⁴⁰ Kolko, D.J., Moser, J.T., Litz, J. *et al.* Promoting awareness and prevention of child sexual victimization using the *Red Flag / Green Flag* program: An evaluation with follow-up. *J Fam Viol* **2**, 11–35 (1987). <https://doi.org/10.1007/BF00976368>

⁴¹ Kolko, D.J., Moser, J.T. & Hughes, J. Classroom training in sexual victimization awareness and prevention skills: An extension of the Red Flag / Green Flag people program. *J Fam Viol* **4**, 25–45 (1989). <https://doi.org/10.1007/BF00985655>

⁴² MacIntyre, D., & Carr, A. (1999). Evaluation of the effectiveness of the Stay Safe primary prevention programme for child sexual abuse. *Child Abuse & Neglect*, **23**, 1307-1325.

⁴³ Merrill, K. G., Knight, L., Namy, S., Allen, E., Naker, D., & Devries, K. M. (2018). Effects of a violence prevention intervention in schools and surrounding communities: Secondary analyses of a cluster randomized control trial and Uganda. *Child Abuse & Neglect*, **84**, 182-195

⁴⁴ McElearney, A., *et al.*, 2018. Cluster randomised controlled trial of ‘whole school’ child maltreatment prevention programme in primary schools in Northern Ireland: study protocol for keeping safe. *BMC public health*, **18** (1), 590. doi:10.1186/s12889-018-5492-8

RCT protocol: the study is still ongoing, so no results are reported	Evaluation of a school-based child maltreatment prevention programme, <i>Keeping Safe</i> , designed by the NSPCC, a charity.
Merrill 2018 RCT, high risk of bias	Uganda (one district). Primary school aged children between 11-14 years. The <i>Good School Toolkit (GST)</i> . Secondary analysis of the <i>GST</i> which is about violence prevention intervention in schools and surrounding communities.

The interventions

All four programmes are delivered to primary school students and their teachers and parents. Two of the programs are personal safety programmes, which seek to increase children’s awareness of the nature of sexual abuse and equip them with the skills to respond. The other two – the *Good School Toolkit* and *Keeping Safe*– are more intensive and intended as ‘whole-school’ approaches to change school culture to reduce acceptance of any form of child abuse. Note that *Keeping Safe* is a protocol for an RCT so no results are available.

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 for a group setting. They address mutual respect, power relations, non-violent discipline techniques, and 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 followed by two classroom training sessions of one and 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 following by classroom implementation of 10-12 sessions of 30-40 minutes each. Teacher (two sessions) and Parent training (one session) includes sexual abuse prevention concepts such as definitions, myths, realities, prevalence and information on typical victim and offender characteristics. It also includes tips on how to identify victims and support them with disclosure and for appropriate referral to legal and social services.

		<p>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 (protocol only)</i>	Northern Ireland	<p>'Whole-school' programme for children aged 4-11 on how to keep safe from any type of maltreatment. Teaching and learning resources are incorporated within 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>

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

Caregivers involved with the *Good School Toolkit* programme were surveyed two years after the programme (the programme itself 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 same scale) compared to controls, although the difference was quite small. No impact was seen on their view of sexual abuse (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). Importantly, 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* applied a 38-item Parents' Knowledge and Attitudes Questionnaire. Significant improvements were recorded for eight of the eighteen items, these items reflecting belief in children's statements, attitudes towards prevention programs, and knowledge about help-seeking. Parents' knowledge and attitude changes were not related to their age or gender.

It is difficult to gauge the impact of *Red Flag / Green Flag People* since Kolko 1987 did not assess the actual knowledge or awareness of parents on child sexual abuse but rather asked them how knowledgeable or aware they thought they were. 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) was significantly higher than controls. The study also says that "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 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 actually decreased for the control group (which made the intervention group scores appear better).

Have the interventions been implemented at scale?

Personal safety programmes such as the *Red Flag / Green Flag People* colouring book are a widely adopted approach, although they vary in duration and intensity. This study of effectiveness, however, was carried out in a small sample.

The *Stay Safe* programme is delivered in most primary schools in the Republic of Ireland.

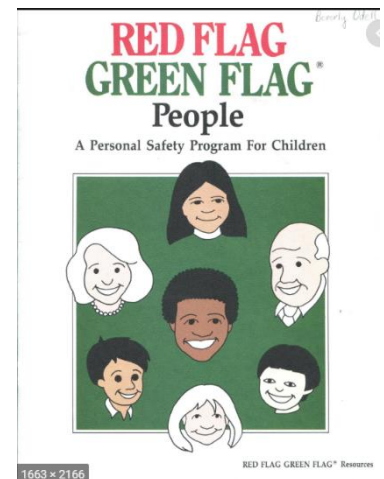
The *Good School Toolkit* is promoted by the NGO Raising Voices, which has implemented it in primary and secondary schools in Uganda and it is being tested in Tanzania. The scale has been limited to date.

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 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.”



What do the interventions cost?

Cost data for the *Good School Toolkit* is reported in a study by Greco et al⁴⁵. Implementing it over 18 months in 21 schools was close to \$400,000. Monitoring and evaluation add another \$50,000 to costs (for monitoring and evaluation for the whole programme). The annual cost to run the programme 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 *Red Flag / Green Flag* programme is noted in the study to be low cost, but actual cost data is not provided. No cost data is available for *Stay Safe*.

⁴⁵ 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.

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

Will the results translate elsewhere?

Probably not. The number of schools included in the studies are quite small. Some of the programmes might have been implemented on a large scale but they have not been evaluated at that level.

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

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 review⁴⁶ of 24 studies of school-based programmes which 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 and sexual abuse. The review did not report on parent / caregiver outcomes.

⁴⁶ Walsh K, Zwi K, Woolfenden S, Shlonksy A. School-based education programmes for the prevention of child sexual abuse. Cochrane Database of Systematic Reviews 2015, Issue 4. Art. No.:CD004380. DOI:10.1002/14651858.CD004380.pub3.

Response / Institutional Safeguarding Practice: Operations

Impact of response-focused interventions to improve institutional operations to safeguard children is promising. More RCTs needed to establish effectiveness.

Evidence status	High risk of bias	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 and day cares can implement various interventions and policies to safeguard the children in their care. Interventions to improve operations to protect children in institutions typically include training for staff or structural interventions (such as improving the ratio of caregivers to children).

While quite a few studies on such interventions are published, very few are RCTs (the EGM has only one). This cell includes two primary studies and one systematic review. Almost all studies trained caregivers to improve their knowledge, attitudes, and practices / behaviours on working with children. The settings for the studies included children’s homes, schools, and day care. Most results suggested moderate improvements in the quality of caregiving (compared to controls) although this was usually only in the short term. One large-scale study from Spain reported increased detection rates of child maltreatment after training of childcare professionals from different disciplines. Results need to be interpreted with caution because of the high risk of bias / low quality of studies.

More RCTs of these interventions are needed to better understand the impact on institutional operations to safeguard children.

The cell has two primary studies (Cerezo 2004⁴⁷; Rheingold 2014⁴⁸) and one systematic review (Hermenau 2017⁴⁹). A full summary of Rheingold 2014 is available in the guidebook.

Contents of the cell

A. Primary Studies	
Cerezo 2004 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

⁴⁷ Cerezo MA, Pons-Salvador G. Improving child maltreatment detection systems: a large-scale case study involving health, social services, and school professionals. *Child Abuse Negl.* 2004;28(11):1153-1169. doi:10.1016/j.chiabu.2004.06.007

⁴⁸ Rheingold, Alyssa & Zajac, Kristyn & Chapman, Jason & Patton, Meghan & de Arellano, Michael & Saunders, Benjamin & Kilpatrick, Dean. (2014). Child Sexual Abuse Prevention Training for Childcare Professionals: An Independent Multi-Site Randomized Controlled Trial of Stewards of Children. *Prevention Science : The Official Journal of the Society for Prevention Research.* 16. 10.1007/s11121-014-0499-6.

⁴⁹ Hermenau, K., Goessmann, K., Rygaard, N. P., Landolt, M. A., & Hecker, T. (2017). Fostering Child Development by Improving Care Quality: A Systematic Review of the Effectiveness of Structural Interventions and Caregiver Trainings in Institutional Care. *Trauma, Violence, & Abuse*, 18(5), 544–561. <https://doi.org/10.1177/1524838016641918>

Rheingold 2015 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 <i>Stewards of Children</i> programme to prevent CSA
B. Systematic Review	
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 such 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 (Cerezo 2004 and Rheingold 2014)

The interventions

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: 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). 181 professionals were trained representing all relevant agencies in the area. One professional was trained for every 926 children living in the region.
- Phase 2: This phase focused on training teachers, psychologists, and support staff in preschools and primary schools. 210 teachers meant to represent all preschools and primary schools and support staff were included in the training.

The training for professionals in both phases included 16-20 hours training over 2-3 days, covering parenting best practices, definition of child maltreatment, dimensions of the problems, physical and behavioural signs of maltreatment and the protocol to follow to report cases to CPS. The trainees were also sensitised to the various aspects of child maltreatment. A form was developed for streamlined referral of potential cases to CPS. The other aspect of the intervention was support provided to trained professionals by a Local Coordination Team (LCT) on reporting cases to CPS. Support was provided through a helpline and in-person visits.

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). The programme involves a two-and-a-half-hour workshop for adults in childcare settings to train them on prevention, recognition, and response to CSA. *Stewards of Children* aims to improve childcare professionals’ ability to prevent CSA from happening (primary prevention) and to recognise and respond to CSA (secondary prevention). The programme is offered in two formats: in-person and interactive web-based training.

The in-person training is a two-and-a-half-hour group training led by a trained facilitator. The training covered the following topics: (i) knowledge on CSA prevalence rates, risks and outcomes; (ii) strategies to reduce CSA opportunities; (iii) methods to bring up CSA with adults and children; (iv) identifying signs of CSA; (v) responding appropriately when a child comes forward with CSA disclosure; (vi) addressing barriers to preventive actions at the individual and organisational level; and (vii) community involvement in CSA prevention. A DVD was used to show experiences of CSA survivors combined with facilitated discussion. The web-based training was delivered over two weeks and is comparable in content and length to the in-person training.

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

The intervention in Cerezo 2004 led to an increase in the number of cases reported to CPS from both its phases. The study reports a tripling of detected cases after the intervention compared to before. The second phase, i.e., training teachers, led to detection of two to three 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. An increase in referral before and after intervention was seen for the first two territories but not for the third territory. The authors thought that this might be due to knowledge of the intervention spreading to the third territory before it was implemented there via mass media, professional networks and professionals moving territories for new jobs. The outcomes were then compared with a different region of the Balearic Islands and the expected increase in referrals was seen.

The results in Rheingold 2014 demonstrate that *Stewards of Children* improved knowledge and behaviours.

- ✓ **Knowledge** about CSA increased. That knowledge declines over the three months after the training (as one would expect), though oddly, the knowledge of the control ('waitlist') group increased during that time, but still not to as high as the trained group.
- ✓ **Attitudes.** Participants' belief in CSA myths was low to begin with so there was little room for improvement. After training, the control group had the better score but at three months there was no difference between groups.
- ✓ **Behaviours.** This also improved, i.e., participants reported having done more of the behaviours three months after the training than did people in the control group. The behaviours most improved were:
 - "limiting the opportunity for other youth and younger youth to have one-to-one interaction". This is significant because juveniles are offenders in more than a third of CSA (Finkelhor et al. 2009⁵⁰) and
 - "sharing with another adult an article, brochure, or other information about CSA prevention". Interestingly, the behaviour of people in the control group also improved during the three months after the training: quite possibly because colleagues who had

⁵⁰ Finkelhor, D., Ormrod, R., & Chaffin, M. (2009). Juveniles who commit sex offenses against minors. Office of Juvenile Justice and Delinquency Prevention. <http://www.ncjrs.gov/pdffiles1/ojjdp/227763.pdf>.

received the training changed their behaviour (making it more normal) and they shared this information with colleagues who had not received the training.

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 had the group trained online. Three months after training, however, there were no differences between the two groups.
- ✓ **Attitudes.** No difference between the group trained in-person vs the group trained online.
- ✓ **Behaviours.** No difference between the group trained in-person vs the group trained online.

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

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.

Rheingold 2014 offers no information on scale of implementation.

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.⁵¹

What do the interventions cost?

Neither study report any cost data.

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

Cerezo 2014 does not mention any specific theory for their programme. The approach adopted was based on the principles of motivational interviewing to overcome barriers and change attitudes on reporting potential cases of maltreatment.

Stewards of Children is not a theory-based prevention programme, but its principles are in line with Finkelhor’s⁵² theory that for CSA to occur, certain preconditions exist. They include: an individual’s tendency to abuse, absence of internal or external inhibitions for the offender, and the offender having

⁵¹ <https://www.d2l.org/get-trained/>

⁵² Finkelhor, D. (1984). *Child sexual abuse: New theory and research*. New York: Free Press.

access to the child. Preventing one or more of these preconditions should reduce the likelihood of CSA. This programme aims to reduce access to children (by offenders) and to increase external barriers for offenders by improving the knowledge, attitudes, and response of adults responsible for childcare.

Are the results generalisable?

We have one large scale study (QED) and one RCT from multiple sites - both from high-income countries (Spain and Canada). Given that the environment for schools, daycare, CPS, hospitals, etc., is reasonably similar in these countries, the findings are likely to translate to similar settings.

How reliable is the evidence?

Not very, as explained below.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study	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

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
Cerezo 2014	Serious	Serious	Moderate	Moderate	Serious	Low	Moderate	Serious

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 looked 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 five 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 it included was wide – from less than four 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) with dependent designs reported a wide-range of effects which the systematic review authors describe as “very large” to “small”. Two 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) with independent designs (no RCTs) reported caregiving quality or attachment. Three studies had caregiver training and structural changes (typically reducing the number of children cared for by each caregiver and improving facilities such as bathrooms) as the intervention; one only had 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 systematic review’s chosen method of analysis). Two other papers on the same intervention (in the same systematic review) 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, more 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 countries. Almost all used caregiver training as the intervention with mostly positive results (although results some were not 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 **low**. 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. They also depended on a different meta-analysis paper for some of their calculations which could add errors to their estimates.

Response / Child Wellbeing: Mental Health

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

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. Very few studies on response interventions and impact on mental health outcomes are available. None are RCTs. Three studies included in two of the systematic reviews were all 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 the last study (from Portugal), more than half of children in institutional care reported a suicide attempt.

More studies of response interventions especially in institutional care settings are needed to understand impact on mental health.

The cell has and three systematic reviews (Hermenau 2017⁵³, Sherr 2017⁵⁴, Radford 2017⁵⁵).

Contents of the cell

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.

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

What are the systematic reviews about?

⁵³ Hermenau, K., Goessmann, K., Rygaard, N. P., Landolt, M. A., & Hecker, T. (2017). Fostering Child Development by Improving Care Quality: A Systematic Review of the Effectiveness of Structural Interventions and Caregiver Trainings in Institutional Care. *Trauma, Violence, & Abuse*, 18(5), 544–561. <https://doi.org/10.1177/1524838016641918>

⁵⁴ Sherr L, Roberts KJ, Gandhi N. Child violence experiences in institutionalised/orphanage care. *Psychol Health Med*. 2017;22(sup1):31-57. doi:10.1080/13548506.2016.1271951

⁵⁵ Radford, Lorraine et al. “Rapid Evidence Assessment: What can be learnt from other jurisdictions about preventing and responding to child sexual abuse.” (2017).

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.

Are results generalisable?

No. There are too few studies to consider generalisability. More studies are needed to answer this question.

How reliable is the evidence?

Not very. The quality of each of the three systematic reviews is rated as **low**. 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.

What else is known from other studies about response interventions and mental health outcomes?

Not much. There are very few studies (and no RCTs) that report on mental health outcomes from response interventions. The available evidence shows mixed results on mental health although it is quite clear that institutional care affects children's mental health adversely. The *Bucharest Early Intervention Project (BEIP)* - with multiple papers on the EGM⁵⁶ - placed children from terrible orphanages in Romania into foster care. Foster care children, predictably, did much better than institutional care children on almost all 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. The earlier the intervention the more beneficial was the intervention. *BEIP* is classified as a treatment intervention on the EGM but it provides insight on an effective intervention for children in institutional care.

⁵⁶The six papers on this trial are: Johnson 2010, Smyke 2010, Bick, 2015, Humphries 2015, Troller-Renfree 2015 and Wade 2018

Treatment/ Child Wellbeing: Mental Health

High-quality foster care which removes children from terrible institutional care and cognitive behavioural therapy (CBT) for sexually abused children 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 three studies assessing the impact 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.

Two papers are written about *The Bucharest Early Intervention Project (BEIP)*, a 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. 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. Both groups were compared with children who had always lived at home with their birth families (never institutionalised group = *NIG*).

Mental health outcomes were measured when children were about eight years old and again when they were 12 years old. They suggest that children in the foster care group (*FCG*) had better outcomes than children who remained in institutional care (care as usual group = *CAUG*) on many (though not all) outcomes. the never institutionalised group (*NIG*) 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.

The other primary study and the systematic review in this cell examined psychotherapy as a treatment for children who had suffered sexual abuse. Cognitive behavioural therapy (CBT) with a focus on trauma is useful in helping children overcome their adverse experiences. The low number of RCTs was identified as a research gap.

The cell has two primary studies (two papers about *BEIP*: Troller-Renfree 2015⁵⁷, Humphreys 2015⁵⁸) and Sullivan 1992⁵⁹, and one systematic review (Radford 2017⁶⁰). A full summary of three *BEIP* papers (Johnson

⁵⁷Troller-Renfree S, McDermott JM, Nelson CA, Zeanah CH, Fox NA. The effects of early foster care intervention on attention biases in previously institutionalized children in Romania. *Dev Sci*. 2015;18(5):713-722. doi:10.1111/desc.12261

⁵⁸Humphreys KL, Gleason MM, Drury SS, et al. Effects of institutional rearing and foster care on psychopathology at age 12 years in Romania: follow-up of an open, randomised controlled trial. *Lancet Psychiatry*. 2015;2(7):625-634. doi:10.1016/S2215-0366(15)00095-4

⁵⁹Sullivan PM, Scanlan JM, Brookhouser PE, Schulte LE, Knutson JF. The effects of psychotherapy on behavior problems of sexually abused deaf children. *Child Abuse Negl*. 1992;16(2):297-307. doi:10.1016/0145-2134(92)90036-q

⁶⁰Radford, Lorraine et al. "Rapid Evidence Assessment: What can be learnt from other jurisdictions about preventing and responding to child sexual abuse." (2017).

2010⁶¹, Smyke 2010⁶², Bick 2015⁶³) is available in the guidebook. There are six papers⁶⁴ in total on the *BEIP* RCT in the EGM.

Contents of the cell

A. Primary Studies	
Troller-Renfree 2015, Humphreys 2015 RCT, moderate risk of bias	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. Two studies from <i>BEIP</i> reporting on mental health outcomes in the long-term (age 8-12 years).
Sullivan 1992 QED, moderate risk of bias	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
B. 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. Included studies in review relevant to this cell are from the US and Canada.

A. Primary Studies (Troller-Renfree 2015, Humphreys 2015 and Sullivan 1992)

The interventions

Troller-Renfree 2015 and Humphreys 2015 are evaluations of *The Bucharest Early Intervention Project (BEIP)*. Under the Ceausescu dictatorship in Romania (till 1989), abandoned children lived in dreadful orphanages⁶⁵. 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.

⁶¹ 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

⁶² 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

⁶³ 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

⁶⁴The six studies on this trial are: Johnson 2010, Smyke 2010, Bick, 2015, Humphries 2015, Troller-Renfree 2015 and Wade 2018

⁶⁵ Weir, K: American Psychological Association. (June,2014). The lasting impact of neglect. Retrieved from <https://www.apa.org/monitor/2014/06/neglect>

BEIP was created in the year 2000 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 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. 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.

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 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.

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

The two papers about *BEIP* studied two slightly different aspects of mental health. Troller-Renfree 2015 looked at attention bias which is considered as a precursor to mental illnesses such as anxiety disorder and depressive disorders. The primary outcome in Humphreys 2015 was that symptom counts for different 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 task⁶⁶' 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.

CAUG (n=50, i.e., the group had 50 people) showed a significant bias towards threat stimuli while FCG (n=55) showed that towards positive stimuli, but when all three groups (CAUG, FCG, NIG) were compared to each other there were no significant differences. NIG (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.

⁶⁶Bradley, B.P., Mogg, K., White, J., Groom, C., & De Bono, J. (1999). Attentional bias for emotional faces in generalized anxiety disorder. *British Journal of Clinical Psychology*, 38 (3), 267–278. doi:10.1348/014466599162845

The size of the threat bias, however, 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)⁶⁷ 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 CAUG (n=55) fulfilled criteria for “any psychiatric disorder,” 27% for those in “stable foster care” i.e., children from FCG who continued in *BEIP* foster care, 43% for those in “disrupted foster care” i.e., children from FCG where the foster care arrangement was changed (for e.g., placement into government foster care, readmitted to institutional care) and 16% for NIG (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 significantly lower for NIG 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 NIG but not significantly so. CAUG children (specifically boys in this group) had significantly higher externalising symptoms than FCG 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 studies took place quite a while after *BEIP* began: these studies were following-up with children in the long term, when they were age 8-12. In fact, the papers study a period which was four years long. 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.

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)

⁶⁷ Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME. NIMH diagnostic interview schedule for children vIV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *J Am Acad Child Adolesc Psychiatry* 2000; **39**: 28–38.

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?

Yes, for *BEIP*. It involved children from all six institutions providing care to abandoned children in Bucharest. 136 children were randomised to *BEIP* or continuing in institutional care.

However, 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.⁶⁸

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 will probably translate to places looking to ramp up support for fostering children.

Results from Sullivan 1992 might not translate elsewhere because it was only in one residential school.

How reliable is the evidence?

Moderately reliable.

⁶⁸ Zeanah, C.H., Nelson, C.A., Fox, N.A., Smyke, A.T., Marshall, P., Parker, S.W., & Koga, S. (2003). Designing research to study the effects of institutionalization on brain and behavioral development: the Bucharest Early Intervention Project. *Development and Psychopathology*, 15 (4), 885–907.

Risk of Bias for Randomised Controlled Trials (RCTs)

Study (Author and year)	Overall risk of bias	Randomisation process ⁶⁹	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

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 Review (Radford 2017)

What is the systematic review 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.

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:

⁶⁹ The various *BEIP* papers all concern one study, so the method of randomisation in them is the same. However, the various papers report different information about the randomisation process, which can lead to them scoring differently on the risk of bias of the randomisation process.

- 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.
- Longer duration of therapy; older age of children and tailoring therapy to an individual's specific needs are factors for increased benefit.

What information is available on cost and cost-effectiveness?

No information is provided on cost or cost-effectiveness.

Are results generalisable?

Probably. The studies are from US and Canada but CBT and other therapeutic approaches are likely available in other countries so the results will apply to such settings.

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.

Radford 2017 did not combine effect sizes statistically, so it is not clear what the overall impact is or if the results were consistent across included studies.

Treatment / Child Wellbeing: Cognitive Functioning

A foster care programme for young children (two to three years old) permanently removed from terrible institutional care in Bucharest, Romania, was effective in improving cognitive development (at 42-54 months). However, by adolescence, foster care children’s cognitive functional outcomes were still lagging behind 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 three primary papers (all on *BEIP*) which report cognitive functioning outcomes – Johnson 2010⁷⁰, Smyke 2010⁷¹ and Wade 2018⁷². A full summary of three *BEIP* papers (Johnson 2010, Smyke 2010, Bick 2015⁷³) is available in the guidebook. There are six papers⁷⁴ 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 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); 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

⁷⁰Johnson 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

⁷¹Smyke 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

⁷²Wade M, Fox NA, Zeanah CH, Nelson CA 3rd. Long-term effects of institutional rearing, foster care, and brain activity on memory and executive functioning. *Proc Natl Acad Sci U S A.* 2019;116(5):1808-1813. doi:10.1073/pnas.1809145116

⁷³Bick 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

⁷⁴The six studies on this trial are: Johnson 2010, Smyke 2010, Bick, 2015, Humphries 2015, Troller-Renfree 2015 and Wade 2018

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).

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> reporting 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).

The intervention

Under the Ceausescu dictatorship in Romania (until 1989), abandoned children lived in dreadful orphanages⁷⁵. 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 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 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.

⁷⁵ Weir, K: American Psychological Association. (June,2014). The lasting impact of neglect. Retrieved from <https://www.apa.org/monitor/2014/06/neglect>

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 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 (LBW) and delayed placement into foster care is clear. Low birth weight children (<2.5kg) placed in foster care after age two had 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). LBW infants in institutional care are especially vulnerable to cognitive deficits. The earlier the intervention (foster care), the better the chances 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 CAU children, 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 (CAU) and children who had always lived with their families at home (NIG). 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 and even income and psychosocial wellness in adulthood. The Cambridge Neuropsychological Test Assessment Battery (CANTAB)⁷⁶ was used to measure these cognitive domains.

⁷⁶ <https://www.cambridgecognition.com/cantab/why-choose-cantab/#:~:text=CANTAB%20accurately%20measures%20cognitive%20function,academic%20institutions%20around%20the%20world.>

1. *Attention and short-term visual memory:* HC performed 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 was no significant differences among groups. However, at age 16, HC did significantly better than the other two groups.
3. *Spatial working memory:* HC were able to strategise significantly better and make less 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 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 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, while it is plausible that children who have always lived at home (HC) will have an advantage compared to children who have ever been in institutional care (FC or IC), one would expect better outcomes for FC compared to IC (which was not the case here).

Have the interventions been implemented at scale?

Yes. *BEIP* involved 136 children from all six institutions providing care to abandoned children in Bucharest.

Which type of organisation delivered the intervention?

The intervention and the associated RCT was 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.⁷⁷

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.

⁷⁷ Zeanah, C.H., Nelson, C.A., Fox, N.A., Smyke, A.T., Marshall, P., Parker, S.W., & Koga, S. (2003). Designing research to study the effects of institutionalization on brain and behavioral development: the Bucharest Early Intervention Project. *Development and Psychopathology*, 15 (4), 885–907.

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
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	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 six papers on the EGM⁷⁸ - 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 was 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).

⁷⁸The six papers on this trial are: Johnson 2010, Smyke 2010, Bick, 2015, Humphries 2015, Troller-Renfree 2015 and Wade 2018

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.

Protocol: Baker-Henningham (2016) (Prevention / cognitive functioning, and adult desistance)

Light cell(s) in which this protocol appears:

Intervention = Prevention; Outcome = Child Wellbeing: Cognitive Functioning (1 study + 1 protocol)

Intervention = Prevention; Outcome = Adult Perpetrator or Offender: Desistance (1 protocol)

Summary: Protocol of universal violence prevention programme for preschool teachers in Jamaica.

{Because this is a protocol, the study has yet to be written up. Therefore, no results are available as yet, and we cannot comment on the quality risk of bias.}

The summary is based on **Baker-Henningham et al. (2016)**⁷⁹ 'Irie Classroom Toolbox: a study protocol for a cluster-randomised trial of a universal violence prevention programme in Jamaican preschools'

The summary in brief

The *Irie* (meaning "all is well" in Jamaican English) *Classroom Toolbox (ICT)* is a universal prevention programme that aims to reduce aggressive behaviours by preschoolers and violent discipline by their teachers. Corporal punishment, although illegal, is common in Jamaican early education institutions. Training preschool teachers can provide them with better skills and support to manage their pupils' problem behaviours without resorting to violence. Early intervention in children can help them develop better social and emotional skills that can benefit them even in adult life. Preschool enrolment in Jamaica is extremely high (97%) which makes it an ideal setting for intervention.

ICT training provides teachers with a low-cost toolkit comprising various options and strategies to address children's behaviour and to teach pupils better social and emotional skills. Training includes five full-day workshops supplemented by monthly in-person coaching for individual teachers and fortnightly encouragement via text messages. The toolkit also includes booklets, story cards and play cards to help teachers implement the plans they develop through their training.

38 preschools in urban neighbourhoods of Kingston, Jamaica were randomly assigned to *ICT* and 38 others were controls.

The primary outcomes are violence by teachers and children's aggressive behaviour at the classroom-level. These will be assessed by independent study personnel who are specifically trained for this task. Other outcomes include quality of the classroom environment, child and teacher mental health and child inhibitory control (a measure of impulse control and self-regulation).

⁷⁹ Baker-Henningham H, Vera-Hernández M, Alderman H, *et al* Irie Classroom Toolbox: a study protocol for a cluster-randomised trial of a universal violence prevention programme in Jamaican preschools *BMJ Open* 2016;6:e012166. doi: 10.1136/bmjopen-2016-012166

What else is known from other studies about the intervention?

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 EGM has two other RCTs focused on reducing violence against children as a form of school discipline albeit in older pupils (primary and secondary). Both reported large reductions in violence against children by teachers although the high risk of bias in the trials reduces our confidence in the findings.

Interaction Competencies with Children for Teachers (ICC-T) was a training workshop for secondary school teachers in Tanzania with a goal of preventing violence and improving teacher-student relationships. The trial found that teacher training reduced self-reported violence by teachers and improves teacher's attitudes away from corporal punishment (Nkuba 2018⁸⁰)

The *Good School Toolkit (GST)* was a comprehensive 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 2015⁸¹). Mental health and educational performance outcomes stayed mostly the same. This trial has five papers in the EGM.

Type of study: RCT protocol. Peer-reviewed journal article.

The intervention

The *Irie Classroom Toolbox (ICT)* is a school-based violence prevention programme to train Jamaican preschool teachers. It aims to reduce aggressive behaviour among preschoolers and violent discipline by teachers. One in eight five to six-year olds in Jamaica have "externalising disorders", i.e., they act out, are aggressive or do not listen to their parents or caregivers. One in five preschoolers from disadvantaged areas of Kingston (capital of Jamaica) have problem behaviours according to their teachers. Corporal punishment although banned in early education institutions in Jamaica is still prevalent.

ICT is based on an American evidence-based intervention for children with behavioural problems called the *Incredible Years Teacher Training Programme*⁸² with modifications to fit the Jamaican context. *ICT* is designed to offer teachers working in resource-poor settings a low-cost toolkit of different behaviour management strategies.

ICT training for preschool teachers is comprised of four modules:

- 1) Creating an emotionally supportive classroom environment,
- 2) Preventing and managing child behaviour problems,
- 3) Teaching social and emotional skills to children,

⁸⁰ 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

⁸¹ 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/>

⁸² Webster-Stratton C. The incredible years training series. Washington DC: Office of Juvenile Justice and Delinquency Prevention, Juvenile Justice Bulletin, 2000

4) Individual and class-wide behaviour planning

The modules provide teachers with various strategies to address child behaviour (e.g., verbal praise, visual cues) and guidance on how to plan lessons and activities around teaching social and emotional skills to children.

Training is run by four facilitators and consists of five full-day workshops (a manual is used to ensure consistency). The training process includes “goal setting, collaborative problem-solving, positive feedback, role playing” to promote teacher commitment and use of new skills.

Monthly coaching in class (also based on a standardised manual) is part of ICT. Coaching in the classroom is meant to provide individual feedback and help tailor strategies to the specific needs of the classroom.

Every two weeks, text messages are sent to teachers to encourage and remind them.

ICT materials provided to teachers include two booklets (one on the guidance for use of the toolbox and an activity book with songs, games and lesson plans), three sets of picture cards (with illustrations on classroom rules, friendly behaviours and emotion faces to help teach important concepts to children) and 14 story cards (with typical classroom situations and guidance to children on how to address them such as the need to share or waiting their turn patiently). Story cards come with discussion guides for teachers. Each classroom also receives educational materials such as books and building blocks to help teachers implement their new skills.

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

No specific theory is mentioned in the protocol. *ICT* is based on the *Incredible Years Teacher Training Programme*, an evidence-based approach to addressing behavioural issues in young children by training teachers and parents.

Has the intervention been implemented at scale?

Unclear. No information in the protocol on scale of implementation.

What does the intervention cost?

An economic evaluation is planned. The analysis will focus on classroom efficiency, i.e., whether the costs of training are offset by reduced disruptions in the classroom (i.e., if the intervention is effective in improving classroom behaviours and saving time). Classroom efficiency will be combined with school attendance and data from other studies in preschoolers to assess long term economic returns on investment.

Costs of any refresher trainings will be factored in based on whether benefits from the training on teacher behaviour are still evident one year after intervention.

Findings will be reported in the future.

The planned trial

Participants: All preschools (with three to six-year olds) in urban areas of Kingston with at least two to four classrooms and an average strength of 10 or more per classroom were eligible. 76 preschools from all eligible schools were randomised to *ICT* or controls (38 preschools each).

All teachers in the intervention arm participated in *ICT* training (114 teachers). Four-year olds from both arms were randomly selected (if they had no obvious disability and had a greater than 50% attendance in the previous two school terms) to assess the impact of the intervention. Note that children do not participate in the actual *ICT* training, but they are the ultimate beneficiaries of the intervention.

Study design: The trial has two groups:

- a) 38 preschools (114 teachers) got *ICT*
- b) 38 preschools (114 teachers) were controls

Baseline assessments took place in the summer of 2015 followed by *ICT* intervention activities in the 2015-16 school year. Follow-up assessments happened in the summer of 2016 and one year later in 2017.

Outcomes:

The outcomes that will be reported are:

- Primary outcomes:
 - a. Teachers' use of violence against children – observed over two different school days
 - b. Class-wide child aggression – observed over one school day
- Secondary outcomes:
 - a. Classroom level measures – prosocial behaviours observed over one day along with ratings using the Classroom Assessment Scoring System (CLASS) by trained observers. CLASS is a validated tool to measure the quality of the classroom environment (emotional support, classroom organisation and instructional support) and predict children's social and academic skills⁸³.
 - b. Teacher mental health– Self-reported measures on depression, self-efficacy, and burnout.
 - c. Child mental health – Child mental health will be measured using the Strengths and Difficulties Questionnaire (SDQ)⁸⁴
 - d. Child inhibitory control – this is a measure of self-regulation, i.e., whether children can suppress their impulses and reduce distractions. The idea is that improved self-regulation developed at a young age can protect against problem behaviours in adult life. Three play tests are used to test this. One example is the 'silly sounds game.' Children are shown

⁸³ Pianta R, La Paro K, Hamre B. Classroom assessment scoring system manual Pre-K. Baltimore, MD: Paul H. Brookes Publishing Co, 2008.

⁸⁴ Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behaviour Checklist: is small beautiful? *J Abnorm Child Psych* 1999;27:17–24.

pictures of a cat and a dog and told that cats say “woof” and dogs “meow.” They are then shown many pictures of cats and dogs and asked what ‘silly sound’ the animal would make.

- e. Child attendance – from school records

Primary outcomes and classroom level measures will be reported by trained personnel. This is quite important since often studies will use self-report by teachers or children which can be biased since they could report “desired” outcomes (for e.g., a teacher who participated in training might be less inclined to report that they use violence for discipline). While having trained observers does not eliminate bias - teachers could change their behaviour because they are being ‘watched’ - but it is more objective than self-report.

Cells in which this protocol appears:

The study relates to prevention:

- Intervention = Prevention; Outcome = Child Wellbeing: Mental Health (17 studies + 6 systematic reviews + 3 protocols)
- Intervention = Prevention; Outcome = Child Wellbeing: Cognitive Functioning (1 study + 1 protocol)
- Intervention = Prevention; Outcome = Disclosure: Disclosure Rates (9 studies + 5 systematic reviews)
- Intervention = Prevention; Outcome = Child Safety: Maltreatment Behaviour (10 studies + 2 systematic reviews + 3 protocols)
- Intervention = Prevention; Outcome = Adult Perpetrator or Offender: Desistance (1 protocol)

Czerwinski (2018) (several heavy cells)

This study does not appear in a light cell. But it appears in some heavy cells, and we had written a summary of it anyway so decided to leave it in this Guidebook.

Light cell(s) in which this study appears:

Intervention = Response; Outcome =Disclosure: Disclosure rates (1 study + 1 systematic review)

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 **Czerwinski et al. (2018)**⁸⁵ ‘Effectiveness of a school-based intervention to prevent child sexual abuse – Evaluation of the German *IGEL* program.’

Evidence status	Moderate risk of bias	Moderate evidence of impact on child knowledge and child mental health. (Rates of disclosure of CSA or actual CSA were not reported.)
------------------------	-----------------------	---

The summary in brief

The *IGEL* programme is a universal child sexual abuse (CSA) prevention programme delivered at the classroom-level to all third graders. The aim is for children to improve their knowledge of, response to and recognition of child sexual abuse. Involvement of teachers and school staff further supports children in these areas. The overall goal is to empower children to disclose any past, current, or future abuse.

In study, eight primary schools in North Rhine-Westphalia, Germany received the *IGEL* programme and four other primary schools were controls. They were not chosen randomly (the rationale for the choice is not given).

As with most school-based CSA prevention programmes, the study measured intermediate outcomes: improving children’s knowledge, courses of action, and recognition of CSA. The study also assessed anxiety and touch aversion (e.g., aversion to parental touching or kissing or aversion to crowds) as possible negative developments from the programme. It did not measure disclosure rates.

The study found that the *IGEL* programme improved knowledge and courses of action / response measures by moderate amounts compared to the control group. Recognition of CSA improved by smaller amount for *IGEL* groups relative to controls. Anxiety levels and touch aversion reported by children or by their parents did not increase indicating *IGEL* programmes did not lead to these negative “side-effects.” Staff awareness outcomes, rates of disclosure of CSA or actual CSA were not reported.

Being from a foreign (non-German) background affected outcomes across all three groups such as lower knowledge scores and higher reports by parents of touch aversion. Girls and boys had similar results in general.

Type of study: QED. Peer-reviewed journal article.

⁸⁵ Czerwinski F, Finne E, Alfes J, Kolip P. Effectiveness of a school-based intervention to prevent child sexual abuse-Evaluation of the German *IGEL* program. *Child Abuse Negl.* 2018;86:109-122. doi:10.1016/j.chiabu.2018.08.023

The intervention

IGEL (German for hedgehog – meant to symbolise an empowered defence even when the abuser is stronger) is a universal programme (i.e., not targeted at any particular group) for primary school children in third grade and for school staff. The main aim is to improve children’s knowledge of sexual abuse as well as skills in responding (‘courses of action’) to abuse especially in terms of informing an adult (disclosure) and seeking help. A second focus of the programme is to increase the preparedness of school personnel in identifying sexual abuse and in responding promptly.

Its intended outcomes are: **Prevention** of child sexual abuse, improving response and **encouraging disclosure**. The programme worked with “third grade primary school children and school personnel”. It also “aims to raise the school personnel’s **awareness** of sexual abuse and enable them to **identify** inappropriate situations and **react** adequately”.

The programme includes seven school sessions for third-grade children focused on multiple topics such as the definition of sexual abuse, the body, body contact, secrets and defence strategies. In each session, the children participate in various (interactive and experiential) exercises and exchange views about the topics, e.g., they discuss situations in which touches or kisses from adults feel like a boundary violation and how to deal with such situations.

Sessions are led by teachers specifically trained for this purpose in two workshops. Teachers also receive ongoing professional support during the programme to raise issues or ask questions. However, no information is given about who trained the teachers, nor how long before the programme, nor how long the seven school sessions are, nor how far apart they are, i.e., the programme’s overall duration. We do not know if the teacher delivering *IGEL* was the child’s normal teacher, another teacher in the school (who the children will have seen) or an external teacher. No information is given about who funded the programme.

Authors refers to a manual for the programme which is publicly available⁸⁶.

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. The *IGEL* programme adopts a comprehensive approach to address child sexual abuse. This child-centred, school-based approach aims to improve knowledge and concepts of child sexual abuse. It empowers children to recognise, resist, and report child sexual abuse. The involvement of teachers and school staff facilitates better support for children.

Has the intervention been implemented at scale?

No information about the scale of implementation is given in the article.

What does the intervention cost?

The study does not report any data on cost.

⁸⁶ <https://www.amazon.com/Pr%C3%A4vention-von-sexualisierter-Gewalt-Primarstufe/dp/3779936003>

The study

Participants: The *IGEL* programme is designed for third graders and school personnel. Close to 290 students and 320 parents / guardians took part in the study. The mean age for children was 8.7 years with just over half of students being female. Nearly half the students were from a foreign background, with the most common foreign country of origin being Turkey (54%) followed by other European countries.

Study design: The study had three groups:

- a) Six schools got the standard *IGEL* package.
- b) Two schools got the standard *IGEL* package and had a one-time visit to the theatre for a preventive intervention at the same time as the *IGEL* programme. This was unexpected but potentially reflects real-world scenarios with parallel events.
- c) Four control schools did not receive *IGEL*. They received an “emergency kit” with information on appropriate handling of children disclosing abuse since they were part of the study.

To address the issue with the two intervention groups, the study compared outcome measures for the two groups as a first step. They found the effects of the intervention to be statistically similar for both groups. Next, they combined the data for both intervention groups for comparison to the control group.

Outcomes: The outcomes measured were:

- Child CSA-related knowledge: tested with a 28 knowledge-related questionnaire on appropriate and inappropriate touch across four areas: good and bad feelings; saying yes / saying no; secrets; limits prohibition. The questionnaire was adapted for German primary schoolchildren.
- Child courses of action / response: tested using a case vignette of a child and their guitar tutor followed by a 15-item questionnaire with possible options for responding to inappropriate situations.
- Recognition of CSA: children were asked to rate seven statements including nudity and sexuality in combination with appropriate and inappropriate behaviour.

The study measured outcomes before intervention began, immediately after intervention and three months after it ended.

Did the intervention work?

The results clearly demonstrate that the intervention improved children’s knowledge, course of action and recognition of CSA. There was no increase in anxiety or touch aversion due to the *IGEL* programme. Gains for both intervention groups, i.e., with or without the theatre-based intervention were similar.

- ✓ **Knowledge** about CSA increased for children who received *IGEL* by a moderate amount compared to controls. The size of the impact increased over the three months after the programme due to knowledge scores of the control group decreasing over time. Children from non-German descent households generally scored lower across groups. Boys too had lower scores than girls across groups.

- ✓ **Courses of action / response** scores when faced with CSA improved moderately for *IGEL* groups compared to controls. The size of the impact remained mostly constant over the three months after the programme. Older children did better on this measure.
- ✓ **Recognition** of CSA improved for all groups. *IGEL* groups did better than controls albeit by a small amount.

Adverse outcomes: Participating in an intervention or even being asked questions about difficult situations could lead to harmful “side-effects.” The study assessed two outcomes: “anxiety” and “touch aversion” to determine if there were any adverse effects.

- a. Anxiety: a German version of the Screen for Child Anxiety Related Disorders (SCARED) a validated scale for children aged 8-12⁸⁷ was used.
- b. Touch aversion: study researchers came up with a set of questions to test if children developed negative attitudes to regular situations such as parental touching and kissing, being touched by other children or teachers and crowds as a consequence of the *IGEL* program.
- c. Parent survey: parents were also surveyed as proxy reports for their children on signs of anxiety and touch aversion.

Anxiety levels and touch aversion reported by children or by their parents reduced over time for all three groups. Boys (compared to girls) and non-German background parents (compared to German-background parents) reported higher touch aversion across groups. However, the overall trend was one of reduction over time.

This suggests that *IGEL* did not cause these adverse effects.

Generalisability

No information is provided about this, and it is not discussed in the study report.

What else is known from other studies about the intervention?

A Cochrane systematic review⁸⁸ of 24 studies of school-based prevention programmes which (i) deliver information about child sexual abuse (ii) strategies to help children avoid it and (iii) encourage them to report abuse found the programmes increase knowledge and skills for dealing with abuse. {This systematic review is on the EGM elsewhere.} There was no evidence that programmes increased children’s anxiety or fear. Children who participated in these programmes may have higher odds of disclosing abuse; however more evaluations are needed to confirm this. No impact was seen on child mental health outcomes or actual sexual abuse. The review did not report parent or caregiver outcomes.

⁸⁷ Essau, C. A., Muris, P., & Ederer, E. M. (2002). Reliability and validity of the Spence Children’s Anxiety Scale and the Screen for Child Anxiety Related Emotional Disorders in German children. *Journal of Behavior Therapy and Experimental Psychiatry*, 33(1), 1–18. [https://doi.org/10.1016/S0005-7916\(02\)00005-8](https://doi.org/10.1016/S0005-7916(02)00005-8).

⁸⁸ Walsh K., Zwi K., Woolfenden S., Shlonsky A. School-based education programmes for the prevention of child sexual abuse. *Cochrane Database of Systematic Reviews* 2015. Issue 4. Art No.: CD004380. DOI:10.1002/14651858. CD004380.pub3.

How reliable is the evidence?

The study is rated as having a moderate 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
Moderate	Moderate	Low	Low	Moderate	Moderate	Moderate	Low

Cells in which this study appears:

The study relates to prevention, and to several outcomes:

- Intervention = Prevention; Outcome = Disclosure: Disclosure Rates (9 studies + 5 systematic reviews)
- Intervention = Response; Outcome = Disclosure: Disclosure Rates (1 study + 1 systematic review)
- Intervention = Prevention; Outcome = Child Wellbeing: Knowledge or Awareness (51 studies + 10 systematic reviews)
- Intervention = Prevention; Outcome = Child Wellbeing: Mental Health (19 studies + 6 systematic reviews)

Devries (2015) Good Schools Toolkit (Prevention / child wellbeing, and cognitive functioning)

Light cell(s) in which this study appears:

Intervention = Prevention; Outcome = Child Wellbeing: Cognitive Functioning (1 study + 1 protocol)

Summary: A universal programme run in primary schools in Uganda to prevent physical violence substantially reduced violence experienced by students from school staff. Results need to be interpreted with caution due to the study's high risk of bias.

The summary is based on **Devries et al. (2015)**⁸⁹ 'The *Good School Toolkit* for reducing physical violence from school staff to primary school students: A cluster-randomised controlled trial in Uganda.'

Evidence status	High risk of bias	Weak evidence of impact on maltreatment behaviour (child safety) mental health, cognitive functioning (child wellbeing) and disclosure.
------------------------	--------------------------	---

The summary in brief

The *Good School Toolkit (GST)*⁹⁰ is a universal programme that aims to prevent physical violence from school staff towards primary school children (aged 11-14: mean age 13.0 years). 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 forward 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.

The EGM contains multiple papers written about this RCT. This paper focuses on the main analysis from the RCT, i.e., impact on physical violence in school. It also reports on mental health, wellbeing and educational performance.

Self-reported violence experienced over the past week and past school term was lower for *GST* students compared to controls. The risk of violence for *GST* students in the past week reduced by 42% compared to controls. School staff in *GST* schools reported reduced use of physical violence compared to control group staff. Mental health and educational performance outcomes were mostly unchanged. The prevalence of corporal punishment at the end of the programme was still high (30%).

⁸⁹ 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/>

⁹⁰ <http://raisingvoices.org/good-school/download-good-school-toolkit/>

More than 400 children were referred to child protective services after the follow-up data were collected. High risk of bias warrants caution when interpreting these findings.

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 Model⁹¹ that theorises that behaviour change goes through "six stages of change: precontemplation, contemplation, preparation, action, maintenance and termination."

⁹¹ 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>

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 al⁹². 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”). While entire schools participated in *GST* only students from classes from fifth to seventh grade (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 girls. 7% reported having some disability and more than half reported eating two or fewer meals in the previous day. 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.

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 (June-July 2012) and a few months after it ended (June-July 2014).

Outcomes: Several studies covering different outcomes from this trial have been published. This study (Devries 2015) reports on the main outcome of physical violence by staff and outcomes on mental health and educational performance.

The outcomes measured were:

⁹² 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.

- Physical violence: Student self-reported physical violence at school in the past week, the past term and school staff self-reported use of physical violence in the past week measured using The International Society for the Prevention of Child Abuse and Neglect Screening Tool – Child Institutional (ICAST-CI)⁹³
- Mental health and wellbeing of students: Safety and wellbeing in schools and mental health status using the Strengths and Difficulty Questionnaire (SDQ)⁹⁴
- Educational performance: a variety of indicators on reading, comprehension in both English and Luganda, spelling and basic maths

Did the intervention work?

- ✓ **Physical violence:** When assessing violence in the past week, the study reports a 42% reduction in the risk of physical violence for *GST* arm students from school staff compared to controls. This large reduction is not surprising because (i) the prevalence of corporal punishment was quite high before the study – more than half of students reported experiencing violence in the past week and (ii) this was a complex intervention with numerous activities that went on for 18 months. However, there are flaws in how the study was carried out which reduces one’s confidence in the reported results.

Other results on physical violence:

- Students reporting that they experienced physical violence at school from school staff in the past week fell from over 50% before *GST* to 31% afterwards, while it remained mostly unchanged for the control arm, i.e., close to 50%.
 - 80% of control arm students said that they experienced violence in the previous term compared to close to 60% in the *GST* arm.
 - The proportion of school staff who participated in *GST* and said that they used physical violence on students in the past week fell by more than half (from 40% to 15%). A significant reduction was seen for control arm school staff too, but it was not as pronounced (43% to 32%).
 - *GST* seemed to work marginally better for male students than female students.
- ✓ **Mental health and Wellbeing and Educational performance:** These measures mostly stayed the same after intervention and between groups.

Adverse outcomes: No adverse effects are reported in the study. However, 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.

⁹³ 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.

⁹⁴ Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 2000;177: 534–39.

Translating elsewhere

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. (This is what the study report says, though it's obviously a rather odd statement given that Africa is huge and diverse.) However, this trial was only conducted in one district of Uganda. More studies are 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.

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	Low risk	High risk	Some concerns

Cells in which this study appears:

The study relates to prevention:

- Intervention = Prevention; Outcome = Child Wellbeing: Mental Health (17 studies + 6 systematic reviews + 3 protocols)
- Intervention = Prevention; Outcome = Child Wellbeing: Cognitive Functioning (1 study + 1 protocol)
- Intervention = Prevention; Outcome = Disclosure: Disclosure Rates (9 studies + 5 systematic reviews)
- 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.

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)**⁹⁵ ‘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.

⁹⁵ 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 review⁹⁶ 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)

⁹⁶ 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>

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)**⁹⁷ ‘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)**⁹⁸ ‘Placement in foster care enhance quality of attachment among young institutionalised children’ and
- **Bick et al. (2015)**⁹⁹ ‘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

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

⁹⁷ 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

⁹⁸ 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

⁹⁹ 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

workers. Social workers aimed to facilitate the establishment of a strong bond between children and their foster carers.

There are six papers¹⁰⁰ 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 orphanages¹⁰¹. 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 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.

¹⁰⁰ The three other studies on this trial (six total) in the EGM are: Humphries 2015, Troller-Renfree 2015 and Wade 2018)

¹⁰¹ Weir, K: American Psychological Association. (June, 2014). The lasting impact of neglect. Retrieved from <https://www.apa.org/monitor/2014/06/neglect>

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:

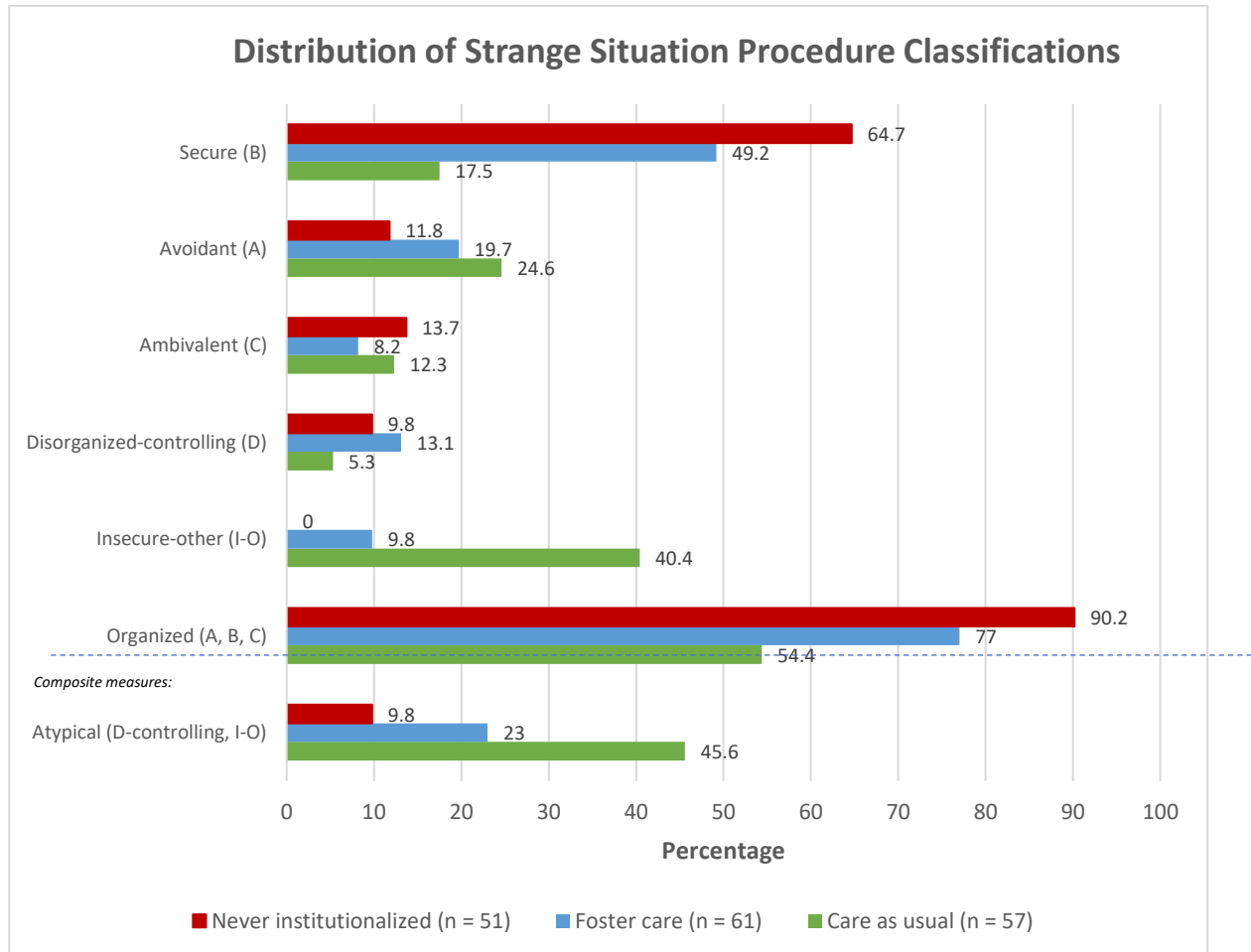
- Had lower height, weight and head-size to start with
 - Better **caregiving quality** (i.e., quality matters) as measured by “caregiving quality scores” developed by the study. Specifically, “positive regard for the child and sensitivity” were positively correlated with height and weight catchup. Conversely, “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 study¹⁰² 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)

¹⁰² 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>

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)**¹⁰³ ‘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)*¹⁰⁴ 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 papers¹⁰⁵ 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.

¹⁰³ 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

¹⁰⁴ <http://raisingvoices.org/good-school/download-good-school-toolkit/>

¹⁰⁵ 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 Model¹⁰⁶ 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 al¹⁰⁷. 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.

¹⁰⁶ 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>

¹⁰⁷ 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)¹⁰⁸ and the WHO Multi-Country Study on Women's Health and Domestic Violence against Women (WHO MCS)¹⁰⁹

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.

¹⁰⁸ 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.

¹⁰⁹ 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)**¹¹⁰ '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.

¹¹⁰ 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 Model¹¹¹ 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)¹¹²

¹¹¹ 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>

¹¹²Devries 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).¹¹³ 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.

¹¹³ 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 2015¹¹⁴). 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)

¹¹⁴ 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/>

Rheingold (2014) (Response / adult caregiver behaviours, knowledge or attitudes)

Light cell(s) in which this study appears:

Intervention = Response; Outcome =Adult Institutional Caregiver: Behaviours, Knowledge or Attitudes (1 study)

Summary: Brief training for childcare professionals, delivered either in-person or via internet, improves child sexual abuse preventive knowledge and behaviour response by childcare professionals.

The summary is based on **Rheingold et al. (2014)**¹¹⁵ 'Child sexual abuse prevention training for childcare professionals: An independent multi-site randomised controlled trial of *Stewards of Children* conducted in the United States.'

Evidence status	Moderate risk of bias	Moderate evidence of impact on adult institutional care provider knowledge, attitudes, and response to child sexual abuse prevention.
------------------------	-----------------------	---

The summary in brief

Stewards of Children is a 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 two- and half-hour session, and via the internet over two weeks.

Most studies of prevention programmes look at interventions with children. This study claims to be one of the first studies of a prevention programme targeting childcare professionals. Note that (as with most prevention programmes), the study measured intermediate outcomes: improving childcare professionals' knowledge, attitudes, and response. The impact on actual prevention of CSA is unclear.

The trial found that both delivery modes improve childcare professionals' knowledge of CSA and preventive behaviours relative to if the professionals receive no training. However, the size of the impact in terms of practice implications is unclear.

Type of study: RCT. Peer-reviewed journal article.

The intervention

Stewards of Children is a child sexual abuse (CSA) prevention programme for childcare professionals. The programme was developed by the US non-profit Darkness to Light (D2L). The programme involves a two-and-a-half-hour workshop for adults in childcare settings to train them on prevention, recognition and response to CSA. *Stewards of Children* aims to improve childcare professionals' ability to prevent CSA from happening (primary prevention) and to recognise and respond to CSA (secondary prevention). The programme is offered in two formats: in-person and interactive web-based training.

¹¹⁵ Rheingold, Alyssa & Zajac, Kristyn & Chapman, Jason & Patton, Meghan & de Arellano, Michael & Saunders, Benjamin & Kilpatrick, Dean. (2014). Child Sexual Abuse Prevention Training for Childcare Professionals: An Independent Multi-Site Randomized Controlled Trial of Stewards of Children. *Prevention Science : The Official Journal of the Society for Prevention Research*. 16. 10.1007/s11121-014-0499-6.

The in-person training was a two-and-a-half-hour group training led by a facilitator who was trained in the programme. The training covered the following topics: (i) knowledge on CSA prevalence rates, risks and outcomes (ii) strategies to reduce CSA opportunities (iii) methods to bring up CSA with adults and children (iv) identifying signs of CSA (v) responding appropriately when a child comes forward with CSA disclosure (vi) addressing barriers to preventive actions at the individual and organisational level and (vii) community involvement in CSA prevention. A DVD was used to show experiences of CSA survivors combined with facilitated discussion.

The web-based training was delivered over two weeks and is comparable in content and length to the in-person training.

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

Stewards of Children is not a theory-based prevention programme, but its principles are in line with Finkelhor's¹¹⁶ theory that for CSA to occur, certain preconditions exist. They include: an individual's tendency to abuse, absence of internal or external inhibitions for the offender, and the offender having access to the child. Preventing one or more of these preconditions should reduce the likelihood of CSA. This programme aims to reduce access to children and to increase external barriers for offenders by improving the knowledge, attitudes, and response of adults responsible for childcare.

Has the intervention been implemented at scale?

There is no information available in the study on whether it was implemented at scale.

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.

The trial

Participants: The trial involved 352 childcare professionals from youth service organisations such as day care centres, schools and churches. They were in cities in three non-contiguous regions across three US states (one site is in the northwest US and the other two in the southeast.) The mean age of participants was 38.9 years with 85% female, 28.4% African American and 34.8% schoolteachers.

Study design: Participants were divided into three groups: (a) a group which got the in-person training, (b) a group which got the training online, and (c) a 'waitlist' group, i.e., which got nothing during the study but would get it later. Each group had a mix of types of professionals (each contained at least one faith-based community, one school, and one community-based programme), so the results are not split out by type of professional such as teachers or clergy (there were not enough people of each profession in each group to do so). It was *individuals* who were randomised to the three groups, not *agencies*.

Outcomes: The outcomes measured were:

¹¹⁶ Finkelhor, D. (1984). *Child sexual abuse: New theory and research*. New York: Free Press.

- **CSA Knowledge.** This was measured using a CSA Knowledge Questionnaire created for this study. It consists of 12 true / false questions about CSA, including prevalence, consequences, risk factors, and preventive steps.
- **CSA Attitudes.** This was measured using the *CSA Myth Scale* (Collings 1997¹¹⁷), which assesses a full range of CSA myths and stereotypes.
- **CSA Prevention Behaviours.** This was measured using a 21-question survey on CSA prevention behaviours that covered: “(a) engaged in primary prevention at work (e.g., prevented adults from having unsupervised time with children); (b) talked to a child about CSA; (c) talked to another adult about CSA; (d) made changes in organisational policies regarding CSA; (e) called a CSA hotline; and (f) reported CSA to authorities. The total score is a count of endorsed behaviours. Individual preventive behaviours were examined on an exploratory basis”. It is items (c), (e) and (f) here which make this study count as response.

The study measured outcomes immediately after the training and then three months later for CSA knowledge and attitudes. CSA prevention behaviours were also measured before the training.

Did the intervention work?

The results clearly demonstrate that the intervention improved knowledge and behaviours.

- ✓ **Knowledge** about CSA increased. That knowledge declined over the three months after the training (as one would expect), though oddly, the knowledge of the control (‘waitlist’) group increased during that time, but still not to as high as the trained group.
- ✓ **Attitudes.** Participants’ belief in CSA myths was low to begin with so there was little room for improvement. After training, the control group had the better score but at three months there was no difference between groups.
- ✓ **Behaviours.** This also improved, i.e., participants reported having done more of the behaviours three months after the training that did people in the control group. The behaviours most improved were:
 - “limiting the opportunity for other youth and younger youth to have one-to-one interaction”. This is significant because juveniles are offenders in more than a third of CSA (Finkelhor et al. 2009¹¹⁸). And
 - “Sharing with another adult an article, brochure, or other information about CSA prevention”. Interestingly, the behaviour of people in the control group also improved during the three months after the training: quite possibly because colleagues who had received the training changed their behaviour (making it more normal) and they shared this information with colleagues who had not received the training.

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

¹¹⁷Collings, S. J. (1997). Development, reliability, and validity of the Child Sexual Abuse Myth Scale. *Journal of Interpersonal Violence*, 12, 665–674.

¹¹⁸ Finkelhor, D., Ormrod, R., & Chaffin, M. (2009). Juveniles who commit sex offenses against minors. Office of Juvenile Justice and Delinquency Prevention. <http://www.ncjrs.gov/pdffiles1/ojjdp/227763.pdf>.

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

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

Adverse outcomes: None was reported.

Generalisability

The trial was conducted in three disparate geographical sites which might imply that the results will translate elsewhere. However, we do not know if this intervention was implemented at scale.

What else is known about the intervention from other studies?

Most systematic reviews assessing interventions to prevent child sexual abuse focus on child-centred or parent-focused strategies. Studies on childcare professional-focused training interventions seem to be lacking in the published literature.

How reliable is the evidence?

The study is 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 this study appears:

The study relates to both prevention and response:

- Intervention = prevention; outcome = safeguarding practice/ operations (5 studies)
- Intervention = response; outcome = safeguarding practice/ operations (3 studies)
- Intervention = prevention; outcome = adult institutional caregiver, behaviours attitudes or knowledge (7 studies)
- Intervention = response; outcome = adult caregiver, behaviours attitudes or knowledge (1 study)

It is one of only three primary studies of response on our EGM.

Taylor (2010) (Prevention / child perpetrator maltreatment behaviour)

Light cell(s) in which this study appears:

Intervention = Prevention; Outcome = Child Perpetrator or Offender: Maltreatment Behaviour

(1 study)

Summary: Two middle school-based dating violence prevention curricula were not effective in reducing violence (in the US). Some improvements in attitudes and knowledge seen but not consistent.

The summary is based on **Taylor et al. (2010)**¹¹⁹ 'The effects of gender violence/harassment prevention programming in middle schools: A randomised experimental evaluation.'

Evidence status	High risk of bias	Weak evidence of impact on maltreatment behaviour (child safety) knowledge or awareness (child wellbeing) and maltreatment behaviour (child perpetrator or offender).
------------------------	--------------------------	---

The summary in brief

Gender violence and sexual harassment (GV/SH) is a prevalent issue among adolescents. GV/SH can cause serious mental and physical health outcomes, especially in teenagers. This intervention implemented two curricula with the aim of reducing GV/SH or "dating violence" among sixth and seventh graders from schools in suburban Cleveland, Ohio (USA).

The first, an interaction-based curriculum, focused on emphasising communication in relationships; setting clear boundaries; friendship in relationships; acceptable and unacceptable behaviours; and bystander roles. The second, a law and justice curriculum, presented facts and knowledge on GV/SH including definitions, relevant laws, and legal consequences.

Curricula were developed in close consultation with school personnel and a local rape crisis centre. Each curriculum consisted of five lessons delivered weekly by a professional from the crisis centre (except in a few instances when classroom teachers delivered the lessons). The curricula and format were tailored for sixth and seventh graders.

Seven schools were involved. 123 classrooms were randomly assigned to three arms. 29 classrooms assigned to each curriculum and the remaining were controls. The students were aged 11-13 years.

No significant differences were seen for dating violence victimisation or sexual harassment between intervention and control arms. Surprisingly, students in intervention arms reported committing significantly *more* violence against their dating partners than control. The study suggests that this might be due to increased sensitivity in the intervention groups due to the programme. Some of the measures of GV/SH attitudes and knowledge improved significantly compared to controls especially for students in the law and justice curriculum group but many of these effects faded with time.

Type of study: RCT. Peer-reviewed journal article.

¹¹⁹ Taylor B, Stein N, Burden F. The effects of gender violence/ harassment prevention programming in middle schools: a randomized experimental evaluation. *Violence Vict.* 2010;25(2):202-223. doi:10.1891/0886-6708.25.2.202

The intervention

Gender violence and sexual harassment (GV/SH) can adversely affect mental and physical health outcomes of teenagers. Two curricula on gender violence and harassment prevention / "dating violence" prevention were implemented for sixth and seventh grader classrooms in seven schools located in the suburbs of Cleveland, Ohio (USA).

One curriculum was interaction-based and focused on "setting and communicating boundaries in relationships; the formation of deliberate relationships / friendships and the continuum between friendship and intimacy; the determination of wanted / unwanted behaviors; and the role of the bystander as intervener." The lessons included here were not straightforward and 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.

Curricula were developed by study authors in partnership with school staff (three school districts) and a local rape crisis centre. All efforts were made to align lessons in the curricula with school health education goals and state standards. Curricula were delivered over five classroom sessions. An experienced professional from a local sexual assault centre taught most of the lessons (in five of the seven school buildings) while the regular classroom teacher taught the curricula in two school buildings (each school building had multiple classrooms that received the curricula). All three curricula instructors were taught by the curriculum developer (same person) and used the same materials.

Ideally the lessons would be spread out over more classroom sessions but competing priorities for schools led to a shortened format.

Curricula were delivered over five weeks with a weekly 40-minute lesson. Though the curricula varied, the first lesson (on establishment of relationship boundaries) was common to both.

Care was also taken to tailor the curricula for sixth and seventh graders with detailed instructions for each lesson. Examples of instructions include setting ground rules ("no swear words"), setting a fixed length of time for an activity and handling questions from students that distract away from the lesson.

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

The curricula 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 toward 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 GV/SH can reduce GV/SH behaviour.

Has the intervention been implemented at scale?

No. This curriculum was developed for this study. Only seven schools (three schools districts) were involved.

What does the intervention cost?

No information is provided on cost.

The trial

Participants: 123 sixth and seventh grade classrooms from seven schools were randomised into intervention arms (one arm for each curriculum) and a control arm. These schools were chosen because they had a large number of sixth and seventh grade classrooms with a diverse student body.

More than 1600 students participated in the trial with just over half being girl students. A quarter of the students were African American, half were white while the rest were Hispanic, Asian, Native American and multiracial or other ethnicities. More than half the students had dated (at least for a week) in the past. A quarter reported exposure to other violence prevention educational programmes. Nearly 30% had experienced dating violence in their lives and 1 in 5 reported being violent (at least once in their lifetime) with their dating partners.

Study design: The trial had three groups:

- c) 29 classrooms were assigned to the interaction-based curriculum
- d) 29 classrooms received the law and justice curriculum
- e) 65 schools served as controls (received standard health education which did not include any GV/SH prevention content)

Outcomes:

The outcomes measured were:

- Sexual and nonsexual violence victimisation, included both prevalence (yes / no) and the incidence (number of times) students experienced violence by their peers or dating partners (the intervention was not focused on peer violence prevention but the study collected this data to assess any extended impact)
- Sexual and nonsexual violence perpetration , included both prevalence (yes / no) and the incidence (number of times) students committed violence against their peers or dating partners
- Sexual harassment victimisation and perpetration, for prevalence and incidence of sexual harassment reported by victims and perpetrators
- Student attitudes towards GV/SH, to assess the impact of intervention on student acceptance and beliefs related to GV/SH. For example, thinking sexual harassment is a girl's fault or believing that GV/SH is not a problem
- Knowledge related to GV/SH, included awareness about state rape laws, definitions of what constitutes abuse and sexual harassment, sexual harassment myths and helpful resources

The study was conducted in 2006-07. Students were surveyed just before randomisation, immediately after the curricula were taught and about five to six months later.

Did the intervention work?

No.

- ✓ **Violence victimisation:** Interaction curriculum students reported lower levels (both prevalence and incidence) of sexual violence by their peers six months after the intervention compared to controls. Note that the intervention was focused on preventing dating violence and not violence by peers. Dating violence victimisation outcomes were not significantly different between intervention groups and controls.
- ✓ **Violence perpetration:** Unexpectedly, students in both intervention arms reported significantly higher violence that they had committed against their dating partners compared to control arm students. The study suggests that this might be due to increased sensitivity to GV/SH in the intervention arm students. For example, many students did not know that sex among minors is legally considered rape.
- ✓ **Sexual harassment:** Both victimisation and perpetration outcomes were not significantly different between intervention groups and controls at any time point.
- ✓ **Attitudes:** Students in the law and justice curriculum scored significantly better in four out of six areas (“appropriate attributions of girls’ fault in GV/SH”; “belief that GV/SH is a serious problem”; positive attitudes toward preventing sexual harassment; disposition about own / other’s personal space) compared to controls immediately after the intervention. However, half of these effects did not last at six months. Interaction curriculum students scored significantly better only on the ‘personal space’ item compared to controls.
- ✓ **Knowledge:** Students’ awareness of state rape laws, definitions of what constitutes abuse and sexual harassment and sexual harassment myths was significantly higher in the law and justice curriculum group (this is expected since knowledge on GV/SH was at the core of that curriculum) compared to controls immediately after the intervention and six months later.

Adverse outcomes: Reported perpetration of violence by students in the intervention arms against their dating partners increased compared to controls.

Generalisability

This trial included only seven schools. We therefore cannot say whether the results will translate to other places.

What else is known about the intervention from other studies?

A Campbell systematic review¹²⁰ of 23 studies of school-based dating and sexual violence prevention programmes found little impact on violent behaviour. Modest improvements were seen in knowledge, attitudes, and conflict resolution skills from these interventions.

¹²⁰ De La Rue L, Polanin JR, Espelage DL, Piggot T. D. School-based Interventions to Reduce Dating and Sexual Violence: A Systematic Review. Campbell Systematic Reviews 2014:7 DOI: 10.4073/csr.2014.7

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	High risk	High risk	Some concerns

Cells in which this study appears:

The study relates to prevention:

- Intervention = Prevention; Outcome = Child Safety: Maltreatment Behaviour (10 studies + 2 systematic reviews + 3 protocols)
- Intervention = Prevention; Outcome = Child Wellbeing: Knowledge or Awareness (50 studies + 1 protocol + 10 systematic reviews)
- Intervention = Prevention; Outcome = Child Perpetrator or Offender: Maltreatment Behaviour (1 study)

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

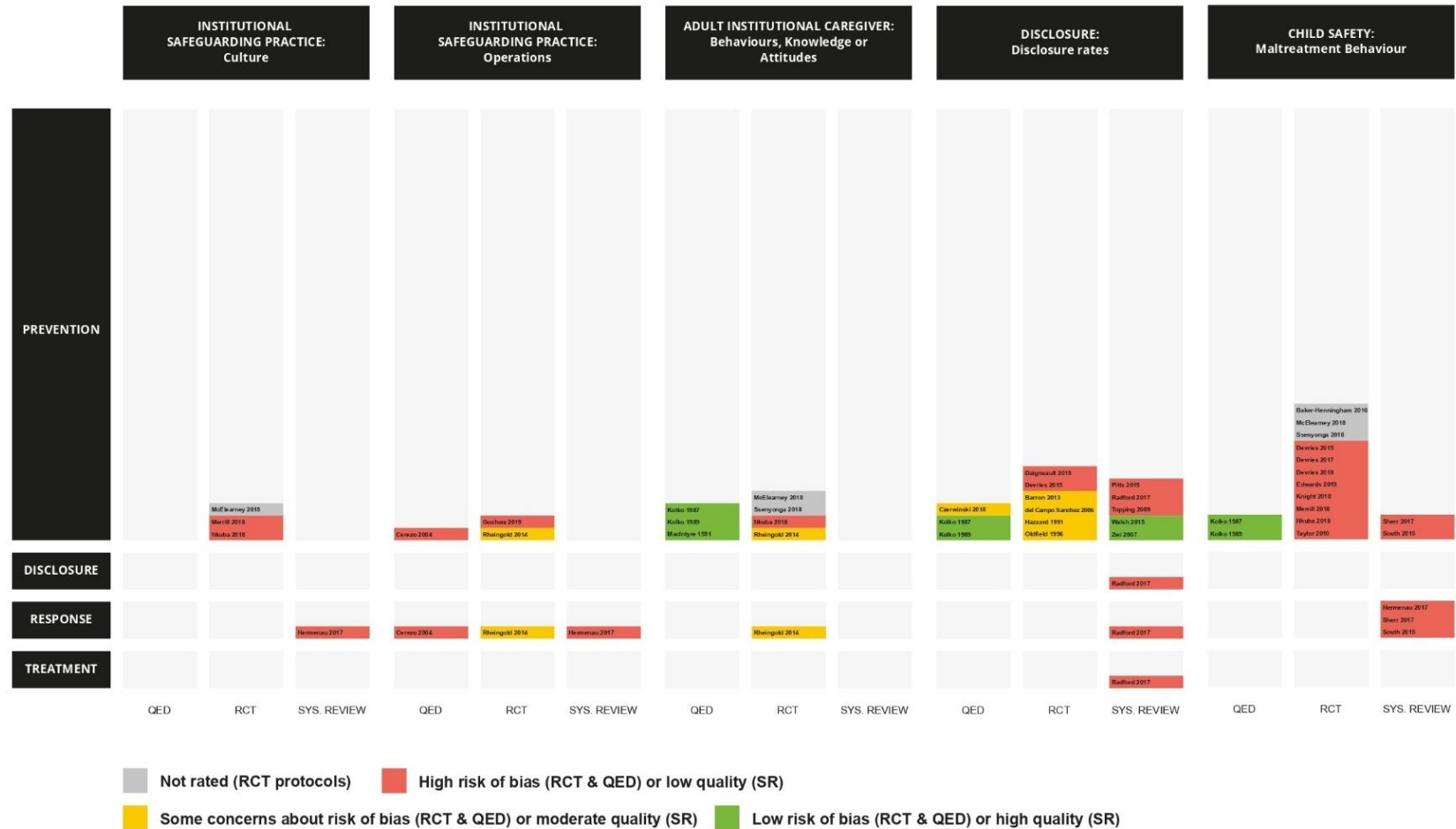
A research report that summarises the evidence (ie., existing studies) on a clearly formulated question according to a predefined protocol, using systematic and explicit methods to identify, select and appraise relevant studies, and to extract, analyse, collate and report their findings. It may or may 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.

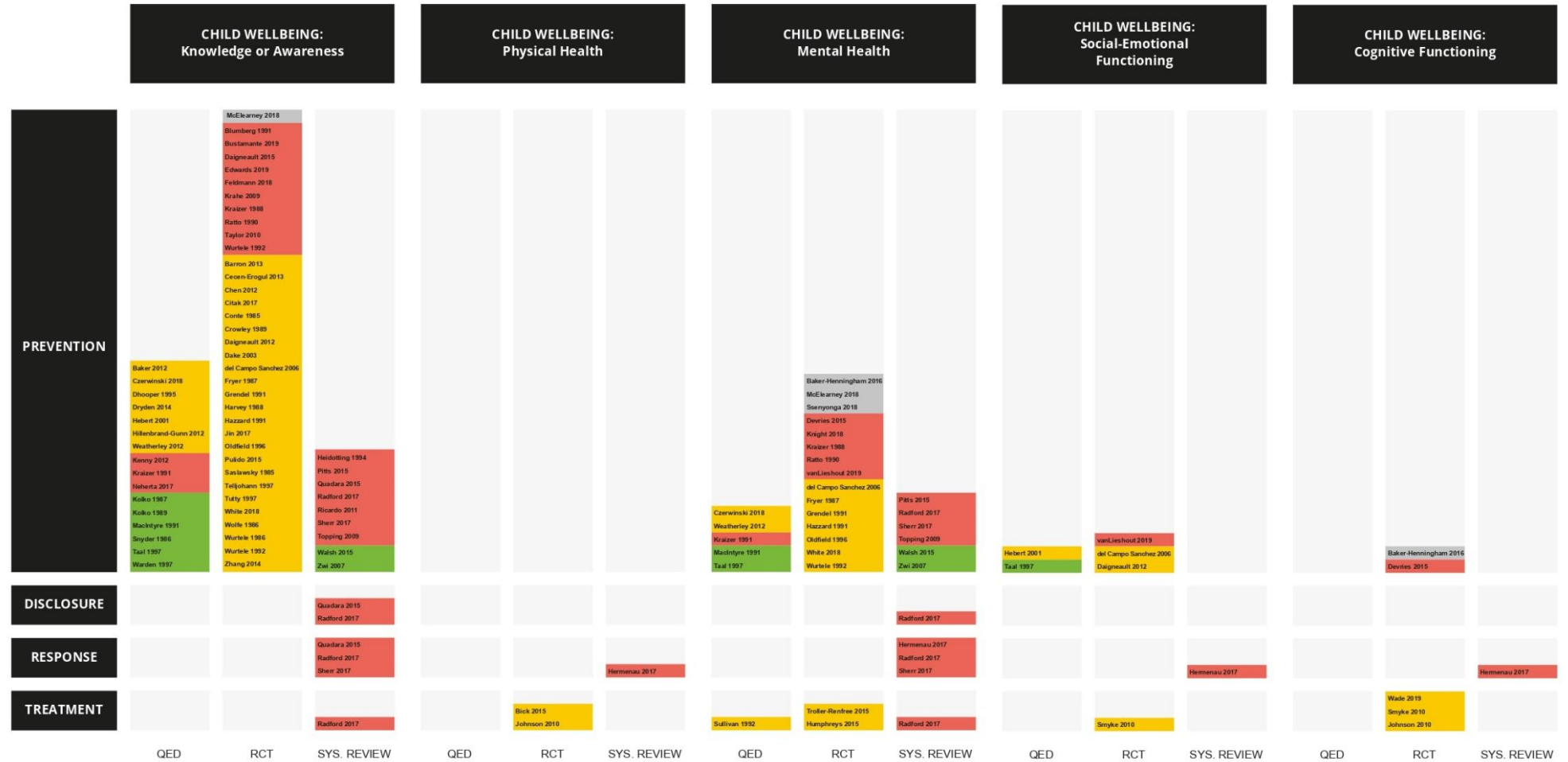
Appendix 1: The Visual Evidence and Gap Map

Page 1 of 3. This map can be downloaded from www.giving-evidence.com/csa. That page also has an interactive, searchable version of this map.

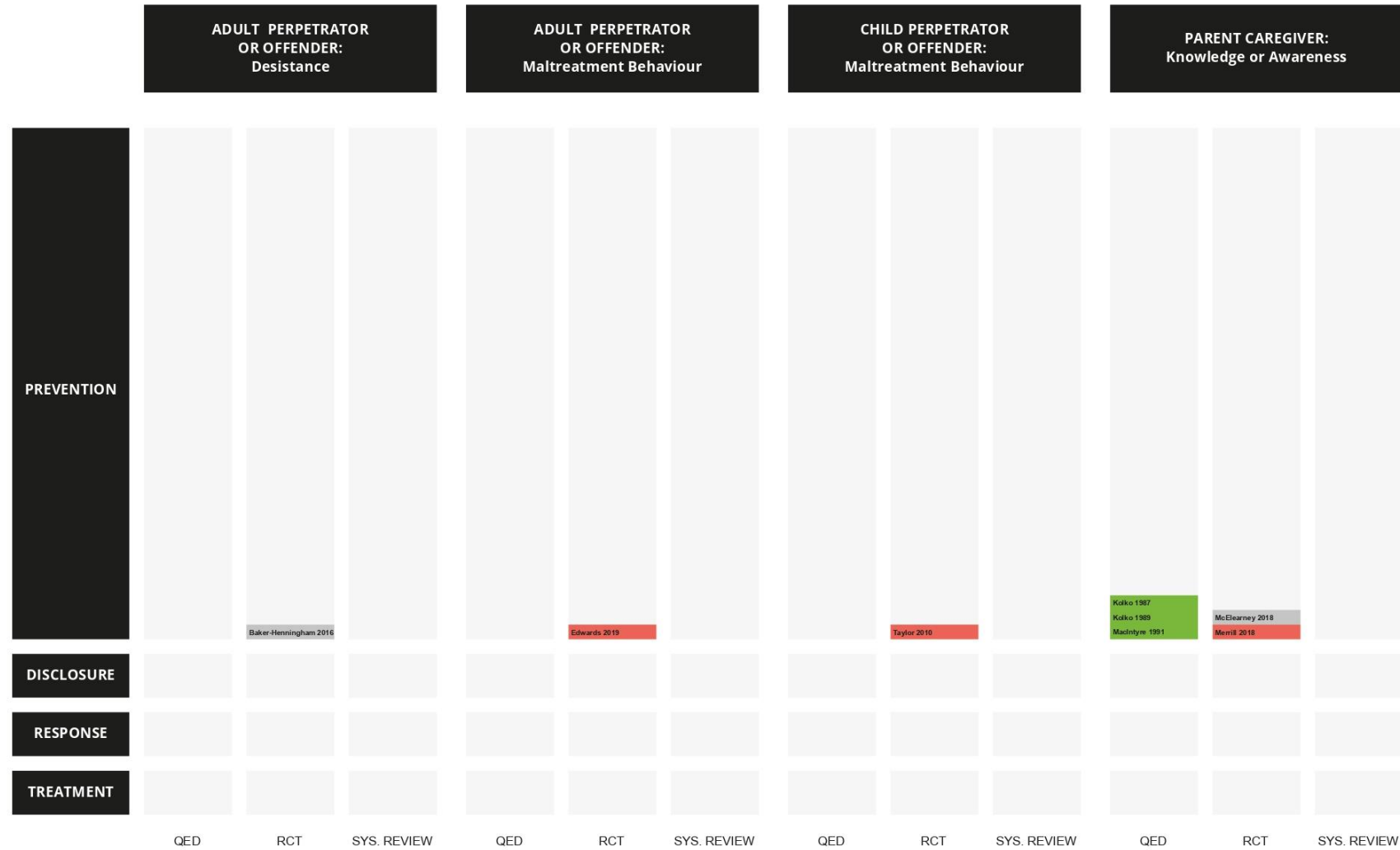


NOTE: We also look for studies which looked for the following outcomes: institutional safeguarding practice (environment); adult perpetrator or offender (recidivism); child perpetrator or offender (desistance); child perpetrator or offender (recidivism); and parent caregiver (knowledge or awareness). We did not find any. Those outcomes would have been shown as columns on this map, and they would have been empty. In the interests of making this map readable, we have removed those columns from this graphic.

Evidence gap map of institutional responses to child maltreatment (Part 2 of 3)



Evidence gap map of institutional responses to child maltreatment (Part 3 of 3)



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; Child Safety: Maltreatment Behaviour</p> <p>3. Response; Child Wellbeing: Physical Health</p> <p>4. Response; Child Wellbeing: Social and Emotional Functioning</p> <p>5. 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 0 - 5</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. Disclosure: Disclosure Rates</p> <p><i>Also in:</i></p> <p>2. Treatment ; Disclosure Rates</p> <p>3. Treatment ; Child Wellbeing: Knowledge or Awareness</p> <p>4. Disclosure; Child Wellbeing: Mental Health</p> <p>5. Disclosure; Child Wellbeing: Knowledge or Awareness</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 0- 5</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>Response; Child Safety: Maltreatment Behaviour</p>	<p>Sherr 2017: Child violence experiences in institutionalised / orphanage care</p>	<p>Low 0 -5</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,</p>

			<p>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 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>Response; Child Safety: Maltreatment Behaviour</p>	<p>South 2015: Scoping review: Evaluations of out-of-home care practice elements that aim to prevent child sexual abuse</p>	<p>Moderate 6 - 11</p>	<p>The aim of this scoping review was to map evaluations of out-of-home care (OOHC) practice elements that aim to prevent child sexual abuse (CSA) in OOHC. It was conducted by the Parenting Research Centre (PRC) and the University of Melbourne for the Royal Commission into Institutional Responses to Child Sexual Abuse. This report describes the methods used to conduct the scoping review and the findings of the scoping review.</p>

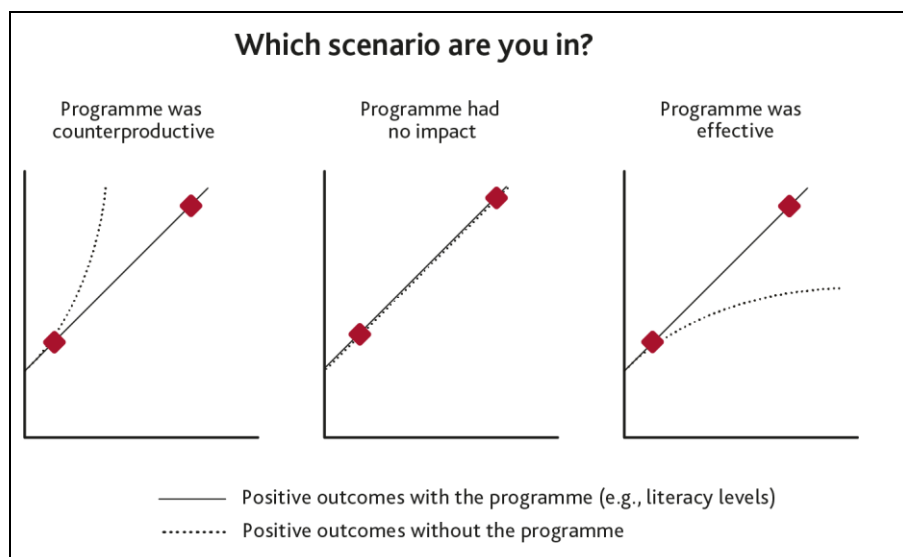
<p>Disclosure; Child Wellbeing: Knowledge or Awareness</p>	<p>Quadara 2015: Conceptualising the prevention of child sexual abuse</p>	<p>Low 0 - 5</p>	<p>Significant numbers of Australian children have experienced neglect, emotional abuse, physical abuse and sexual abuse. The adverse, long-term consequences of these experiences are well demonstrated in the research literature and well recognised by the policy and practice communities. It is also now well recognised that responding to abuse and neglect after it has been detected is only one aspect of prevention. Indeed, the National Framework for Protecting Australia’s Children 2009–2020 (Council of Australian Governments [COAG], 2009b), which provides a long-term strategy for children’s wellbeing and safety, and advocates a shift in focus from statutory tertiary responses to locating child abuse prevention and child safety within a public health model. In this approach, primary prevention and universal supports for all families are the central strategies out of which more intensive interventions should flow. In other words, the focus needs to be on preventing abuse and neglect before it occurs.</p>
---	---	------------------	---

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 the relevant information from the studies, in the Guidebook.

The information that we have included in the 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. 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 really 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 never yet 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 researches 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 Uganda¹²¹ found that they helped

¹²¹ <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 documents. 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.

The quality assessments in the Guidebook are taken from the EGM: we didn't re-do them for the Guidebook. As explained in the EGM documents, the assessments were:

- RCTs were assessed on their risk of bias (RoB) using the Cochrane tool¹²²
- Non-RCT primary studies (i.e., quasi-experimental designs, QEDs) were assessed using ROBINS II tool¹²³
- Systematic reviews were assessed using the AMSTAR2 tool¹²⁴.

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.

¹²² 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).

¹²³ 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

¹²⁴ 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.

Appendix 4: About Porticus, Giving Evidence and Campbell Collaboration

Porticus is an international organisation managing and developing the philanthropic programmes of charitable entities established by Brenninkmeijer family entrepreneurs. Porticus is involved with and fund a broad range of social service activities, including to both faith-based organisations, and organisations unrelated to religious institutions.

Porticus commissioned this EGM and Guidebook to further support its own and others' ongoing work to enhance organisational safeguarding.

Porticus has supported, and currently supports, efforts among its grantees to improve organisational safeguarding of children. Specifically, Porticus:

- requires from all its grantees to have a safeguarding policy
- works with some grantees to further develop interventions that can make the organisations safer.

These projects are conducted in collaboration with both faith-based organisations and non-faith-based organisations. To ensure that all standards for the production of a Campbell EGM are met, Porticus was not involved in any technical steps taken to produce the EGM or this Guidebook, including information retrieval, data analysis and reporting of findings.

Giving Evidence is a consultancy and campaign, which enables and encourages charitable *giving* based on sound *evidence*.

Through consultancy, Giving Evidence helps donors and charities in many countries to understand their impact and to raise it. Through campaigning, thought-leadership and meta-research, we show what evidence is available and what remains needed, what it says, and where the quality and infrastructure of evidence need improving. We have advised many donors in many countries on many issues.

Giving Evidence was founded by Caroline Fiennes, a former award-winning charity CEO, and now Visiting Fellow at Cambridge University's Centre for Strategic Philanthropy. She wrote the *How To Give It* column in the Financial Times for three years, the first column about philanthropy in any major newspaper globally. She is author of the acclaimed book *It Ain't What You Give, It's The Way That You Give It*, which is a guide for donors. She has also written in *Freakonomics*, the Daily Mail and spoken at TED, and is one of the few people whose work has appeared in both *OK! Magazine* and the scientific journal *Nature*.

The Campbell Collaboration is an international social science research network that produces high quality, open and policy-relevant evidence syntheses, plain language summaries and policy briefs. It promotes positive social and economic change through the production and use of systematic reviews and other evidence synthesis for evidence-based policy and practice. It exists to help people make well-informed decisions about social and behavioural interventions. As a largely voluntary organisation, Campbell recognises that it is essential to have transparent policies that set out roles and responsibilities and also guide the work of the organization.

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